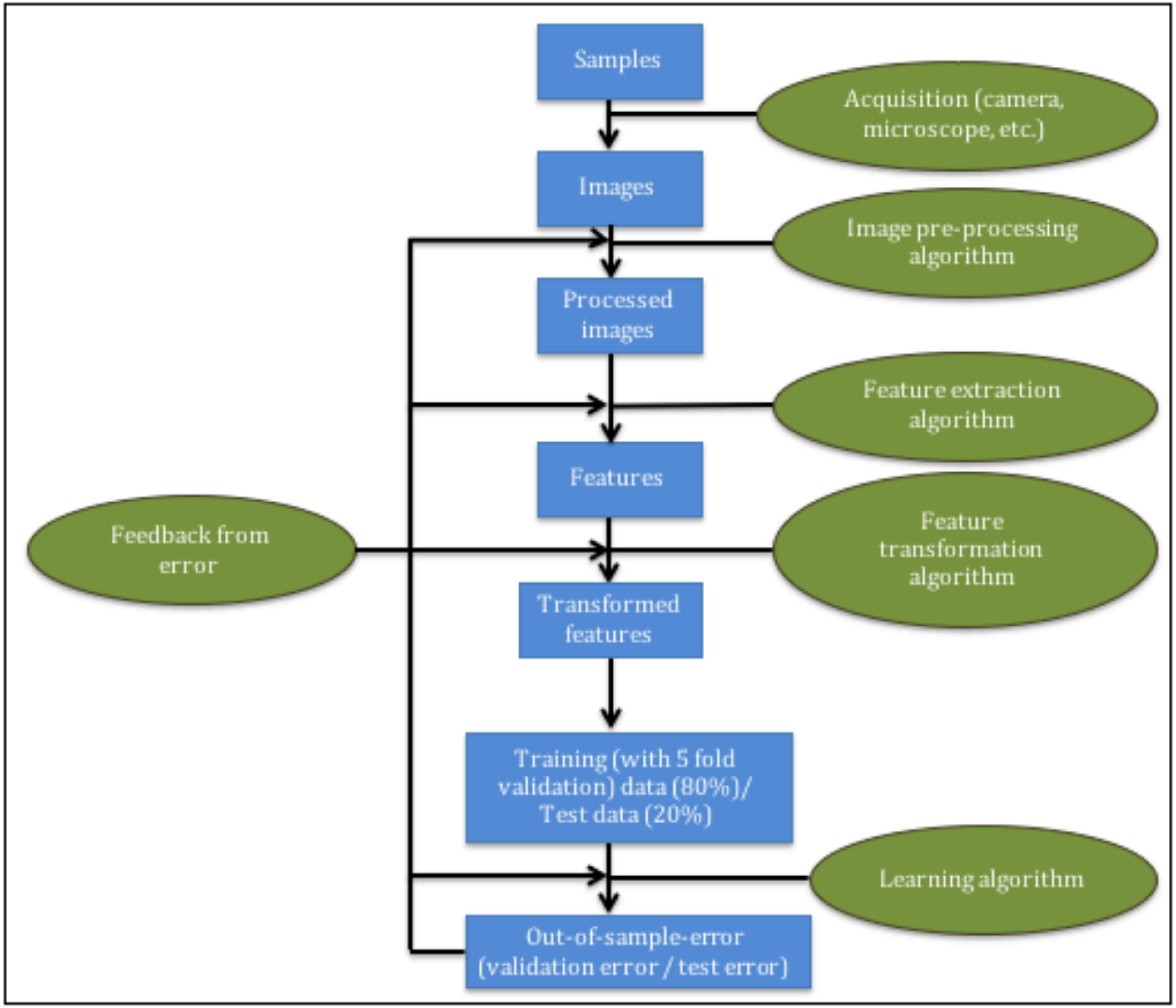
Reducing image classification error and quantifying the contributions of components in a learning pipeline

Outline

- Problem definition and motivation
- Our contributions
    1. Reduction of classification error
    2. Quantification of the contribution of components in a learning pipeline
- Conclusion and future work
- References

Image classification pipeline



Problem definition and motivation

- *Motivation*: Classification error occurs due to the accumulation of error from the combination of the components in the pipeline.

- *Problem*: Reduce the classification error and quantify the error contribution with respect to different components of an image classification pipeline

Outline

- Problem definition and motivation
- Our contributions
    1. Reduction of classification error
    2. Quantification of the contribution of components in a learning pipeline
- Conclusion and future work
- References

Our contributions

1. Reduction of classification error

    1. Local reduction of error by modification of a particular component of the pipeline [*Blood vessel characterization using virtual 3D models and convolutional neural networks in fluorescence microscopy (ISBI 2017)]*

    2. Global reduction of error by optimizing the pipeline as a whole [*Image driven machine learning methods for microstructure recognition (Journal of Computational Material Science)]*

2. Quantification of contribution from components of the pipeline

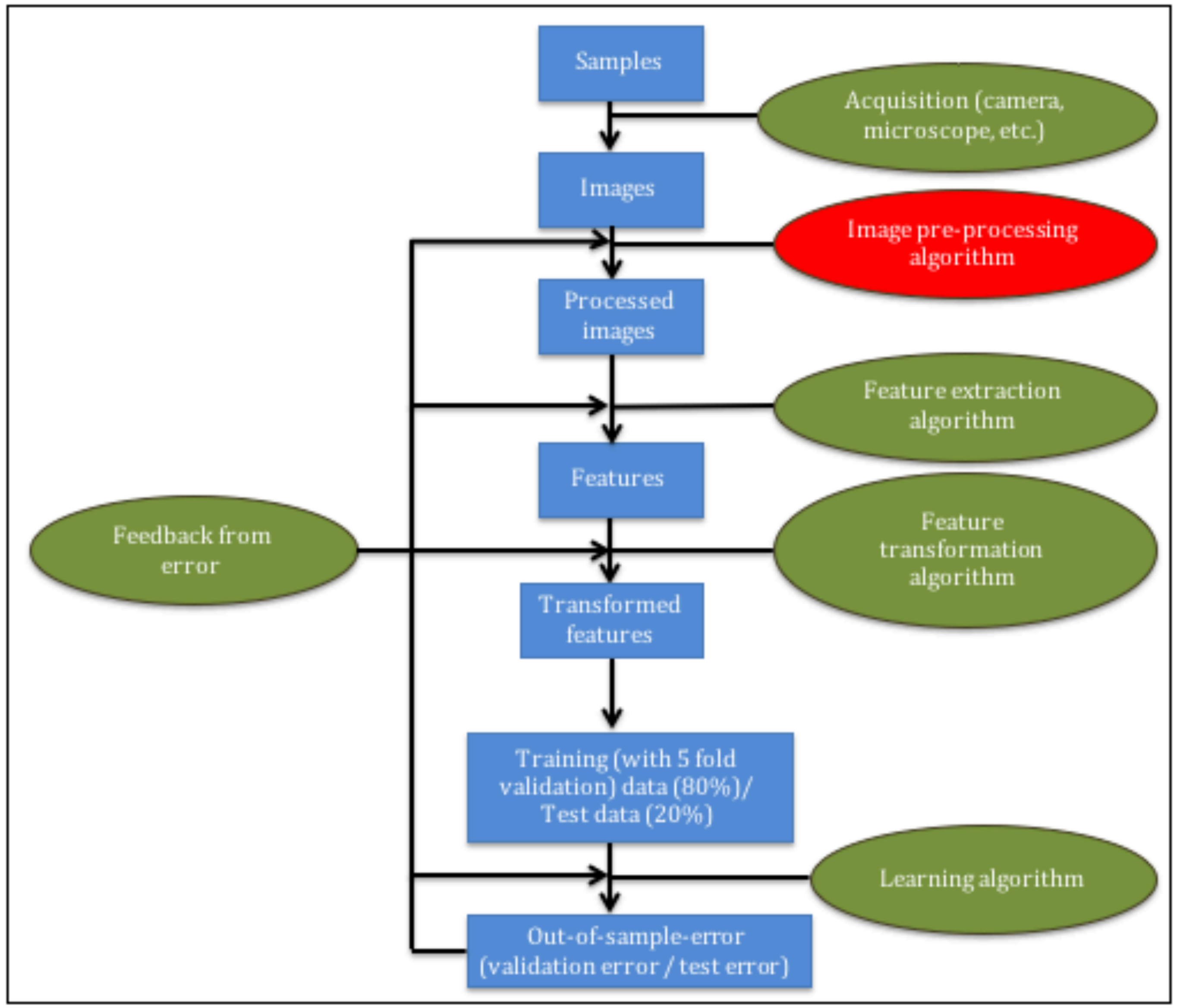
    1. Quantification of the contribution based on the data [*A machine learning based approach to quantifying noise in medical images (SPIE 2016)*]

    2. Quantification of contributions based on the components [*Quantifying error contributions of computational steps, algorithms and hyperparameter choices in image classification pipelines (Submitted to ICDM 2018)*]

Outline

- Problem definition and motivation
- Our contributions
    1. Reduction of classification error
    2. Quantification of the contribution of components in a learning pipeline
- Conclusion and future work
- References

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." *(ISBI 2017), 2017 IEEE 14th International Symposium on Biomedical Imaging*.

Introduction

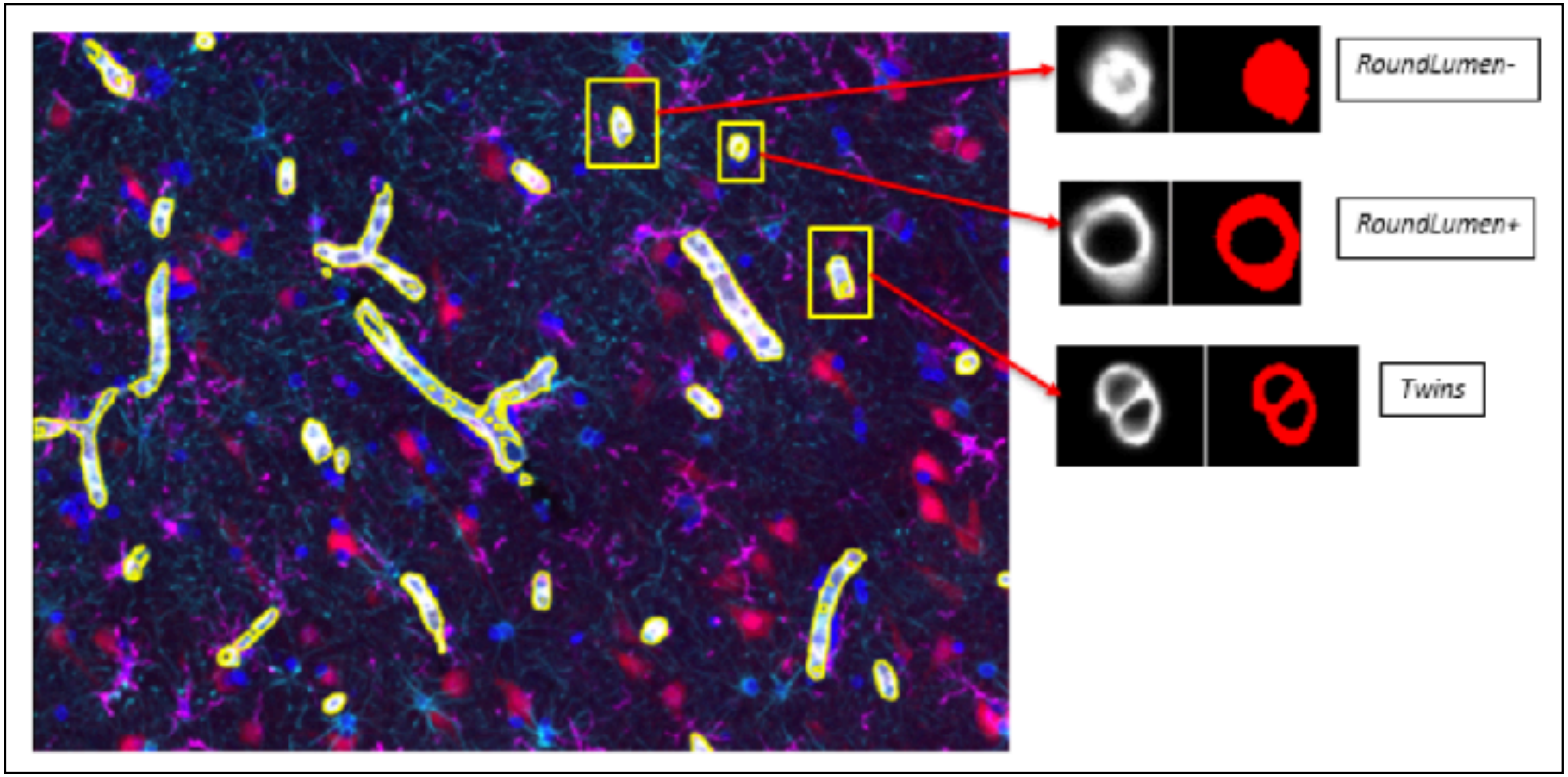
- Problem: Reduce classification error of blood vessel characterization by performing data augmentation using artificial parametric 3D models of vasculature.

