



OPEN ACCESS

EDITED BY

Fernando De La Prieta,
University of Salamanca, Spain

REVIEWED BY

Reza Rahimian,
Laval University, Canada
Giorgio Gosti,
Italian Institute of Technology (IIT),
Italy
Shan Gao,
Rensselaer Polytechnic Institute,
United States

*CORRESPONDENCE

Lopamudra Mukherjee
mukherjl@uww.edu
Kevin W. Eliceiri
eliceiri@wisc.edu

RECEIVED 08 September 2022

ACCEPTED 30 November 2022

PUBLISHED 16 December 2022

CITATION

Mukherjee L, Sagar MAK, Ouellette JN,
Watters JJ and Eliceiri KW (2022) A
deep learning framework for
classifying microglia activation state
using morphology and intrinsic
fluorescence lifetime data.
Front. Neuroinform. 16:1040008.
doi: 10.3389/fninf.2022.1040008

COPYRIGHT

© 2022 Mukherjee, Sagar, Ouellette,
Watters and Eliceiri. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

A deep learning framework for classifying microglia activation state using morphology and intrinsic fluorescence lifetime data

Lopamudra Mukherjee^{1*}, Md Abdul Kader Sagar²,
Jonathan N. Ouellette², Jyoti J. Watters³ and
Kevin W. Eliceiri^{2,4,5*}

¹Department of Computer Science, University of Wisconsin, Whitewater, WI, United States,

²Laboratory for Optical and Computational Instrumentation, Department of Biomedical Engineering, University of Wisconsin Madison, Madison, WI, United States, ³Department of Comparative Biosciences, University of Wisconsin Madison, Madison, WI, United States, ⁴Morgridge Institute for Research, Madison, WI, United States, ⁵Department of Medical Physics, University of Wisconsin Madison, Madison, WI, United States

Microglia are the immune cell in the central nervous system (CNS) and exist in a surveillant state characterized by a ramified form in the healthy brain. In response to brain injury or disease including neurodegenerative diseases, they become activated and change their morphology. Due to known correlation between this activation and neuroinflammation, there is great interest in improved approaches for studying microglial activation in the context of CNS disease mechanisms. One classic approach has utilized Microglia's morphology as one of the key indicators of its activation and correlated with its functional state. More recently microglial activation has been shown to have intrinsic NADH metabolic signatures that are detectable *via* fluorescence lifetime imaging (FLIM). Despite the promise of morphology and metabolism as key fingerprints of microglial function, they has not been analyzed together due to lack of an appropriate computational framework. Here we present a deep neural network to study the effect of both morphology and FLIM metabolic signatures toward identifying its activation status. Our model is tested on 1,000+ cells (ground truth generated using LPS treatment) and provides a state-of-the-art framework to identify microglial activation and its role in neurodegenerative diseases.

KEYWORDS

microglia activation state, deep learning, fluorescence lifetime, morphology, LSTM

1. Introduction

Microglia are Central Nervous System (CNS) resident macrophages that play important roles in many neuropathologies (Watters et al., 2005; Garden and Möller, 2006; Tambuyzer et al., 2009; Charles et al., 2011). They are involved in brain development, response to injury and infection as well as maintenance of the healthy

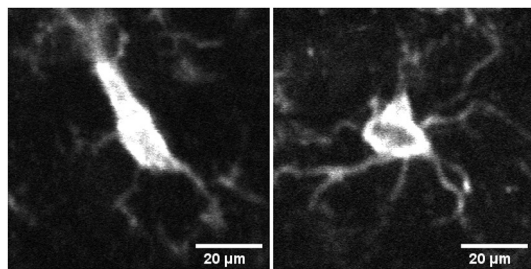


FIGURE 1
Image of microglia in activated (**left**) and surveillant/resting (**right**) state from mid brain section of mouse brain tissue samples.

neural microenvironment. Due to their central role in so many CNS processes and neurodegenerative diseases, it is important to understand the function of microglia in a number of scenarios including microglia activation. In this respect, accurate quantitative imaging and computational tools are needed to identify morphological signatures specific to microglia and understand how they correlate with their activation status.

Microglia react quickly to changes in their environment by exhibiting morphological changes. For example, the change in microglia's activation status is reflected in its gradual morphological transformation from highly ramified into less ramified often amoeboid state (see [Figure 1](#)). [Table 1](#) summarizes the recent literature of morphological classification of microglia. These morphological changes are often closely related to their functional states, and for this reason, microglial morphology is often utilized to infer their activation status, and to study their involvement in virtually all brain diseases ([Heindl et al., 2018](#)). Until recently, available microscopic methods have been unable to capture the extent of these changes in an automated manner, relying mostly on manual assessments which can be prone to error. In the last few years, studies such as ([Zanier et al., 2015](#); [Leyh et al., 2021](#)) have performed microglia morphology classification using machine learning methods by using carefully chosen shape features. However, because the commonly used feature set is typically limited to a few, manually chosen shape parameters, there may be selection bias which can compromise the resulting classifications.

Furthermore, given the important role of microglia plays in all neural diseases, accurate tools for detecting their function beyond morphological alterations are also necessary. In this respect, it has been shown by [Sagar et al. \(2020b\)](#) that microglia metabolic state have unique metabolic fluxes which can be detected by changes in reduced nicotinamide adenine dinucleotide (NADH) *via* fluorescence lifetime imaging microscopy (FLIM) ([Lakowicz et al., 1992a, 1992b](#)). See [Mechawar et al. \(2022\)](#) for a survey of proinflammatory, anti-inflammatory and metabolic pathways in microglia as well as

TABLE 1 Table summarizing a few methods related to classification of microglia phenotypes from literature.

Torres-Platas et al. (2014)	Identified four main phenotypes based on morphology: ramified, primed, reactive and amoeboid in humans
Reemst et al. (2022)	Identified two morphological subtypes in PBS-injected mice, a small cell soma and long branched ramifications or a larger cell soma and thicker, branched ramifications
Leyh et al. (2021)	Identified four microglial morphologies, ramified, rod-like, activated and amoeboid microglia using a mouse model of ischemic stroke using machine learning
Verdonk et al. (2016)	Using the knock-in mouse model, found that in resting state, microglial cells were distributed in four microglial sub-populations with a regional pattern and specific behavior.

Additional review can be found in [Schwabenland et al. \(2021\)](#).

[Rahimian et al. \(2021\)](#) to understand the diversity of microglial phenotype and function in psychiatric diseases. FLIM can probe the cellular microenvironment of the fluorescent NADH in a label-free manner which does not change the metabolic signature. Furthermore, recent machine learning methods have shown that FLIM based lifetime data capturing the metabolic alterations, can be used to both differentiate microglia from other CNS cell-types using deep learning approaches ([Sagar et al., 2020a](#); [Mukherjee et al., 2021](#)) as well as identify their activation status ([Sagar et al., 2020b](#)). These studies show promise in the use of computational tools for analyzing FLIM data to understand the functional role of microglia in the CNS.

It has been recently shown that different microglia functional phenotypes are associated with both distinct metabolic pathways as well as specific morphological changes. [Voloboueva et al. \(2013\)](#) and [Orihuela et al. \(2016\)](#). However a complete computational study investigating both the role of morphology as well as metabolism toward identifying activation state of microglia has not been performed until now. In this paper, we propose an unified deep neural network method to study the effect of both morphology as well as lifetime toward identifying the activation status of microglia. The framework of the model is shown in [Figure 2](#). First, we extract a number of shape features from segmented images of microglia to characterize their morphological properties. This along with the lifetime data from the same images is inputted to a neural network, consisting of Long Short Term Memory (LSTM) ([Yu et al., 2019](#)) and convolutional sub-networks, which process each type of data, which are subsequently combined to yield a joint classification framework that can distinguish between activated and resting microglial cells. Experimental evaluations show that this leads to a highly accurate network compared to

morphology or lifetime considered standalone and can classify the activation state correctly across a number of samples. This combined morphology-FLIM metabolism deep learning framework provides a computationally efficient approach to identify the activation state of microglia automatically, which can be useful to analyze its role in neurodegenerative diseases.

