

## Captions for figures in the supporting material.

**Fig. S1:** a,b) Single cells within plane c) cell through plane d,e) two cells f) 3 connected cells g, h, i) showing cells bounded within blood vessels, with some at bifurcations; scale bars are 25  $\mu\text{m}$ . Cropped from images with width of 300  $\mu\text{m}$  (no MPIO labeled MSCs were cropped from view).

**Fig. S2:** Similar to the concept of eigen faces, Principle Component Analysis (PCA) was utilized to extract eigen spot shapes using all of the  $9 \times 9$  spot patches in the training set. The top PCA component for the spot patches obtained on three labeled rats in  $G_A$  are shown here. An iteratively increasing threshold is then applied on the values of these top PCA components to

extract different binary patches that are utilized as filters to capture the shape and intensity information on spot patches.

**Fig. S3:** Binary shape filters are obtained using the top PCA components (shown in Fig. S2.). By iteratively increasing the threshold values from dark to light intensities, PCA components can result in different binary shapes. Domain experts agree that these binary filter patches represent many frequent shapes of the actual spots. All these patches are rotated and translated to obtain a large set of different shape filters. These filters are convolved with each candidate patch and the computed response is taken as a feature. A large set of these responses comprehensively captures the shape and intensity information of spot patches.

**Fig. S4:** Visual representation of context features for two context patches: While learning the definition of a spot, it may also be useful for a classifier to learn more about its surrounding context. Therefore, to capture the appearance of the patch’s context, a larger patch of  $21 \times 21$  was extracted around the center of a candidate patch  $x_i$ . Two well-known appearance descriptors in computer vision were then used to extract features: (a) Histogram of Gradients (HoG) and (b) Gist. A visual representation of HoG for two context patches is shown here. Red lines here indicate the different directions of intensity gradients whereas the lengths of the lines determine their magnitude.

**Fig. S5:** CNN architecture used in this work. The network takes a  $9 \times 9$  image patch as input for a classification task  $i$ . Each composite layer  $L^k$ ,

$k \in \{1, 2, 3\}$ , is composed of a convolutional layer  $C^{ik}$  which produces the feature maps  $F^{ik}$  and a non-linear gating function  $\beta$  producing the transformed feature maps  $F_{\beta}^{ik}$ . After passing through the composite layers, the net passes through the fully connected layer  $L^{fc}$  which produces the output. The *softmax* function is then applied to the output. Note that in the context of this work, this architecture represent the model  $M$ . The weights of all the filters across its processing layers are learned using the training data.

**Fig. S6:** Some convolution filters learned by the  $2^{nd}$  composite layers of our deep learning approach. Usually, each filter acts as a neuron and due to co-adaptation between a large number of such neurons, highly sophisticated features are extracted that can potentially model spot shape, intensity, texture etc.

**Fig. S7:** Transformed, average images for different source datasets.

# Intelligent and automatic *in vivo* detection and quantification of transplanted cells in MRI (Supporting Material)

Muhammad Jamal Afridi<sup>1</sup>, Arun Ross<sup>1</sup>, Xiaoming Liu<sup>1</sup>, Margaret F. Bennewitz<sup>2</sup>,  
Dorela D. Shuboni<sup>3</sup>, Erik M. Shapiro<sup>3</sup>

1. Department of Computer Science and Engineering, Michigan State University, USA
2. Vascular Medicine Institute, University of Pittsburgh, USA
3. Department of Radiology, Michigan State University, USA

Correspondence to:

Erik M. Shapiro

Department of Radiology, Michigan State University

Radiology Building, Michigan State University

846 Service Road, East Lansing, MI 48824

*Erik.Shapiro@radiology.msu.edu*

*Submitted to Magnetic Resonance in Medicine*

## Supporting Material - Figure S1

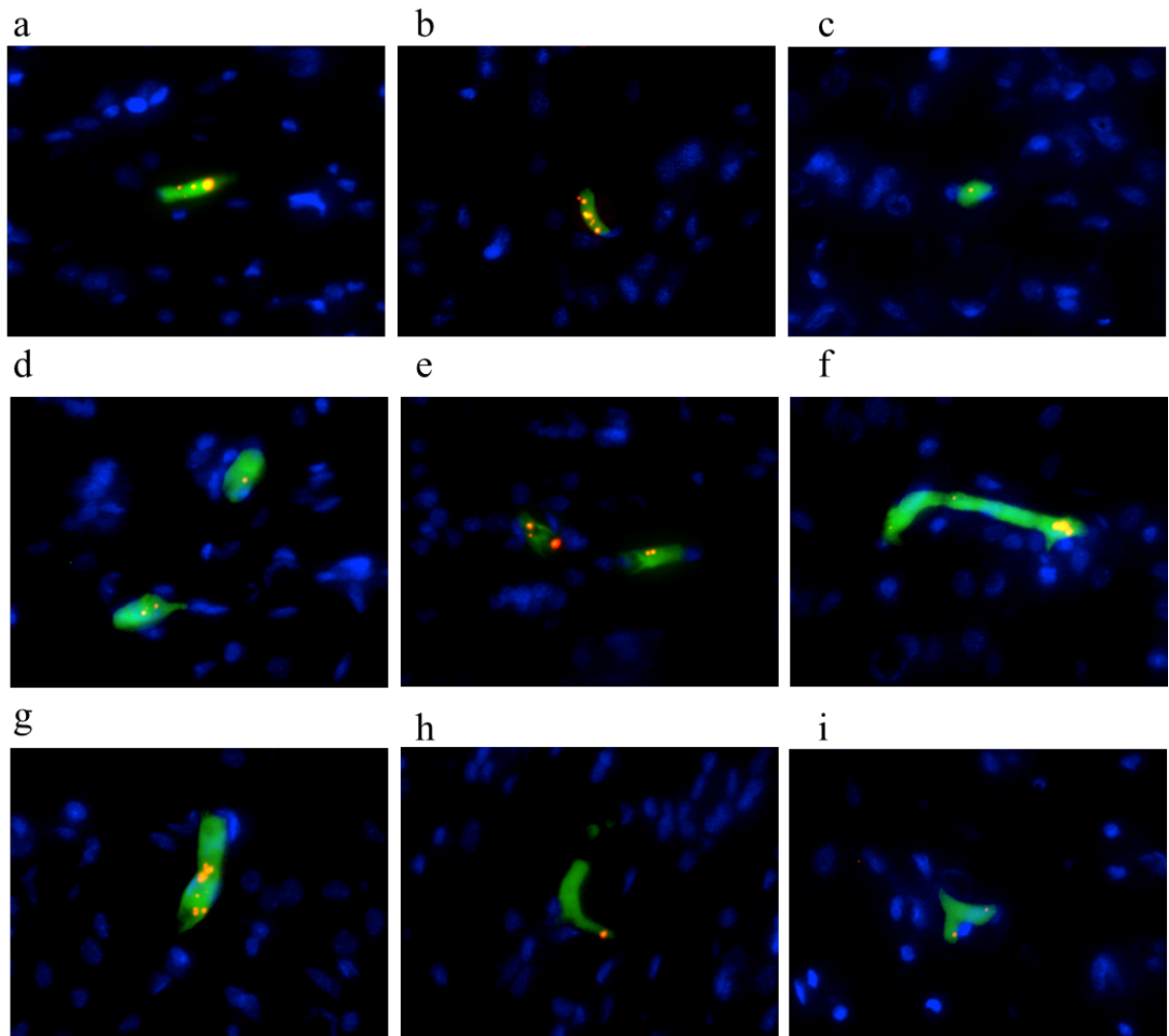


Figure 1: a,b) Single cells within plane c) cell through plane d,e) two cells f) 3 connected cells g, h, i) showing cells bounded within blood vessels, with some at bifurcations; scale bars are 25 μm. Cropped from images with width of 300 μm (no MPIO labeled MSCs were cropped from view).

## Supporting Material - Figure S2

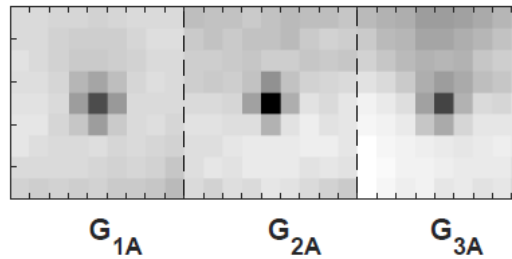


Figure 2: Similar to the concept of eigen faces, Principle Component Analysis (PCA) was utilized to extract eigen spot shapes using all of the  $9 \times 9$  spot patches in the training set. The top PCA component for the spot patches obtained on three labeled rats in  $G_A$  are shown here. An iteratively increasing threshold is then applied on the values of these top PCA components to extract different binary patches that are utilized as filters to capture the shape and intensity information on spot patches.