- 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

Data

Depiction of the different morphologies in the data

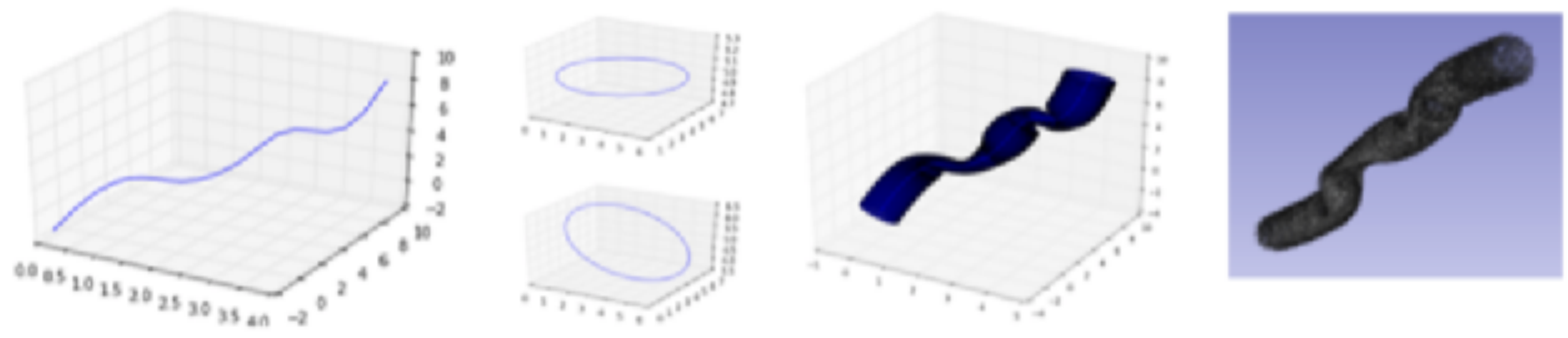


Distribution of the morphologies

RoundLumen-	689
Roundlumen+	3427
Twins	266
Total	4382

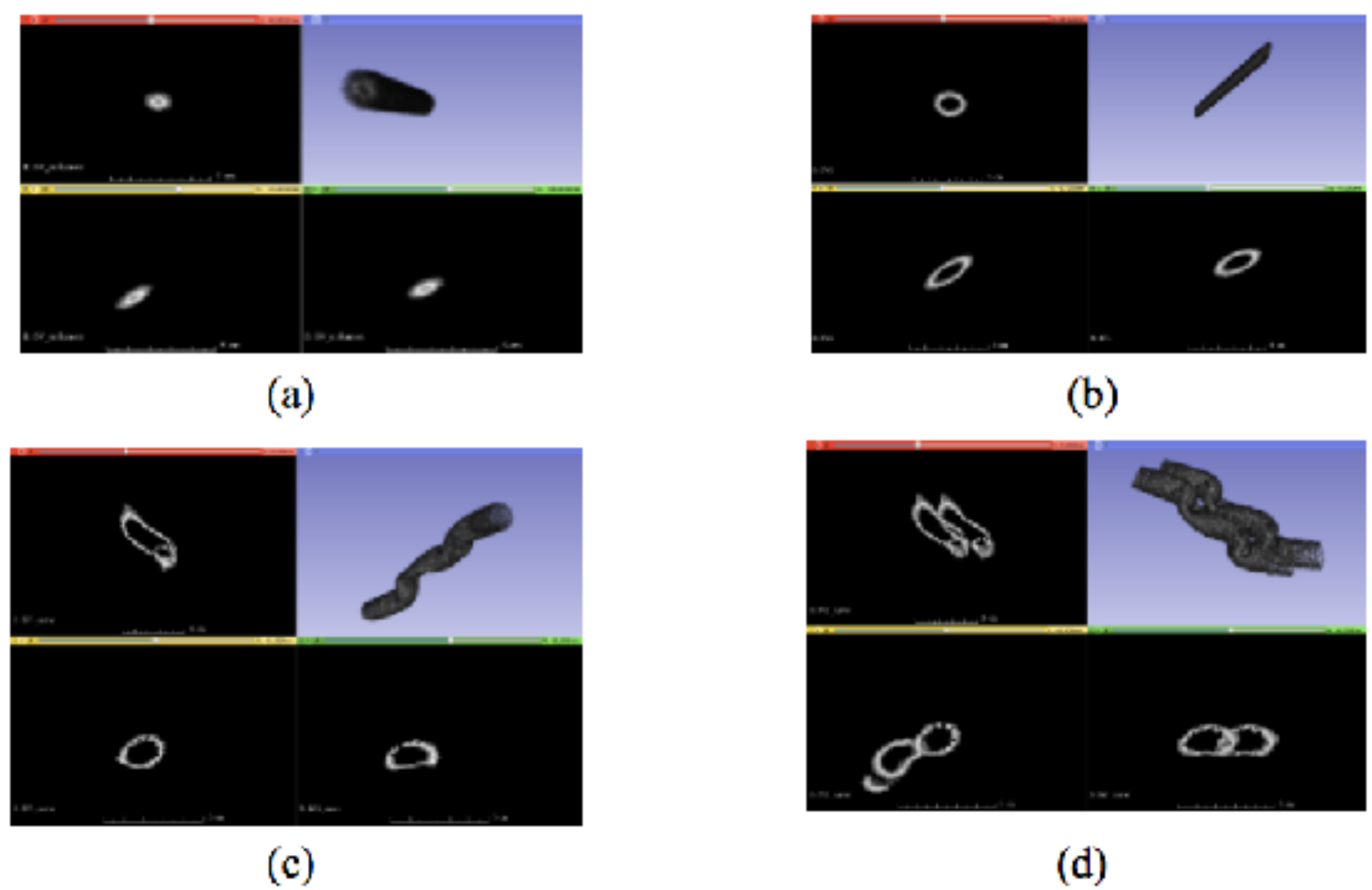
Artificial 3D model of vasculature

Development of the 3D virtual model



(a) (b) (c) (d)

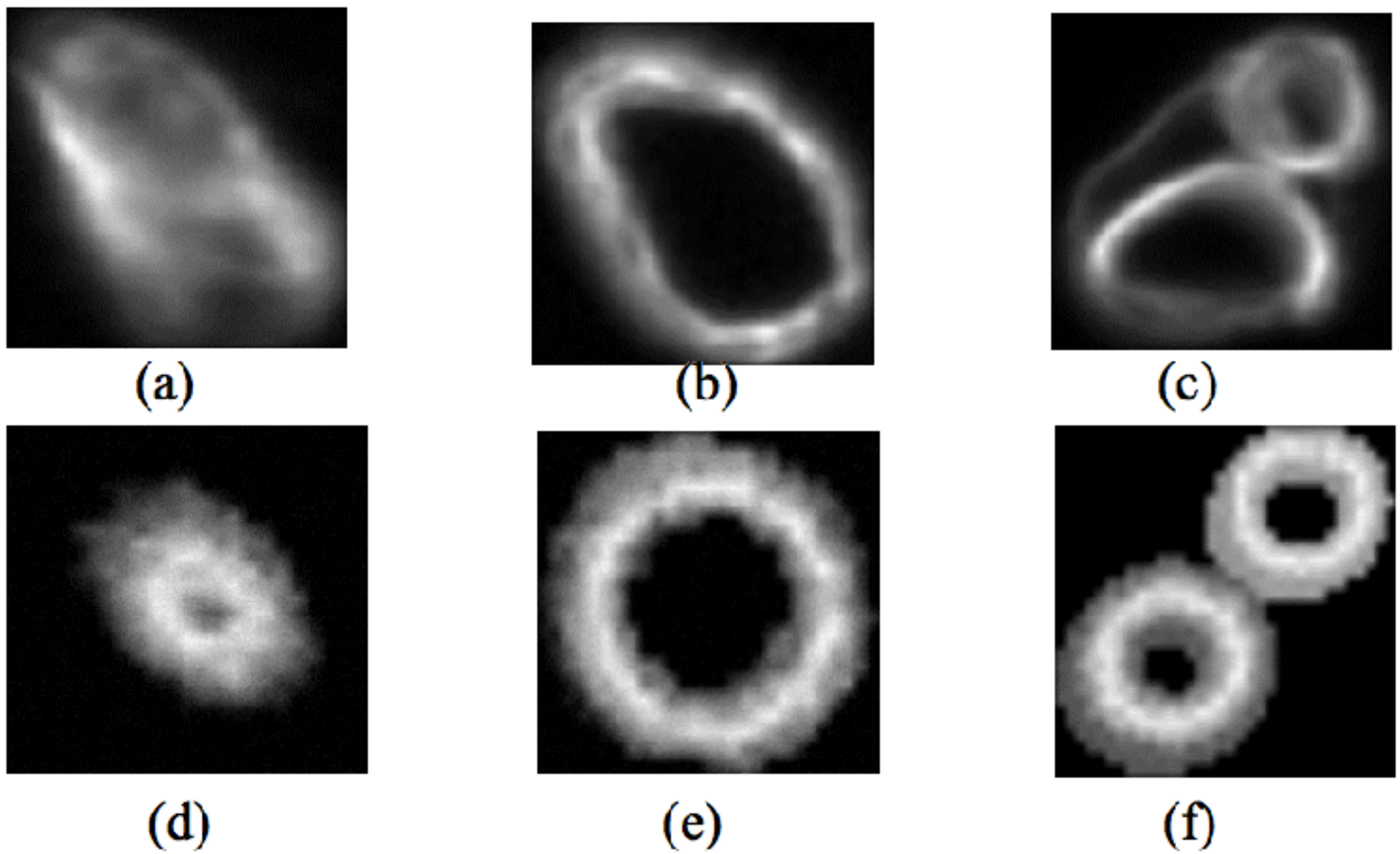
2D slices of the 3D model on 3D slicer



(a) (b) (c) (d)