2. Methods

The proposed model combines CNN for learning features from the morphology data with LSTM for learning temporal dependencies in the lifetime data, to derive the classification of activated microglia from resting. We describe the details of each approach next.

2.1. Morphological features and network

The morphology of microglia is one of their more outstanding characteristics. Microglia remain in a *resting or surveillant state* in the normal brain, but upon the detection of any brain lesion or injury, they obtain an “activated” state which displays more inflammatory features. This change in state also manifests in a change of morphological characteristics—from a more ramified structure in resting state to a more amoeboid shape in the activated state. Quantifying such changes can be the key to identifying the activation state of microglia. Here, we aim to capture the changes in morphological characteristics of microglia, by obtaining a number of shape features, which are then concatenated and passed as input to a convolutional neural network (CNN). The CNN transforms the input features to a different feature space, which is more conducive for learning the classes. To obtain the input shape features, we segmented the microglia from background using the WEKA segmentation toolbox in Fiji (Arganda-Carreras et al., 2017). We used a diverse set of features from moment based features to contour and transform based features as well as shape signature based 1D features. We describe these features next.

2.1.1. Zernike invariant moment features

The Zernike moment (Khotanzad and Hong, 1990; Hwang and Kim, 2006) is a type of the orthogonal invariant (to translation, rotation and scale) moment on the unit sphere and is the most commonly used in image shape feature extraction and description. These are designed to capture both global and geometric information about objects of interest in the image. To compute the Zernike moments of an image, the range of the image is first mapped to the unit sphere with its origin at the image’s center. The pixels falling outside the unit sphere are not used in the computation process. Let $f(r, \theta)$ be the two

dimensional image intensity function using polar coordinates, then Z_{mn} , the Zernike moment of order m and repetition n is denoted by

$$Z_{mn} = \frac{m}{\pi} \sum_r \sum_{\theta} f(r, \theta) V_{mn}^*(r, \theta) \quad (1)$$

Where $V_{mn}^*(r, \theta)$ is the complex conjugate of Zernike polynomial $V_{mn}(r, \theta)$, defined as follows Liu et al. (2007):

$$V_{mn}(r, \theta) = R_{mn}(\theta) \exp(\sqrt{-1}r\theta) \quad (2)$$

Where m and n are nonnegative integers with $n \geq m \geq 0$ and the orthogonal radial polynomial $R_{mn}(\theta)$ is given as

$$R_{nm}(\theta) = \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k! \left(\frac{n+m}{2} - k\right)! \left(\frac{n-m}{2} - k\right)!} \theta^{n-2k} \quad (3)$$

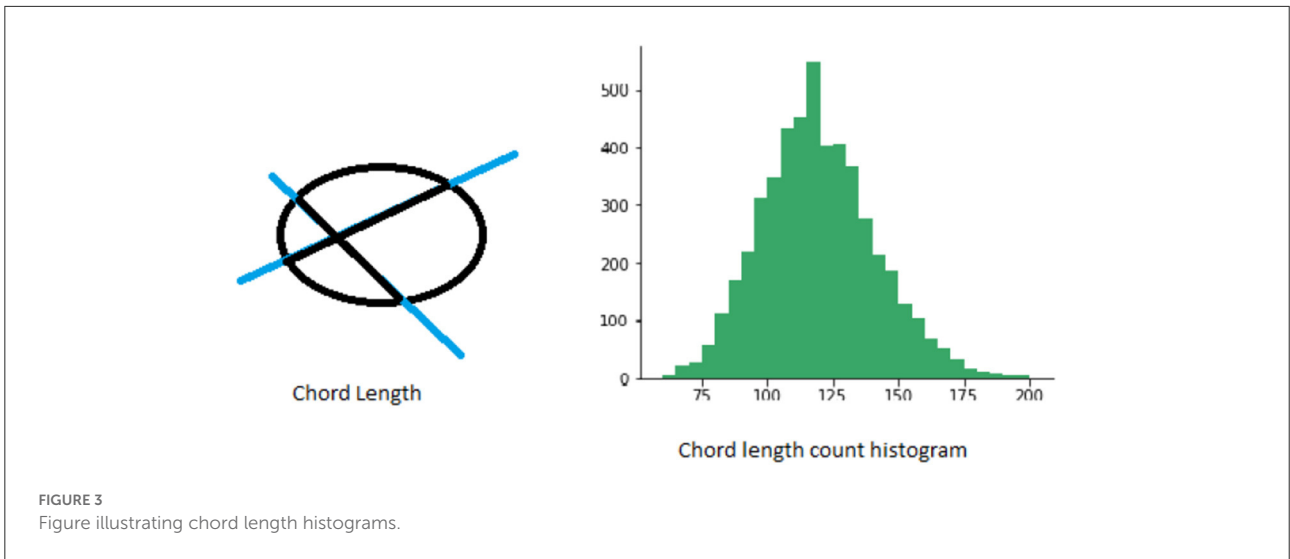
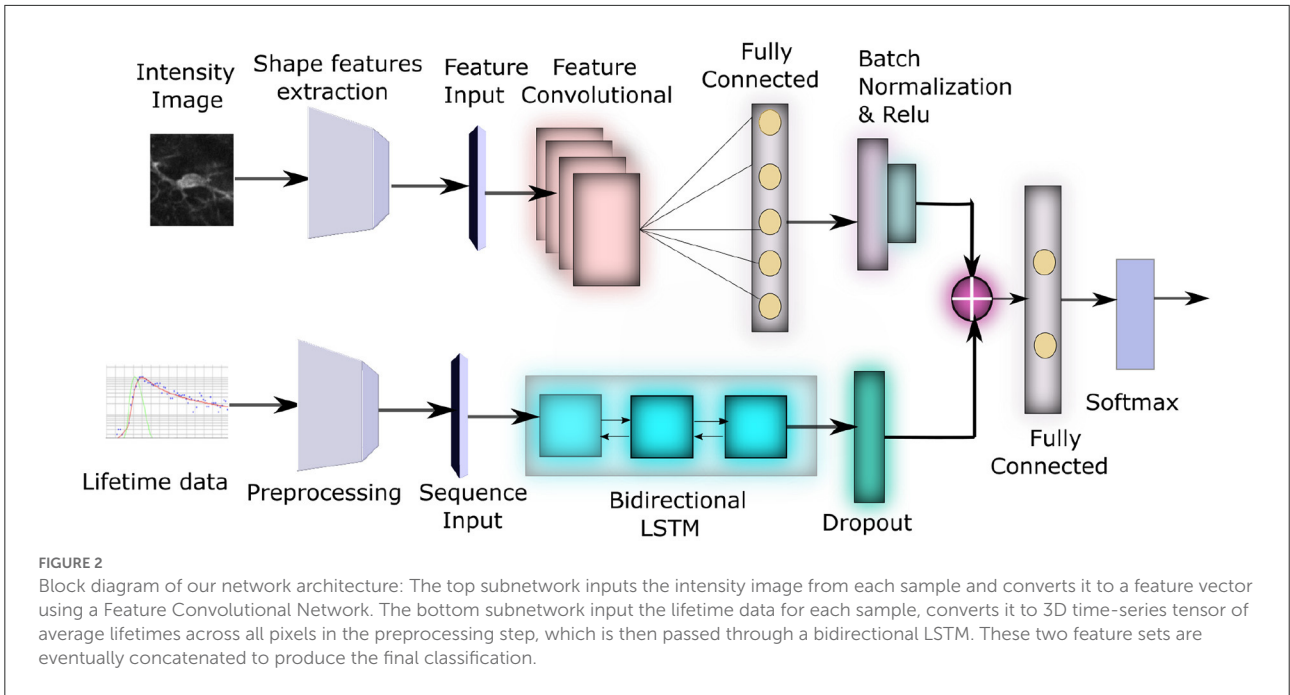
Note that the different order of the Zernike moments can be computed *via* (Equation 1) by varying the order or keeping the order fixed and varying the repetition. It has been shown that the lower-order Zernike moments are useful to represent the whole shape of the image whereas the high-order Zernike moments can describe the details. In our approach, we compute the top 10 Zernike moments.

