



# Semi-Supervised Learning in Medical Images Through Graph-Embedded Random Forest

Lin Gu<sup>1,2</sup>, Xiaowei Zhang<sup>3</sup>, Shaodi You<sup>4</sup>, Shen Zhao<sup>5</sup>, Zhenzhong Liu<sup>6,7\*</sup> and Tatsuya Harada<sup>1,2</sup>

<sup>1</sup>RIKEN AIP, Tokyo, Japan, <sup>2</sup>Research Center for Advanced Science and Technology (RCAST), The University of Tokyo, Tokyo, Japan, <sup>3</sup>Bioinformatics Institute (BI), A\*STAR, Singapore, Singapore, <sup>4</sup>Faculty of Science, Institute of Informatics, University of Amsterdam, Amsterdam, Netherlands, <sup>5</sup>Department of Medical Physics, Western University, London, ON, Canada, <sup>6</sup>Tianjin Key Laboratory for Advanced Mechatronic System Design and Intelligent Control, School of Mechanical Engineering, Tianjin University of Technology, Tianjin, China, <sup>7</sup>National Demonstration Center for Experimental Mechanical and Electrical Engineering Education, Tianjin University of Technology, Tianjin, China

One major challenge in medical imaging analysis is the lack of label and annotation which usually requires medical knowledge and training. This issue is particularly serious in the brain image analysis such as the analysis of retinal vasculature, which directly reflects the vascular condition of Central Nervous System (CNS). In this paper, we present a novel semi-supervised learning algorithm to boost the performance of random forest under limited labeled data by exploiting the local structure of unlabeled data. We identify the key bottleneck of random forest to be the information gain calculation and replace it with a graph-embedded entropy which is more reliable for insufficient labeled data scenario. By properly modifying the training process of standard random forest, our algorithm significantly improves the performance while preserving the virtue of random forest such as low computational burden and robustness over over-fitting. Our method has shown a superior performance on both medical imaging analysis and machine learning benchmarks.

**Keywords:** vessel segmentation, semi-supervised learning, manifold learning, central nervous system (CNS), retinal image

## OPEN ACCESS

### Edited by:

Heye Zhang,  
Sun Yat-sen University, China

### Reviewed by:

Chenxi Huang,  
Xiamen University, China  
Guang Yang,  
Imperial College London,  
United Kingdom

### \*Correspondence:

Zhenzhong Liu  
zliu@email.tjut.edu.cn

**Received:** 01 September 2020

**Accepted:** 23 September 2020

**Published:** 10 November 2020

### Citation:

Gu L, Zhang X, You S, Zhao S, Liu Z and Harada T (2020) Semi-Supervised Learning in Medical Images Through Graph-Embedded Random Forest. *Front. Neuroinform.* 14:601829. doi: 10.3389/fninf.2020.601829

## 1. INTRODUCTION

Machine learning has been widely applied to analyze medical images such as an image of the brain. For example, the automatic segmentation of brain tumor (Soltaninejad et al., 2018) could help predict Patient Survival from MRI data. However, traditional methods usually require a large number of diagnosed examples. Collecting raw data during routine screening is possible but making annotations and diagnoses for them is costly and time-consuming for medical experts. To deal with this challenge, we propose a novel graph-embedded semi-supervised algorithm that makes use of the unlabeled data to boost the performance of the random forest. We specifically evaluate the proposed method on both a neuronal image and the retinal image analysis that is highly related to diabetic retinopathy (DR) (Niu et al., 2019) and Alzheimer's Disease (AD) (Liao et al., 2018), and make the following specific contributions:

1. We empirically validate that the performance bottleneck of random forest under limited training samples is the biased information gain calculation.
2. We propose a new semi-supervised entropy calculation by incorporating local structure of unlabeled data.
3. We propose a novel semi-supervised random forest which shows advantage performance of the state-of-the-art in both medical imaging analysis and machine learning benchmarks.

Among various supervised algorithms, random forest or random decision trees (Breiman et al., 1984; Criminisi et al., 2012) are one of the state-of-the-art machine learning algorithms for medical imaging applications. Despite its robustness and efficiency, its performance relies heavily on sufficiently labeled training data. However, annotating a large amount of medical data is time-consuming and requires domain knowledge. To alleviate the challenge of having enough labeled data, a class of learning methods named semi-supervised learning (SSL) (Joachims, 1999; Zhu et al., 2003; Belkin and Niyogi, 2004; Zhou et al., 2004; Chapelle et al., 2006; Zhu, 2006) were proposed to leverage unlabeled data to improve the performance. Leistner et al. (2009) proposed a semi-supervised random forest which maximizes the data margin via deterministic annealing (DA). Liu et al. (2015) showed that the splitting strategy appears to be the bottleneck of performance in a random forest. The authors estimate the unlabeled data through kernel density estimation (KDE) on the projected subspace, and when constructing the internal node, they progressively refine the splitting function with the acquired labels through KDE until it converges. Without explicit affinity relation, CoForest (Li and Zhou, 2007) iteratively guesses the unlabeled data with the rest of the trees in the forest and then uses the new labeled data to refine the tree. Semi-supervised based super-pixel (Gu et al., 2017) has proved to be effective in the segmentation of both a retinal image and a neuronal image.