Natural and artificial data

Examples of the vessel classes *RoundLumen-*, *RoundLumen+* and *Twins*



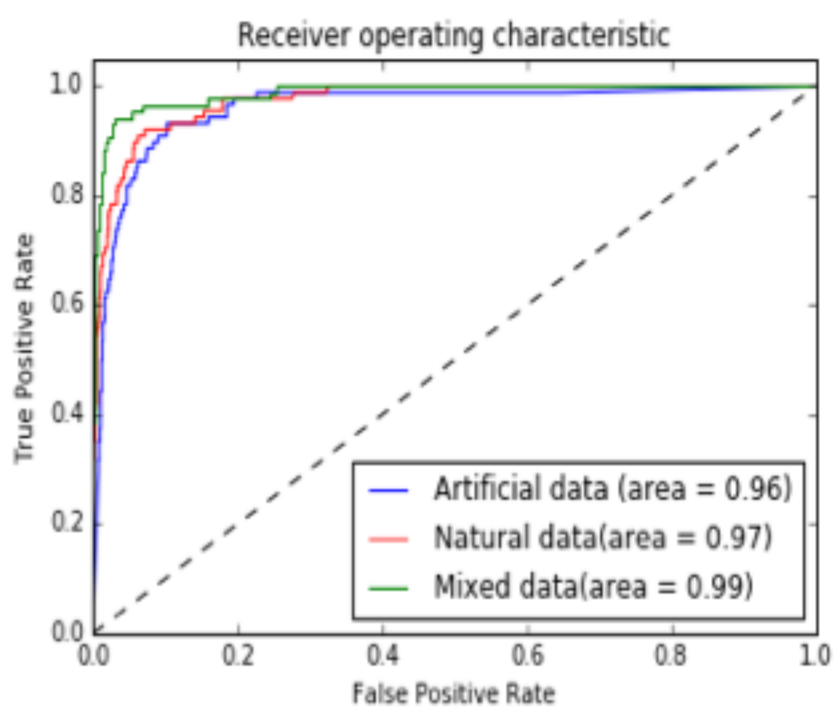
Results of task 1 (*RoundLumen vs Twins*) :

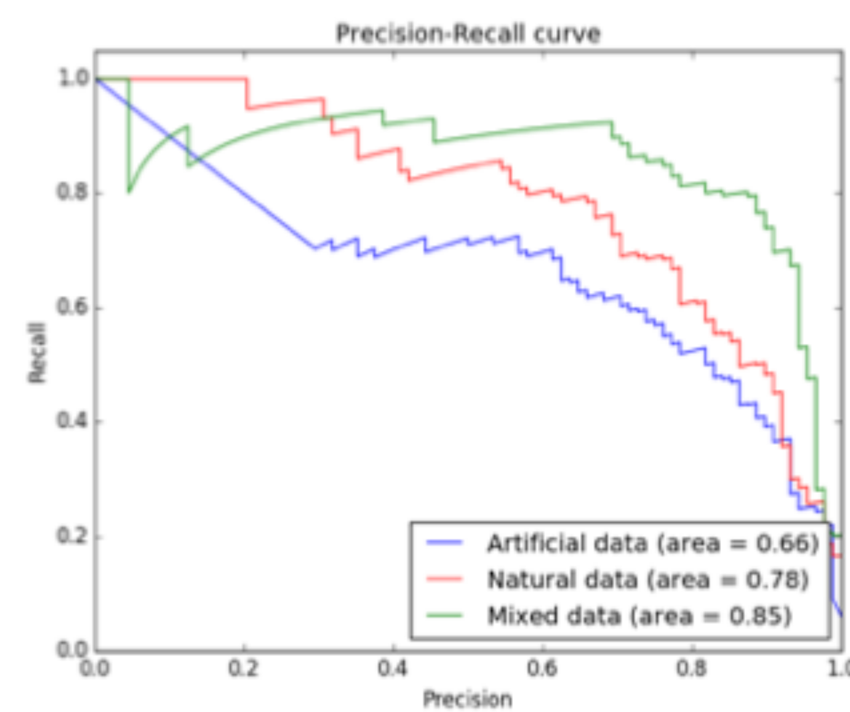
Mixed (natural + artificial) data performs the best in terms of classification

Classification metrics

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

ROC and PR curves





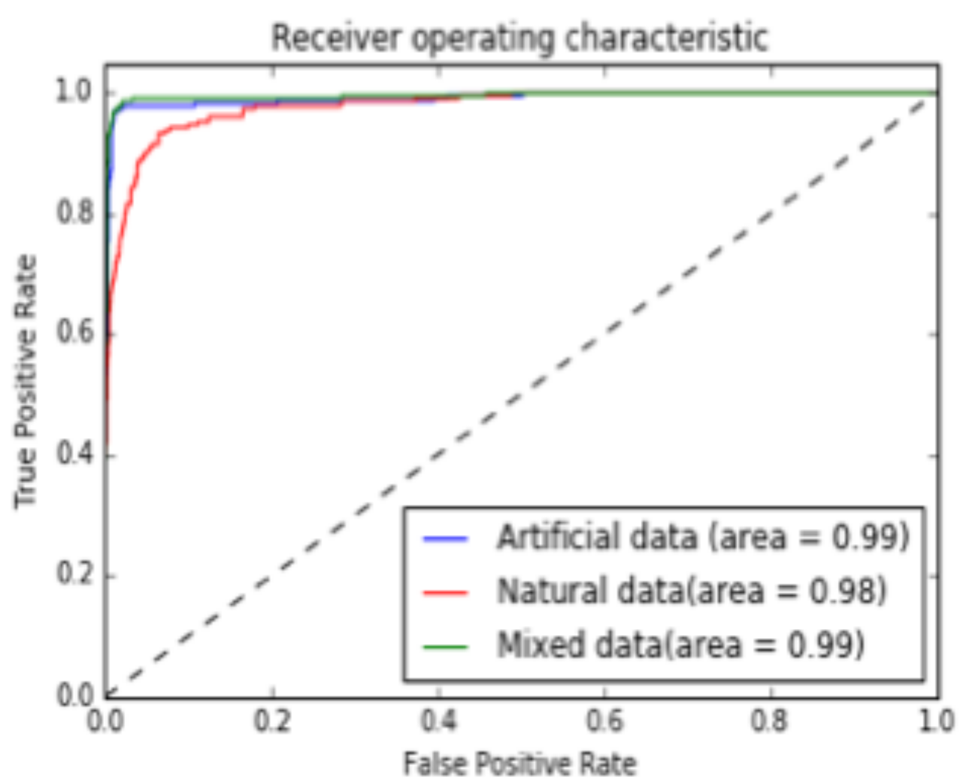
Results of task 2 (*RoundLumen-* vs *RoundLumen+*)

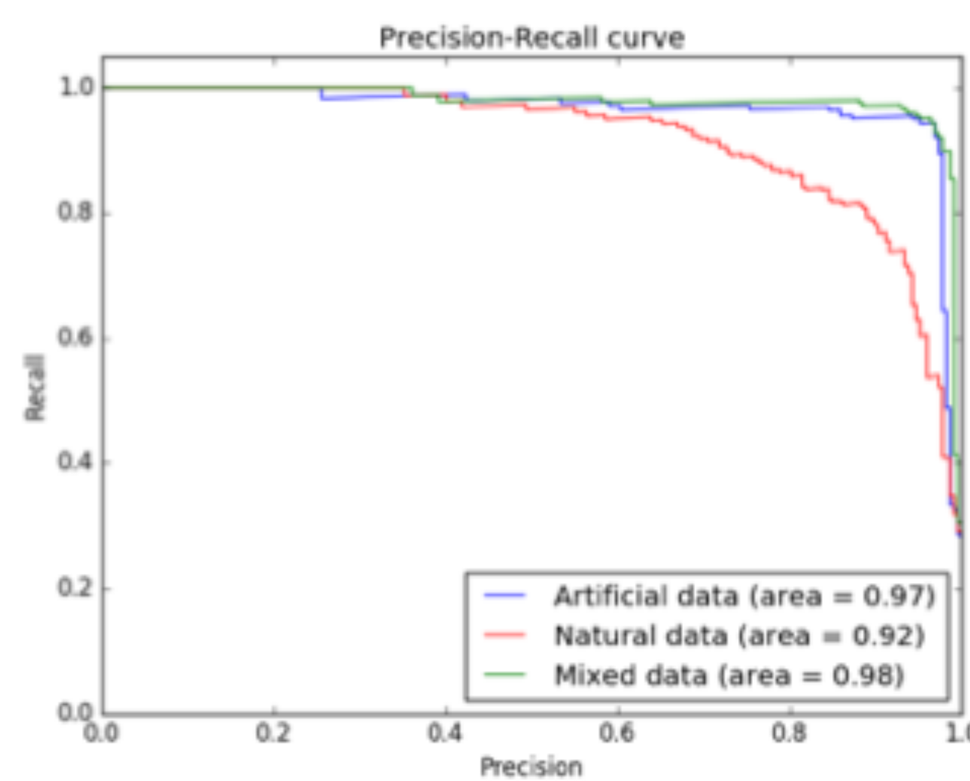
Mixed (natural + artificial) data performs the best in terms of classification

Classification metrics

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

ROC and PR curves



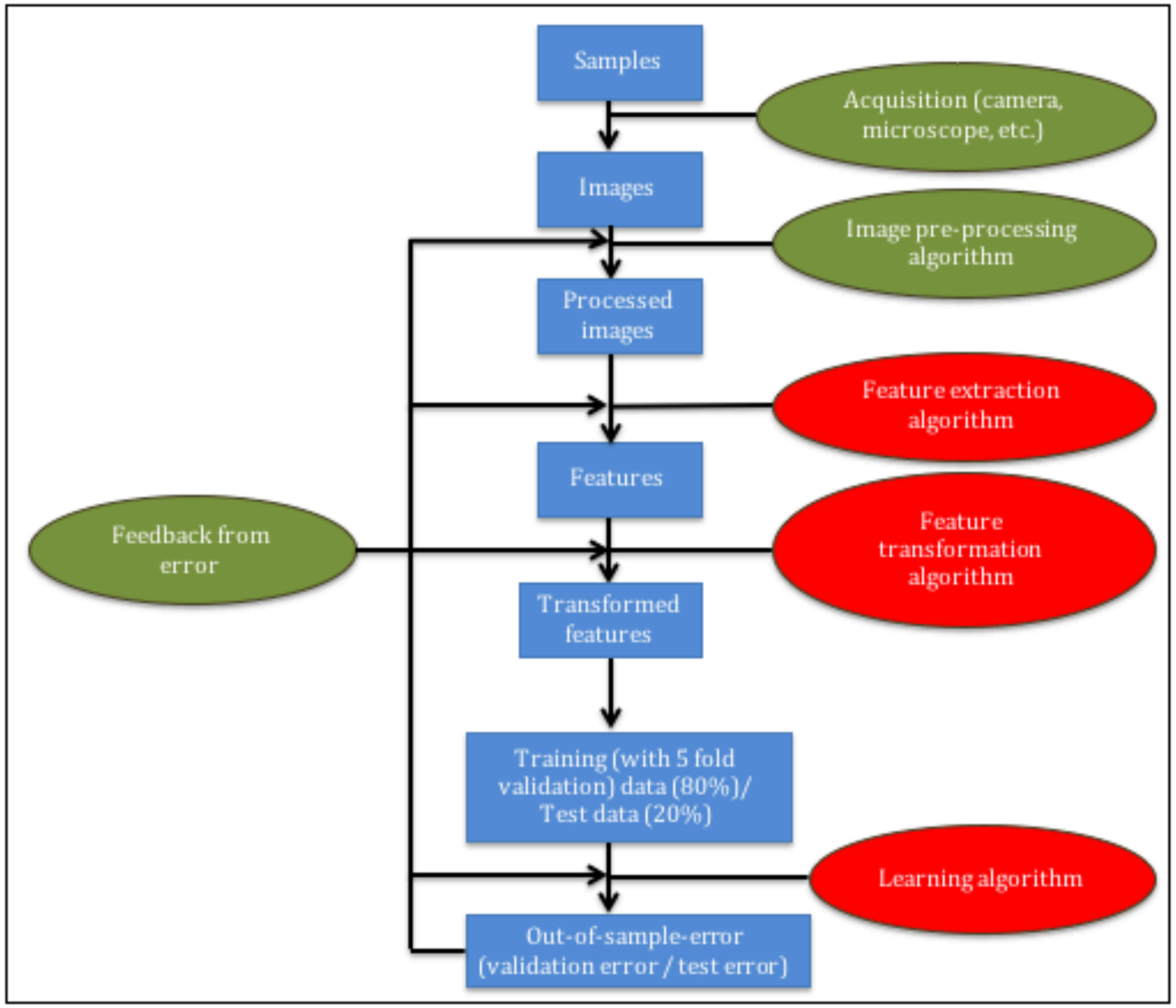


Discussion

- Mixture of natural and artificial data increases the classification F1-score by an average of 0.28 over the two classification tasks.

