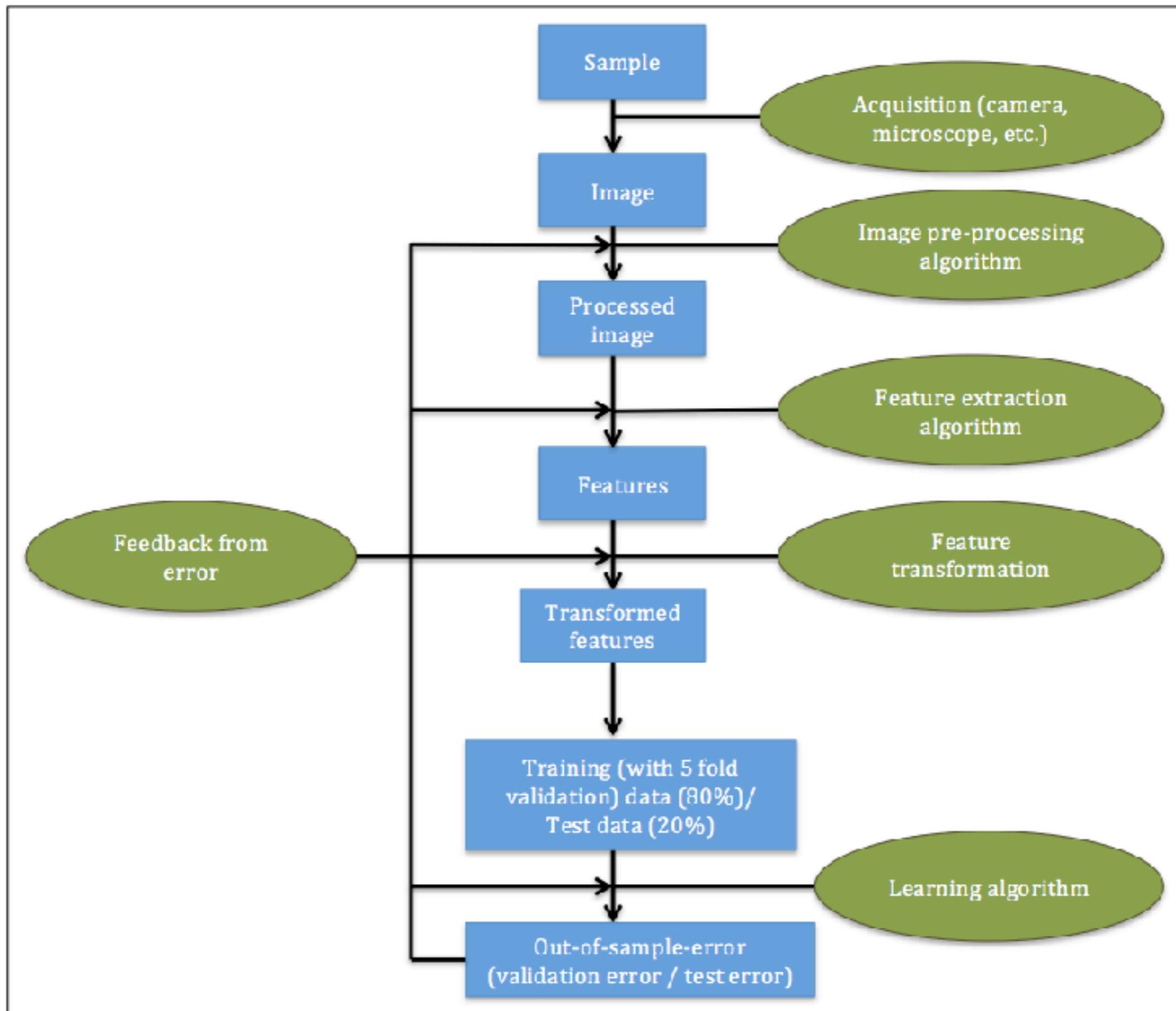


Optimization and quantification of error in image classification pipelines for noisy scientific image datasets

Outline

- Problem definition and motivation
- Our contributions
- Minimization of classification error in image classification pipelines
 - Minimization of classification error in blood vessel morphology characterization using artificial parametric 3D models
 - Minimization of classification error in microstructure characterization using exhaustive grid search
- Quantification of classification error in image classification pipelines
 - Machine learning based approach to quantify noise in medical images
 - Quantification of error contribution from image classification pipelines using methods for algorithm selection and hyper-parameter optimization
- Conclusion and future work
- References

Image classification pipeline



Problem definition and motivation

- *Motivation:* Classification error does not occur only due to the learning algorithm. It is caused by all the components of an image classification pipeline.
- *Problem:* Minimize and quantify the classification error from different components of an image classification pipeline

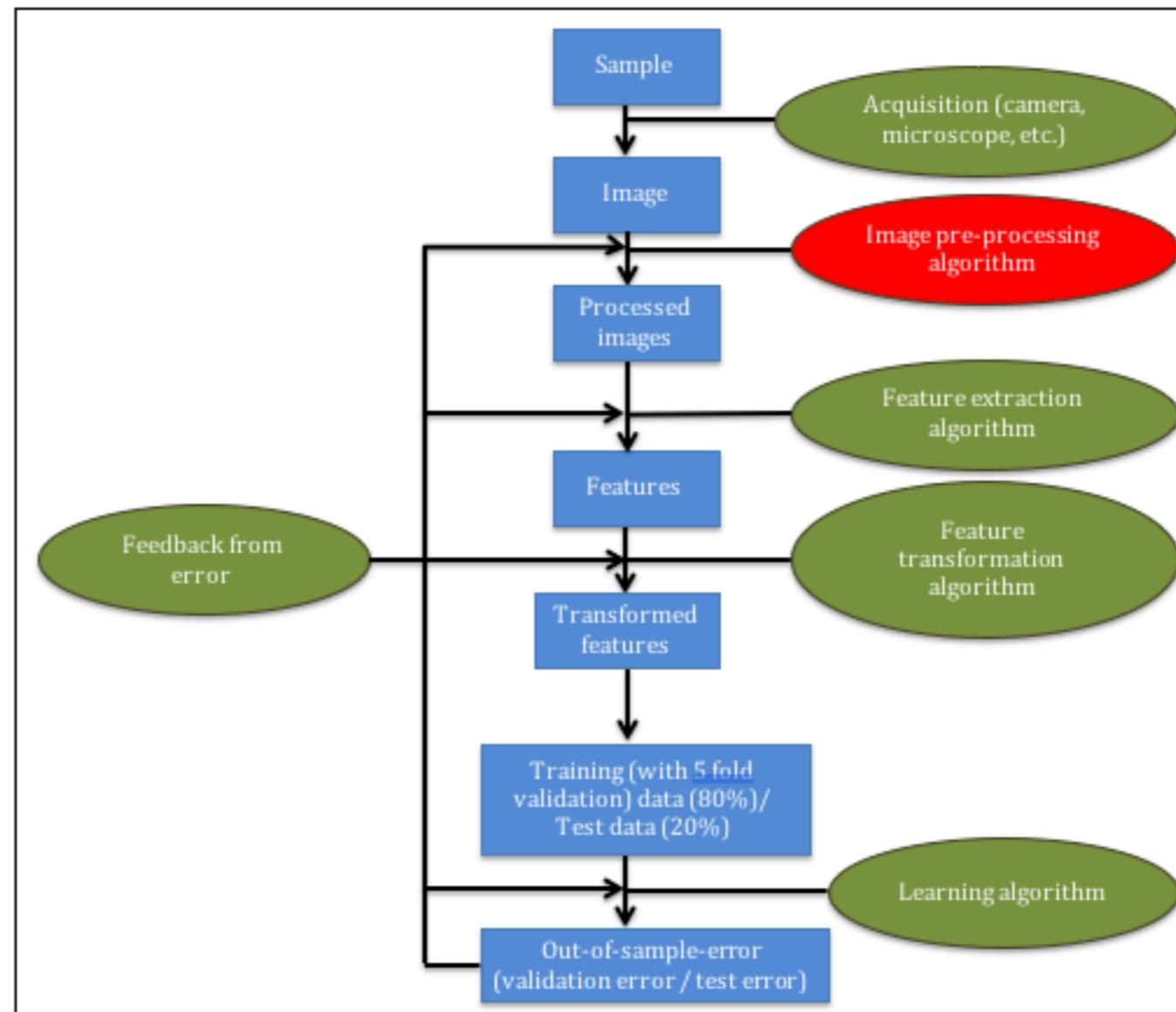
Our contributions

Our contributions

- Minimization of image classification error from different components of the pipeline
 - Minimization of error by modification of a particular component of the pipeline (*Blood vessel morphology characterization using artificial parametric 3D models*)
 - Minimization of error by optimizing the pipeline as a whole (*Microstructure characterization using exhaustive grid search*)
- Quantification of error from different components of the pipeline
 - Quantification the quality of the data (*A machine learning based approach to quantifying noise in medical images*)
 - *Quantification of error contributions from computational steps, algorithms and hyper-parameters in the pipeline*

Minimization of error in image classification pipelines

Blood vessel morphology characterization using artificial parametric 3D models



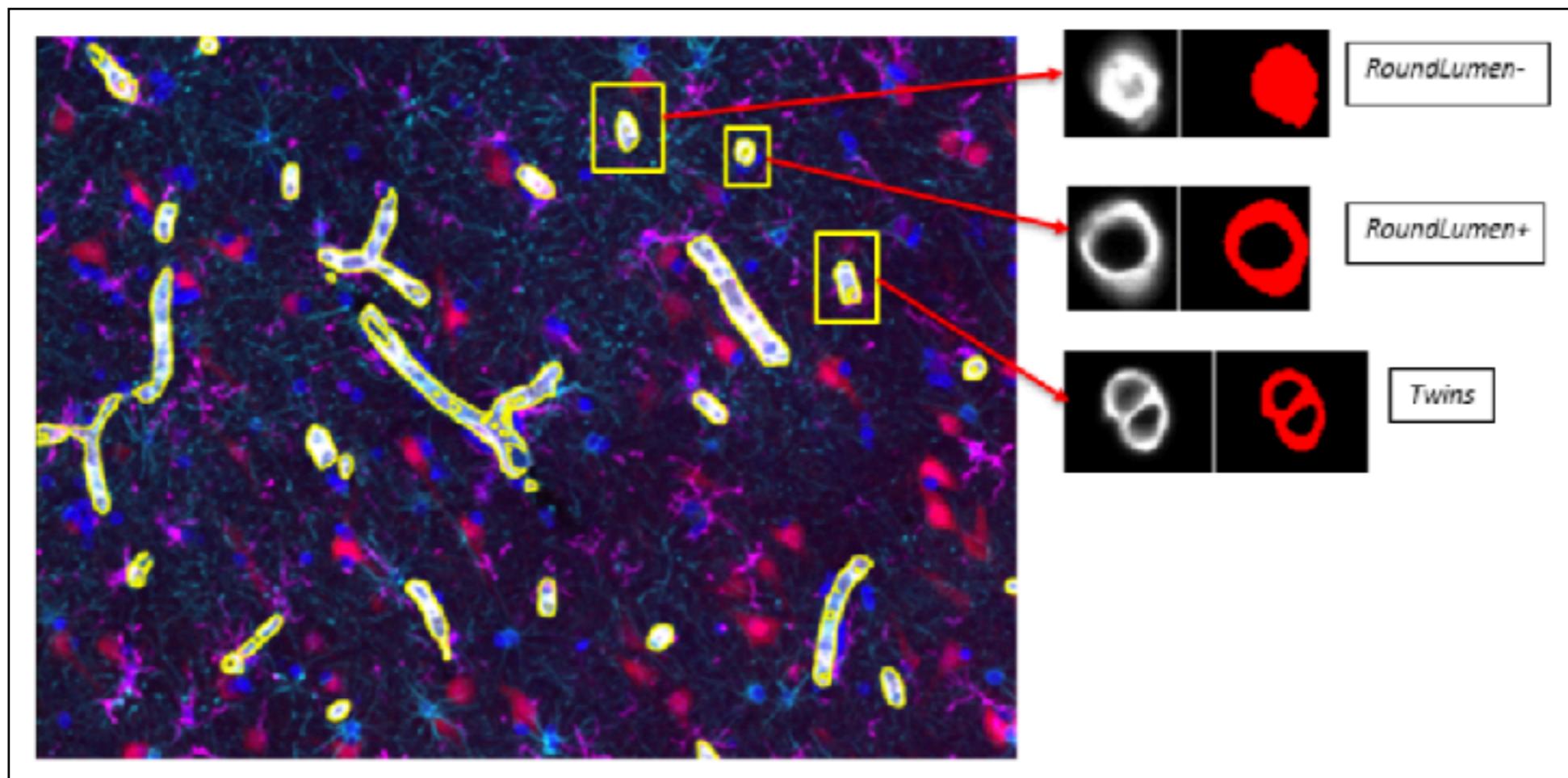
- Chowdhury, Aritra, et al. "Blood vessel characterization using virtual 3D models and convolutional neural networks in fluorescence microscopy." *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017.

Introduction

- Problem: Minimize classification error of blood vessel characterization by performing data augmentation using artificial parametric 3D models of vasculature.
 - Two classification tasks: Single blood vessels (*RoundLumen*) vs Double blood vessels (*Twins*), vessels with lumen (*RoundLumen+*) vs vessels without lumen (*RoundLumen-*)
 - Pre-trained convolutional neural networks (*AlexNet* trained on *ImageNet*) was used as the feature extraction algorithm and logistic regression was used as the classification algorithm in this work
-
- A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.
 - J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large- scale hierarchical image database,” in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 248–255.

Data

Depiction of the different morphologies in the natural data with respect to a multichannel image, overlaid with different protein markers. The three types of morphologies analyzed in this study is represented on the right.