Following the research line of a previous semi-supervised random forest (RF), we identify that RF's performance bottleneck, under insufficient data, is the biased information gain calculation when selecting an optimal splitting parameter (shown as blue in **Figure 1**). Therefore, as illustrated in red in **Figure 1**, we slightly modified the training procedure of RF to relieve this bias. We replace the original information gain with our novel graph-embedded entropy which exploits the data structure of unlabeled data. Specifically, we first use both labeled and unlabeled data to construct a graph whose weights measure local similarity among data and then minimize a loss function that sums the supervised loss over labeled data and a graph Laplacian regularization term. From the optimal solution, we can get label information of unlabeled data which is utilized to estimate a more accurate information gain for node splitting. Since a major part of training and the whole testing remains unchanged, our graph-embedded random forest could significantly improve the performance without losing the virtue of a standard random forest such as low computational burden and robustness over over-fitting.

## 2. ANALYSIS OF PERFORMANCE BOTTLENECK

Let us first review the construction of the random forest (Breiman et al., 1984) to figure out why random forest fails under limited training data. A random forest is an ensemble of decision trees:  $\{t_1, t_2, \dots, t_T\}$ , of which an individual tree is independently trained and tested.

**Training Procedure:** Each decision tree  $t$ , as illustrated in **Figure 1**, learns to classify a training sample  $x \in X$  to the corresponding label  $y$  by recursively branching it to the left or right child until reaching a leaf node. In particular, each node is associated with a binary split function  $h(x_i, w, \tau)$ , e.g., oblique linear split function

$$h(x_i, w, \tau) = [\langle w, x_i \rangle < \tau], \quad (1)$$

where  $[\cdot]$  is an indicative function and  $\tau$  is a scalar threshold.  $w \in R^d$  serves as a feature weight parameter that projects the high dimension data  $x \in R^d$  to a one dimensional subspace.

Given a candidate splitting function  $h(x, w_j, \tau_j)$ , its splitting quality is measured by information gain  $G(w_j, \tau_j)$ . In practice, given the training data  $X$  and their labels  $Y$ , the construction of the splitting node, as illustrated in the left side of **Figure 1**, comprises the following three stages:

**Algorithm 1:** Training of node splitting.

- 1: Randomly generates a set of feature subspace candidates  $\{w_j\}$
- 2: For each  $w_j$ , find the optimal  $\tau_j^* = \arg \max_{\tau} G(w_j, \tau, X, Y)$  that best splits the data.
- 3: Among all  $\{w_j, \tau_j^*\}$ , pick the one with largest information gain:  $j^* = \arg \max_j G(w_j, \tau_j^*, X, Y)$