2.1.2. Chord length histogram

Chord length histogram analysis (Agimelen et al., 2016) corresponds to finding the distribution of all chord lengths in different directions in a given shape. The chords are defined by the parts of lines within the contour of a binary shape. For each boundary point p , its chord length function is the shortest distance between p and another boundary point \hat{p} such that line $p\hat{p}$ is perpendicular to the tangent vector at p . The chord length function is invariant to translation and its centroid is not biased by boundary noise. A shape can be represented by a discrete set of chords sampled from its contour. The number and length of chords obtained in different directions is generally not the same. Therefore, one way to capture this information more effectively is to calculate the distribution of chord lengths in the same direction in a spatial histogram. See Figure 3 for an illustration.

2.1.3. Elliptical Fourier features

This method is based on computing Fourier coefficients to describe a closed contour by a function. First the boundary is extracted from the segmented image of the object. Then the contour edge is encoded using Freeman encoding of closed contour, which yields a chain code, where each integer represents an oriented vector in a specific direction. The length of each of these vectors and their projections on the x and the y axes are computed, which is then used in the calculation of the elliptic harmonics. Elliptical Fourier analysis then approximate



a closed contour as a sum of elliptic harmonics. Let $x(t)$ be the x projection of the complete contour. Then,

$$\beta_n = \frac{2}{T} \int_0^T x(t) \sin\left(\frac{2n\pi t}{T}\right) dt$$

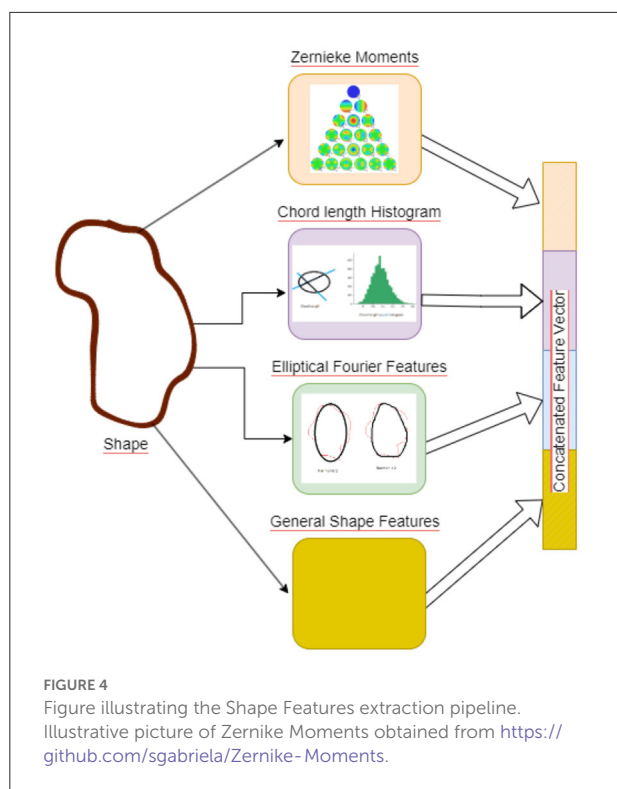
$$x(t) = x_0 + \sum_{n=1}^N x_n(t)$$

$$x_0 = \frac{1}{T} \int_0^T x(t) dt$$

$$x_n(t) = \alpha_n \cos\left(\frac{2n\pi t}{T}\right) + \beta_n \sin\left(\frac{2n\pi t}{T}\right)$$

$$\alpha_n = \frac{2}{T} \int_0^T x(t) \cos\left(\frac{2n\pi t}{T}\right) dt$$

Here T is the period of $x(t)$. We followed the strategy of [Kuhl and Giardina \(1982\)](#) who used four Fourier coefficients for each of N harmonics. Then an inverse process is employed to identify the closed contour as k elements, which can be found as a function of the N harmonics. In addition, phase shift is employed so that the representation does not depend on starting point. See [Kuhl and Giardina \(1982\)](#) and [Ballaro et al. \(2002\)](#) for details regarding specific computations.



2.1.4. General shape features

In addition to the above, we also computed a number of 1D general shape features, which describe the high-level geometric properties of the objects. These include area, perimeter, eccentricity, major and minor axis length, bounding box and equivalent diameter. Such features can be used as filters to eliminate false positives and are generally combined with other shape descriptors (as in our case) to discriminate shapes.

2.1.5. Network for learning shape features

Shape features are concatenated to generate a 67 dimensional feature vector, see Figure 4, the features are then mean centered and passed through a feature convolutional neural network. The first layer in this network is a feature input layer. This is followed by a Feature Convolution Network as described by Hu (2021) which extends the idea of convolution to tabular feature data. Normally the convolution process is applied to the spatial region of an image using a kernel. To extend this to tabular features, we use the method proposed by Hu (2021) for combining pairs of features to create new features.

$$\phi_{ij}(x) = \alpha_i x_i + \beta_j x_j \quad (4)$$

Where x_i (and x_j) are the i th (and j th) dimensional feature and α_i and β_j are the weights of the kernel. There is one kernel parameter (α and β) for each feature in the convolution operation. Therefore, if the number of features are n , this

produces a feature map of size $C(n, 2)$. The network outputs ϕ as the output of feature convolution layer, which is then added with x , the original features, making it a function of both original and convolved features. In our network, this is followed by a Fully Connected (FC) layer (with 50 hidden units), a batch normalization layer, and Rectified Linear activation (ReLU) layer. The output of the ReLU layer is concatenated with the output of the LSTM to form a joint network.

2.2. Lifetime data and LSTM subnetwork

Fluorescence Lifetime image microscopy (FLIM) (Lakowicz et al., 1992a, 1992b) is an imaging technique which is used to visualize physiological properties in living cells by measuring the time a molecule (excited by a photon) remains in its excited state on an average before returning to its ground state and emitting a photon. This lifetime of the molecule is then calculated using an exponential decay function and is used to localize specific fluorophores. Also, FLIM can use intrinsic fluorescence of NADH to probe cellular micro-environment in a label-free manner which does not alter the cellular signature. FLIM related monitoring of microglial metabolic alterations can be used to both differentiate microglia from other CNS cell-types (Sagar et al., 2020a; Mukherjee et al., 2021) as well as identify their functional state (Sagar et al., 2020b).

We first describe the mathematical representation of the fluorescent decay curve obtained by FLIM, followed the architecture of the LSTM subnetwork used to process this data. Assume a single image is under consideration of size $m \times m$. The measured fluorescence intensity decay data (Y_t) for a given lifetime component t is given by the convolution of the tissue fluorescence response signal (I_t) with the excitation light pulse [part of the instrument response (F_t)] along with some additive noise (ϵ_t). This relationship can be written as

$$Y_t = I_t \otimes F_t + \epsilon_t \quad (5)$$

Where \otimes represents convolution of the two signals and Y_t (as well as I_t and ϵ_t) are of size $m \times m$. The function I_t can be approximated as a multi-exponential decay function, with multiple components. Here we use bi-exponential decay function to represent the signal, because in practice, it is enough to closely approximate the signal and can be written as

$$I_t \sim a_1 \exp\left(\frac{-t}{\tau_1}\right) + a_2 \exp\left(\frac{-t}{\tau_2}\right) \quad (6)$$

Where a_1 , a_2 and τ_1 , τ_2 represent the amplitude and the lifetimes for each exponential sub-function in this bi-exponential function. We now extend this data to multiple images by using the superscript i to denote the image number. The lifetime data $\{Y_t\}^i$ (without the noise) is collected for the same lifetime component $t \in t_1, \dots, t_p$ for all images $i \in 1 \dots n$

in our dataset. $\{Y_t\}^i$ is then vectorized and binned into h bins to generate a vector $\{X_t\}^i$ in R^h , which forms the input to the following Long Short Term Memory or LSTM network. The same number of bins are used across all images, making it feasible to compare them as multiple time-series data. The target class label y is of size n . An intuitive way to understand this data is that there are h dimensional features for each image which evolve over lifetime components and can be used to understand group differences between the two classes under consideration.