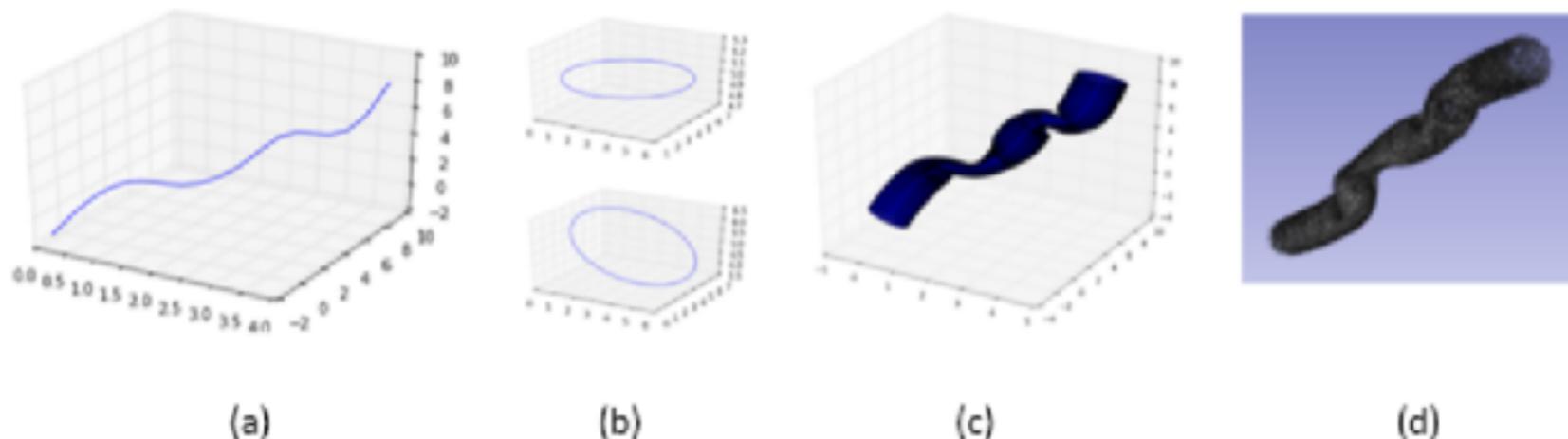


Distribution of vascular morphologies

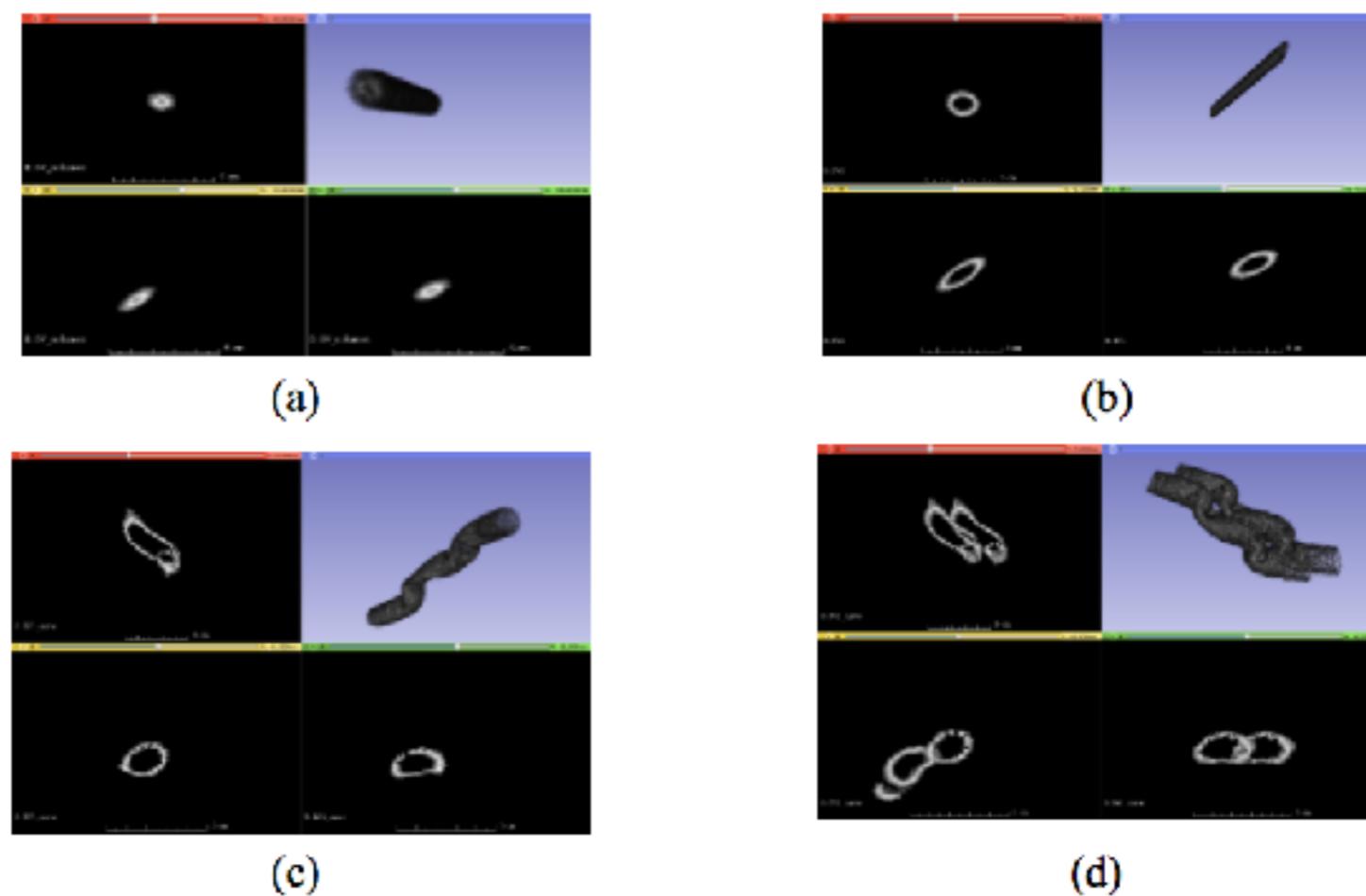
<i>RoundLumen-</i>	689
<i>Roundlumen+</i>	3427
<i>Twins</i>	266
Total	4382

Artificial 3D model of vasculature

Development of the 3D virtual model

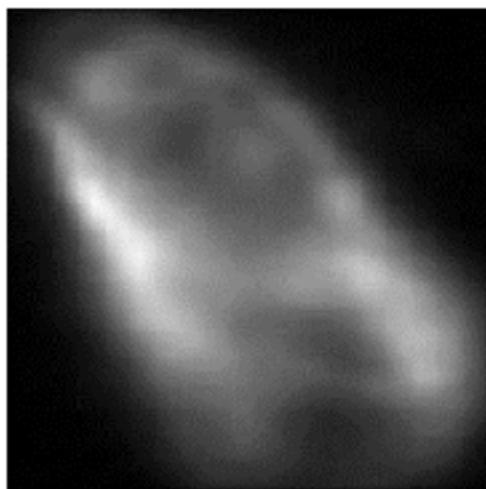


3D virtual models and their corresponding projections along different planes of view
(a) Linear model of *RoundLumen-* (b) Linear model of *RoundLumen+* (c) Non-linear model of *RoundLumen+* (d) Non-linear model of *Twins*

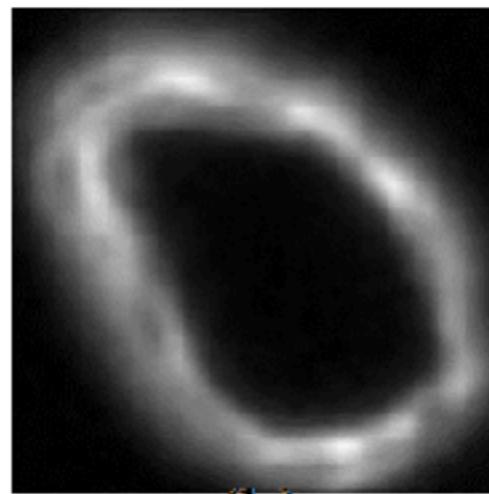


Natural and artificial data

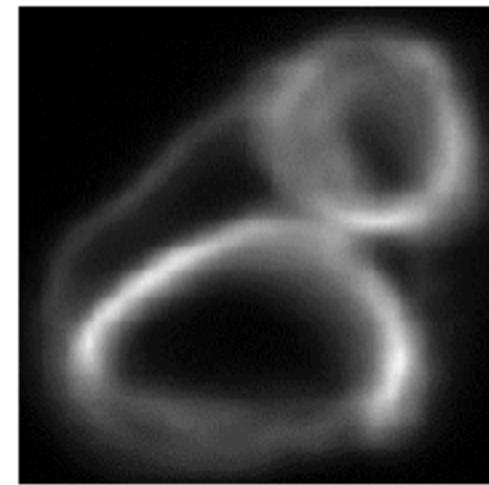
Examples of vessel classes *RoundLumen-* (a/d), *RoundLumen+* (b/e) and *Twins* (c/f) for natural (a/b/c) and virtual data (d/e/f)



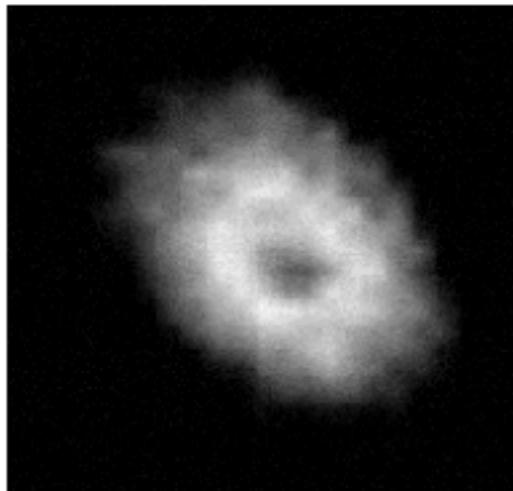
(a)



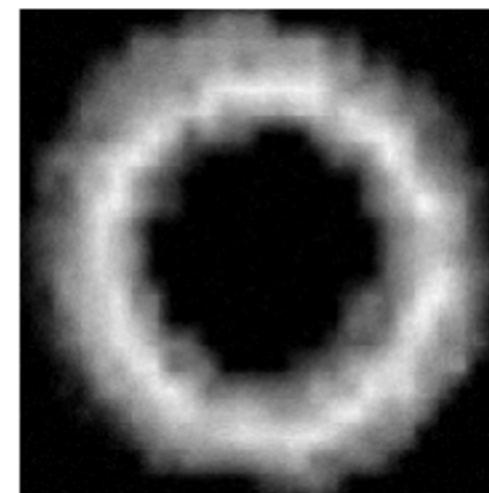
(b)



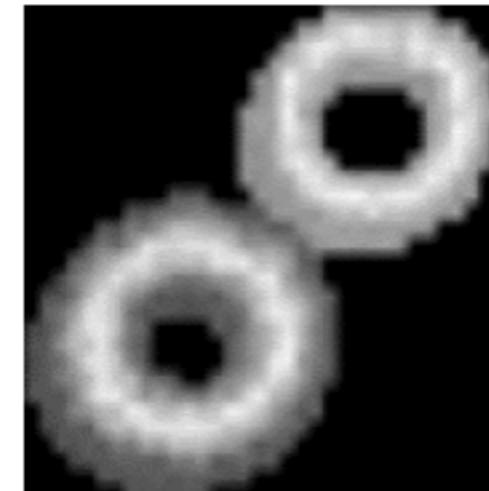
(c)



(d)



(e)



(f)

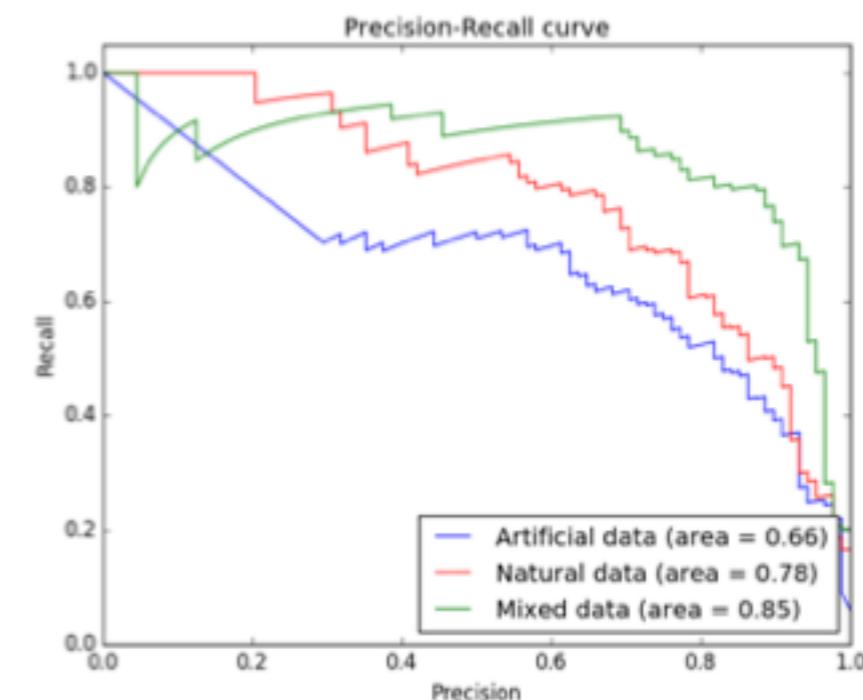
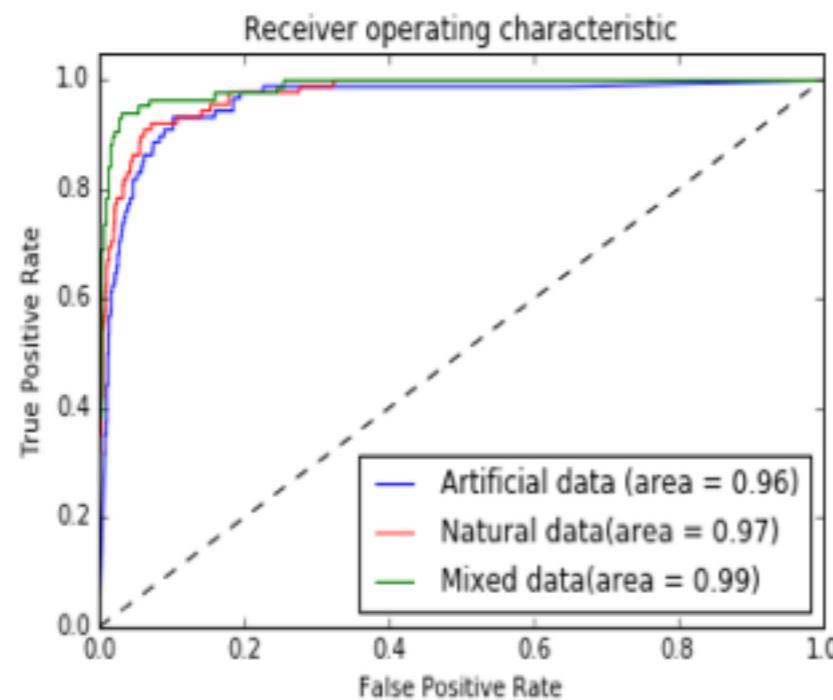
Results of task 1(*RoundLumen* vs *Twins*) :

Mixed (natural + artificial) data performs the best in terms of classification

Results of binary classification between *RoundLumen* and *Twin*

Data	Accuracy	f1-score	Precision	Recall
<i>Artificial</i>	92.81	59.36	45.24	86.36
<i>Natural</i>	96.34	71.03	68.42	73.86
<i>Mixed</i>	97.71	81.76	79.57	84.01

Plots of the receiver operating characteristics (ROC) curve and the precision recall (PR) curve of the classification between *RoundLumen* and *Twins* along with the area under the curves (AUC) for the three experiments denoted as legends in the plots. From the nature of the curves, and the values of the AUC, we conclude that combining the using *mixed* data performs better than using the *Natural* data or the *Artificial* data in isolation.



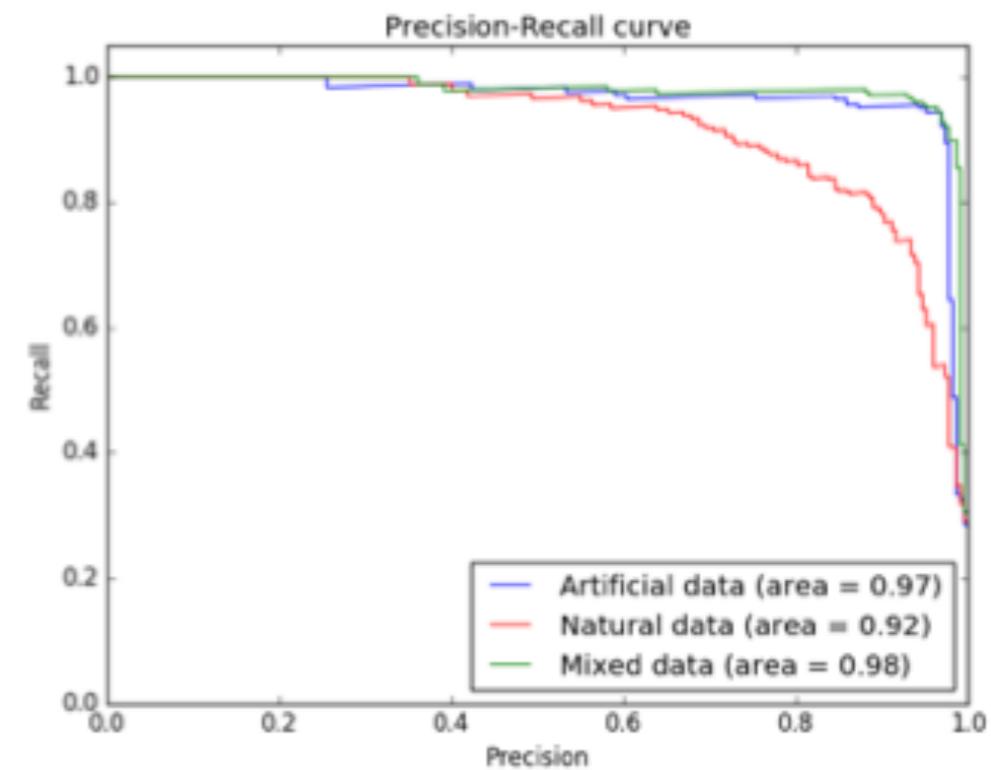
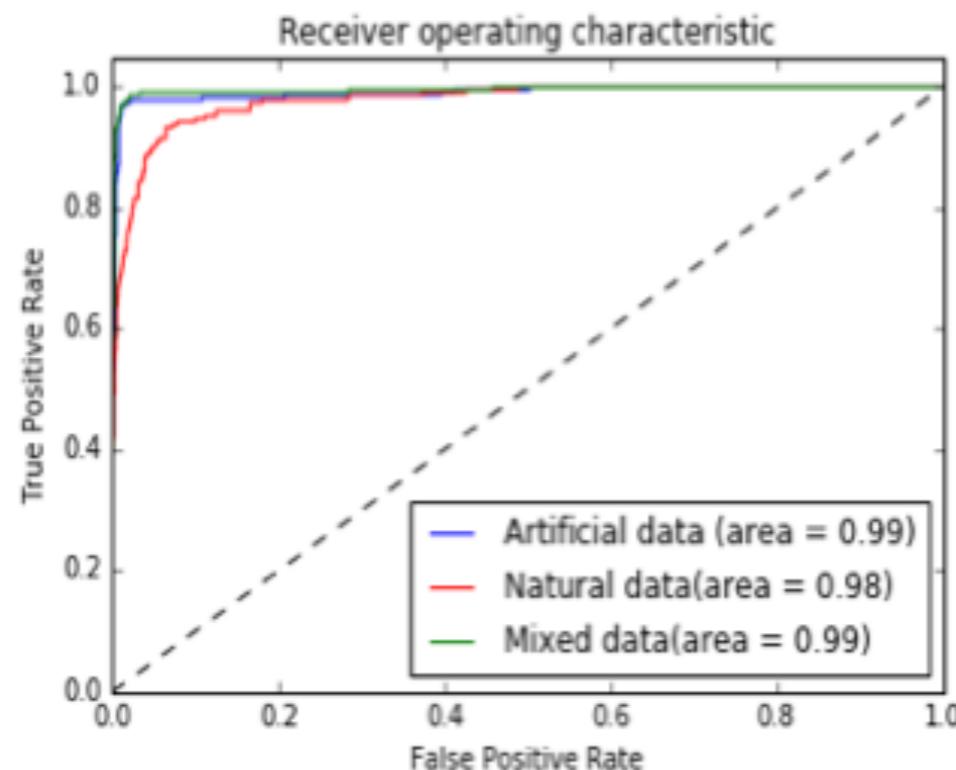
Results of task 2 (*RoundLumen-* vs *RoundLumen+*)