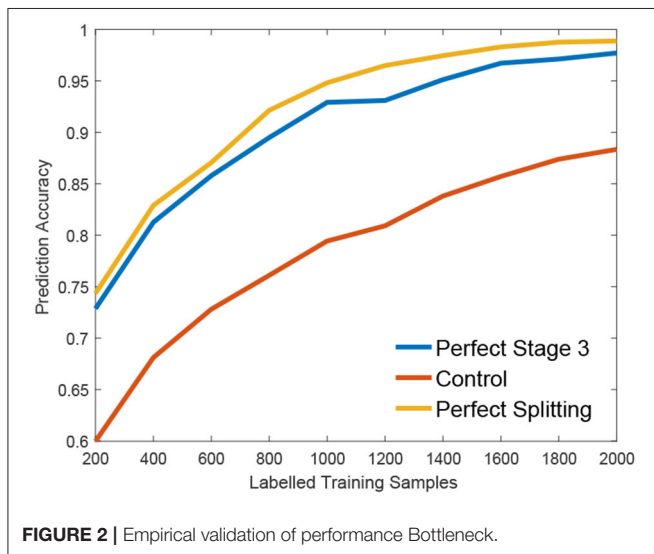
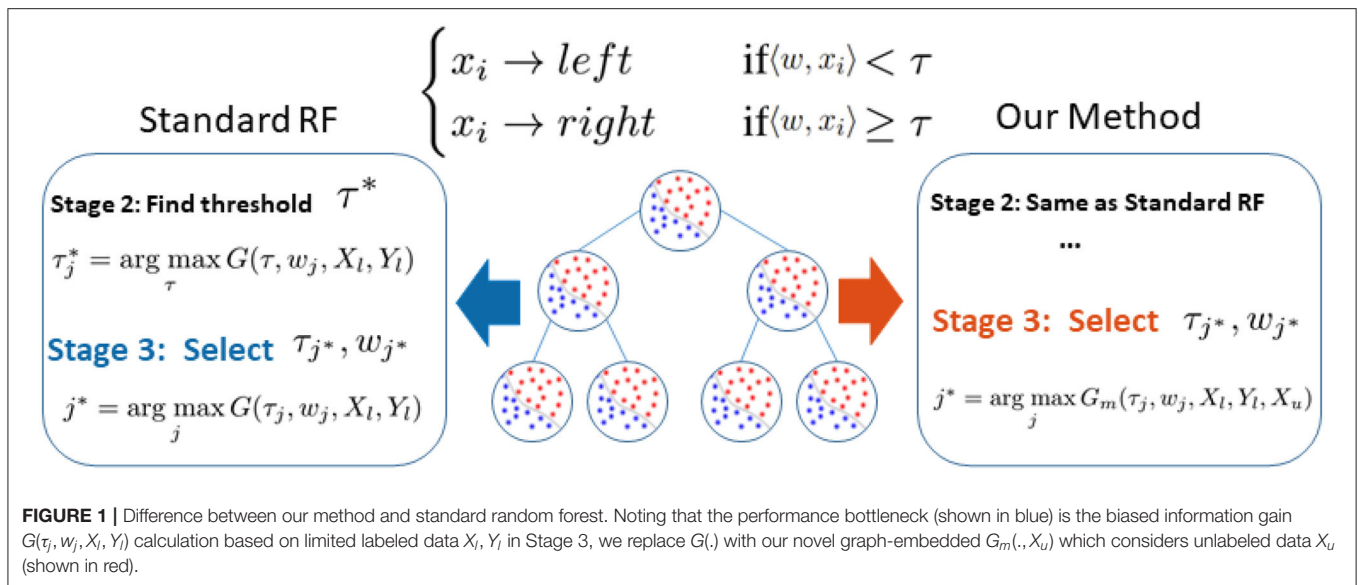
Through the above stages, each split node is associated with a splitting function  $h(x, w, \tau)$  that best splits the training data.

**Testing Procedure:** When testing data  $x$ , the trained random forest predicts the probability of its label by averaging the ensemble prediction as  $\hat{p}(y|x) = \sum_t p_t(y|x)$ , where  $p_t(y|x)$  denotes the empirical label distribution of the training samples that reach leaf node of tree  $t$ .

### 2.1. Performance Bottleneck Under Insufficient Data

According to the study of Liu et al. (2015), insufficient training data would impact the performance of RF in three ways (Liu et al., 2015): (1) limited forest depth; (2) inaccurate prediction model of leaf nodes; (3) sub-optimal splitting strategy. Among them, Liu et al. (2015) identified that (1) is inevitable, and (2) is solvable with their proposed strategy. In this paper, we further improve the method by tackling (3).

We claim that the performance bottleneck of random forest is its sub-optimal splitting strategy in Algorithm 1. To empirically support this claim, we build three random forests, similar to Liu et al. (2015), for comparison: the first one, the **Control** is trained with a small size of a training set S1 as control; the second one, the



**Perfect Stage 3** is constructed with the same training set  $S_1$  but its node splitting uses a large training set  $S_2$  to select the optimal parameter in stage 3 of Algorithm 1, to simulate the case that random forest selects the optimal parameter of stage 3 with full information; the third one, the **Perfect Splitting** is constructed with  $S_1$  while  $S_2$  was used for both Stage 2 and 3 of Algorithm 1.

Following the protocol of Liu et al. (2015), each random forest comprises 100 trees and the same entropy gain is adopted as the splitting criterion. We evaluate three random forests on Madelon (Guyon et al., 2004), a widely used machine learning benchmark. As shown in **Figure 2**, Perfect Stage 3, which only uses the full information to select the best parameter set, significantly improves the performance compared to the control group. Interestingly, the Perfect Splitting one, which utilizes the full information for both optimal parameter proposing (Stage 2)

and optimal parameter decision (Stage 3), only makes a subtle improvement compared to Perfect Stage 3.