We now discuss the network architecture to process this data. First the 3D tensor X is passed through a sequence input layer. The next layer in the network is a bidirectional Long Short Term Memory (LSTM) (Yu et al., 2019) network. LSTM networks can capture long term dependencies in temporal data and has been successfully used for a number of time series classification problems. LSTM (similar to Recurrent Neural Networks, RNN; Sherstinsky, 2020) contains loops in its architecture which allows it to memorize previous states such that the network can effectively process temporal data. A typical LSTM layer consists of a set of recurrently connected blocks, known as memory blocks. Figure 5 shows the design of a typical LSTM unit or memory block. Each block contains one or more recurrently connected memory cell (c^t) and three multiplicative units — the input (i^t), output (o^t), and forget gates (f^t) which regulate the extent to which data is propagated through the LSTM unit. The operations inside an LSTM block can be formulated by the following:

$$\begin{aligned}
 f_t &= \rho(W_f x_t + U_f h_{t-1} + b_f) \\
 i_t &= \rho(W_i x_t + U_i h_{t-1} + b_i) \\
 o_t &= \rho(W_o x_t + U_o h_{t-1} + b_o) \\
 \hat{c}_t &= \phi(W_c x_t + U_c h_{t-1} + b_f) \\
 c_t &= f_t \circ c_{t-1} + i_t \circ \hat{c}_t \\
 h_t &= o_t \circ \phi(c_t)
 \end{aligned}
 \tag{7}$$

Here ρ and ϕ are activation functions, \circ denote the element wise product operation, x_t is the input vector, W and U are weights and h_t is hidden state vector also known as output vector of the LSTM unit. Since the values of histogram for a certain lifetime component can be dependent of both past and future lifetime components, we use a Bidirectional LSTM (BLSTM), which includes both a forward and backward layer of LSTMs. Both the forward and backward layer outputs are calculated by using the standard LSTM update (Equation 7). Then BLSTM connects the two hidden layers to the same output layer. More details about BLSTMs can be found in Graves and Schmidhuber (2005) and Cui et al. (2020). The final layer of this subnetwork is a dropout layer, which randomly sets 50% of the input to 0.

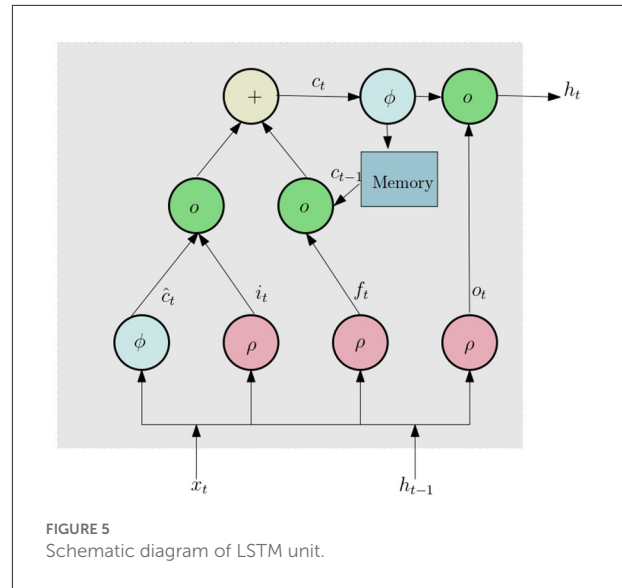


FIGURE 5 Schematic diagram of LSTM unit.

2.3. Joint feature-LSTM network

The outputs of the morphological features subnetwork ($\text{ReLU}(\text{BN}(W_c \phi))$, where W_c is the activation weights for the fully connected layer in that subnetwork) and the LSTM subnetwork ($\delta(h_t)$ where h_t is the output of the final LSTM unit and the function $\delta(\cdot)$ implements the dropout) is concatenated to produce a joint feature map. This is followed by a fully connected layer with 2 hidden unit (activation function W_f) corresponding to each of the classes, a batch normalization layer (denoted by $\text{BN}(x) = \frac{x - E(x)}{\sqrt{\text{Var}(x)}}$) and finally a Softmax layer with cross entropy loss $\text{loss}(\sigma(y, \hat{y})) = -\frac{1}{n} \sum_{i=1}^n \hat{y}_i \log(y_i)$ where \hat{y} is the ground truth).

$$z = \text{concatenate}(\text{ReLU}(\text{BN}(W_c \phi)) + \delta(h_t)) \tag{8}$$

$$y = \arg \min(\sigma(\text{BN}(W_f z), \hat{y})) \tag{9}$$

3. Materials

We conducted our experiments on two datasets of microglial cells containing both intensity images and corresponding lifetimes. The first data collected from mouse brain tissue samples consisting of 826 different cells, and the lifetime data collected across 256 time bins for each. We will refer to this as Dataset₁. The second dataset is a collection of 506 samples from frontal cortex and hippocampus of mice, here the lifetime data is collected across 64 time bins. We refer to this dataset as Dataset₂. For each dataset, we utilized the Becker&Hickl TCSPC (Time correlated single photon counting) software (Becker,

2021) to produce various parameters that includes the lifetime parameters and intensity images. We briefly elaborate on the acquisition details of the dataset in next section.

3.1. Tissue preparation and imaging

All animals were maintained in an AAALAC-accredited animal care facility with a 12-h light/dark cycle regime and had free access to food and water. All experiments were approved by the University of Wisconsin-Madison Institutional Animal Care and Use Committee (protocol V005173; exp. 5/16/2024).

3.1.1. Dataset₁ preparation and immunohistochemistry

One hundred micrometer thick coronal slices were generated from the fixed brains of 6–8 weeks old young adult male C57BL/6J and CX3CR1-GFP mice (Jackson Labs), to obtain the FLIM images. The mice were euthanized by isoflurane overdose and transcardially perfused with 30 ml of ice-cold PBS. This was followed by a second perfusion with an ice-cold solution of 4% PFA in PBS. After this, the brains were dissected and acutely post-fixed in 4% PFA prior to putting them into 30% sucrose in PBS overnight at 4°C until they sank. Brains were then stored in 15% sucrose/HBSS at –20°C prior to sectioning.

Immunohistochemistry: 100 μm thick coronal sections from the midbrain region of each brain was prepared using a Leica Vibratome. For immunohistochemical staining, two slices from each animal were used. These were washed with 0.3% TritonX-100 in PBS at room temperature, before incubating in blocking buffer (1% BSA, 0.3% 2 h also at room temperature). The slices were incubated with anti-Iba1 antibodies (1 : 1,000; Wako Catalog No. 019-19741) in blocking buffer at 4°C overnight without light. This was followed by washing the slices at room temperature with 0.3% TritonX-100 in PBS. After this, the slices were incubated in the dark for 2 h with AlexaFluor594 anti-rabbit IgG antibodies (1 : 200) in blocking buffer, at room temperature. Slices were then washed with 0.3% TritonX-100 in PBS and mounted on 1mm slides using Cytoaseal60 mounting medium. Finally the mounted sections were stored at room temperature and protected from light until imaging was done.

3.1.2. Dataset₂ preparation and immunohistochemistry

Dataset₂, our second dataset, comprises CX3CR1-GFP mice (stock no 005582) ordered from Jackson laboratories (Bar Harbor, ME, USA). Mice were divided into two treatment groups and injected intraperitoneally with either 1 mg/kg lipopolysaccharide (LPS; Sigma-Aldrich) diluted in sterile Hanks Buffered Salt Solution (HBSS; Corning) or with vehicle (sterile

1xHBSS). There were 5 mice in each of the LPS treated, and vehicle treated groups. Animals were euthanized 16 h following vehicle or LPS injections and intracardially perfused with ice cold 1X Phosphate Buffered Saline (PBS) solution (30 mL per mouse). Mice were then perfused again with ice-cold 4% paraformaldehyde (PFA) in 1xPBS, pH 7.4. Brains were dissected intact, post-fixed for 24 h in a solution of 4% PFA in 1xPBS, then moved to 15% sucrose/1xHBSS (all performed at 4°C and protected from light).