- Data augmentation using artificial parametric 3D models can be 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 (with highest accuracy) of algorithms to characterize microstructures.

- Two classification tasks: dendrites vs non-dendrites, longitudinal dendrites vs transverse dendrites.

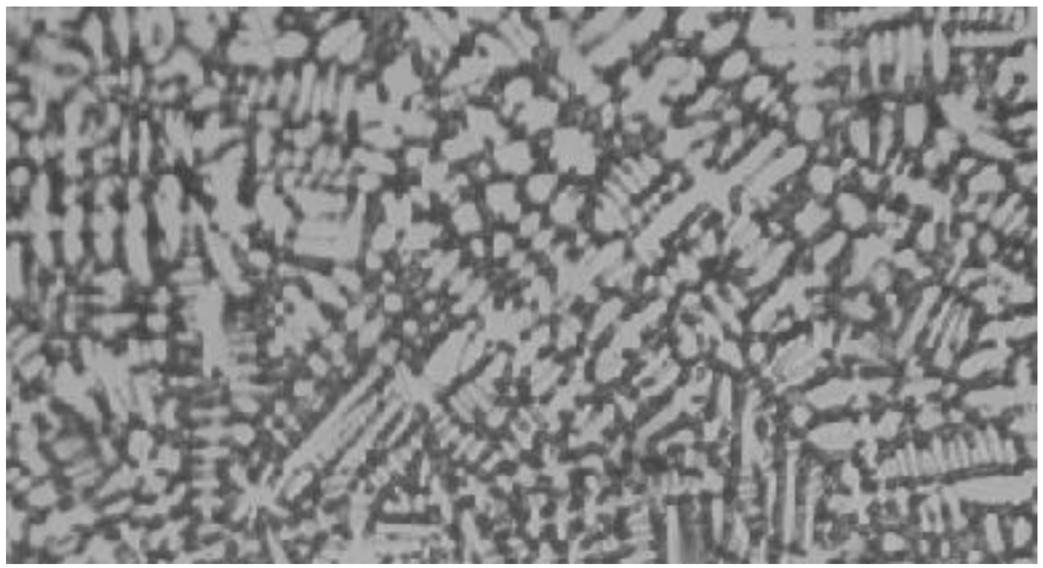
- Reduction of error in image classification pipeline as a whole by performing exhaustive grid search over an image classification pipeline.

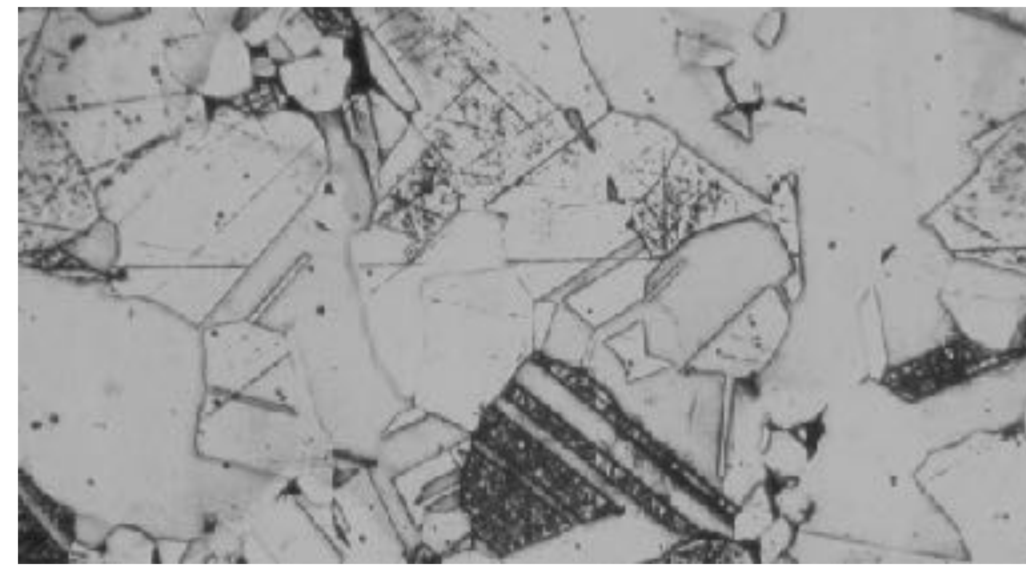
Classification tasks

Dendrite

Non-dendrite

Task 1 →

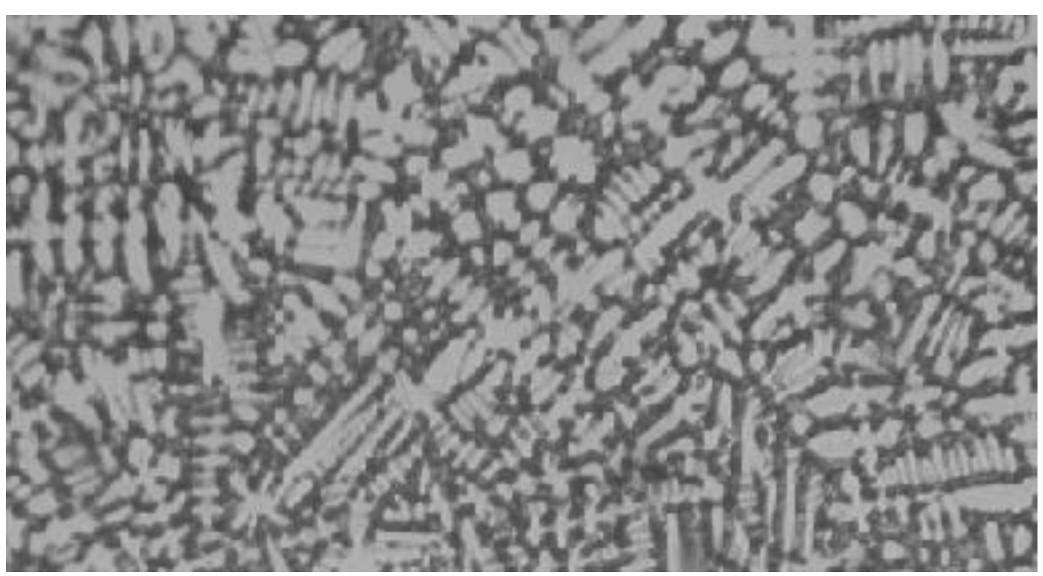




Longitudinal dendrite

Transverse dendrite

Task 2 →



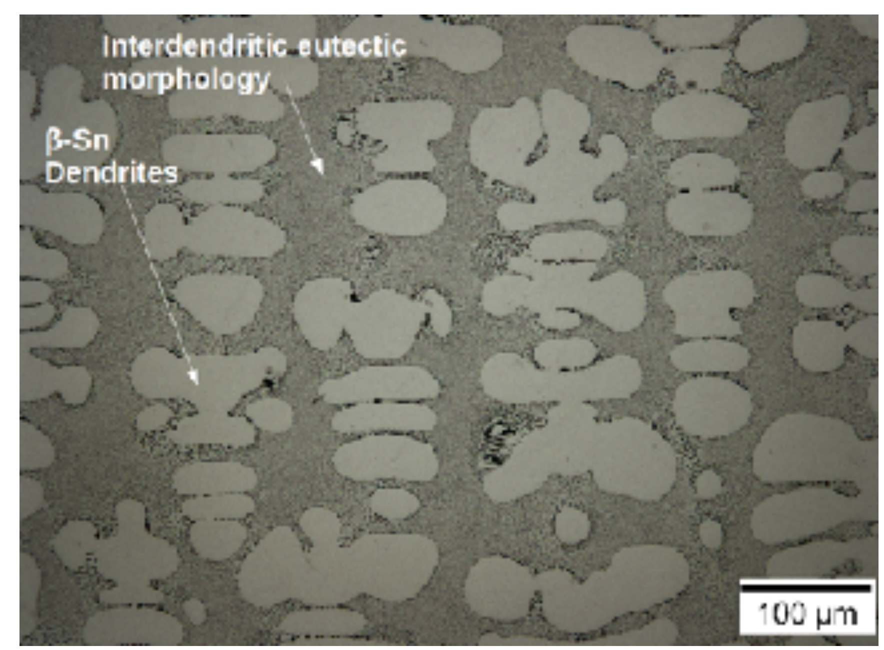
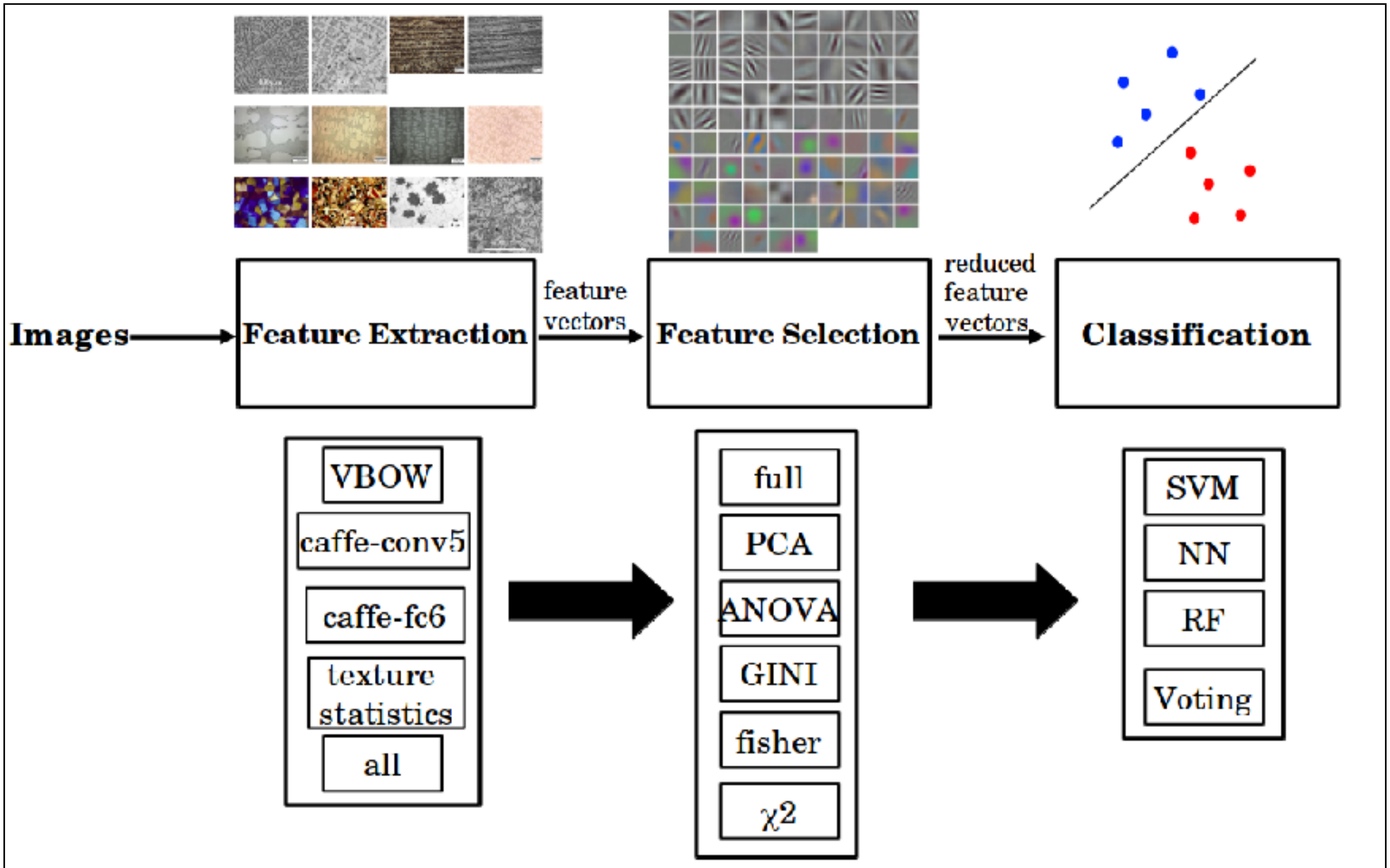