From **Figure 2**, we found that Stage 3, optimal parameter selection, is the performance bottleneck of the splitting node construction, which is also the keystone of random forest construction (Liu et al., 2015). When deciding the optimal parameter, random forest often fails to find the best one as its information gain calculation  $g(w, \tau)$  is biased under insufficient training data. Interestingly, insufficient data has a smaller effect on the Stage 2, parameter proposal. Motivated by this observation, we propose a new information calculation which exploits unlabeled data to make a better parameter selection in Stage 3 of Algorithm 1.

### 3. GRAPH-EMBEDDED REPRESENTATION OF INFORMATION GAIN

In the previous section, we show that gain estimation appears to be the performance bottleneck of random forests. Empirically, we show that more label information helps to obtain more accurate gain estimation. This encourages us to consider the possibility of mining label information from unlabeled data through structural connections between labeled and unlabeled data. In particular, we perform a graph-based semi-supervised learning to get label information of unlabeled data, and compute information gain from both labeled and unlabeled data. To achieve a better gain estimation, we embed all data into a graph. Moreover, we assume the underlying structure of all data form a manifold, and compute data similarity based on the assumption.

Let  $l$  and  $u$  be the number of labeled and unlabeled instances, respectively. Let  $X_l = [x_1, \dots, x_l]^T \in \mathbb{R}^{d \times l}$  be the matrix of feature vectors of labeled instances, and  $X_u = [x_{l+1}, \dots, x_{l+u}]^T \in \mathbb{R}^{d \times u}$  be the matrix of unlabeled instances. To accommodate label information, we define a label matrix  $Y \in \mathbb{R}^{(l+u) \times K}$  (assuming there are  $K$  class labels available), with

each entry  $Y_{ik}$  containing 1 provided the  $i$ -th data belongs to  $X_l$  and is labeled with class  $k$ , and 0 otherwise. Besides, we define  $Y_l$  as a submatrix of  $Y$  corresponding to the labeled data,  $y_i \in \mathbb{R}^K$  as the  $i$ -th row of  $Y$  corresponding to  $x_i$ , and  $y_l \in \mathbb{R}^l$  as the vector of class labels for  $X_l$ .

Based on both labeled and unlabeled instances, our purpose is to learn a mapping  $f: \mathbb{R}^d \rightarrow \mathbb{R}^K$  and predict the label of instance  $x$  as  $k^* := \arg \max_k f_k(x)$ . Many semi-supervised learning algorithms use the following regularized framework

$$\sum_{i=1}^l \text{loss}(y_i, f(x_i)) + \lambda \sum_{i,j=l+1, i \neq j}^{l+u} s(x_i, x_j) \|f(x_i) - f(x_j)\|_2^2,$$

where  $\text{loss}()$  is a loss function and  $s(x_i, x_j)$  is a similarity function. In this paper, we apply the idea of graph embedding to learn  $f$ . We construct a graph  $\mathcal{G} = (V, E, W)$ , where each node in  $V$  denotes a training instance and  $W \in \mathbb{R}^{(l+u) \times (l+u)}$  denotes a symmetric weight matrix.  $W$  is computed as follows: for each point find  $t$  nearest neighbors, and  $W_{ij} = \exp(-\|x_i - x_j\|_2^2 / \sigma^2)$  if  $(x_i, x_j)$  are neighbors, 0 otherwise. Such construction of graph implicitly assumes that all data resides on some manifold and exploits local structure. Based on the graph embedding, we propose to minimize

$$\mathcal{L}(\{f_i\}) = \frac{1}{2} \left( \sum_{i=1}^{l+u} \|f_i - y_i\|_2^2 + \lambda \sum_{i,j=1}^{l+u} W_{ij} \left\| \frac{f_i}{\sqrt{D_{ii}}} - \frac{f_j}{\sqrt{D_{jj}}} \right\|_2^2 \right), \tag{2}$$

where  $D$  is a diagonal matrix with its  $D_{ii}$  equal to the sum of the  $i$ -th row of  $W$ . Let  $F^* = [f_1^*, \dots, f_{l+u}^*] = \arg \min_{\{f_i\}} \mathcal{L}(\{f_i\})$  be the optimal solution, it has been shown in Zhou et al. (2004) that

$$F^* = ((1 + \lambda)I - \lambda D^{-1/2} W D^{-1/2})^{-1} Y. \tag{3}$$

Based on the learned functions  $F^*$ , we can predict the label information of  $X_u$  and then utilize such information to estimate more accurate information gain. Specifically, we let  $\hat{y}_u$  denotes the predicted label of  $X_u$ , and for node  $S$  we compute Gini index  $G_m(S) = \sum_{k=1}^K p_k(1 - p_k)$ , where

$$p_k = \frac{1}{|S|} \left( \sum_{x_i \in S, 1 \leq i \leq l} \mathbb{1}_{\{(y_l)_i=k\}} + \sum_{x_i \in S, l+1 \leq i \leq l+u} \mathbb{1}_{\{(\hat{y}_u)_i=k\}} \right)$$