Each brain was then cut into 100μm thick coronal sections using a Leica vibratome. Slices were collected from regions of the brain containing frontal cortex and hippocampus for imaging (4 slices from each region). Slices were then mounted on 1-mm slides using Cytoaseal604 (Richard-Allan Scientific, Kalamazoo, Michigan) mounting medium and 1.5 coverslips. The Cytoaseal60 was allowed to cure for 24 h before sealing the edge of the slides with clear nail polish. Mounted sections were stored at room temperature, protected from light until imaging.

3.1.3. Multiphoton lifetime imaging

The fluorescence lifetime (Lakowicz et al., 1992a) and multiphoton imaging (Denk et al., 1990) was performed on a custom multiphoton laser scanning system (built around an inverted Nikon Eclipse TE2000U) at the Laboratory for Optical and Computational Instrumentation (LOCI) in Madison, Wisconsin (Yan et al., 2006). A 20x air objective (Nikon Plan Apo VC, 0.75 NA) (Melville, NY, USA) was used for all imaging. The data was collected using an excitation wavelength of 740 nm for NAD(P)H, and the emission was filtered at 457/50 nm (Semrock, Rochester, NY) for the spectral peak. To identify the microglia, Iba1 (Ito et al., 1998) was used as the primary binding protein, whereas AlexaFluor594 was used as secondary protein. For generating the intensity images from FLIM, excitation was set at 810 nm, and a 615/20 (Semrock, Rochester, NY) bandpass emission filter was used for emission. For processing, we used Becker and Hickl time domain FLIM imaging software where decays curves are built with TCSPC (Time Correlated Single Photon Counting) electronics. FLIM images of 256 × 256 pixels (or 64 × 64 pixels) were collected with 120 s collection (for Dataset₁) and 90 s (for Dataset₂) using SPC-150 Photon Counting Electronics (Becker and Hickl GmbH, Berlin, Germany) and Hamamatsu H7422P-40 GaAsP photomultiplier tube (Hamamatsu Photonics, Bridgewater, NJ). To determine the Instrumentation Response Function (IRF), we used urea crystals with a 370/10 bandpass emission filter (Semrock, Rochester, NY) and was measured during each imaging session. In Dataset₁, about 20 neighboring FOVs were chosen randomly, and the average lifetime value and free NADH ratio was calculated based on masking. For Dataset₂, our second dataset, the entire region was imaged (Frontal Cortex or Hippocampus) and we selected the FOVs which that contained sufficiently bright microglia.

4. Experiments

We describe our evaluations in the following way: (1) First we obtain classification accuracy and other metrics using five-fold cross validation to quantitatively measure how well our model performs for the task of distinguishing activated microglia from resting in both datasets. (2) Secondly, we study the contribution of lifetime data in the joint model by comparing it against the performance of the network which uses only the shape features. (3) Third, we evaluate the performance of our model as a classifier by comparing it to other well-known classification schemes such as SVM, KNN and random forest. (4) Next, we evaluate the utility of each individual shape features that constitute our feature set and their combinations toward classification by comparing the classification results on a number of different subset of shape features. (5) We then study how the data is correlated across the two feature sets (shape and lifetime) by the observing their projection on canonical components found by CCA. (6) Finally, we evaluate the effect of training set size and model parameters toward the accuracy and efficacy of learning. Our models were implemented using Matlab's deep learning toolbox and Stochastic Gradient Descent (SGD) was employed to do the optimization. We discuss these issues next.

4.1. Evaluation of efficacy of joint morphology-lifetime based model

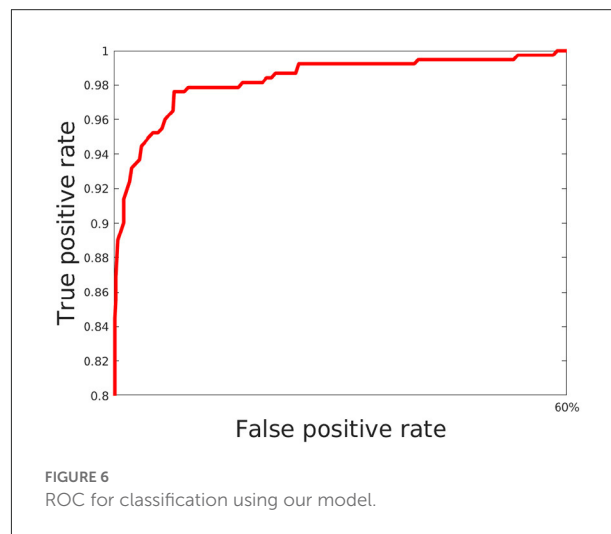
Here, we train out model using 5 fold cross validation (this was performed by dividing the data into 5 equal parts and using 4 parts for training and one part for evaluation/validation of the model's performance. To choose the model parameters, such cross validation experiments were repeated for various settings of parameter values such as epoch and the setting with best accuracy was chosen). We use 6 different classification metrics to evaluate the quality of the classification with respect to ground truth. These include the Accuracy (Acc), Precision (prec), Recall (rec), Jaccard coefficients (JI), Area Under the Curve (AUC) as well as the ROC curve which plots true positive rate with respect to false positive rate (Figure 6). Results shown in Table 2 shows these metrics by averaging across all cross validation runs for the testing set. Analysis of these results show our model performs very well across all metrics and has both high accuracy as well as precision. This indicates that it is successfully able to distinguish activated microglia from resting for both datasets.

4.2. Evaluation of utility of lifetime data

In this section, we study the impact of the lifetime data in the joint model. To do this, we use a separate network consisting of only of the subnetwork described in Section 2.1, which processes

TABLE 2 Results showing performance of the joint lifetime+feature network on various metrics.

Metric	Dataset ₁	Dataset ₂
Accuracy	0.964	0.977
Precision	0.986	0.996
Recall	0.95	0.961
Jaccard index	0.938	0.958
AUC	0.966	0.981



the shape feature data. We use similar 5-fold cross validation studies to obtain the accuracy results for both training and test and compare it to the results from the joint network in Table 3. The results show that while the shape based network performs well, accuracy improves by 2 – 3%, when lifetime data is included, which indicates it is a useful feature for this purpose.

We also studied the utility of LSTM network in processing the lifetime data as a time series by comparing with the fitted life time data τ [given as $\tau = \text{mean}(a_1\tau_1 + a_2\tau_2)$] where the variables have the same meaning as in Equation 6. Since this converts lifetime into a feature, it can simply be concatenated to the existing feature set and then our subnetwork in Section 2.1 is used for learning. The results did not show any significant improvements compared to the features used without τ (Table 3, last two rows). This shows that using the individual values a_1 , a_2 , τ_1 and τ_2 for each lifetime component t is more useful for this purpose than a summary of the lifetime (using τ).

4.3. Comparison with other classifiers

Here, we discuss the performance of other classifiers on our datasets. We ran similar 5-fold cross validation studies by training SVM, KNN (with $K = 3$), an ensemble classifier

TABLE 3 Results showing comparison of the shape+lifetime network with only feature based network.

	Metric	Dataset ₁	Dataset ₂
Shape + Lifetime	Train accuracy	0.971	0.981
	Test accuracy	0.972	0.978
Shape	Train accuracy	0.9472	0.977
	Test accuracy	0.9418	0.975

TABLE 4 Comparison with other classifiers.