## Supporting Material - Figure S3

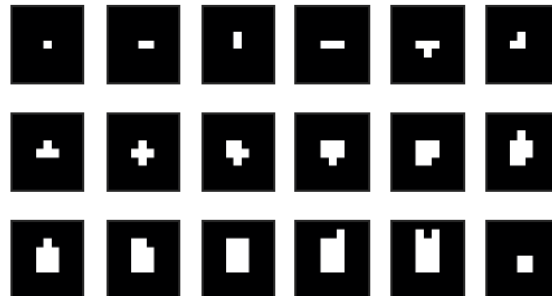


Figure 3: Binary shape filters are obtained using the top PCA components (shown in Fig. S2.). By iteratively increasing the threshold values from dark to light intensities, PCA components can result in different binary shapes. Domain experts agree that these binary filter patches represent many frequent shapes of the actual spots. All these patches are rotated and translated to obtain a large set of different shape filters. These filters are convolved with each candidate patch and the computed response is taken as a feature. A large set of these responses comprehensively captures the shape and intensity information of spot patches.

## Supporting Material - Figure S4

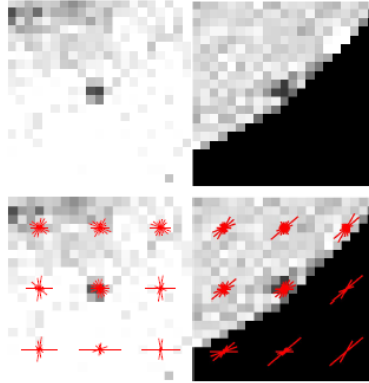


Figure 4: Visual representation of context features for two context patches: While learning the definition of a spot, it may also be useful for a classifier to learn more about its surrounding context. Therefore, to capture the appearance of the patch's context, a larger patch of  $21 \times 21$  was extracted around the center of a candidate patch  $x_i$ . Two well-known appearance descriptors in computer vision were then used to extract features: (a) Histogram of Gradients (HoG) and (b) Gist. A visual representation of HoG for two context patches is shown here. Red lines here indicate the different directions of intensity gradients whereas the lengths of the lines determine their magnitude.

## Supporting Material - Figure S5

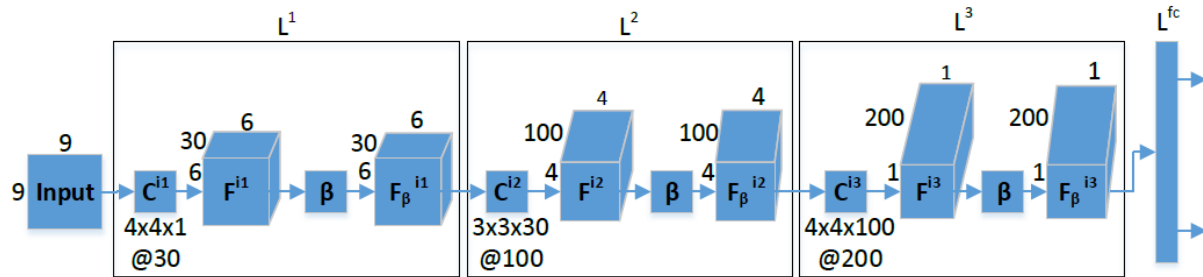


Figure 5: CNN architecture used in this work. The network takes a  $9 \times 9$  image patch as input for a classification task  $i$ . Each composite layer  $L^k$ ,  $k \in \{1, 2, 3\}$ , is composed of a convolutional layer  $C^{ik}$  which produces the feature maps  $F^{ik}$  and a non-linear gating function  $\beta$  producing the transformed feature maps  $F_{\beta}^{ik}$ . After passing through the composite layers, the net passes through the fully connected layer  $L^{fc}$  which produces the output. The *softmax* function is then applied to the output. Note that in the context of this work, this architecture represent the model  $M$ . The weights of all the filters across its processing layers are learned using the training data.

## Supporting Material - Figure S6

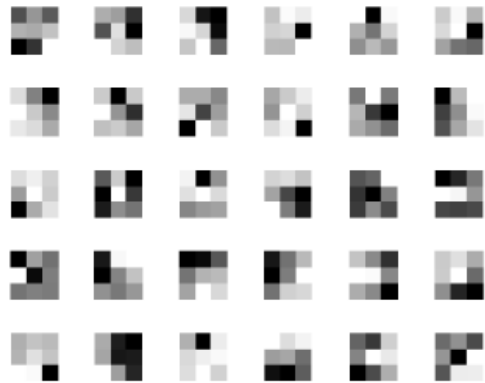


Figure 6: Some convolution filters learned by the 2nd composite layers of our deep learning approach. Usually, each filter acts as a neuron and due to co-adaptation between a large number of such neurons, highly sophisticated features are extracted that can potentially model spot shape, intensity, texture etc.