Mixed (natural + artificial) data performs the best in terms of classification

Results of binary classification between *RoundLumen-* and *RoundLumen+*

Data	Accuracy	f1-score	Precision	Recall
<i>Artificial</i>	98.38	99.02	99.38	98.67
<i>Natural</i>	96.34	71.03	68.42	73.86
<i>Mixed</i>	98.60	99.16	99.29	99.03

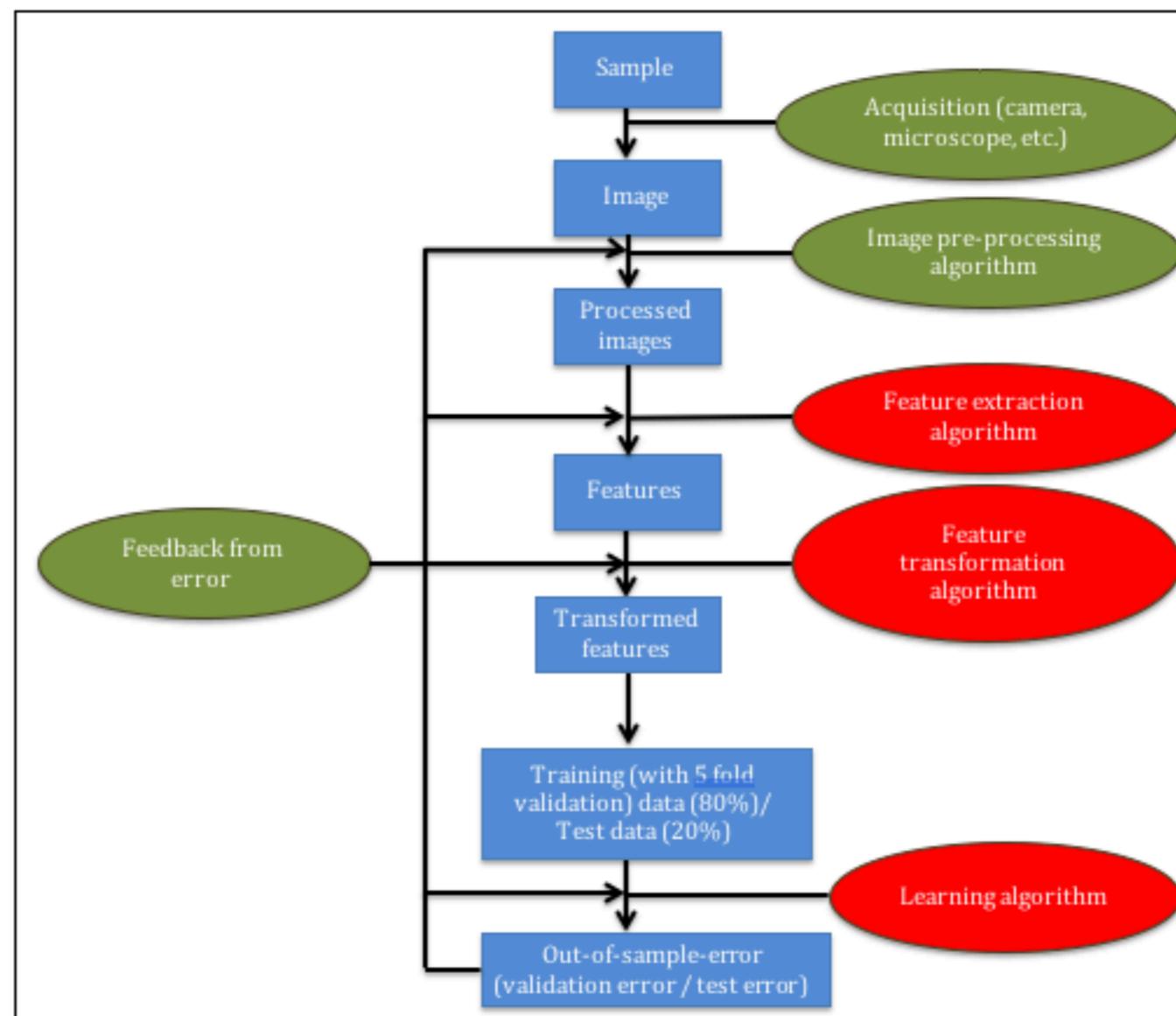
Plots of the receiver operating characteristics (ROC) curve and the precision recall (PR) curve of the classification between *RoundLumen+* and *RoundLumen-* along with the area under the curves (AUC) for the three experiments denoted as legends in the plots. From the nature of the curves, and the values of the AUC, we conclude that combining the using *mixed* data performs better than using the *Natural* data or the *Artificial* data in isolation.



Discussion

- Pre-trained convolutional neural networks maybe used to characterize blood vessel morphologies
- Mixture of natural and artificial data increases the classification accuracy of blood vessel characterization.
- Data augmentation using artificial parametric 3D models maybe used to reduce the error of classification.

Microstructure characterization using exhaustive grid search



- Chowdhury, Aritra, et al. "Image driven machine learning methods for microstructure recognition." *Computational Materials Science* 123 (2016): 176-187.

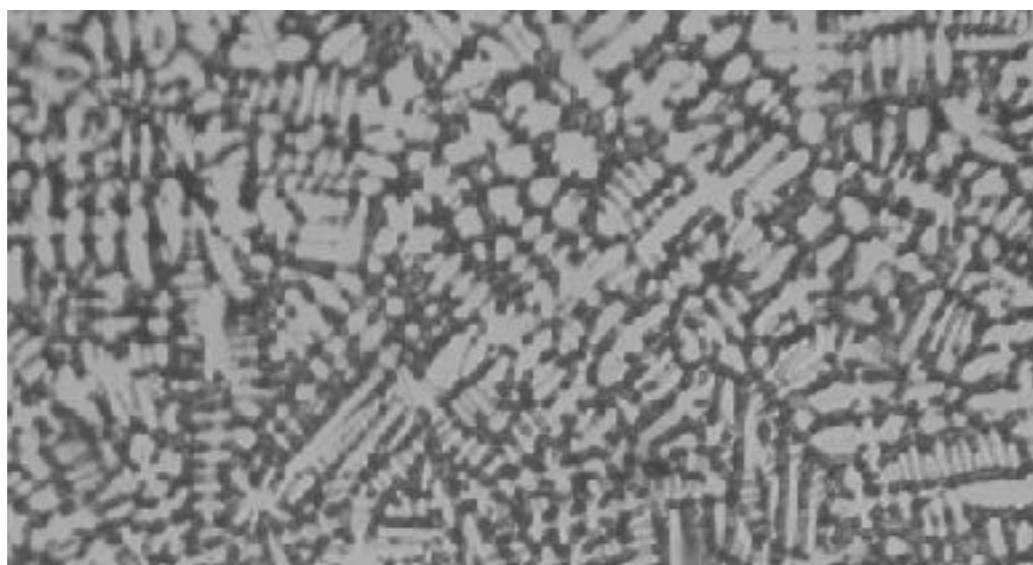
Introduction

- Problem: Find the best configuration of algorithms to characterize microstructures.
- Two classification tasks: dendrites vs non-dendrites, longitudinal dendrites vs transverse dendrites.
- Minimization of error in image classification pipeline as a whole by performing exhaustive grid search over the pipeline.

Classification tasks

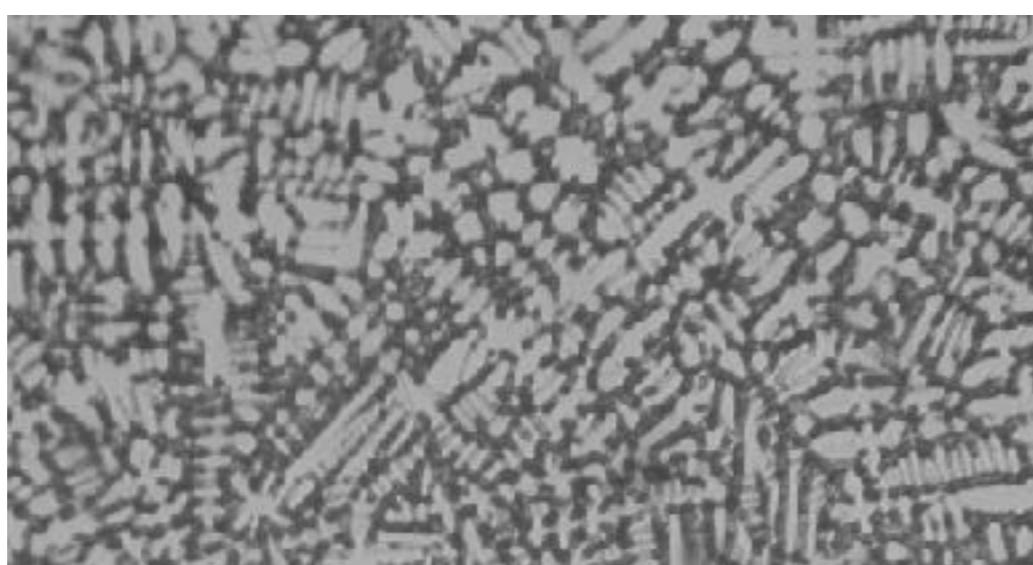
Task 1 →

Dendrite

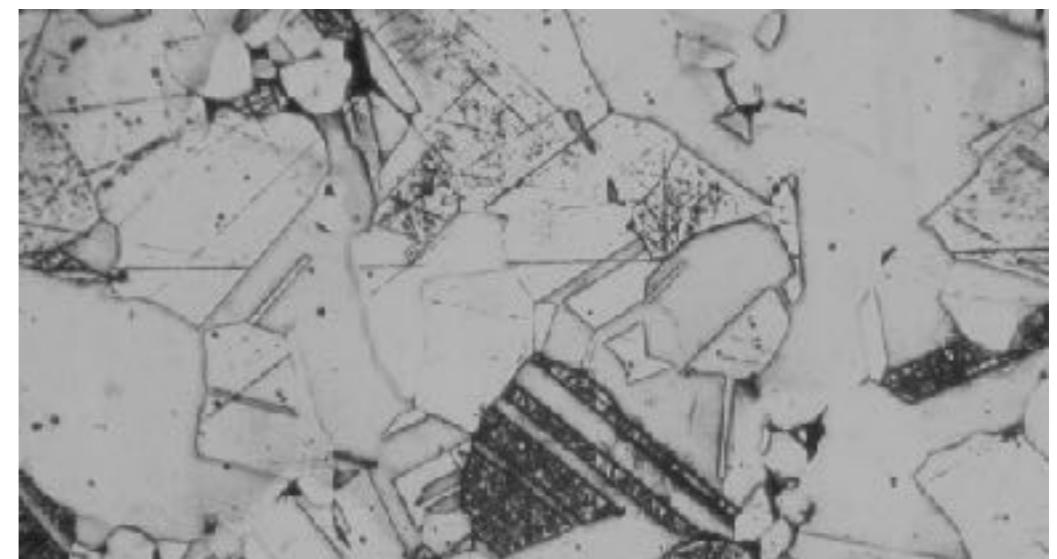


Task 2 →

Longitudinal dendrite



Non-dendrite



Transverse dendrite

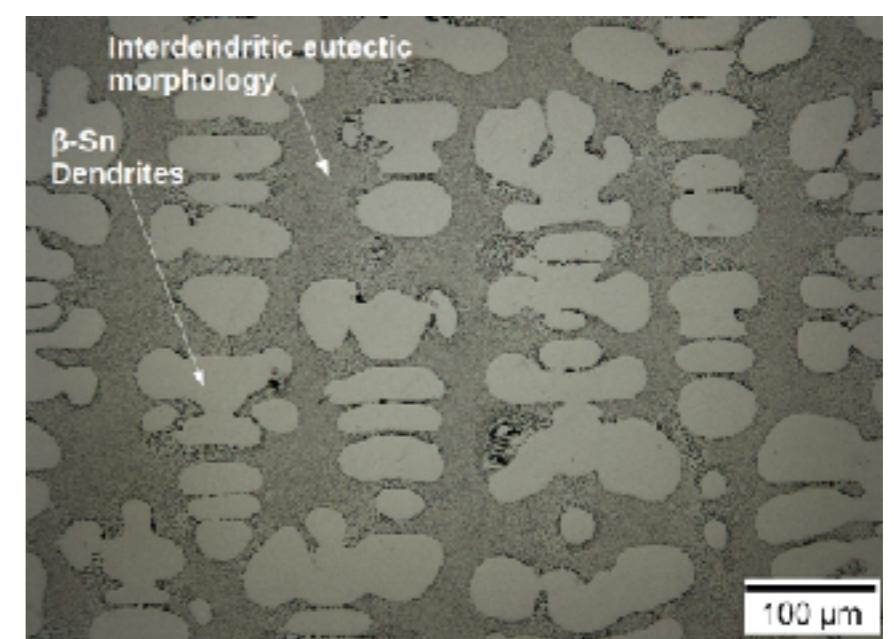
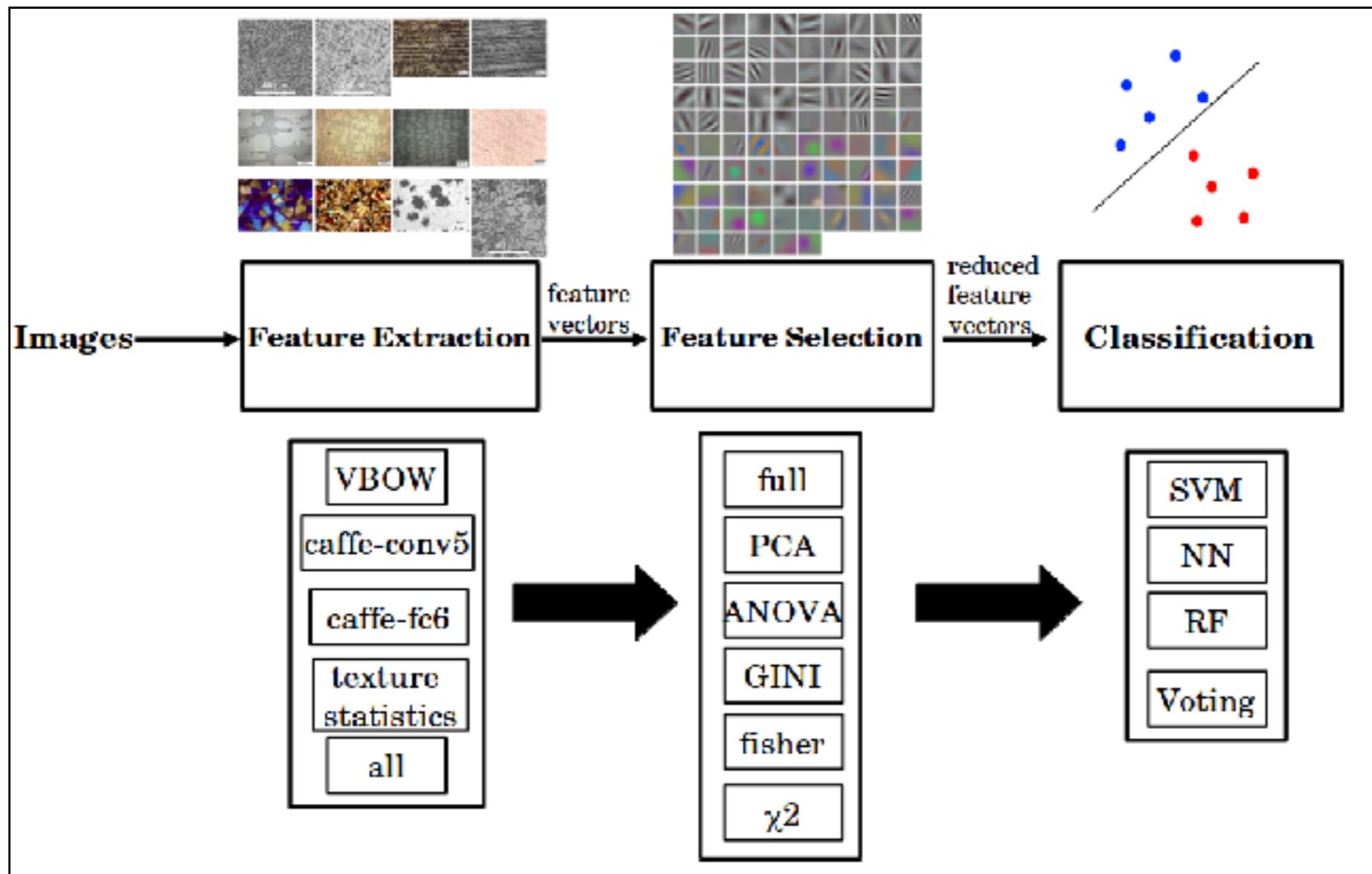