Image classification pipeline

Exhaustive grid search is performed over the following pipeline for the two classification tasks



Results of the exhaustive grid search experiments

Pre-trained CNNs are best able to characterize the microstructures

Best configuration for task 1 (*dendrites* vs *non-dendrites*)

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

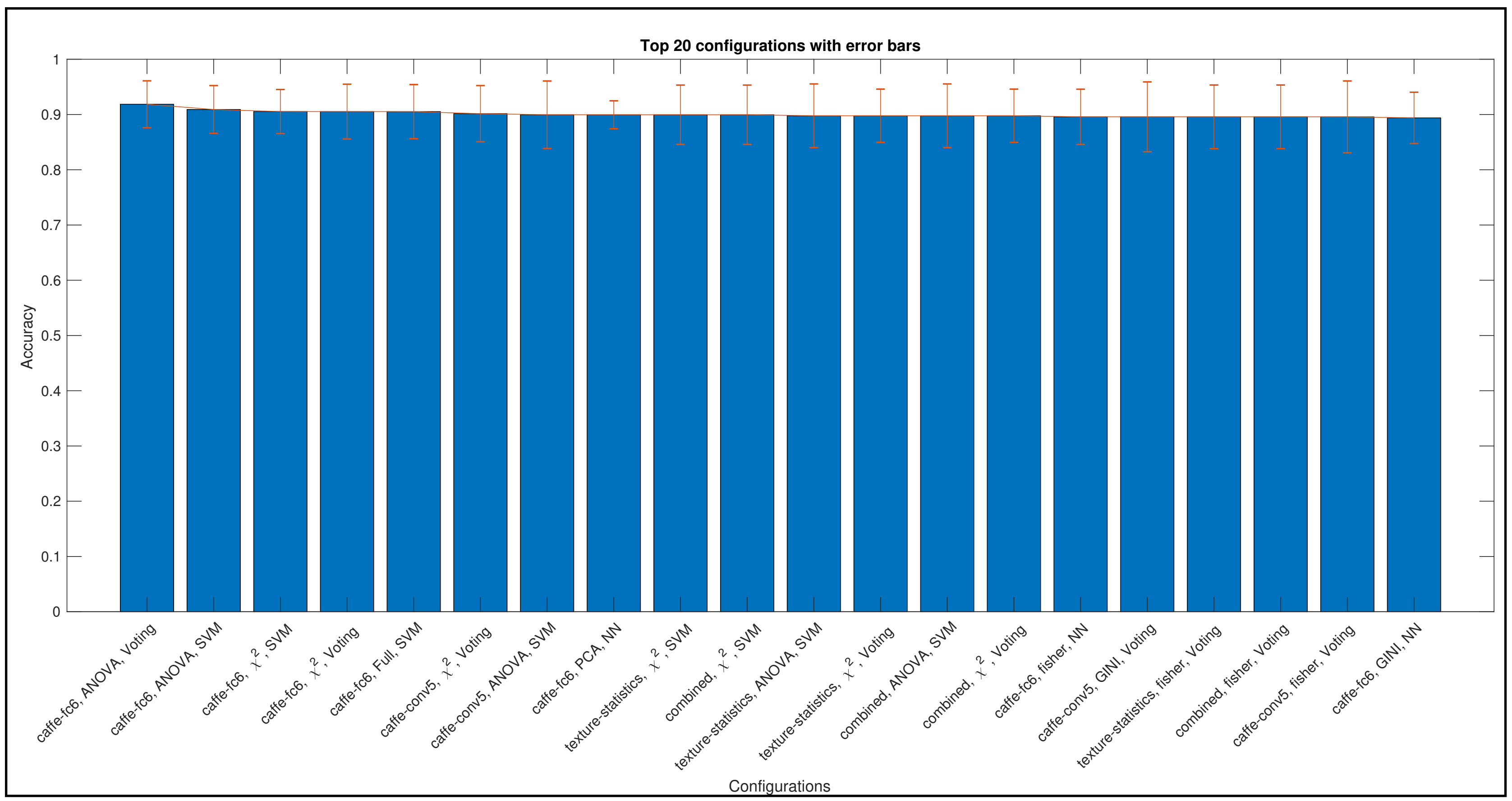
Best configuration for task 2 (*longitudinal* vs *transverse dendrites*)

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

Analysis of configurations of task 1

caffe-fc6 **is best able to distinguish between** ***dendritic*** **and** ***non-dendritic*** **microstructures**

Top 20 configurations with error bars



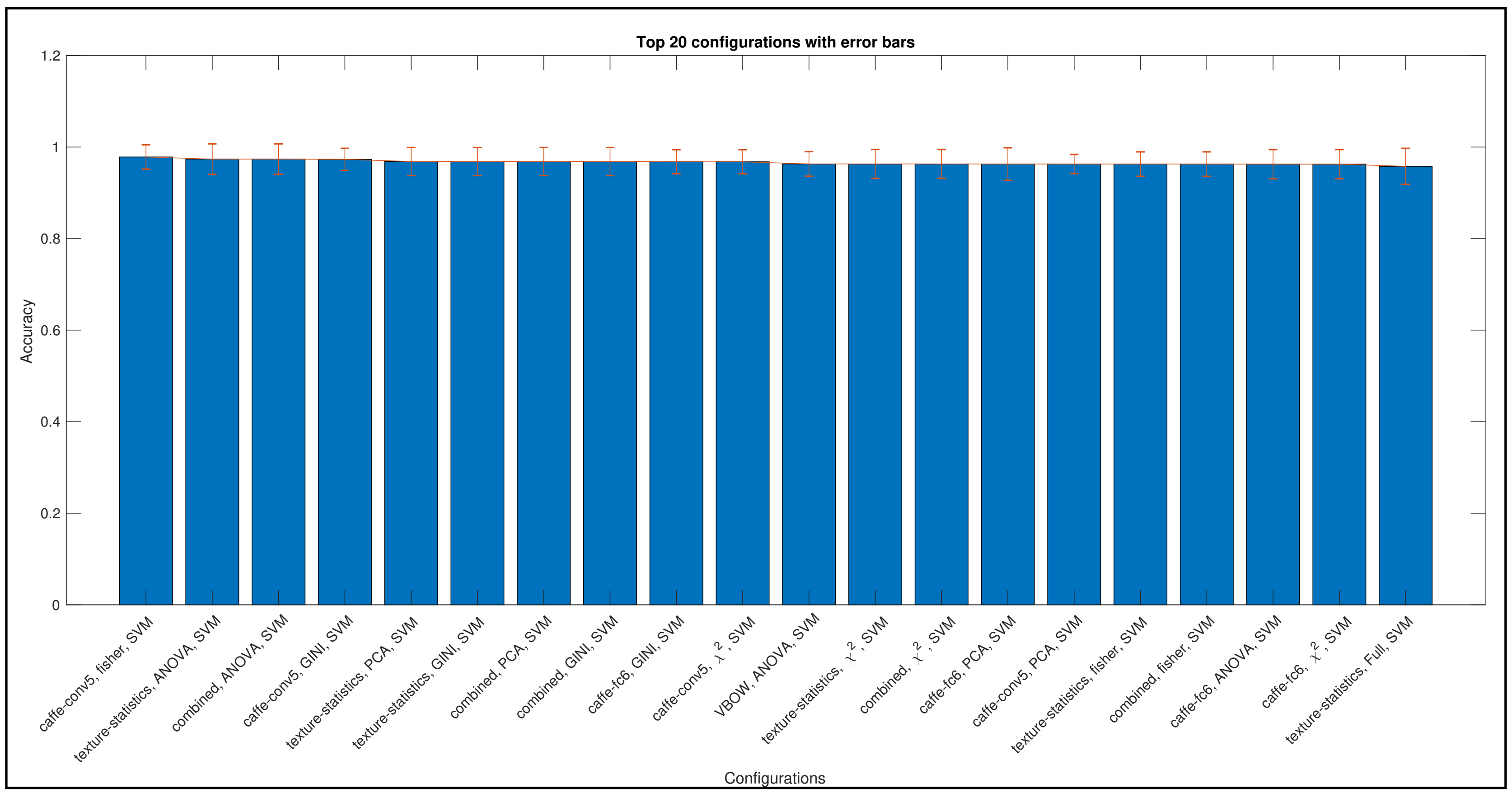
Average rank of algorithms

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

Analysis of configurations of task 2

***caffe-fc6* is best able to distinguish between *longitudinal* and *transverse* dendrites**

Top 20 configurations with error bars



Average rank of algorithms

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 optimizing the image classification pipeline as a whole using exhaustive grid search over algorithms and hyper-parameters.

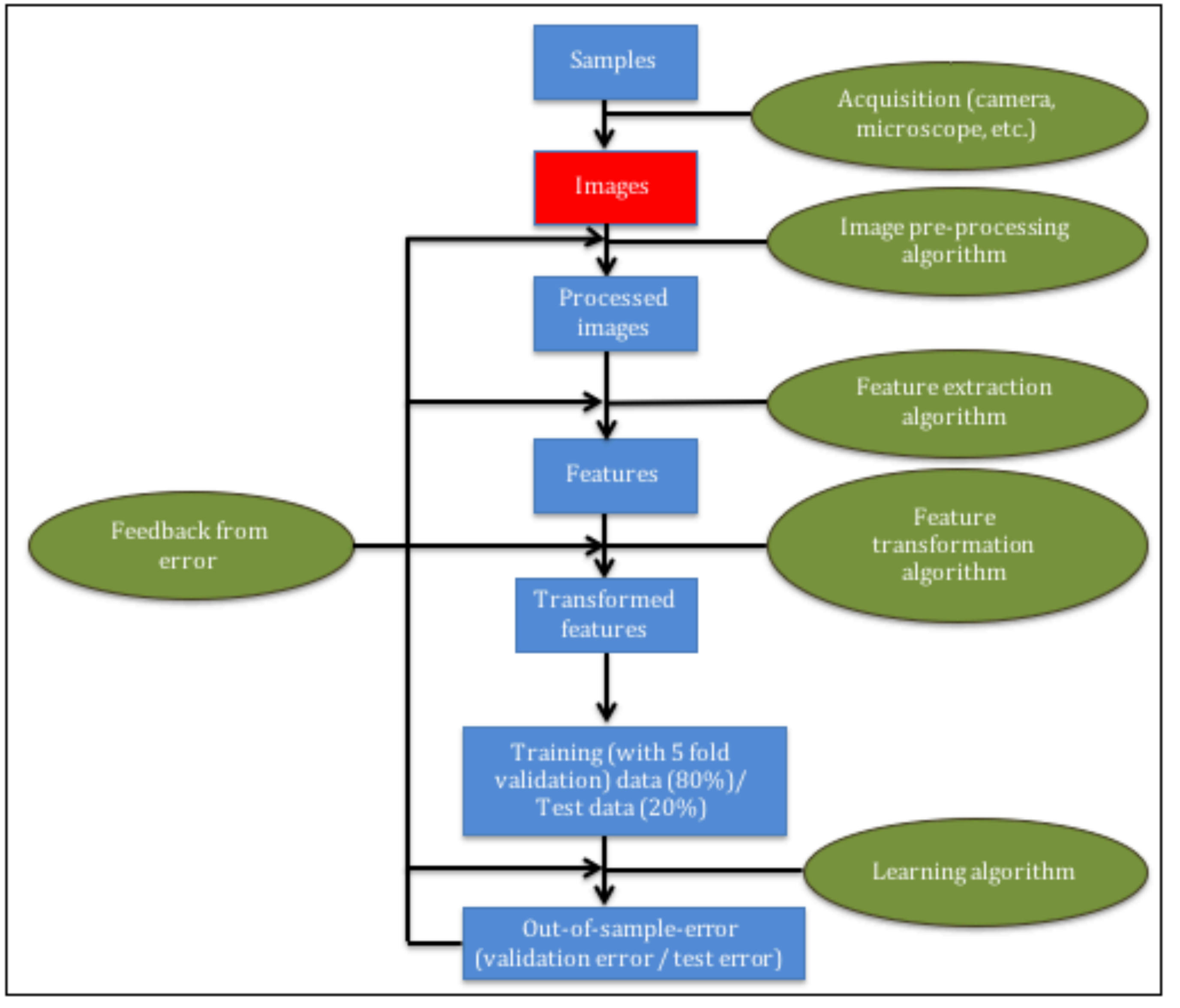
- Pre-trained neural networks (*caffe-fc6*) is able to best characterize and distinguish microstructural features.