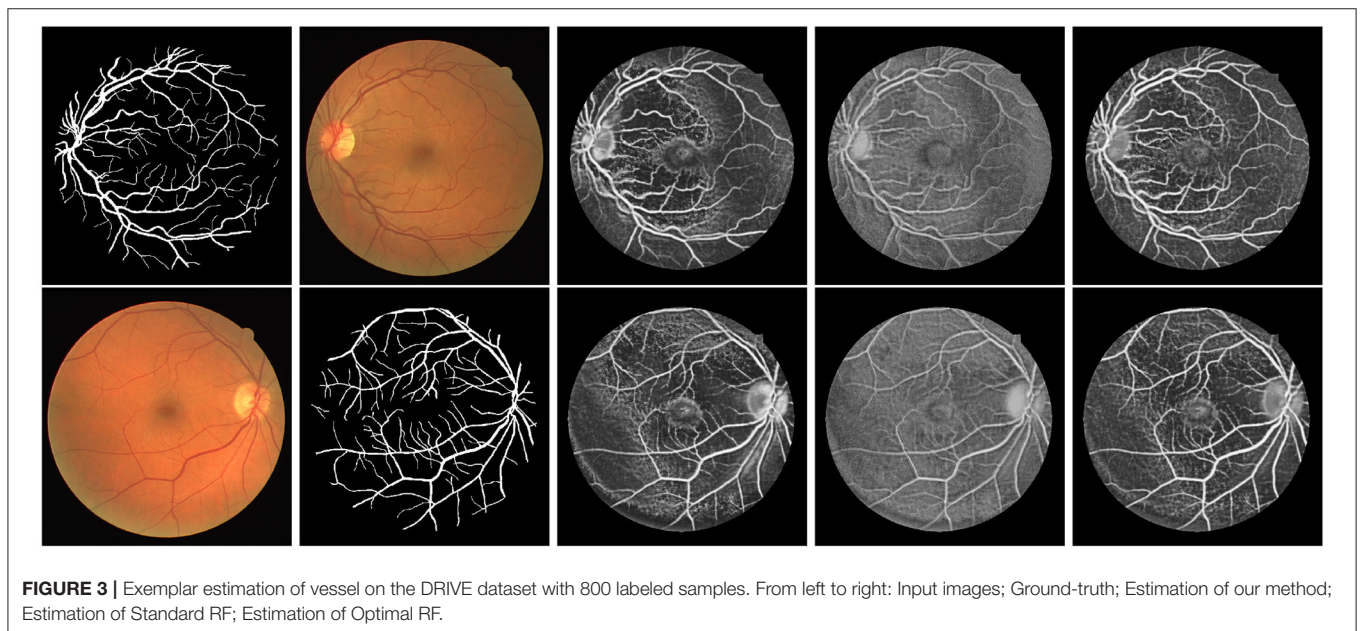
is the proportion of data from class  $k$ . Note that we utilize information from both labeled and unlabeled data to compute the Gini index. For each node, we estimate information gain as

$$G_m(w, \tau, X_l, Y_l, X_u) = G_m(S) - (|S_l|G_m(S_l) + |S_u|G_m(S_u)) / |S|, \tag{4}$$

where  $S_l$  and  $S_u$  are left and right child nodes, respectively.

### 4. CONSTRUCTION OF SEMI-SUPERVISED RANDOM FOREST

In our framework, we preserve the major structure of the standard random forest where the testing stage is exactly the same as the standard one. As illustrated in the right part of **Figure 1**, we only make a small modification in stage 3 of Algorithm 1 where the splitting efficiency is now evaluated by our novel graph-embedded based information gain  $G_m(\tau_j, w_j, X_l, Y_l, X_u)$  from Equation (4). Specifically, we leave stage 2 unchanged that the threshold  $\tau$  of each subspace candidate  $w$  is still based on standard information gain such as the Gini index. Now with a set of parameter candidates  $w, \tau$ , the stage 3 calculates the corresponding manifold based information score  $\hat{g}(w, \tau)$  instead and select the optimal one through  $\max_{w_j, \tau_j} \hat{g}(w_j, \tau_j)$ .