Image classification pipeline



- A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.
- J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large- scale hierarchical image database,” in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 248–255.
- Navneet Dalal, Bill Triggs, Histograms of oriented gradients for human detection, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol. 1, IEEE, 2005, pp. 886–893.
- George H. Dunteman, Principal Components Analysis, vol. 69, Sage Publications, Inc., 1989.
- Richard G. Lomax, Debbie L. Hahs-Vaughn, Statistical Concepts: A second Course, Routledge, 2013.
- Mark A. Hall, Correlation-based Feature Selection for Machine Learning PhD Thesis, The University of Waikato, 1999.
- Ron Kohavi, George H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1) (1997) 273–324.
- Huan Liu, Rudy Setiono, Chi2: feature selection and discretization of numeric attributes, in: Proceedings of the Seventh International Conference on Tools with Artificial Intelligence, 1995, pp. 388–391.
- Thierry Denoeux, A k-nearest neighbor classification rule based on Dempster–Shafer theory, *IEEE Trans. Syst., Man Cybernet.* 25 (5) (1995) 804–813

Results of Task 1 (*dendrites vs non-dendrites*)

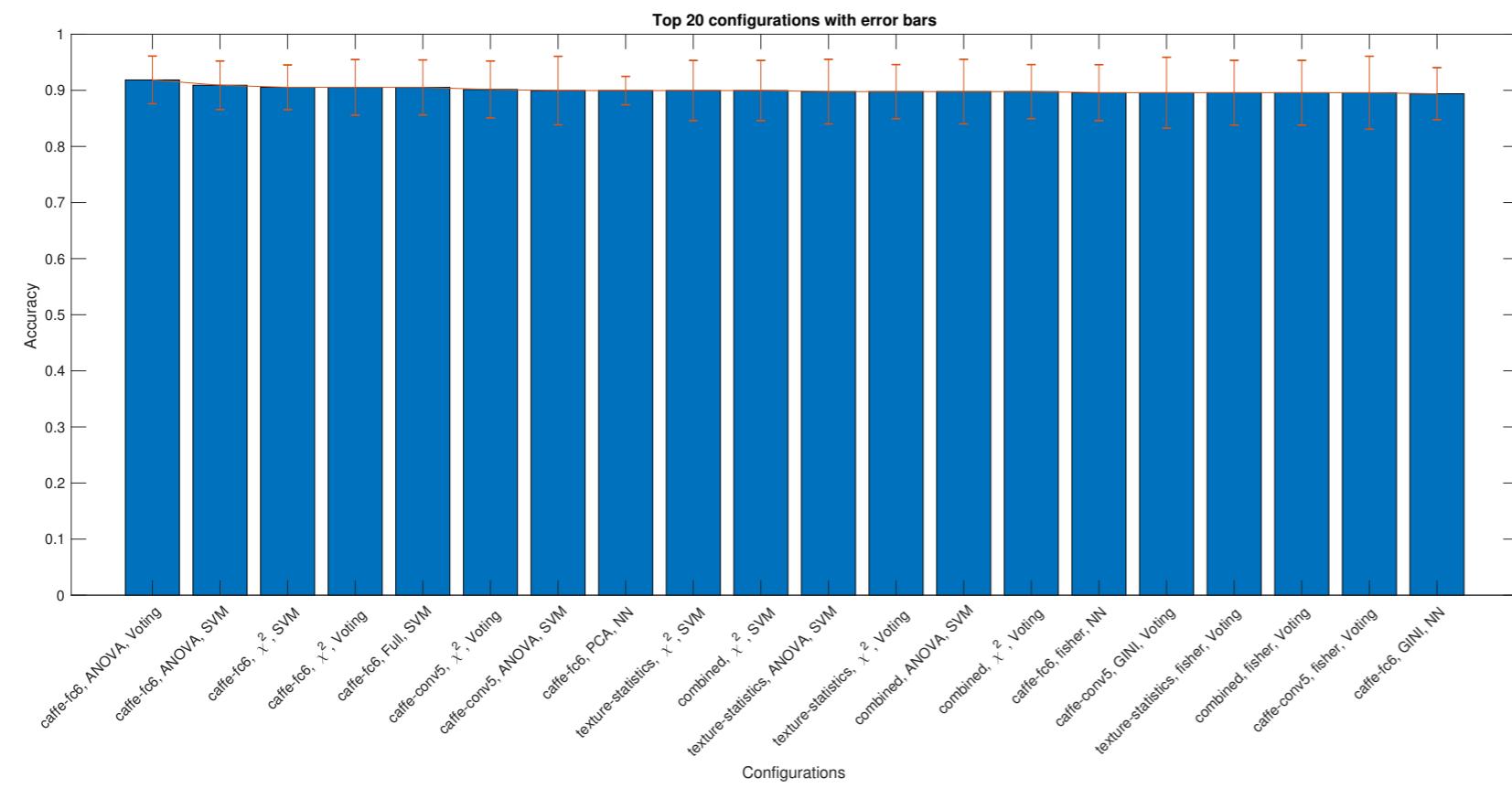
Pre-trained convolutional neural networks (*caffe-fc6*) is able to characterise microstructures well to distinguish between dendrites and non-dendrites

Best configuration with the maximum mean classification accuracy

Task	Feature Extraction	Feature Selection	Classifier	Accuracy
1	caffe-fc6	ANOVA	Voting	91.85 ± 4.25 %

Top 20 configurations of algorithms in Task 1 with error bars representing one standard deviation. There is no significant difference in the accuracies in the different configurations. Most of the feature extraction algorithms in the top 20 configurations are pre-trained CNNs (*caffe-fc6* or *caffe-conv5*)

Average rank of the algorithms in Task 1 with respect to feature extraction, dimensionality reduction and classification. The average rank of an algorithm quantifies it's position in the sorted list of configurations.



Feature extraction	Average rank
caffe-fc6	47.82
texture-statistics	61.46
combined	64.46
caffe-conv5	72.39
VBOW	106.36

Dimensionality reduction	Average rank
χ^2	54.45
fisher	58.05
PCA	60.95
ANOVA	61.80
GINI	66.10
Full	66.55

Classification	Average rank
SVM	54.57
Voting	60.8
RF	81.40
NN	85.23

Results of Task 2 (*longitudinal vs transverse cross-sections*)

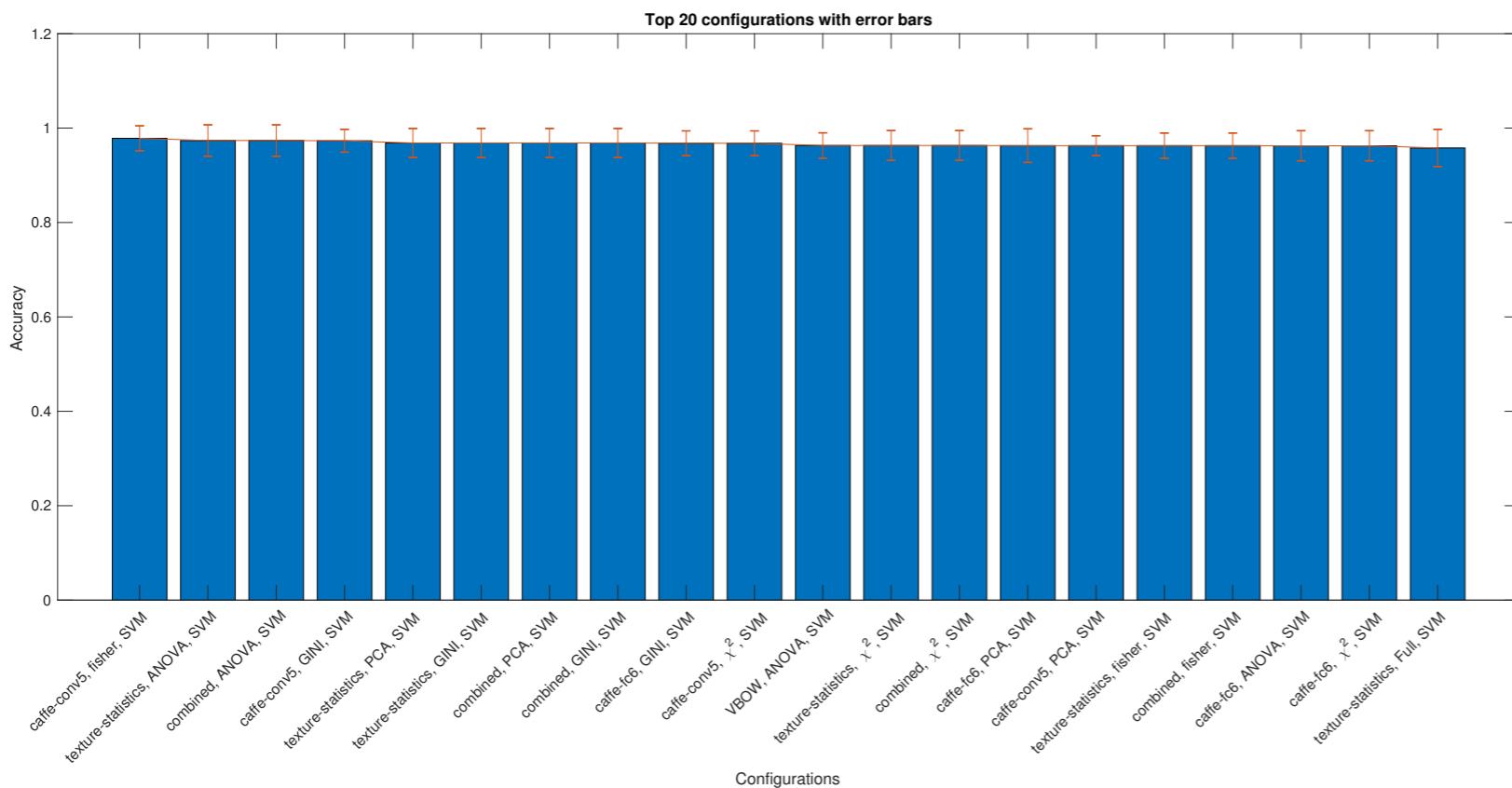
Pre-trained convolutional neural networks (*caffe-fc6*) is able to characterize microstructures well to distinguish between longitudinal and transverse dendrites

Best configuration with the maximum mean classification accuracy

Task	Feature Extraction	Feature Selection	Classifier	Accuracy
2	caffe-conv5	Fisher	SVM-L	97.84 ± 2.65 %

Top 20 configurations of algorithms in Task 2 with error bars representing one standard deviation. There is no significant difference in the accuracies in the different configurations. Most of the feature extraction algorithms in the top 20 configurations are pre-trained CNNs (*caffe-fc6* or *caffe-conv5*)

Average rank of the algorithms in Task 2 with respect to feature extraction, dimensionality reduction and classification. The average rank of an algorithm quantifies it's position in the sorted list of configurations.



Feature extraction	Average rank
caffe-fc6	47.64
texture-statistics	58.64
VBOW	70.5
combined	81.82
caffe-conv5	93.89

Dimensionality reduction	Average rank
PCA	54.45
χ^2	58.05
ANOVA	61.80
GINI	66.10
fisher	66.55
Full	69.6

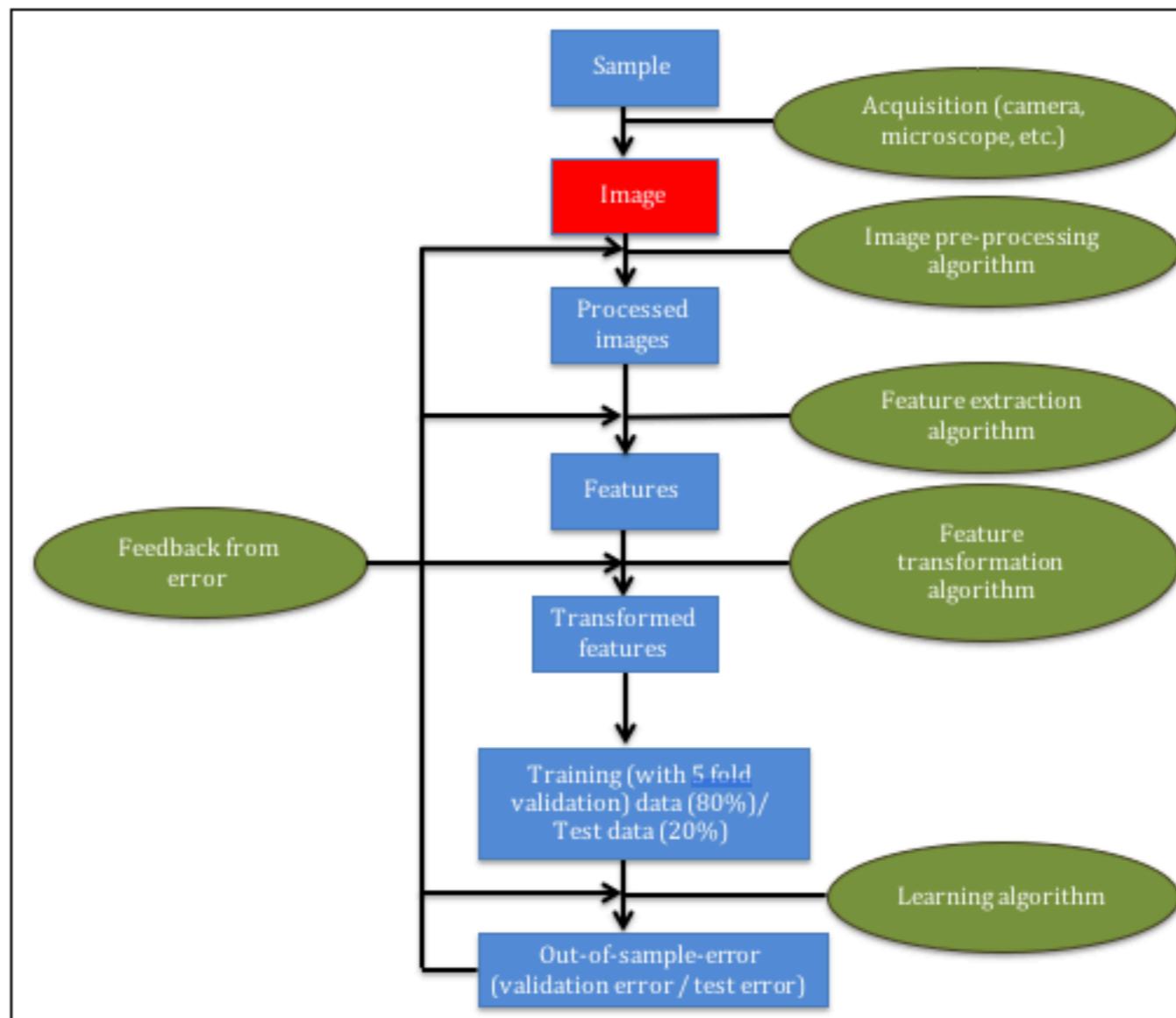
Classification	Average rank
SVM	31.22
NN	71.6
Voting	75.20
RF	103.97

Discussion

- The best configuration maybe found by minimizing the image classification pipeline as a whole using exhaustive grid search over algorithms and hyper-parameters.
- Pre-trained neural networks (*caffe-fc6*) maybe used to characterize and distinguish microstructural features.
- Grid search and combined algorithm selection and hyperparamater optimization based methods can be used to minimize classification error in image classification tasks in material science and other domains.