- 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.

Outline

- Problem definition and motivation
- Our contributions
    1. Reduction of classification error
    2. Quantification of the contribution of components in a learning pipeline
- Conclusion and future work
- References

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

- Problem: Quantify the the quality of the data (individual images and dataset) or the contribution of the data to the pipeline.

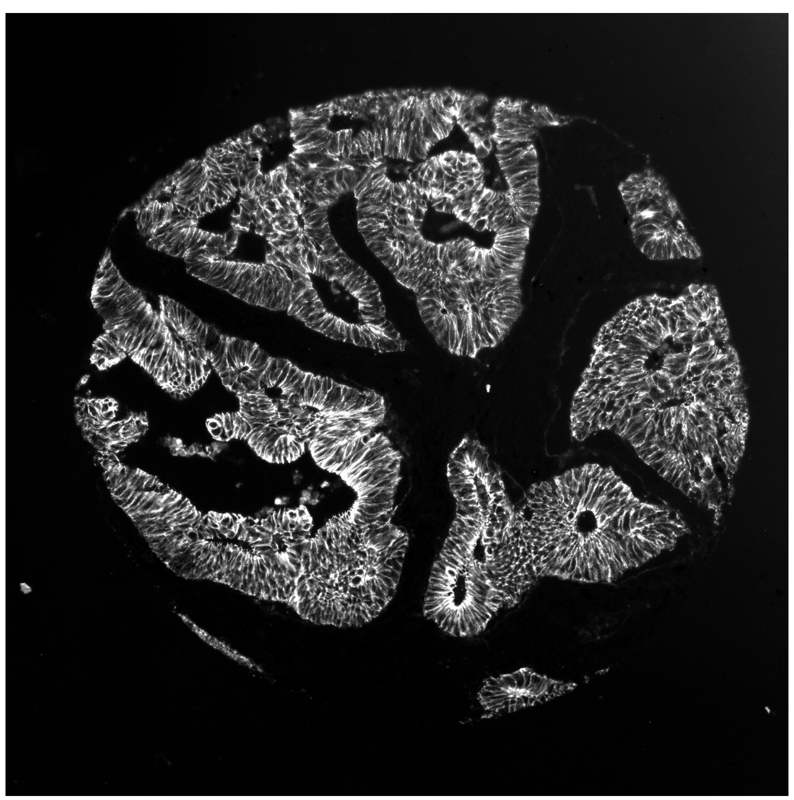
- A machine learning based score is used to quantify the quality of an image or dataset.

Data

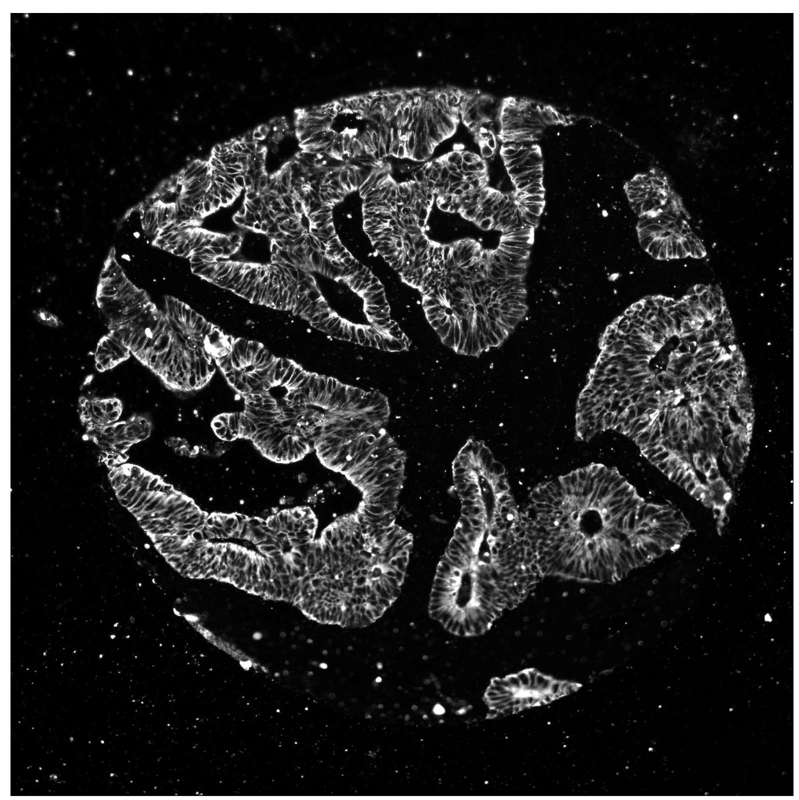
- Markers E_cad, CK15 and pck26 were used for this analysis.

- Images were annotated as *good* (high signal) or *bad* (low signal) by a pathologist.

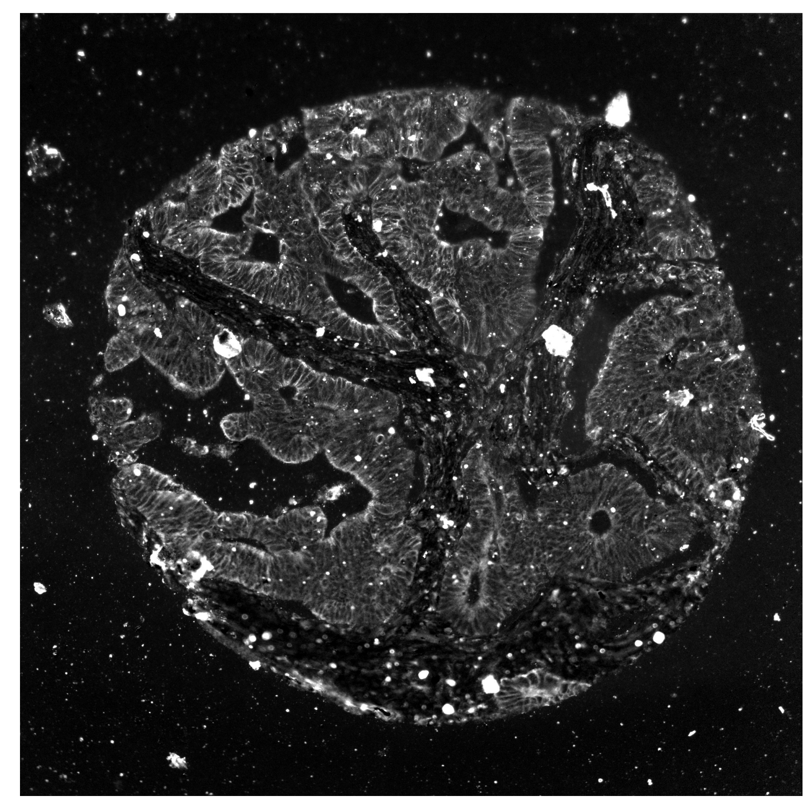
E_cad



pck26



CK15



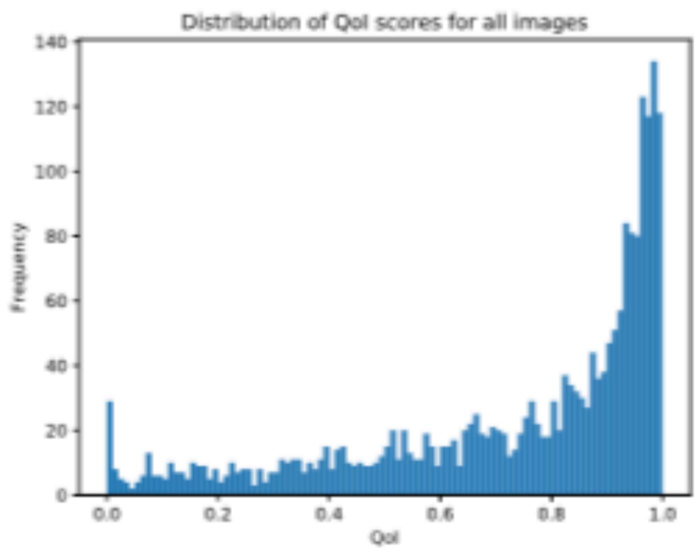
Methods

- Haralick texture features were used for feature extraction
- Synthetic minority oversampling technique (SMOTE) was used to perform data balancing.
- PCA was used to reduce the dimensionality of the dataset from 13 to 5 features (capturing 95% of the variance).
- Logistic regression was used because it learns the probabilities of classification directly.
- The *QoI* score is defined as the probability that an image is from the *good* class. It is given by the following equation

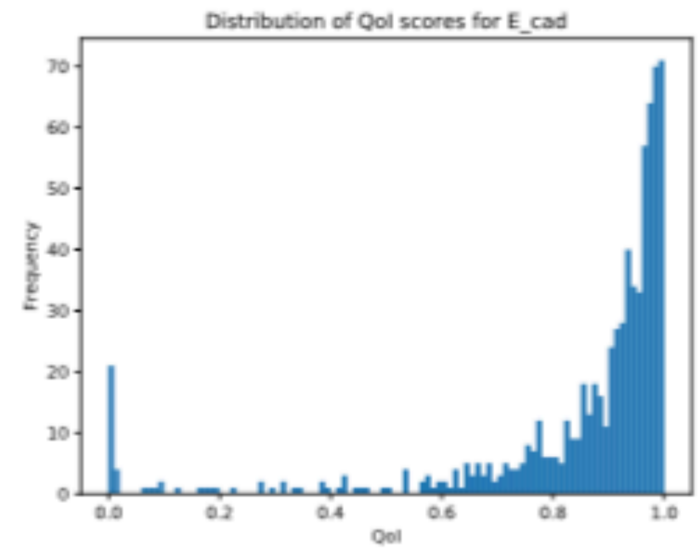
$$\boxed{S_i \;=\; p_{i1}}$$

Distribution of QoI scores for the markers

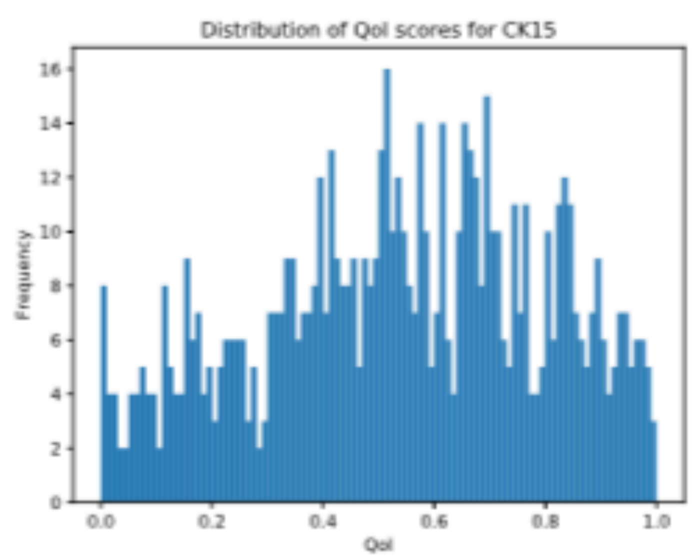
Distribution of the scores is able to quantify the perceived difference in quality between the E_cad, pck26 and CK15 markers with average QoI scores of 0.85, 0.82 and 0.54 respectively.