Method	Dataset ₁	Dataset ₂
Ours	0.972	0.978
SVM	0.8848	0.9644
KNN	0.9304	0.9605
Ensemble	0.9406	0.9723
DNN	0.9479	0.9684

and a feed-forward deep neural network (DNN) with two hidden layers each followed by a leaky ReLU and a softmax function as the output layer on both datasets. The SVM was trained with a Radial Basis Function (RBF) kernel whereas the ensemble classifier uses boosting to aggregate of 100 individual classification trees. Since there is no easy way to incorporate 3D time-series data similar to the format we are using in our network, in these classifiers, they were trained on shape features alone. We report on these results in Table 4. It shows that our model outperforms the other classifiers especially on Dataset₁. Note that we should not compare the classifier scores across these two datasets, since we found that in general Dataset₂ is easier to classify compared to Dataset₁ as is evidenced by the higher accuracy and other metrics for some of the comparable methods used here.

4.4. Evaluation of individual shape features and their combinations

This experiment is aimed at validating the choice of shape feature set described earlier by comparing its performance with other several feature extraction approaches and their concatenation. To do this, we not only run the joint shape+lifetime model on General shape features (GenShape), Moment based features (Mom), Chord length histograms (ChordLen) and Elliptical Fourier features (ElliFou) individually, but also on various combinations of these features listed in Table 5. These features have been described earlier in Section 2.1. We see that Chord length histogram are the best individual feature whereas Moment based features perform the worst. The combinations of these feature gradually

improve the performance and the best results are obtained when all four feature sets are concatenated to yield the feature set.

4.5. Correlation of lifetime and feature data

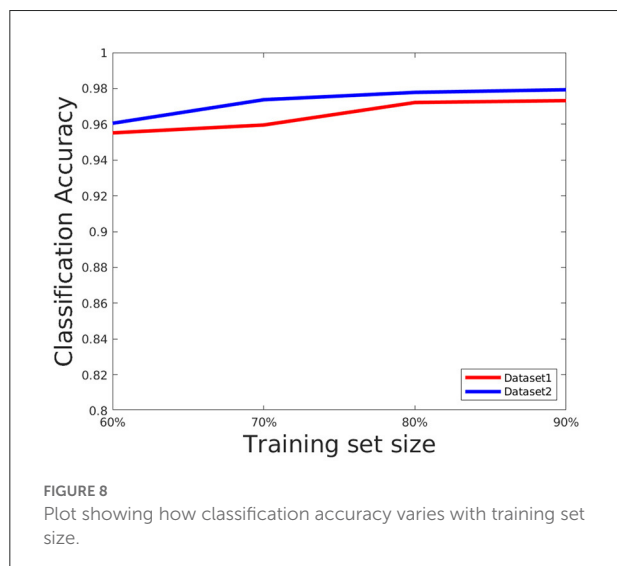
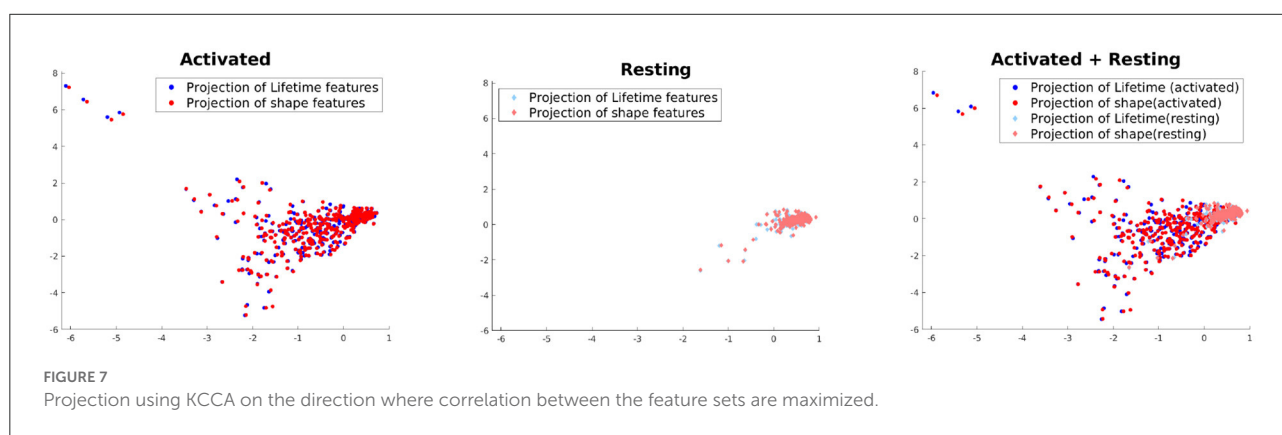
This experiment is not directly related to studying the performance of our model but aimed at understanding to what extent the shape and lifetime data used in this paper are correlated to each other across the two activation classes. For this purpose, we use Kernel Canonical Correlation Analysis (KCCA) (Hardoon et al., 2004) to find a common subspace where both of these feature sets are the most correlated. To apply KCCA, we need to construct kernels for both these feature sets. For the lifetime data, we compute the Wasserstein distance (a distance functions used for probability distributions) (Rüschendorf, 1985) of the histograms at each lifetime component for two samples i and j , which is then summed up across all lifetime components, to yield a distance d_{ij} . The kernel is then constructed as $K_{ij}^L = \exp(-td_{ij})$, where one can tune if needed the bandwidth parameter t according to the learning task. For the features, kernel K^F is simply the gaussian kernel. We use the KCCA approach of Hardoon et al. (2004) to compute two directions of maximum correlation between these feature sets and project the data onto these subspaces. The results can be seen in Figure 7. It shows that the variance in the data is higher and the correlation between the feature sets are lower in case of activated vs. resting. This seems to intuitively explain the multiple morphological states which all can be evident in the activated state of microglia.

4.6. Evaluation of training set size network parameters

The network performance depends on various parameters such as learning rate, number of epochs used in training as well as size of the training set. We briefly discuss these issues here. We set the number of epochs to 50. Higher epochs lead to higher training and test accuracy but the running time in training is higher. The learning rate has been set to 0.0001, this value is also determined empirically as a trade-off of accuracy vs. running time. To evaluate the effect of the training set, we choose the training set size from the following set: {60, 70, 80, and 90%} of the overall data, and the rest of the data is considered for testing and evaluate the accuracy in each setting. The results are shown in Figure 8. As can be seen from this figure, the accuracy improves for a larger training size but change in magnitude is relatively small. This shows the model can give

TABLE 5 Comparison of different feature subsets.

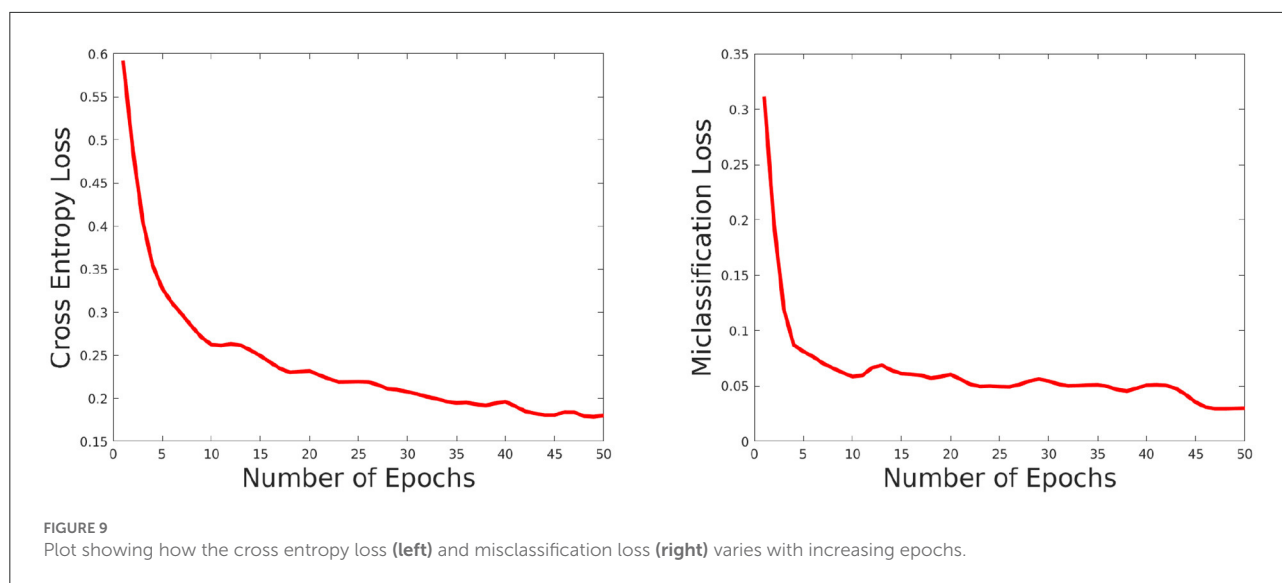
Feature set	Dataset ₁	Dataset ₂	Feature set	Dataset ₁	Dataset ₂
GenShape	0.86	0.845	Mom	0.545	0.562
ChordLen	0.96	0.95	ElliFou	0.918	0.8535
GenShape+Mom	0.957	0.958	GenShape+ChordLen	0.961	0.950
GenShape+ElliFou	0.963	0.9683	Mom+ChordLen	0.962	0.948
ChordLen+ElliFou	0.961	0.956	Mom+ElliFou	0.64	0.697
Mom+ChordLen+ElliFou	0.961	0.9624	GenShape+ChordLen+Mom	0.961	0.954
GenShape+ChordLen + ElliFou	0.958	0.960	All	0.964	0.977