Quantification of error in image classification pipelines

A machine learning based approach to quantify noise in medical images



- Chowdhury, Aritra, et al. "A machine learning approach to quantifying noise in medical images." *Medical Imaging 2016: Digital Pathology*. Vol. 9791. International Society for Optics and Photonics, 2016.

Introduction

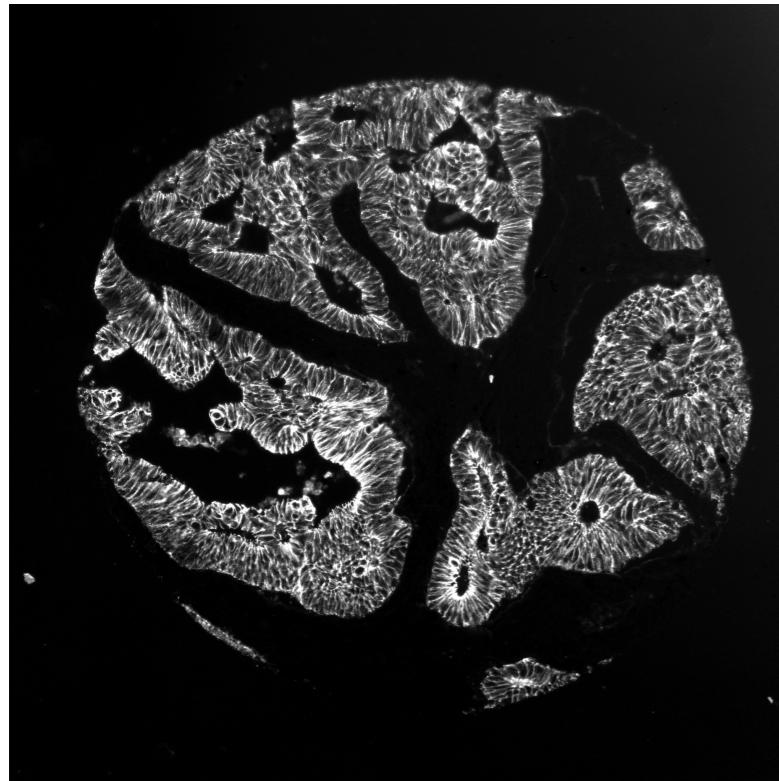
- *Quality of Image (QoI)* score quantifies the amount of noise in an image.
- The score may also be used to quantify the quality of a protein marker or a dataset.
- Haralick texture features, SMOTE, PCA and logistic regression is used to formulate the *QoI*.

- R. M. Haralick, K. Shanmugam et al., “Textural features for image classification,” IEEE Transactions on systems, man, and cybernetics, no. 6, pp. 610–621, 1973.
- N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” Journal of artificial intelligence research, vol. 16, pp. 321–357, 2002.
- George H. Dunteman, Principal Components Analysis, vol. 69, Sage Publications, Inc., 1989.

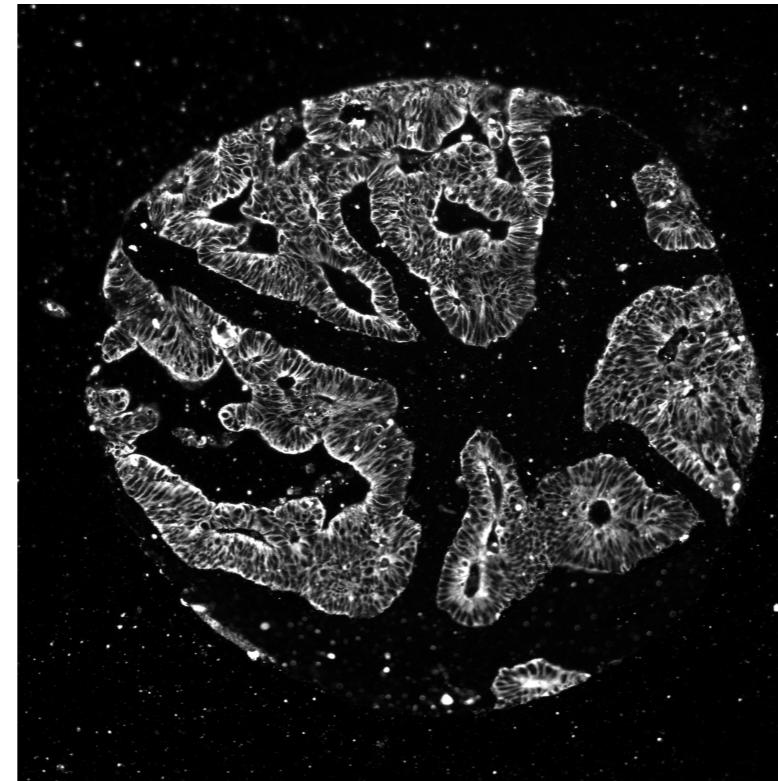
Data

- Markers E_cad, CK15 and pck26 were used for this analysis.
- Images were annotated as *good* (high signal) or *bad* (low signal) by pathologist.

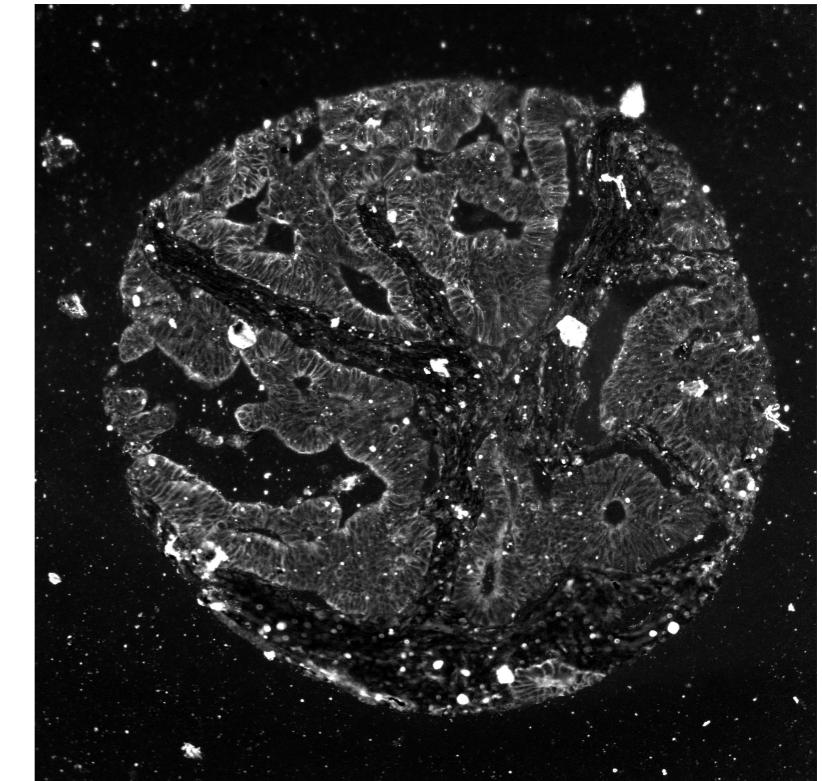
E_cad



pck26



CK15



Methods

- Haralick texture features were used for feature extraction
- A set of 13 texture features were computed based on the gray level co-occurrence matrix and the intensity information of the image.

$$G_{ij} = \begin{bmatrix} p(1, 1) & p(1, 2) & \cdots & p(1, N_g) \\ p(2, 1) & p(2, 2) & \cdots & p(2, N_g) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_g, 1) & p(N_g, 2) & \cdots & p(N_g, N_g) \end{bmatrix}$$

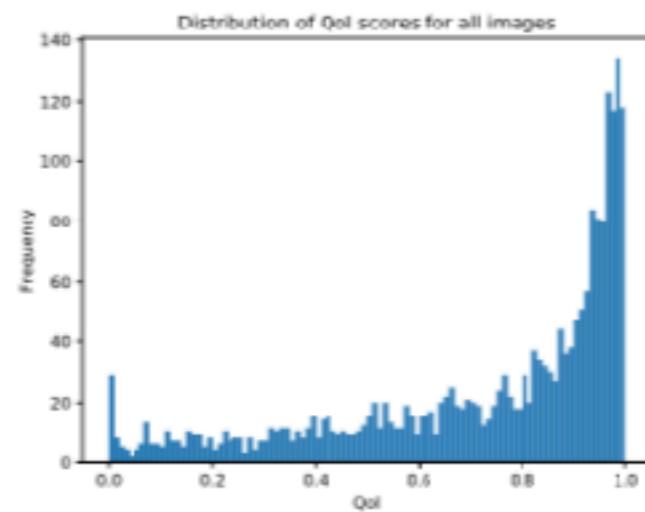
- The *QoI* score is defined as the probability that an image is from the *good* class. It is given by the following equation

$$S_i = p_{i1}$$

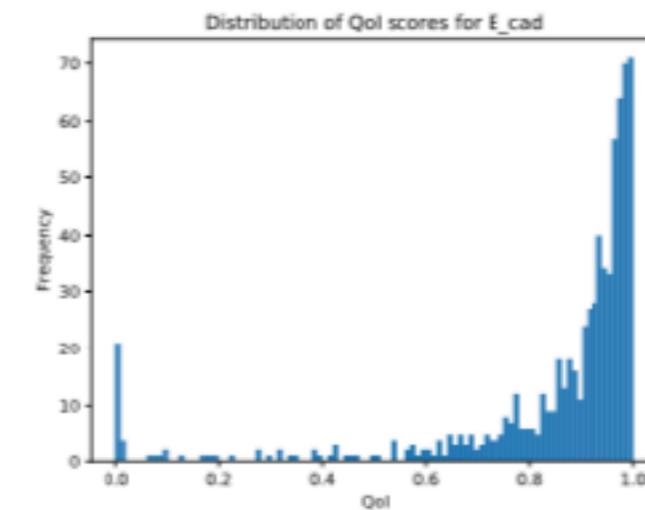
Results

Distribution of QoI scores is able to quantify the perceived difference in quality between the E_cad, pck26 and CK15 markers

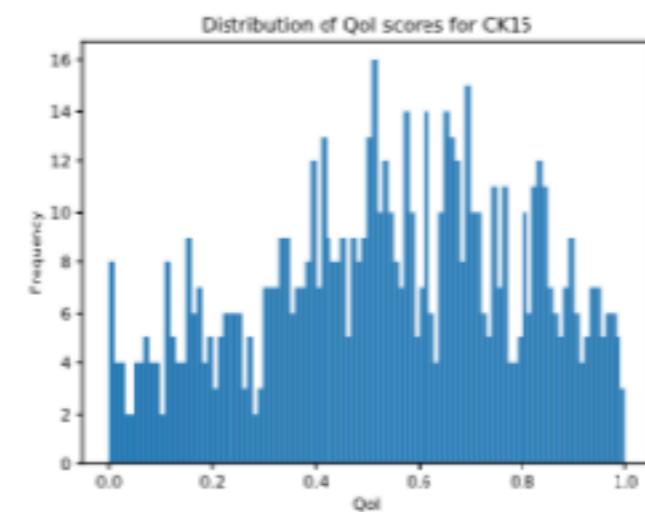
Distribution of *QoI* scores of the images. According to the pathologist, the perceived quality of the images from the E_cad and pck26 marker is good and the images from the CK15 marker has low signal and high noise in general. This is reflected in the distribution of these markers



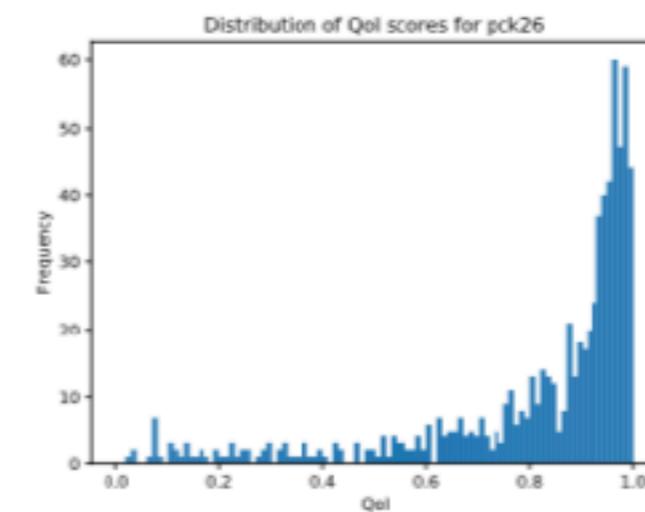
(a) Distribution of scores on all the images



(b) Distribution of scores from E_cad marker.



(c) Distribution of scores from CK15 marker.



(d) Distribution of scores from pck26 marker

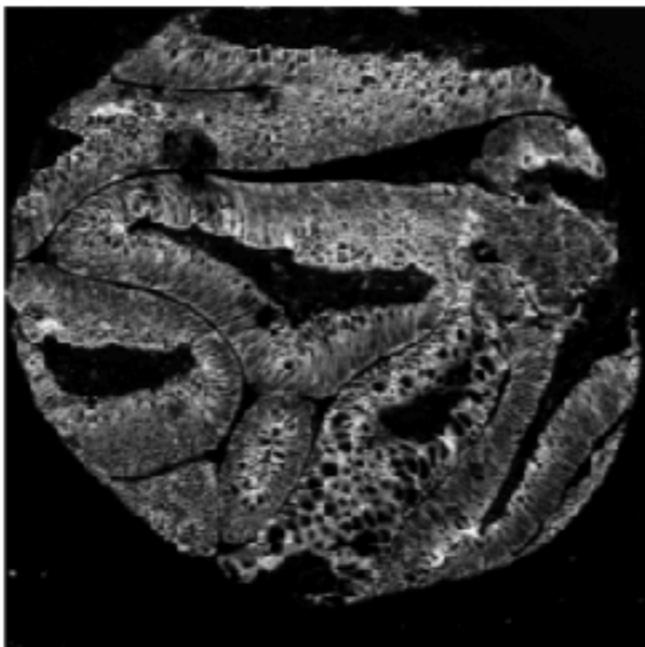
Results

Distribution of QoI scores is able to quantify the perceived difference in quality between the E_cad, pck26 and CK15 markers

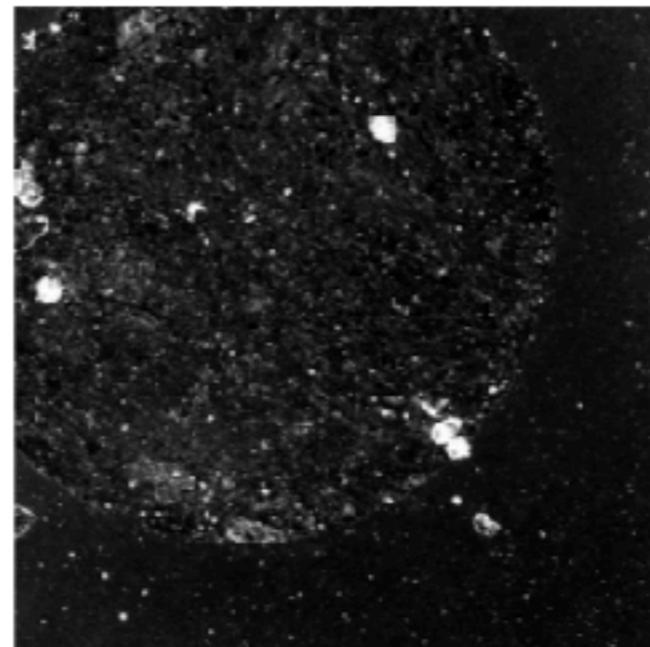
Percentage of the number of images that fall in the *good*(0.67- 1), *bad* (0 - 0.33) and *ugly* (0.33 - 0.67) regions of the QoI score with respect to the 3 markers. The percentage values clearly quantify the claim that E_cad is the least noisy followed by pck26 and CK15.