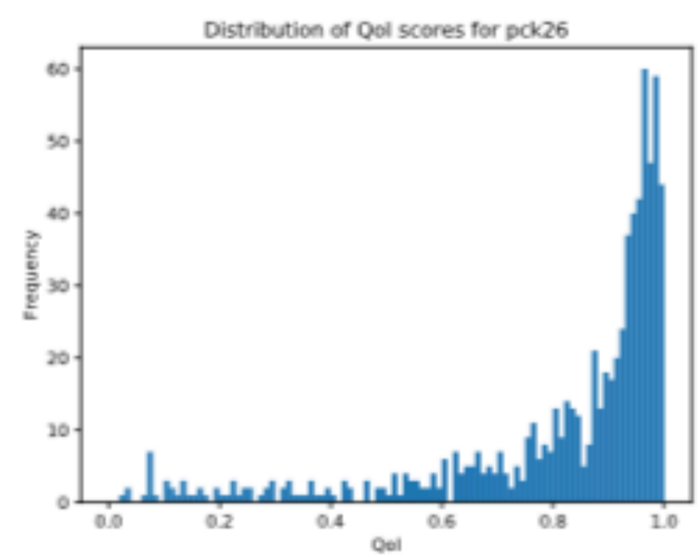
(a) Distribution of scores on all the images



(b) Distribution of scores from E_cad marker.



(c) Distribution of scores from CK15 marker.

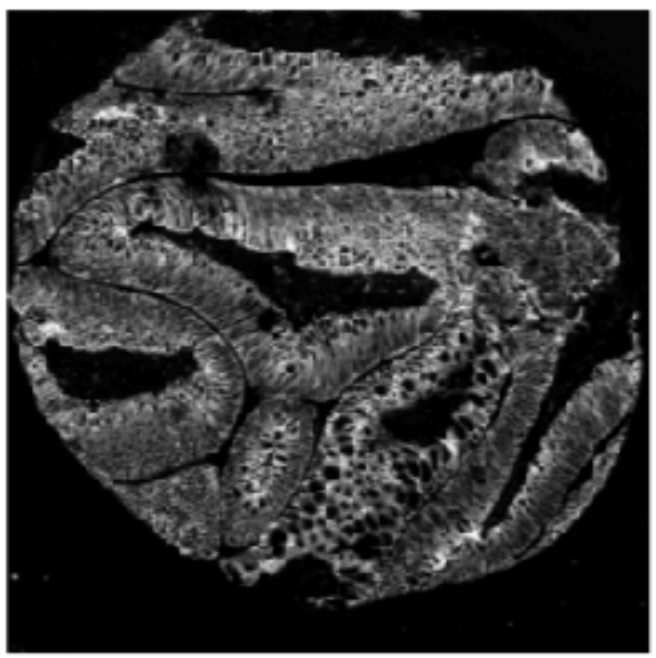


(d) Distribution of scores from pck26 marker

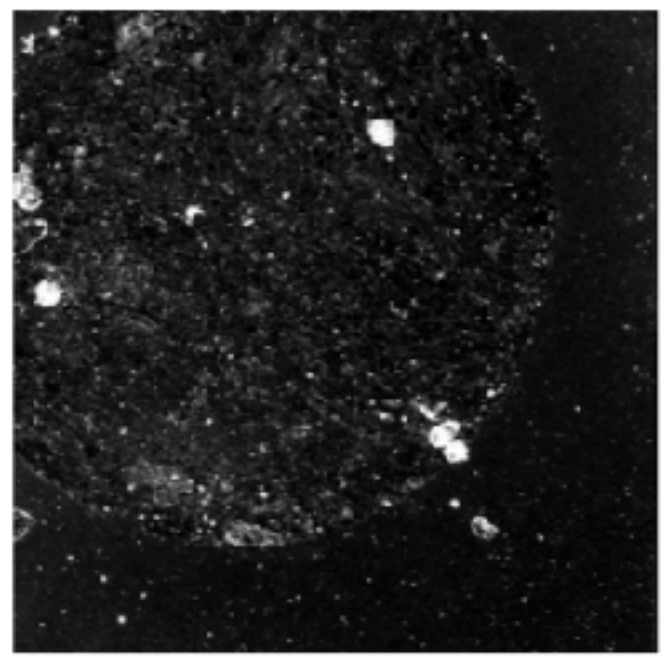
Possible application of the QoI score

The *QoI* score can be used in a data-driven approach to filter a dataset

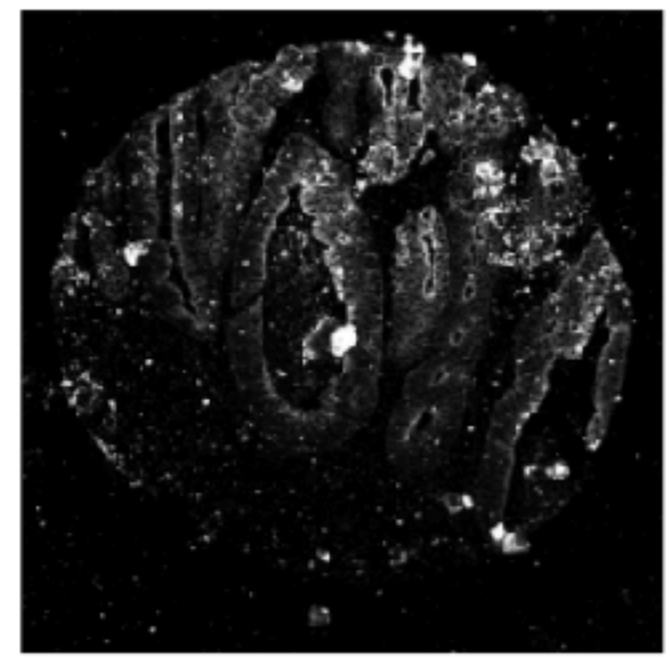
Examples of *good*, *bad* and *ugly* images based on the QoI score.



(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

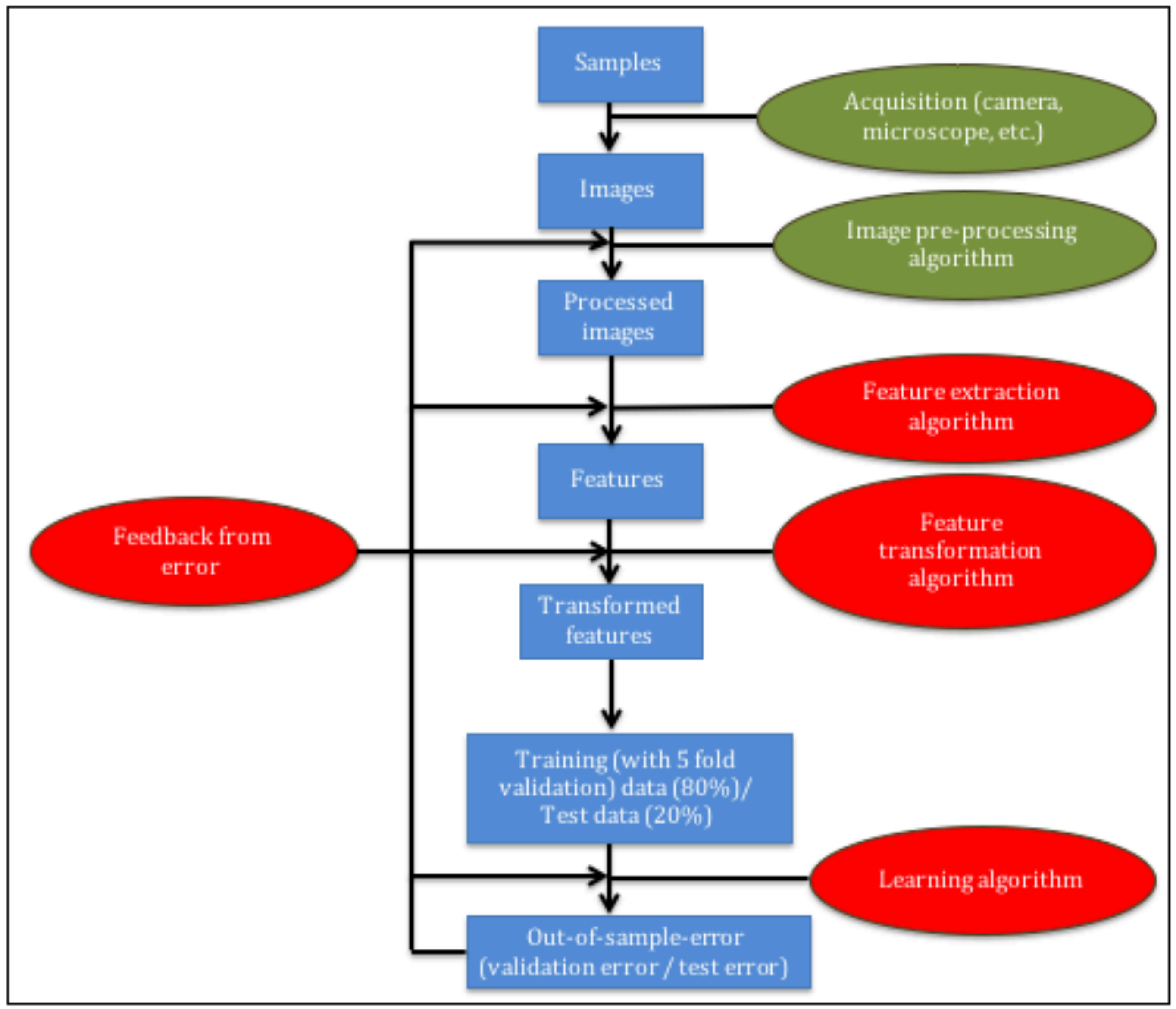
Discussion

- 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 hyper-parameters



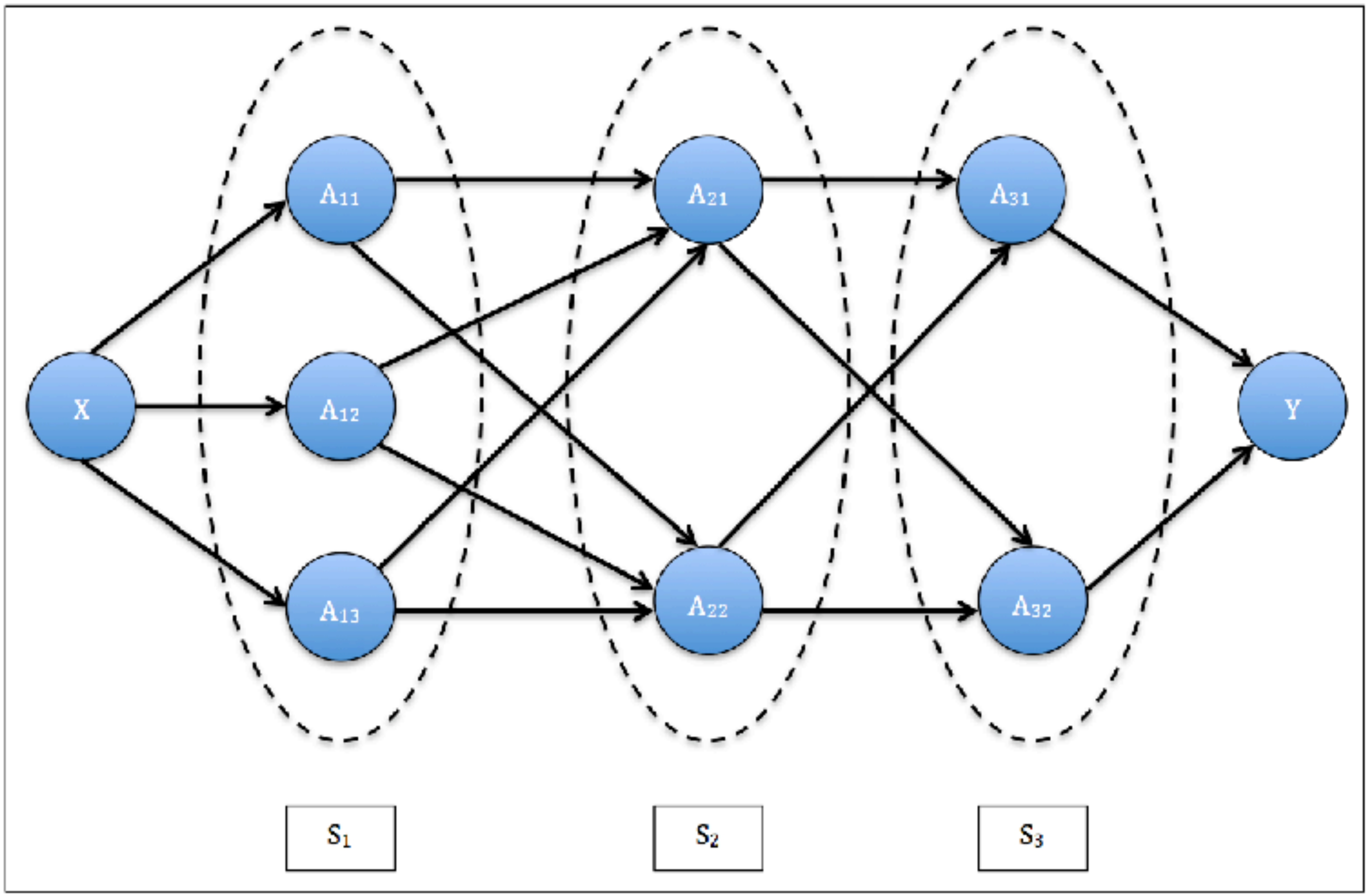
Chowdhury, Aritra, et al. "Quantifying error contributions of computational steps, algorithms and hyperparameter choices in image classification pipelines" *IEEE International Conference on Data Mining (ICDM) 2018* (Submitted)