## Supporting Material - Figure S7

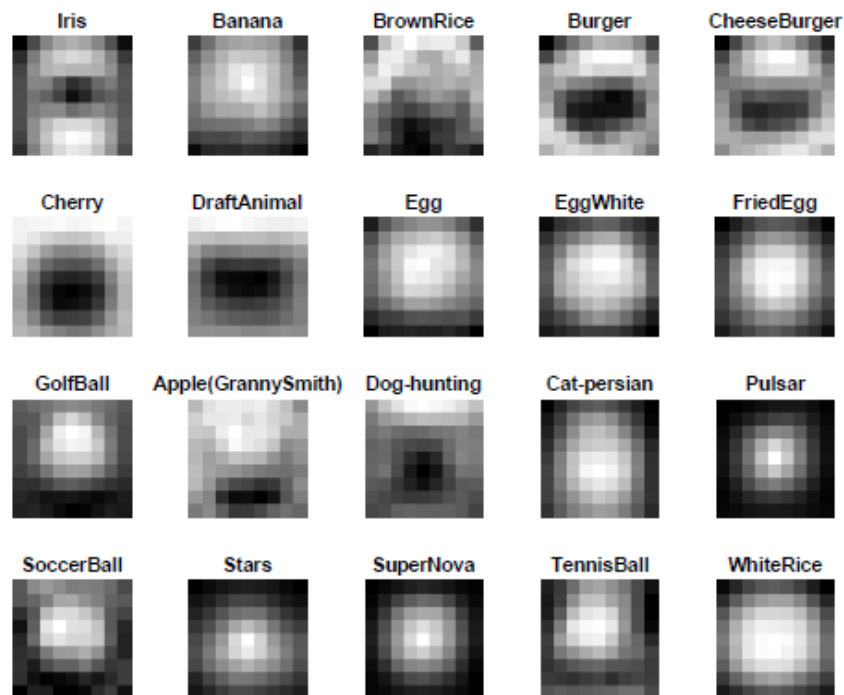


Figure 7: Transformed, average images for different source datasets.



# Supporting Material - Methods

## Cell preparation

A model of single, dispersed cells in the rat brain was created by intracardiac injection of magnetically labeled MSCs. Animal experiments were approved by the Yale University and Michigan State University Animal Care and Use Committees. MSCs were maintained in low glucose DMEM without L-glutamine, supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine, and 1% penicillin/streptomycin. The cells were grown until 80% confluent in T25 culture flasks. Prior to injection, MSCs were labeled with fluorescent 1.63 micron MPIO (Bangs Laboratory, Flash Red) by overnight incubation at a concentration of 40 beads per cell ( $\sim 44$  pg Fe/cell), followed by washing in PBS to remove uninternalized particles. This labeling scheme has previously been validated to label cells to a level of  $\sim 10$ -15 pg Fe/cell (37). Control cells were labeled with media only. Unlabeled and MPIO labeled MSCs were also stained with carboxyfluorescein succinimidyl ester (CFSE), a fluorescent dye. In brief, cells were incubated for 5 min in a concentration of 0.05mM CFSE and then washed twice with PBS containing 5% HI FBS after centrifugation at 300g for 5 min. Prior to injection, the cells were then diluted in 0.9% saline into two concentrations ( $n=3$ /group), which would deliver 200,000 cells (Set  $G_A$ ) or 250,000 cells (Set  $G_B$ ) per 200uL injection.

## Animal preparation

Adult Fischer rats were anesthetized with 3% isoflurane in 90% oxygen/10% medical air and orally intubated. A mechanical ventilating unit provided regulated breathing at 65 breaths/minute. Respiratory patterns and end tidal  $\text{CO}_2$  were monitored. After being placed in the supine position, rats were shaved along the left side of the chest, from the sternum downwards and prepped using aseptic technique. No skin incisions were made; rather, cells were injected noninvasively. 200,000 or 250,000 MPIO labeled MSCs or unlabeled MSCs (200  $\mu\text{L}$ ) were injected slowly into the left ventricle of the rodent heart with a 1 ml syringe attached to a 30-gauge needle. The needle was held at a 45-degree angle, placed ventrally, just left of the sternum, and lowered into the intercostal space between the 6<sup>th</sup> and 7<sup>th</sup> ribs. Noticeable blood flow into the syringe was used to gauge placement of the needle into the left ventricle. Animals remained under anesthesia, with ventilated breathing for 1–1.5 hours before being imaged on MRI