Marker	good (%)	bad (%)	ugly (%)
E_cad	43.14	16.33	10.54
pck26	40.15	19.52	18.61
CK15	16.71	64.14	70.85

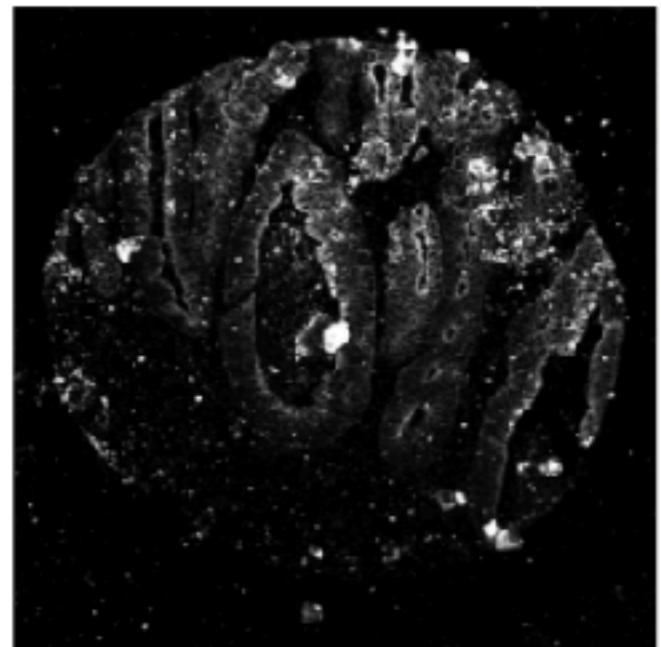
Examples of *good*, *bad* and *ugly* images based on the *QoI* score.



(a) A *good* image from E_cad marker with a *QoI* of 0.9945



(b) A *bad* image from CK15 marker with a *QoI* of 0.0077

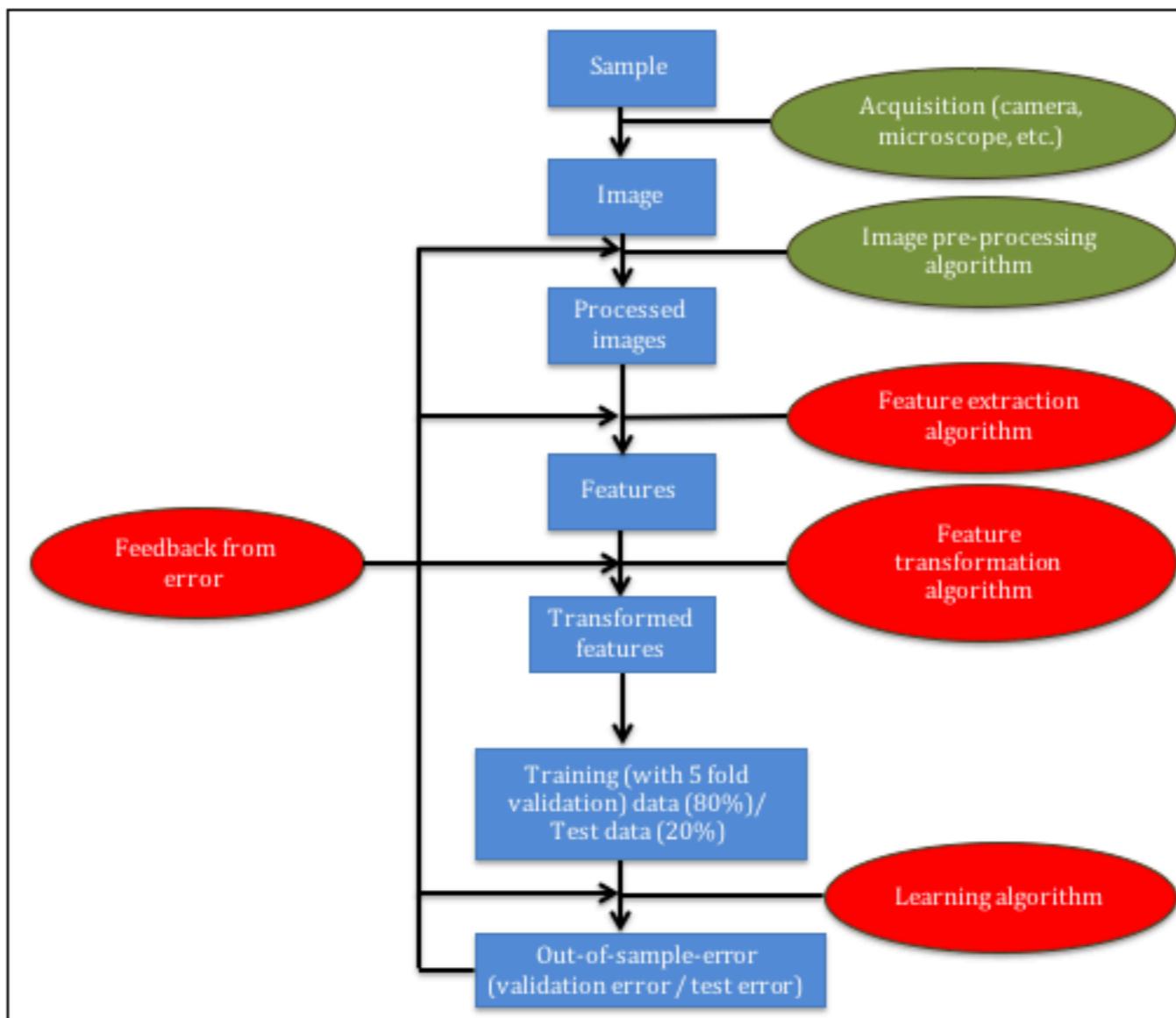


(c) A *ugly* image from pck26 marker with a *QoI* of 0.5262

Discussion

- The QoI score maybe used to quantify the perceived quality of an image.
- The QoI score maybe used to filter images or markers from a dataset.
- This can be used as a pre-processing step to perform further analysis of medical images.

Quantification of error contribution from computational steps, algorithms and hyperparameters



- Chowdhury, Aritra, et al. "Algorithm selection and hyperparameter optimization based quantification of error contribution in image classification pipelines." *IEEE International Conference on Data Mining (ICDM) 2018* (Submitted)

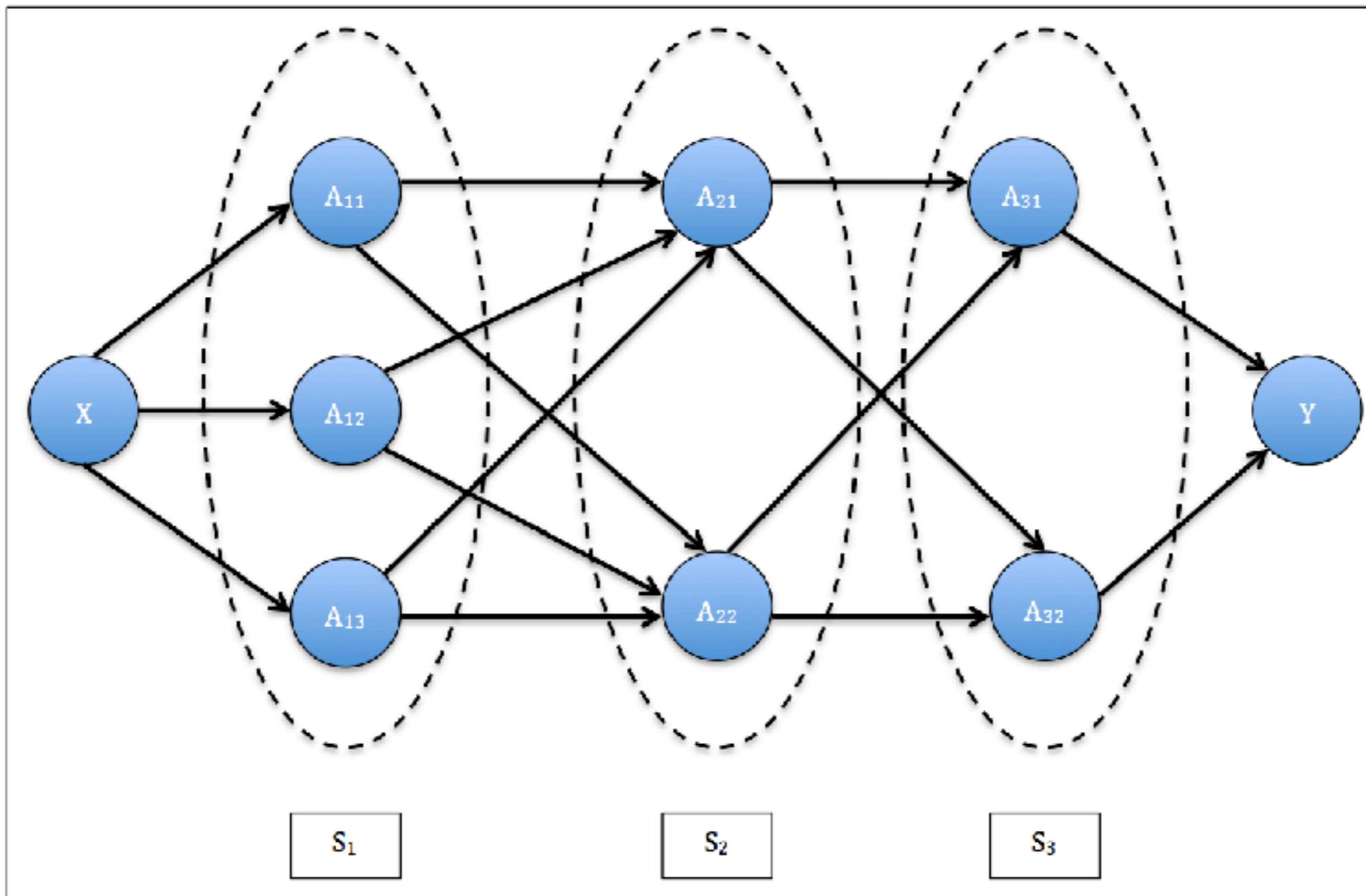
Introduction

- Contribution of error from different components of image classification pipeline - steps, algorithms, hyper-parameters.
- Provides data scientists and domain experts with insights about the pipeline in terms of which components are important for the performance of the pipeline.
- Hyper-parameter optimization methods and algorithms are used to quantify error contributions - grid search, random search, Bayesian optimization.

- J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” Journal of Machine Learning Research, vol. 13, no. Feb, pp. 281–305, 2012.
- J. Snoek, H. Larochelle, and R. P. Adams, “Practical bayesian optimization of machine learning algorithms,” in Advances in neural information processing systems, 2012, pp. 2951–2959.

Image classification pipeline used in problem

Representation of the image classification pipeline as a directed acyclic graph used in this work. This is an instantiation of the generalized data analytic pipeline

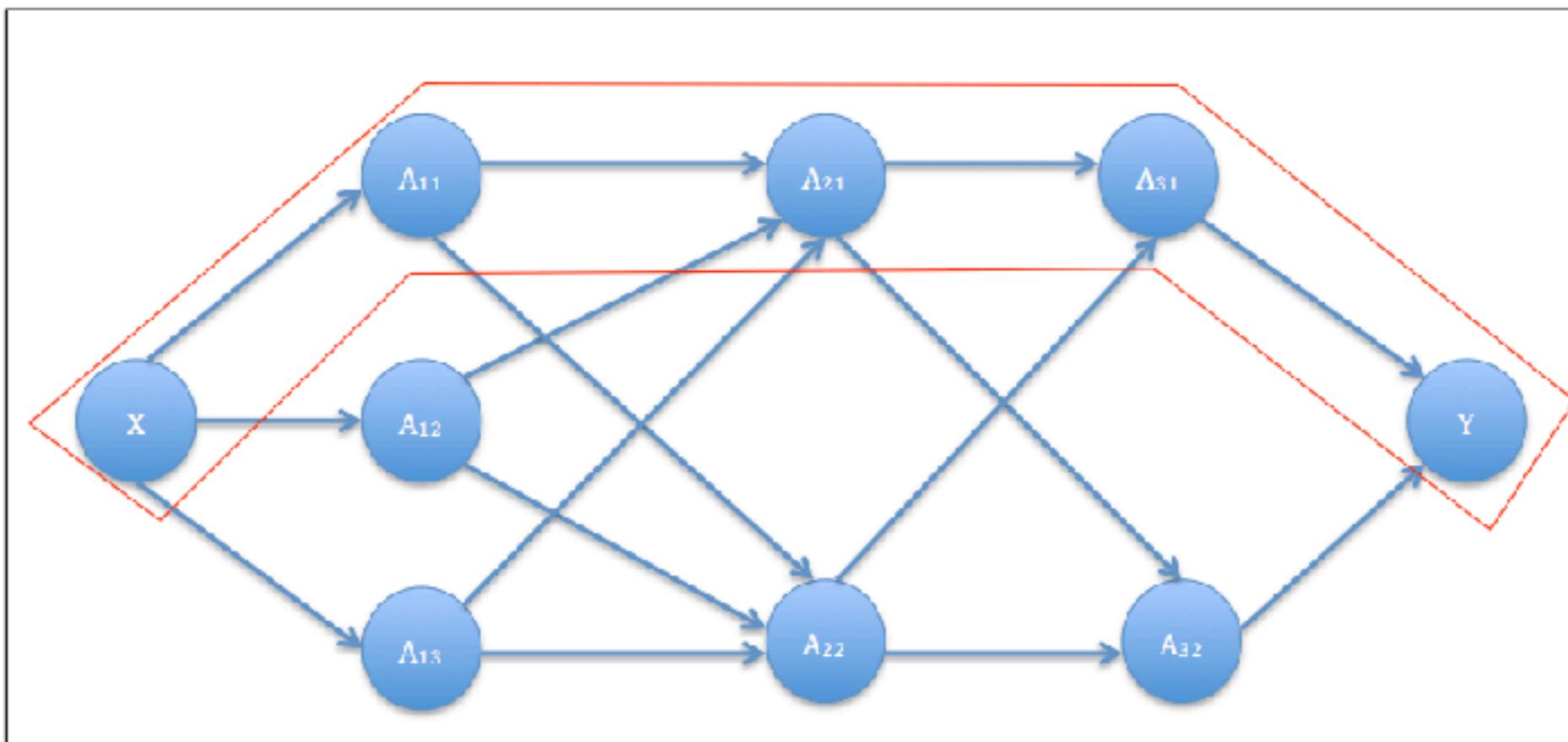


Hyper-parameter optimization (HPO)

Let the n hyperparameters in a path be denoted as $\theta_1, \theta_2, \dots, \theta_n$, and let $\Theta_1, \Theta_2, \dots, \Theta_n$ be their respective domains. The hyperparameter space of the path is $\Theta = \Theta_1 \times \Theta_2 \times \dots \times \Theta_n$.