good performance even with a smaller training set. We also plotted the cross entropy loss and misclassification loss with respect to epochs in Figure 9. This shows that both losses show relatively small change in magnitude after about 20 epochs, so the number of epochs can be reduced without a substantial change in accuracy of the network.

5. Discussion

We have shown that our classification method of microglial activation status performs well by using morphology and lifetime metabolism data together. This study provides improved tools for researching microglia morphological changes and their corresponding responses that are associated with changes in the CNS microenvironment. There have been several recent developments investigating intensity-based data from tissue immunohistochemistry (Heindl et al., 2018) to analyze morphological shifts during the microglial inflammatory response. But to our knowledge, this is the first study that incorporates deep learning approaches to study microglia morphology in order to fully automate extraction of morphological information. This method can be further adapted to study microglia in various situations to characterize their activation state without requiring intensity-based morphological analyzes. In the future, we aim to study microglial activation in response to neurodegenerative disease, such as Alzheimer's disease, in which microglial inflammation is a hallmark. We postulate that more accurate classification of microglia functional states will lead to better prediction of early onset of neurodegenerative disease. Given how physiological inflammation affects immune cells differently; we believe our study could allow differentiation of microglia and macrophages or other CNS cell types. The novelty of combining microglial



morphology with FLIM NADH measures adds an additional layer of tools with which to identify a microglia-specific cellular signature. This automated deep learning approach can also help distinguish the localized distribution of activated microglia to detect local inflammation.

In addition to the above mentioned future directions, below, we discuss some possible extensions to the methodological side of this approach as well as some challenges associated with this method.

Challenges: One of the greatest challenges which we inherit from other methods for automatically detecting cell morphology from fluorescence microscopy data is the uneven fluorescence detection across a field of view. Furthermore, because the shape feature detection module is dependent on the upstream segmentation of microglia, an error in the segmentation process can affect the quality of the morphological features obtained. Furthermore, the morphology analysis described above relies on segmentation of a two-dimensional projection that represents three dimensional volume, which is inherently a lossy transformation. In terms of computational issues, the time required for training increases with the size of the dataset, even though it was not an issue with the dataset used in this paper.

5.1. Possible extensions

1. To verify our classification method with regard to using shape features, we analyzed several morphological parameters obtained from 4 different types of shape features. Although this is more extensive than the types of features used in previous studies, in the context of microglia (Zanier et al., 2015; Fernández-Arjona et al., 2019), there are a lot more shape features that could also be used. Examples of such shape

features include but are not limited to shapelets, wavelets, and Bag of contour fragments (BCF) (Wang et al., 2014). Since shape features are generated independently from our neural network, any of the above mentioned features can be easily integrated and tested with our model.

2. In this paper, we have focused on only two functional states of microglia (surveillant and reactive), however in terms of morphology alone, microglia are often classified into several different morphological classes (Leyh et al., 2021). It would be interesting to extend our model to a multiclass classification framework to study how lifetime parameters correlate with different morphological classes.
3. As an additional step, we can study microglia from serial sections for 3D reconstruction of human brain tissue. For this, we would need to acquire volumetric features, but the rest of the algorithm can be applied without adjustments.

6. Conclusion

Using fluorescence lifetime imaging, here we propose an efficient approach to characterize microglial function/activation state, using a wide variety of shape features (far beyond what is commonly used in literature), together with metabolic characteristics of microglia. Our results show that this tandem analysis results in highly accurate predictions of microglial activation states, delivering state of the art results on over 1,000 cell samples. Our proposed method for identifying microglial activation status using deep learning methods provides an unbiased, objective and computationally efficient approach that can serve as a useful tool for characterizing microglial functional and morphological transformations in the diseased CNS of animal models and humans alike. It also provides a baseline which can be extended by future studies that aim to apply deep

learning algorithms toward identifying microglial subtypes and assess their accuracy.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The animal study was reviewed and approved by University of Wisconsin-Madison Institutional Animal Care and Use Committee.

Author contributions

LM and MS contributed to the design and implementation of the study. JO is responsible for acquiring the dataset under the supervision of JW. KE is a senior author who contributed to developing of the main ideas and wrote sections of the paper. All authors contributed to manuscript revision, read, and approved the submitted version.