**FIGURE 3** | Exemplar estimation of vessel on the DRIVE dataset with 800 labeled samples. From left to right: Input images; Ground-truth; Estimation of our method; Estimation of Standard RF; Estimation of Optimal RF.

## 5. EXPERIMENTS

We evaluate our method on both 2D, and 3D brain related medical image segmentation tasks as well as two machine learning benchmarks.

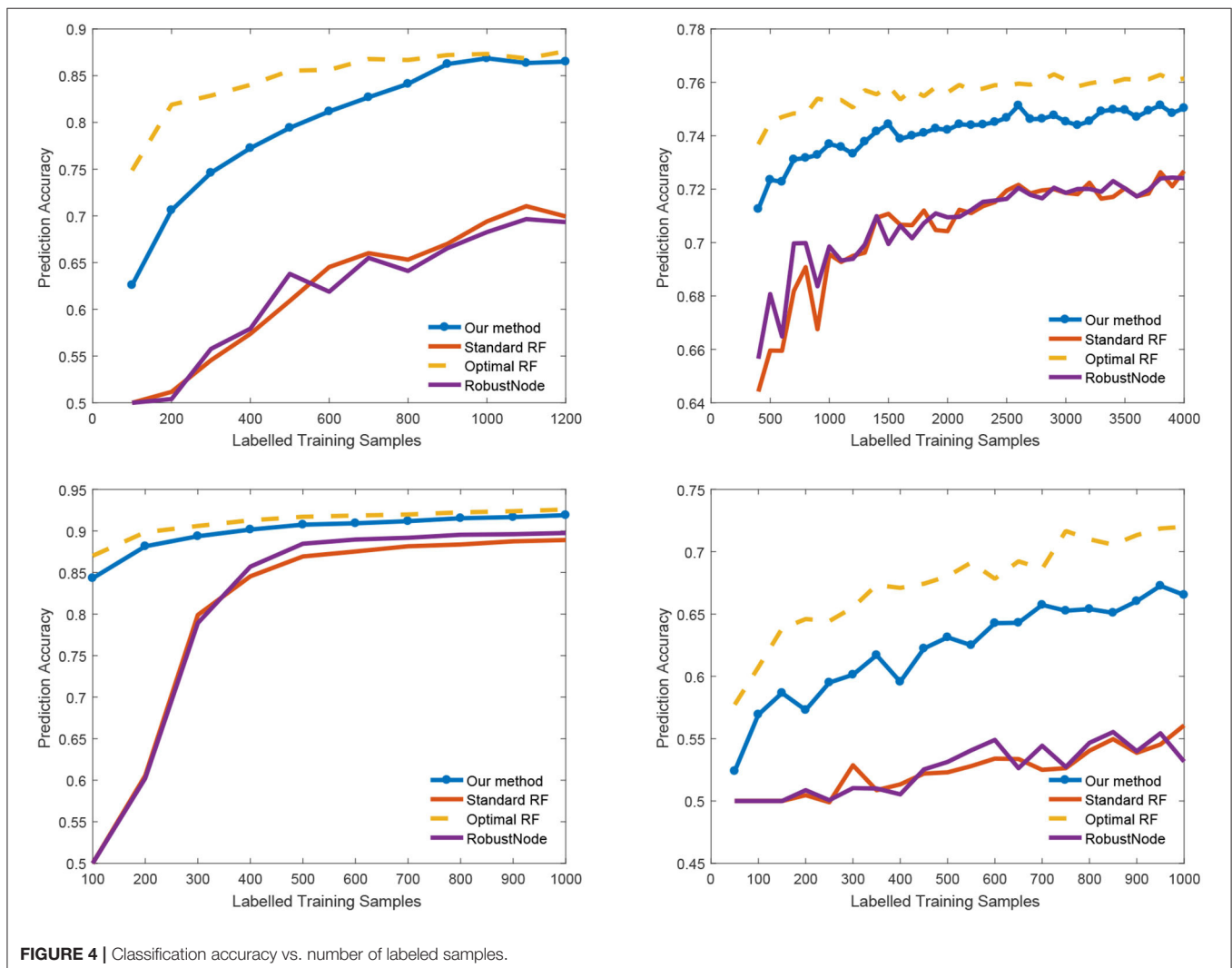
The retinal vessel, a part of the Central Nervous System (CNS), directly reflects the vascular condition of CNS. The accurate segmentation of vessels is important for this analysis. Much progress has been made based on either random forest (Gu et al., 2017) or deep learning (Liu et al., 2019). The DRIVE dataset (Staal et al., 2004) is a widely used 2D retinal vessel segmentation dataset that comprises of 20 training images and 20 testing ones. Each image is a  $768 \times 584$  color image along with manual segmentation. For the image, we extract two types of widely used features: 1, local patch  $x_1 \in R^{15 \times 15 \times 3}$  of target. 2,  $x_2 \in R^{4 \times 7 \times 3}$  Gabor wavelets (Soares et al., 2006). We also investigate the single neuron segmentation in a brain image. BigNeuron project<sup>1</sup> (Peng et al., 2015) is a 3D neuronal dataset