When trained with $\theta \in \Theta$ on data D_{train} , the validation error is denoted as $\mathcal{L}(\theta, D_{train}, D_{valid})$. Using k -fold cross-validation, the hyperparameter optimization problem for a dataset D is to minimize:

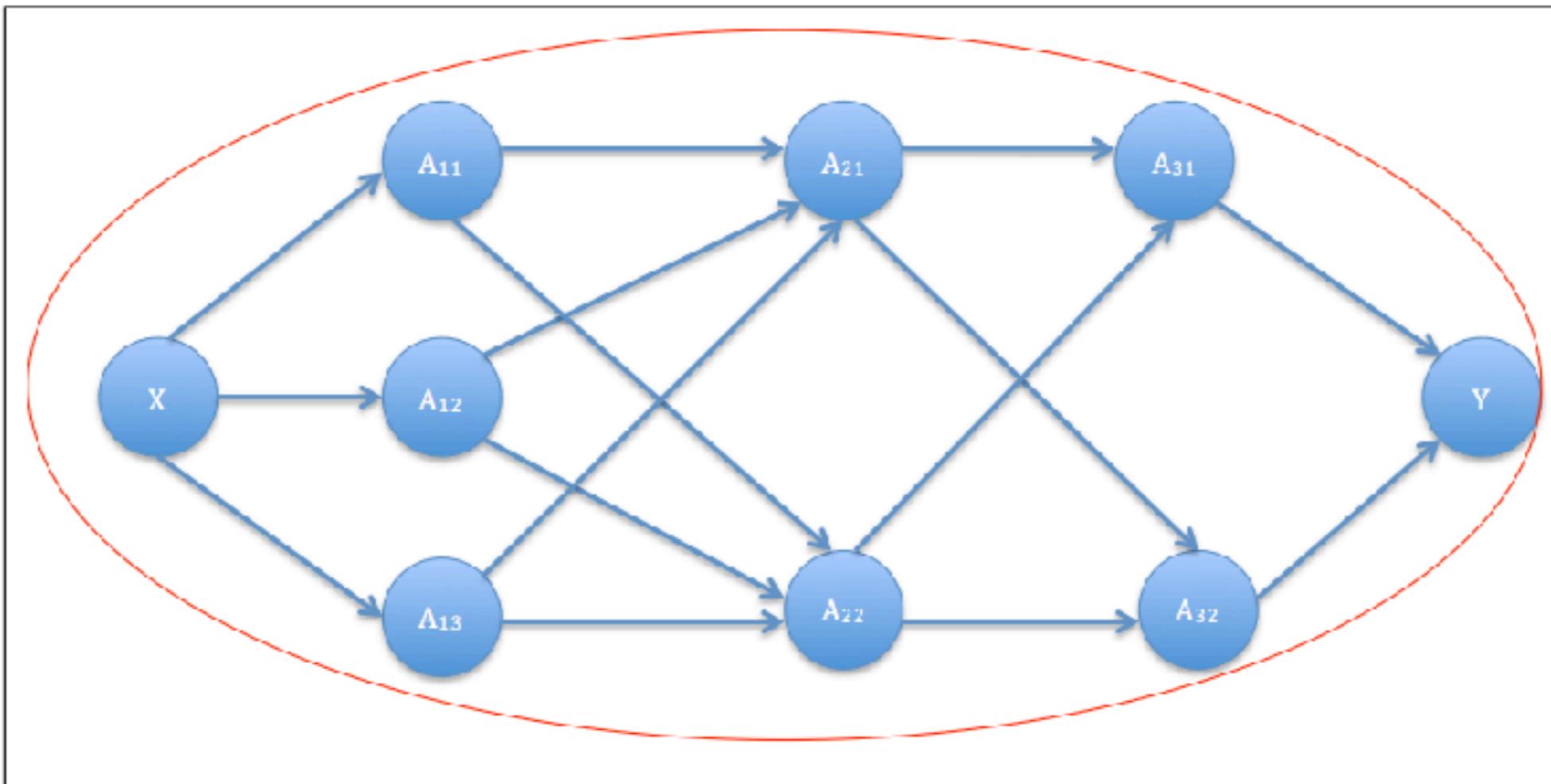
$$f^D(\theta) = \frac{1}{k} \sum_{i=1}^k \mathcal{L}(\theta, D_{train}^{(i)}, D_{valid}^{(i)})$$



Combined algorithm selection and hyperparameter optimization (CASH)

Let there be n computational steps in the pipeline. Each step i in the pipeline consists of algorithms $A_i(\Theta_i)$, where $A_i(\Theta_i) = \{A_{i1}(\theta_{i1}), \dots, A_{im_i}(\theta_{im_i})\}$, m_i is the number of algorithms in step i , A_{ij} represents the j -th algorithm in step i , and θ_{ij} represents the set of hyperparameters corresponding to A_{ij} . The entire space of algorithms and hyperparameters is therefore given by $\mathcal{A} = A_1(\Theta_1) \times A_2(\Theta_2) \times \dots \times A_n(\Theta_n)$. The objective function to be minimized for CASH is given by

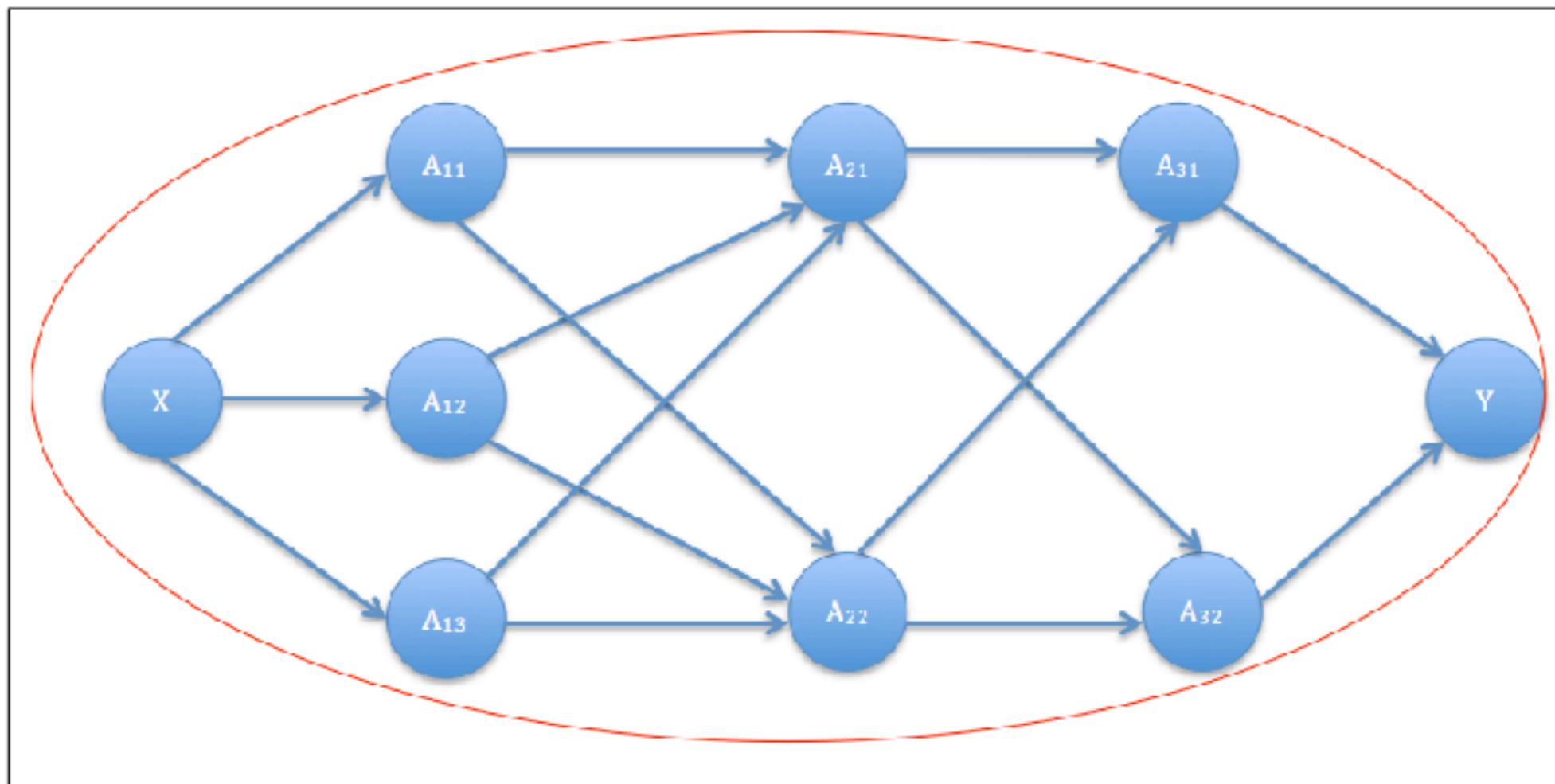
$$f^D(A) = \frac{1}{k} \sum_{i=1}^k \mathcal{L}(A, D_{train}^{(i)}, D_{valid}^{(i)})$$



Error contribution from computational steps using the agnostic methodology

Let n be the number of steps in the pipeline. Each step in the pipeline is denoted as S_i . $|S_i|$ is the number of algorithms in step i . A_{ij} denotes the j -th algorithm in the i -th step. E^* represents the minimum validation error found after optimization of the entire pipeline (using the CASH framework). $E_{A_{ij}}^*$ is the minimum validation error found with A_{ij} as the only algorithm in step i . For $i = 1, \dots, n, j = 1, \dots, |S_i|$,

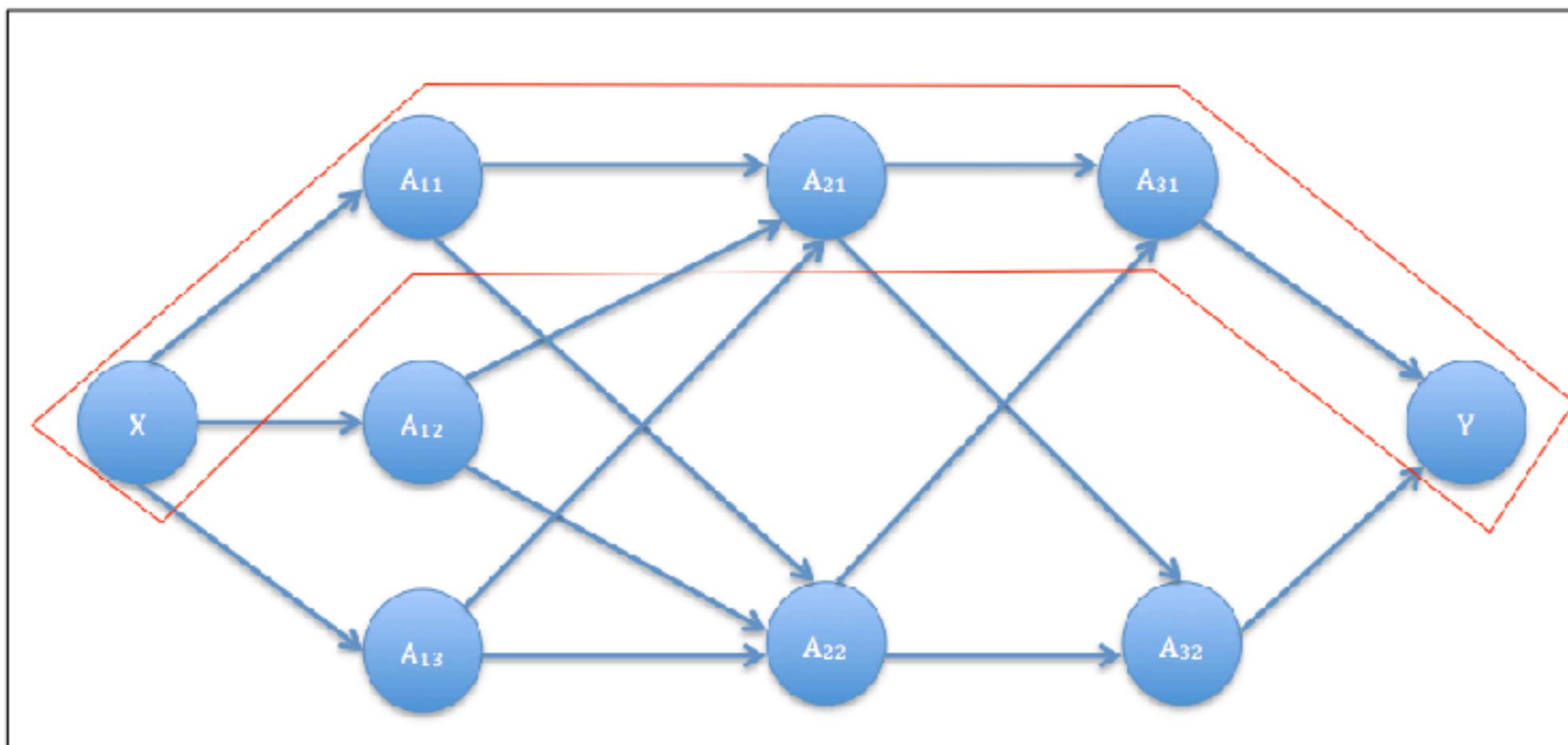
$$EC_{S_i}^* = \frac{1}{|S_i|} \sum_{z=1}^{|S_i|} E_{A_{iz}}^* - E^*,$$



Error contribution from algorithms using the *agnostic* methodology

For, $i = 1, \dots, n, j = 1, \dots, |\theta_{ij}|$, $|\theta_{ij}|$ represents the number of hyperparameter configurations of A_{ij} , $E_{A_{ij}}^z$ * is the minimum error obtained with the z -th configuration of θ_{ij} and $E_{A_{ij}^p}^*$ is the minimum error found over the path p that consists of algorithm A_{ij} .

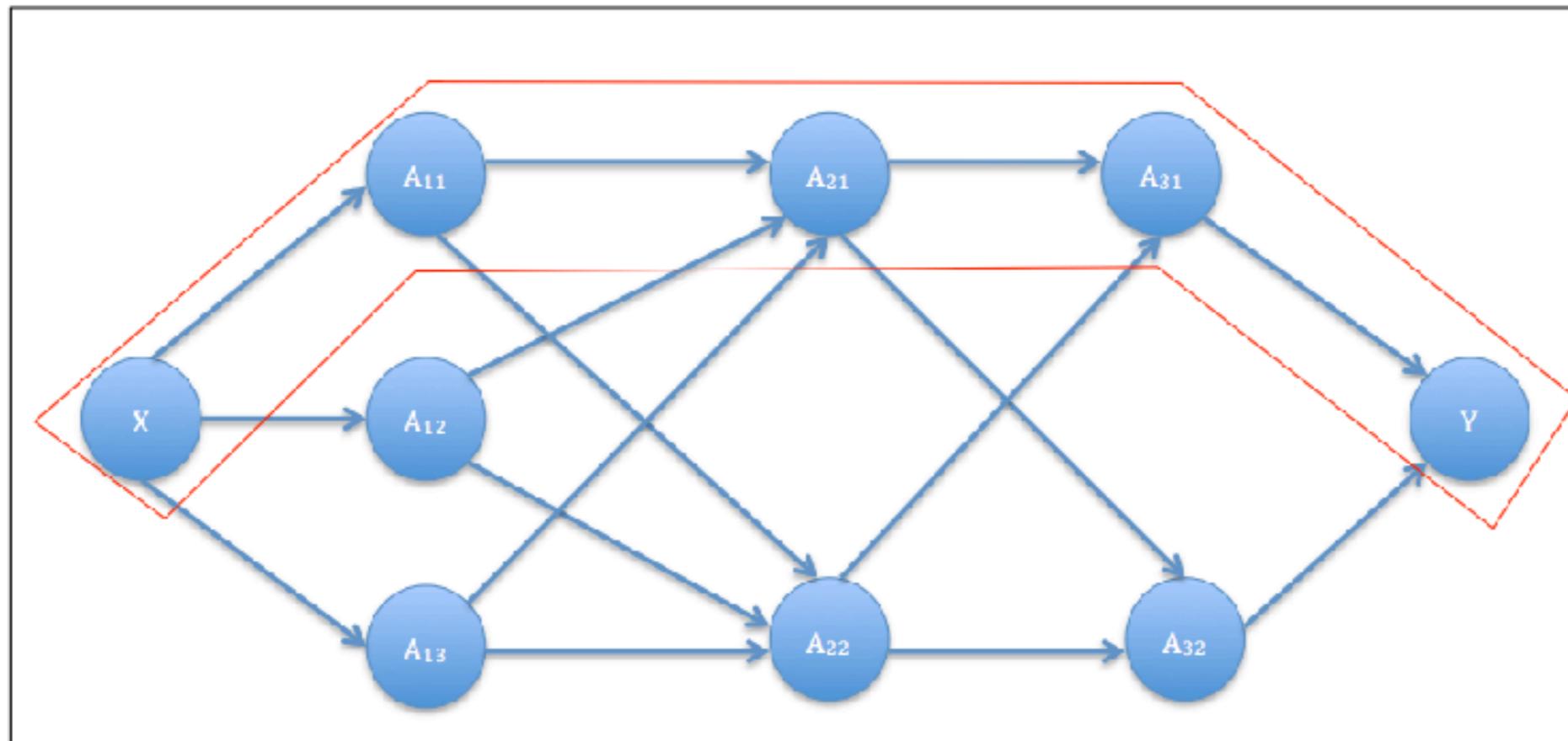
$$EC_{A_{ij}}^* = \frac{1}{|\theta_{ij}|} \sum_{z=1}^{|\theta_{ij}|} E_{A_{ij}}^z - E_{A_{ij}^p}^*,$$