References

- Agimelen, O. S., Jawor-Baczynska, A., McGinty, J., Dziejewicz, J., Tachtatzis, C., Cleary, A., et al. (2016). Integration of in situ imaging and chord length distribution measurements for estimation of particle size and shape. *Chem. Eng. Sci.* 144, 87–100. doi: 10.1016/j.ces.2016.01.007
- Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A., et al. (2017). Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* 33, 2424–2426. doi: 10.1093/bioinformatics/btx180
- Ballaro, B., Reas, P., and Tegolo, D. (2002). “Elliptical fourier descriptors for shape retrieval in biological images,” in *Conferences on Electronics, Control and Signal*.
- Becker, W. (2021). *The bh TCSPC Handbook*. Available online at: www.becker-hickl.com
- Charles, N. A., Holland, E. C., Gilbertson, R., Glass, R., and Kettenmann, H. (2011). The brain tumor microenvironment. *Glia* 59, 1169–1180. doi: 10.1002/glia.21136
- Cui, Z., Ke, R., Pu, Z., and Wang, Y. (2020). Stacked bidirectional and unidirectional lstm recurrent neural network for forecasting network-wide traffic state with missing values. *Transport. Res. C Emerg. Technol.* 118, 102674. doi: 10.1016/j.trc.2020.102674
- Denk, W., Strickler, J., and Webb, W. (1990). Two-photon laser scanning fluorescence microscopy. *Science* 248, 73–76. doi: 10.1126/science.2321027
- Fernández-Arjona, M. D. M., Grondona, J. M., Fernández-Llebrez, P., and López-Avalos, M. D. (2019). Microglial morphometric parameters correlate with the expression level of il-1 β , and allow identifying different activated morphotypes. *Front. Cell. Neurosci.* 13, 472. doi: 10.3389/fncel.2019.00472
- Garden, G. A., and Möller, T. (2006). Microglia biology in health and disease. *J. Neuroimmune Pharmacol.* 1, 127–137. doi: 10.1007/s11481-006-9015-5
- Graves, A., and Schmidhuber, J. (2005). “Framewise phoneme classification with bidirectional lstm networks,” in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005, Vol. 4* (Montreal, QC: IEEE), 2047–2052.
- Hardoon, D. R., Szedmak, S., and Shawe-Taylor, J. (2004). Canonical correlation analysis: an overview with application to learning methods. *Neural Comput.* 16, 2639–2664. doi: 10.1162/0899766042321814
- Heindl, S., Gesierich, B., Benakis, C., Llovera, G., Duering, M., and Liesz, A. (2018). Automated morphological analysis of microglia after stroke. *Front. Cell. Neurosci.* 12, 106. doi: 10.3389/fncel.2018.00106
- Hu, H. (2021). “Feature convolutional networks,” in *Proceedings of The 13th Asian Conference on Machine Learning, volume 157 of em Proceedings of Machine Learning Research*, eds V. N. Balasubramanian and I. Tsang, 830–839.
- Hwang, S.-K., and Kim, W.-Y. (2006). A novel approach to the fast computation of zernike moments. *Pattern Recognit.* 39, 2065–2076. doi: 10.1016/j.patcog.2006.03.004
- Ito, D., Imai, Y., Ohsawa, K., Nakajima, K., Fukuuchi, Y., and Kohsaka, S. (1998). Microglia-specific localisation of a novel calcium binding protein, iba1. *Mol. Brain Res.* 57, 1–9. doi: 10.1016/S0169-328X(98)00040-0
- Khotanzad, A., and Hong, Y. H. (1990). Invariant image recognition by zernike moments. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 489–497. doi: 10.1109/34.55109
- Kuhl, F. P., and Giardina, C. R. (1982). Elliptic fourier features of a closed contour. *Comput. Graphics Image Process.* 18, 236–258. doi: 10.1016/0146-664X(82)90034-X
- Lakowicz, J. R., Szmajcinski, H., Nowaczyk, K., Berndt, K. W., and Johnson, M. (1992a). Fluorescence lifetime imaging. *Anal. Biochem.* 202, 316–330. doi: 10.1016/0003-2697(92)90112-K
- Lakowicz, J. R., Szmajcinski, H., Nowaczyk, K., and Johnson, M. L. (1992b). Fluorescence lifetime imaging of free and protein-bound NADH. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1271–1275. doi: 10.1073/pnas.89.4.1271
- Leyh, J., Paeschke, S., Mages, B., Michalski, D., Nowicki, M., Bechmann, I., et al. (2021). Classification of microglial morphological phenotypes using machine learning. *Front. Cell. Neurosci.* 15, 701673. doi: 10.3389/fncel.2021.701673

Funding

This research is funded by NIH Grant Nos. [R01NS085226 (JW)] and [R01CA199996, P41GM135019, and U54CA268069 (KE)]. This research is supported by the UW Laboratory for Optical and Computational Instrumentation, the Morgridge Institute for Research, the UW Carbone Cancer Center.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Liu, M., He, Y., and Ye, B. (2007). Image zernike moments shape feature evaluation based on image reconstruction. *Geospatial Inf. Sci.* 10, 191–195. doi: 10.1007/s11806-007-0060-x
- Mechawar, N., Rahimian, R., Belliveau, C., and Chen, R. (2022). Microglial inflammatory-metabolic pathways and their potential therapeutic implication in major depressive disorder. *Front. Psychiatry* 13, 871997. doi: 10.3389/fpsyt.2022.871997
- Mukherjee, L., Sagar, M. A. K., Ouellette, J. N., Watters, J. J., and Eliceiri, K. W. (2021). Joint regression-classification deep learning framework for analyzing fluorescence lifetime images using nadh and fad. *Biomed. Opt. Express* 12, 2703–2719. doi: 10.1364/BOE.417108
- Orihuela, R., McPherson, C. A., and Harry, G. J. (2016). Microglial m1/m2 polarization and metabolic states. *Br. J. Pharmacol.* 173, 649–665. doi: 10.1111/bph.13139
- Rahimian, R., Wakid, M., O’Leary, L. A., and Mechawar, N. (2021). The emerging tale of microglia in psychiatric disorders. *Neurosci. Biobehav. Rev.* 131, 1–29. doi: 10.1016/j.neubiorev.2021.09.023
- Reemst, K., Kracht, L., Kotah, J. M., Rahimian, R., van Irsen, A. A., Sotomayor, G. C., et al. (2022). Early-life stress lastingly impacts microglial transcriptome and function under basal and immune-challenged conditions. *bioRxiv*. doi: 10.1101/2022.07.13.499949
- Rüschendorf, L. (1985). The wasserstein distance and approximation theorems. *Probabil. Theory Related Fields* 70, 117–129. doi: 10.1007/BF00532240
- Sagar, M. A. K., Cheng, K. P., Ouellette, J. N., Williams, J. C., Watters, J. J., and Eliceiri, K. W. (2020a). Machine learning methods for fluorescence lifetime imaging (flim) based label-free detection of microglia. *Front. Neurosci.* 14, 931. doi: 10.3389/fnins.2020.00931
- Sagar, M. A. K., Ouellette, J. N., Cheng, K. P., Williams, J. C., Watters, J. J., and Eliceiri, K. W. (2020b). Microglia activation visualization via fluorescence lifetime imaging microscopy of intrinsically fluorescent metabolic cofactors. *Neurophotonics* 7, 1–14. doi: 10.1117/1.NPH.7.3.035003
- Schwabenland, M., Brück, W., Priller, J., Stadelmann, C., Lassmann, H., and Prinz, M. (2021). Analyzing microglial phenotypes across neuropathologies: a practical guide. *Acta Neuropathol.* 142, 923–936. doi: 10.1007/s00401-021-02370-8
- Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D* 404, 132306. doi: 10.1016/j.physd.2019.132306
- Tambuyzer, B. R., Ponsaerts, P., and Nouwen, E. J. (2009). Microglia: gatekeepers of central nervous system immunology. *J. Leukoc. Biol.* 85, 352–370. doi: 10.1189/jlb.0608385
- Torres-Platas, S. G., Cruceanu, C., Chen, G. G., Turecki, G., and Mechawar, N. (2014). Evidence for increased microglial priming and macrophage recruitment in the dorsal anterior cingulate white matter of depressed suicides. *Brain Behav. Immun.* 42, 50–59. doi: 10.1016/j.bb.2014.05.007
- Verdonk, F., Roux, P., Flamant, P., Fiette, L., Bozza, F. A., Simard, S., et al. (2016). Phenotypic clustering: a novel method for microglial morphology analysis. *J. Neuroinflammation.* 13, 1–16. doi: 10.1186/s12974-016-0614-7
- Voloboueva, L. A., Emery, J. F., Sun, X., and Giffard, R. G. (2013). Inflammatory response of microglial bv-2 cells includes a glycolytic shift and is modulated by mitochondrial glucose-regulated protein 75/mortalin. *FEBS Lett.* 587, 756–762. doi: 10.1016/j.febslet.2013.01.067
- Wang, X., Feng, B., Bai, X., Liu, W., and Latecki, L. J. (2014). Bag of contour fragments for robust shape classification. *Pattern Recognit.* 47, 2116–2125. doi: 10.1016/j.patcog.2013.12.008
- Watters, J. J., Schartner, J. M., and Badie, B. (2005). Microglia function in brain tumors. *J. Neurosci. Res.* 81, 447–455. doi: 10.1002/jnr.20485
- Yan, L., Rueden, C. T., White, J. G., and Eliceiri, K. W. (2006). Applications of combined spectral lifetime microscopy for biology. *BioTechniques*, 41, 249–257. doi: 10.2144/000112251
- Yu, Y., Si, X., Hu, C., and Zhang, J. (2019). A review of recurrent neural networks: Lstm cells and network architectures. *Neural Comput.* 31, 1235–1270. doi: 10.1162/neco_a_01199
- Zanier, E. R., Fumagalli, S., Perego, C., Pischiutta, F., and De Simoni, M.-G. (2015). Shape descriptors of the “never resting” microglia in three different acute brain injury models in mice. *Intensive Care Med. Exp.* 3, 1–18. doi: 10.1186/s40635-015-0039-0