## *In vivo* MRI

MRI of live rats was performed on a 7.0T Bruker Biospin or 11.7T Varian system, 1–1.5 hr after intracardiac injection of unlabeled or MPIO labeled MSCs. Separate transmit-only volume and receive-only surface coils were used. During MRI, rats remained mechanically ventilated at 65 breaths/min, under 3% isoflurane anesthesia in 90% oxygen/10% medical air. End tidal  $\text{CO}_2$  and respiratory patterns were monitored.  $T_2^*$  weighted 3D gradient echo MRI was performed at 100 micron resolution using the following imaging parameters: TR = 30 ms and TEs = 10 (7.0T) or 8 (11.7T) ms.

## Histology

Immediately following MRI, rats were transcardially perfused with warm 0.9% saline, followed by ice cold 4% paraformaldehyde solutions of pH 6.5 and pH 9.5. Brains were excised and post-fixed in 4% paraformaldehyde with 10% sucrose at 4°C. After sinking, brains were rinsed in PBS, patted dry, placed in a plastic cryomold, and immersed in TissueTek embedding compound. After a few minutes of equilibration, brains were quickly frozen on dry ice and stored at -80°C. A cryostat was used to cut 16 $\mu$ m sections of the brains. Slides of interest were hydrated in PBS for 10 minutes, and mounted with ProLong Gold antifade with DAPI. Fluorescent microscopic images were acquired with the GFP filter for MSCs, Texas Red filter for MPIOs, and the UV filter for DAPI. Images of cells were taken with an approximate width of 300 $\mu$ m across. To determine the percentage of images containing 1, 2, or 3 cells, the number of MPIO labeled cells in each image was counted for a total of 70 images. Each image was cropped to a FOV of 100  $\mu$ m across for a closer view of the cells to display in Figure S1; no MPIO labeled MSCs were removed from view by cropping.

# Supporting Material - Source CNN Selection

This section explains how the terms  $J_i$  and  $U_i$ , utilized in the CNN selection criterion, are computed.

As mentioned in the paper, the proposed CNN architecture can be seen as a sequence of functions and therefor for a specific task  $i$ , it can written as:

$$M_i = (f_{ui} \circ f_{(u-1)i} \circ f_{(u-2)i} \circ \dots \circ f_{1i}). \quad [1]$$

where,  $f_{(u-1)i}$  denotes the fully connected layer whose output is subjected to a softmax function in  $f_{ui}$ .  $f_{(u-2)i}$  represents a non-linear activation function whereas  $f_{(u-3)i}$  denotes the deepest convolutional layer.

The CNN learned on small MRI data  $M$  already captures the information available in the target's training data  $X$ . Therefore the goal of transferring weights (also called learned feature representations) from a source CNN is to capture that information which is *not* already accounted for by  $M$  and may be relevant in distinguishing a spot patch from non-spot patches. Intuitively, a source ranking criterion can be based on :

1. how *different* is the learned information (CNN weights) of a source CNN  $M_i$  from the target CNN  $M$ ? If the source CNN  $M_i$  provides the exact same information as  $M$  then a transfer from such a source may not result in any performance improvement. Hence difference between the learned weights of the two CNNs is a desirable property. Previous research shows that the feature representations (weights) learned in the initial convolutional layers are more generic and may be the same irrespective of the learning task. However, the deeper convolutional layer are more task specific. Therefore, the difference between two CNNs, i.e.,  $M$  and  $M_i$  is denoted by  $J_i$  which is given by

$$J_i = ||W_{(u-3)x} - W_{(u-3)i}||_2^2. \quad [2]$$

where  $W_{(u-3)i}$  denotes a vector containing all the weights in the deepest convolutional layer  $f_{(u-3)i}$  of the CNN  $M_i$ . Similarly,  $W_{(u-3)x}$  denotes the deepest convolutional layer weights of the CNN  $M$ .

2. how discriminating a source CNN  $M_i$  is for the target task? This determines how relevant a source CNN is to the target task despite its difference with  $M$ . This can be measured by computing  $U_i$  for each source CNN  $M_i$ .

$$U_i = ||Y - M_i(X)||_2^2. \quad [3]$$

where  $Y$  denotes a vector of labels on training data  $X$  and  $M_i(X)$  represents the vector of predicted outputs resulting when the source CNN  $M_i$  operates on the target's limited training data  $X$  as a *test* set.