Error contribution from hyper parameters using the *agnostic* methodology

For, $i = 1, \dots, n, j = 1, \dots, |\theta_{ij}|$, $k = \text{number of hyper-parameters of algorithm } A_{ij}$. $|\theta_{ijk}|$ represents the number of configurations of θ_{ijk} , $E_{\theta_{ijk}}^z$ * is the minimum error obtained with the z -th configuration of θ_{ijk} and $E_{A_{ij}^p}^*$ is the minimum error found over the path p that consists of algorithm A_{ij} .

$$EC_{\theta_{ijk}}^* = \frac{1}{|\theta_{ijk}|} \sum_{z=1}^{|\theta_{ijk}|} E_{\theta_{ijk}}^z - E_{A_{ij}^p}^*,$$



Experimental settings

Pipeline

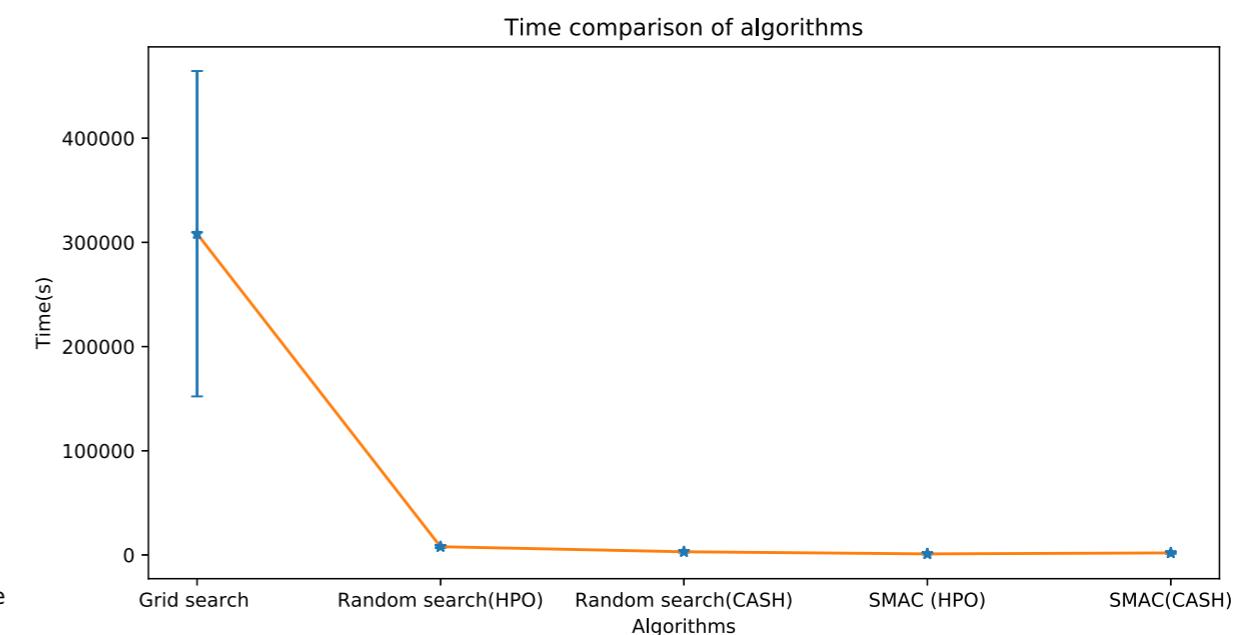
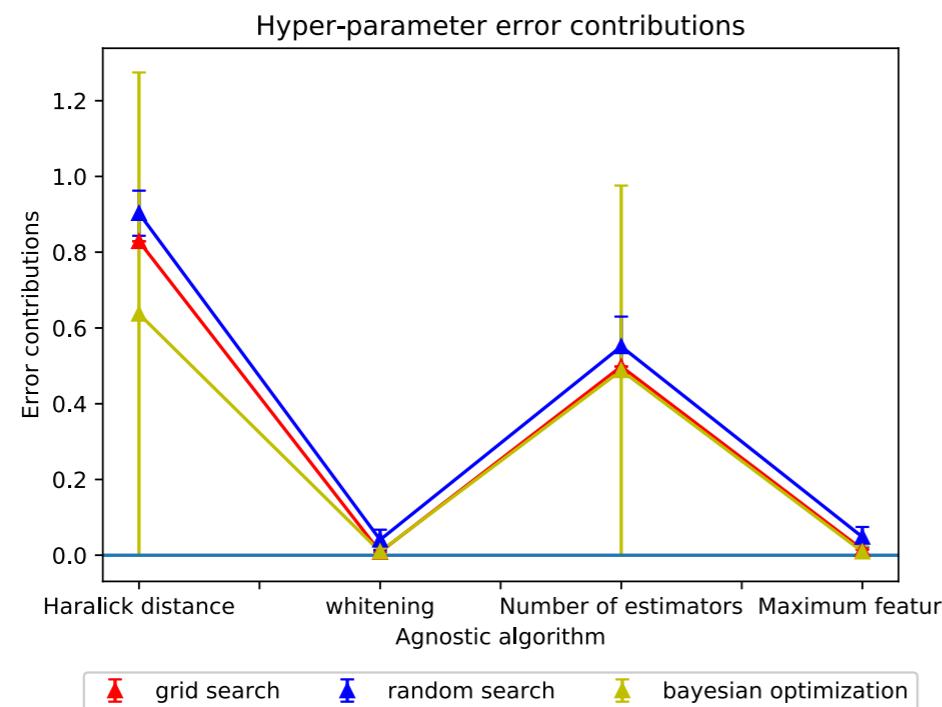
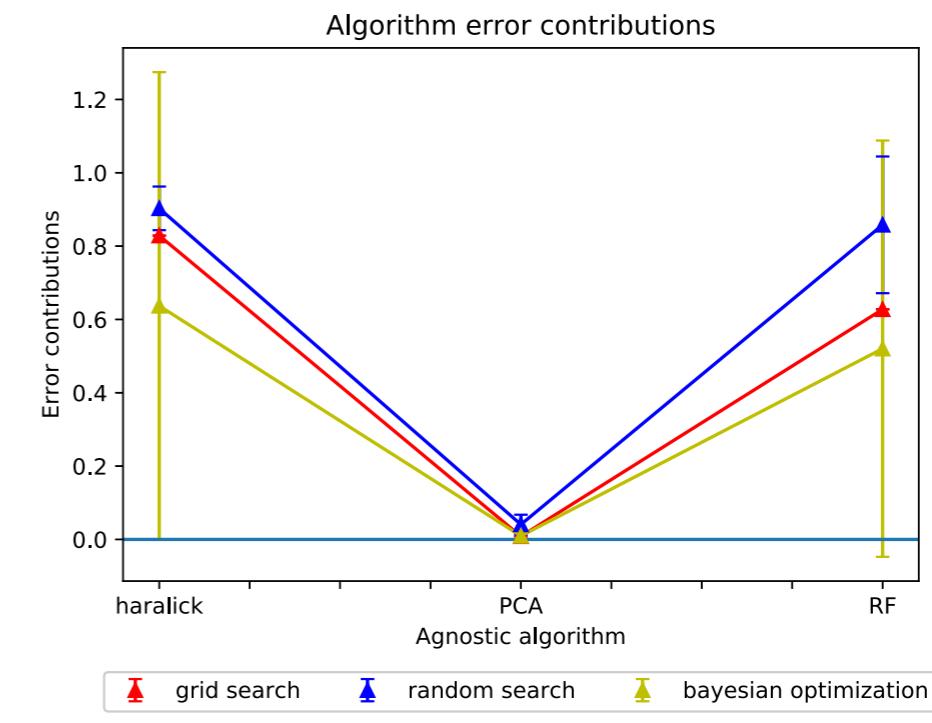
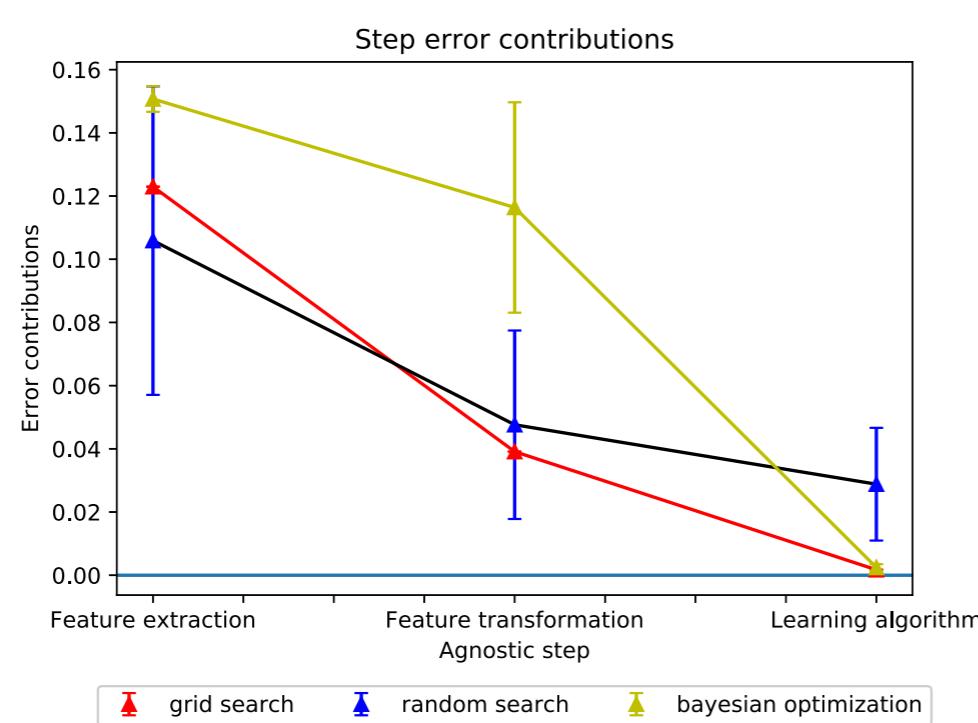
Step	$A_{ij}(\theta_{ij})$	Definition
Feature extraction	$A_{11}(\theta_{11})$	Haralick texture features (<i>distance</i>)
	$A_{12}(\theta_{12})$	Pre-trained CNN trained on ImageNet database with VGG16 network
	$A_{13}(\theta_{13})$	Pre-trained CNN trained on ImageNet database with Inception network
Feature transformation	$A_{21}(\theta_{21})$	PCA (<i>whitening</i>)
	$A_{22}(\theta_{22})$	ISOMAP (<i>number of neighbors, number of components</i>)
Learning algorithms	$A_{31}(\theta_{31})$	Random forests (<i>number of trees, maximum features</i>)
	$A_{32}(\theta_{32})$	SVM (C, γ)

Datasets

Dataset (notation)	Distribution of classes
Breast cancer (<i>breast</i>)	<i>benign</i> : 151, <i>in-situ</i> : 93, <i>invasive</i> : 202
Brain cancer (<i>brain</i>)	<i>glioma</i> : 16, <i>healthy</i> : 210, <i>inflammation</i> : 107
Material science 1 (<i>matsc1</i>)	<i>dendrites</i> : 441, <i>non-dendrites</i> : 132
Material science 2 (<i>matsc2</i>)	<i>transverse</i> : 393, <i>longitudinal</i> : 48

Random search quantifies the error contribution from steps, algorithms and hyper-parameters accurately and efficiently

Random search is able to quantify the error contributions from the components of the pipeline accurately and efficiently



Discussion

- We propose a method to quantify the contributions of components in an image classification pipeline in terms of the error.
- HPO and CASH methods maybe used to quantify error contribution and importance of components (steps, algorithms and hyper-parameters)
- Random search is able to quantify the contributions accurately and efficiently based on the results.

Conclusion

- The error observed in image classification pipelines is due to the components of the pipeline and not just the error due to the learning algorithm.
- The error maybe reduced or minimized by :
 - Modifying or optimizing the individual components of the pipeline.
 - Modifying or optimizing the components of the pipeline as a whole.
- The quality of the image classification pipeline can be estimated by :
 - Quantifying the quality of the data using a machine learning based score
 - Quantifying the contributions of the components of the pipeline (steps, algorithms and hyper-parameters)

Future work

- Parametric 3D models can be used to redress data imbalance in other domains.
- Exhaustive grid search and other CASH or HPO methods maybe used for optimizing image classification pipelines in other domains.
- The Quality of Image (QoI) score maybe used to filter datasets into *good* data and *bad* data using a data driven approach.
- The *agnostic* methodology maybe used for quantifying the error contributions in end-to-end learning frameworks for images and other sources of data.

References

- Chowdhury, Aritra, et al. "Blood vessel characterization using virtual 3D models and convolutional neural networks in fluorescence microscopy." *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017.
- Chowdhury, Aritra, et al. "Image driven machine learning methods for microstructure recognition." *Computational Materials Science* 123 (2016): 176-187.
- Chowdhury, Aritra, et al. "A machine learning approach to quantifying noise in medical images." *Medical Imaging 2016: Digital Pathology*. Vol. 9791. International Society for Optics and Photonics, 2016.
- Chowdhury, Aritra, et al. "Algorithm selection and hyperparameter optimization based quantification of error contribution in image classification pipelines." *IEEE International Conference on Data Mining (ICDM) 2018* (Submitted)

References

- Navneet Dalal, Bill Triggs, Histograms of oriented gradients for human detection, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol. 1, IEEE, 2005, pp. 886–893.
- George H. Duntzman, Principal Components Analysis, vol. 69, Sage Publications, Inc., 1989.
- Richard G. Lomax, Debbie L. Hahs-Vaughn, Statistical Concepts: A second Course, Routledge, 2013.
- Mark A. Hall, Correlation-based Feature Selection for Machine Learning PhD Thesis, The University of Waikato, 1999.

References

- Ron Kohavi, George H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1) (1997) 273–324.
- Huan Liu, Rudy Setiono, Chi2: feature selection and discretization of numeric attributes, in: *Proceedings of the Seventh International Conference on Tools with Artificial Intelligence*, 1995, pp. 388–391.
- Thierry Denoeux, A k-nearest neighbor classification rule based on Dempster–Shafer theory, *IEEE Trans. Syst., Man Cybernet.* 25 (5) (1995) 804–813.
- A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

References

- J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large- scale hierarchical image database,” in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 248–255.
- R. M. Haralick, K. Shanmugam et al., “Textural features for image classification,” IEEE Transactions on systems, man, and cybernetics, no. 6, pp. 610–621, 1973.
- N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” Journal of artificial intelligence research, vol. 16, pp. 321–357, 2002.
- J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” Journal of Machine Learning Research, vol. 13, no. Feb, pp. 281–305, 2012.

References

- J. Snoek, H. Larochelle, and R. P. Adams, “Practical bayesian optimization of machine learning algorithms,” in Advances in neural information processing systems, 2012, pp. 2951–2959.