with ground truth annotation from experts. For BigNeuron data, we manually picked 13 images among which a random 10 were used for training while the rest were left for testing, because this dataset is designed for tracing rather than segmentation. For example, some annotation is visibly thinner than the actual neuron. Furthermore, the image may contain multiple neurons but only one is properly annotated. For both datasets, we randomly collected 40,000 (20,000 positive and 20,000 negative) samples from the training and testing sets, respectively. For 3D data, our feature is  $x_1 \in R^{15 \times 15 \times 7}$  local cube similar to the setting of Gu et al. (2017).

Apart from the medical imaging, we also demonstrate the generality of our method on two binary machine learning benchmark, IJCNN1 (Prokhorov, 2001) and Madelon (Guyon et al., 2004), in Libsvm Repository (Chang and Lin, 2011).

During the evaluation, we randomly selected a certain number  $n$  of labeled samples from the whole training set while leaving the rest unlabeled. Standard Random Forest (RF) is trained with  $n$  labeled training data only. Our method and RobustNode (Liu et al., 2015) are trained with both labeled data and unlabeled

<sup>1</sup><https://www.alleninstitute.org/bigneuron/about/>



**TABLE 1** | Classification accuracy (represented in percentage %) on different dataset.

	Drive	Big neuron	IJCNN1	Madelon
Our method	79.42	74.16	89.36	59.57
Standard RF	60.90	70.93	79.89	51.33
Robust node RF	63.79	70.99	78.92	50.53
Optimal RF	85.53	75.55	91.29	67.10

We show the accuracy on the training sample of 400 (DRIVE), 1,500 (Big Neuron), 300 (IJCNN1), and 400 (Madelon).

data. For reference, we also compared it with Optimal RF which is trained with labeled data as a standard RF. However, its node splitting is supervised with the whole training samples and their label. Optimal RF indicates the upper bound for all of semi-supervised learning algorithms.

### 5.1. Medical Imaging Segmentation

First, we illustrate the visual performance of segmentation in **Figure 3**. The estimated score is the possibility of the vessel given by the individual method. Our algorithm has consistently improved the estimation compared to the standard RF.

### 5.2. Quantitative Analysis

We also report the classification accuracy with respect to the number of labeled data in **Figure 4**, **Table 1**. We compared our method with alternatives on both medical imaging segmentation and machine learning benchmarks. **Figure 4** shows that our algorithm significantly outperformed alternative methods. Specifically, in the DRIVE dataset, our algorithm approaches the upper bound at 1,000 labeled samples. In the IJCNN1 dataset, our method quickly approaches the optimal one while the alternatives take 400 samples to approach.

## REFERENCES

- Belkin, M., and Niyogi, P. (2004). Semi-supervised learning on Riemannian manifolds. *Mach. Learn.* 56, 209–239. doi: 10.1023/B:MACH.0000033120.25363.1e
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification And Regression Trees*. CRC Press.
- Chang, C.-C., and Lin, C.-J. (2011). Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2:27. doi: 10.1145/1961189.1961199
- Chapelle, O., Scholkopf, B., and Zien, A. (2006). *Semi-Supervised Learning*. London: MIT Press. doi: 10.7551/mitpress/9780262033589.001.0001
- Criminisi, A., Shotton, J., and Konukoglu, E. (2012). Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Found. Trends. Comput. Graph. Vis.* 7, 81–227. doi: 10.1561/06000000035
- Gu, L., Zhang, X., Zhao, H., Li, H., and Cheng, L. (2017). Segment 2D and 3D filaments by learning structured and contextual features. *IEEE Trans. Med. Imaging* 36, 596–606. doi: 10.1109/TMI.2016.2623357
- Gu, L., Zheng, Y., Bise, R., Sato, I., Imanishi, N., and Aiso, S. (2017). “Semi-supervised learning for biomedical image segmentation via forest oriented super pixels (voxels),” in *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, eds M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne (Quebec City, QC: Springer International Publishing), 702–710. doi: 10.1007/978-3-319-66182-7\_80

## 6. CONCLUSION

In this paper, we propose a novel semi-supervised random forest to tackle the challenging problem of the lacking annotation in the analysis of medical imaging such as a brain image. Observing that the bottleneck of the standard random forest is the biased information gain estimation, we replaced it with a novel graph-embedded entropy which incorporates information from both labeled and unlabeled data. Empirical results show that our information gain is more reliable than the one used in traditional random forest under insufficient labeled data. By slightly modifying the training process of the standard random forest, our algorithm significantly improves the performance while preserving the virtue of the random forest. Our method has shown a superior performance with very limited data in both brain imaging analysis and machine learning benchmarks.

## DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This research was supported by JST, ACT-X Grant Number JPMJAX190D, Japan and the National Natural Science Foundation of China (Grant No. 61873188).

- Guyon, I., Gunn, S., Hur, A. B., and Dror, G. (2004). “Result analysis of the NIPS 2003 feature selection challenge,” in *Proceedings of the 17th International Conference on Neural Information Processing Systems, NIPS’04* (Cambridge, MA: MIT Press), 545–552.
- Joachims, T. (1999). “Transductive inference for text classification using support vector machines,” in *Proceedings of the Sixteenth International Conference on Machine Learning, ICML ’99* (San Francisco, CA: Morgan Kaufmann Publishers Inc.), 200–209.
- Leistner, C., Saffari, A., Santner, J., and Bischof, H. (2009). “Semi-supervised random forests,” in *2009 IEEE 12th International Conference on Computer Vision (Kyoto)*, 506–513. doi: 10.1109/ICCV.2009.5459198
- Li, M., and Zhou, Z. H. (2007). Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. *IEEE Trans. Syst. Man Cybern.* 37, 1088–1098. doi: 10.1109/TSMCA.2007.904745
- Liao, H., Zhu, Z., and Peng, Y. (2018). Potential utility of retinal imaging for Alzheimer’s disease: a review. *Front. Aging Neurosci.* 10:188. doi: 10.3389/fnagi.2018.00188
- Liu, B., Gu, L., and Lu, F. (2019). “Unsupervised ensemble strategy for retinal vessel segmentation,” in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2019*, eds D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P. T. Yap, and A. Khan (Shenzhen: Springer International Publishing), 111–119. doi: 10.1007/978-3-030-32239-7\_13

- Liu, X., Song, M., Tao, D., Liu, Z., Zhang, L., Chen, C., et al. (2015). Random forest construction with robust semisupervised node splitting. *IEEE Trans. Image Process.* 24, 471–483. doi: 10.1109/TIP.2014.2378017
- Niu, Y., Gu, L., Lu, F., Lv, F., Wang, Z., Sato, I., et al. (2019). “Pathological evidence exploration in deep retinal image diagnosis,” in *AAAI conference on artificial intelligence (AAAI)* (Honolulu: AAAI Press), 1093–1101. doi: 10.1609/aaai.v33i01.33011093
- Peng, H., Hawrylycz, M., Roskams, J., Hill, S., Spruston, N., Meijering, E., et al. (2015). Bigneuron: large-scale 3D neuron reconstruction from optical microscopy images. *Neuron* 87, 252–256. doi: 10.1016/j.neuron.2015.06.036
- Prokhorov, D. (2001). *IJCNN 2001 Neural Network Competition*. Washington, DC: IJCNN2001.
- Soares, J. V. B., Leandro, J. J. G., Cesar, R. M., Jelinek, H. F., and Cree, M. J. (2006). Retinal vessel segmentation using the 2-D gabor wavelet and supervised classification. *IEEE Trans. Med. Imaging* 25, 1214–1222. doi: 10.1109/TMI.2006.879967
- Soltaninejad, M., Zhang, L., Lambrou, T., Yang, G., Allinson, N., and Ye, X. (2018). “MRI brain tumor segmentation and patient survival prediction using random forests and fully convolutional networks,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, eds A. Crimi, S. Bakas, H. Kuijff, B. Menze, and M. Reyes (Cham: Springer International Publishing), 204–215. doi: 10.1007/978-3-319-75238-9\_18
- Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A., and van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* 23, 501–509. doi: 10.1109/TMI.2004.825627
- Zhou, D., Bousquet, O., Lal, T., Weston, J., and Scholkopf, B. (2004). “Learning with local and global consistency,” in *NIPS* (Vancouver, BC).
- Zhu, X. (2006). *Semi-Supervised Learning Literature Survey*. Technical report, University of Wisconsin-Madison.
- Zhu, X., Ghahramani, Z., Lafferty, J., et al. (2003). “Semi-supervised learning using Gaussian fields and harmonic functions,” in *ICML, Vol. 3* (Washington, DC), 912–919.

**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.

Copyright © 2020 Gu, Zhang, You, Zhao, Liu and Harada. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.