Introduction

- Problem: Quantify the contribution of components (steps, algorithms and hyperparameters) to the pipeline.
- We propose a method denoted as the *agnostic* methodology to quantify the contributions.
- Hyper-parameter optimization methods and algorithms are used to quantify error contributions - grid search, random search, Bayesian optimization.

Image classification pipeline used in problem

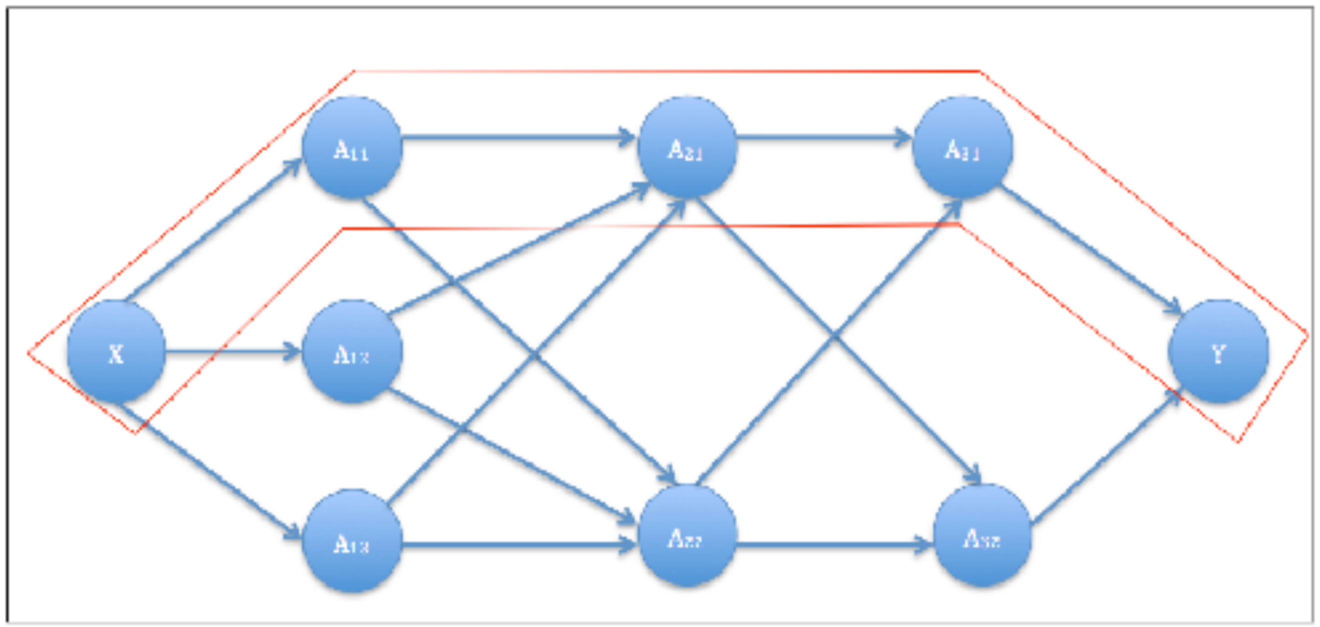
The image classification pipeline denoted as a feed-forward network



Optimization frameworks

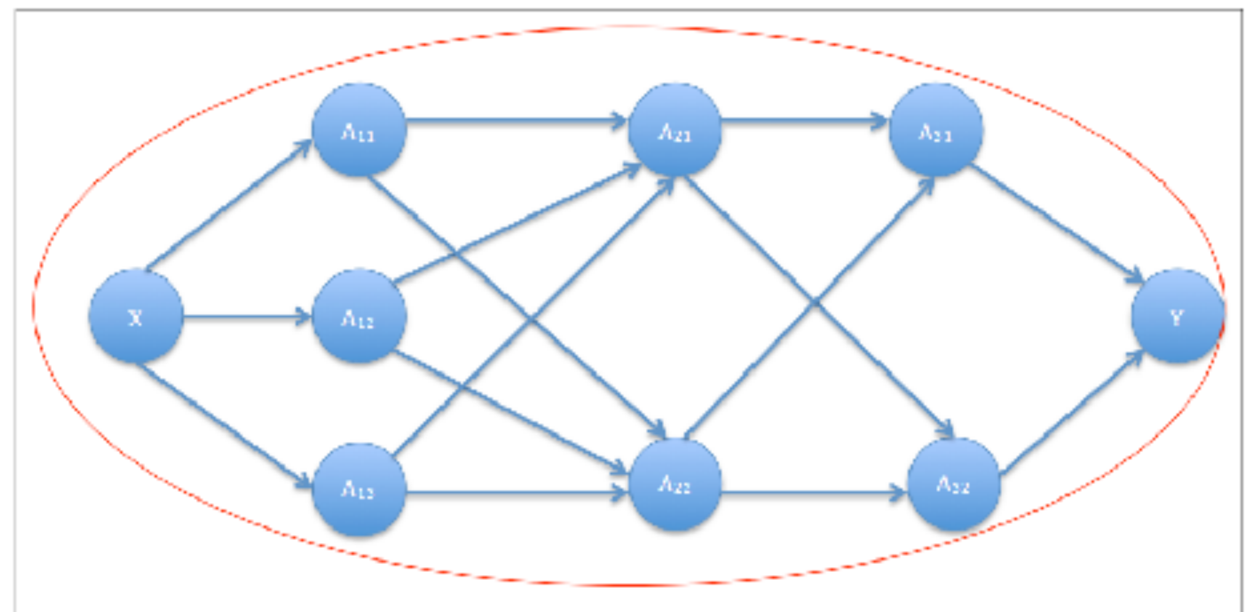
Hyper-parameter optimization (HPO)

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



Combined algorithm selection and hyper parameter optimization (CASH)

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



where,

$\mathcal{L}$ = Validation error

$\theta \in \Theta = \Theta_1 \times \Theta_2 \times \ldots \times \Theta_n$

$A \in \mathcal{A} = A_1(\Theta_1) \times A_2(\Theta_2) \times \ldots \times A_n(\Theta_n)$

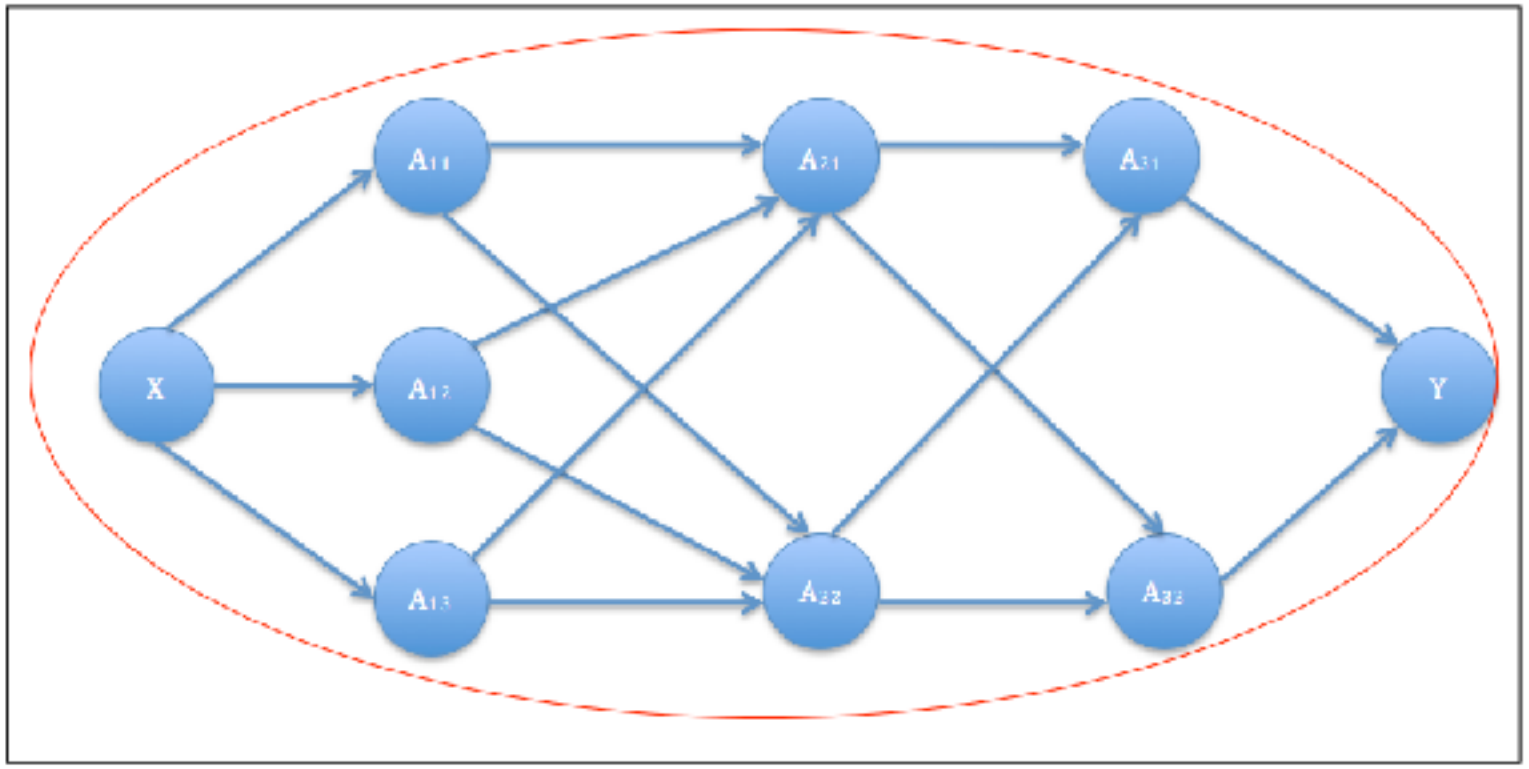
$D_{train}^{(i)}$ = Training set on the i-th fold

$D_{valid}^{(i)}$ = Validation set on the i-th fold

Error contribution from computational steps using the *agnostic* methodology

$$\boxed{EC^*_{S_i} = \frac{1}{|S_i|}\sum_{z=1}^{|S_i|} E^*_{A_{iz}} - E^*,}$$

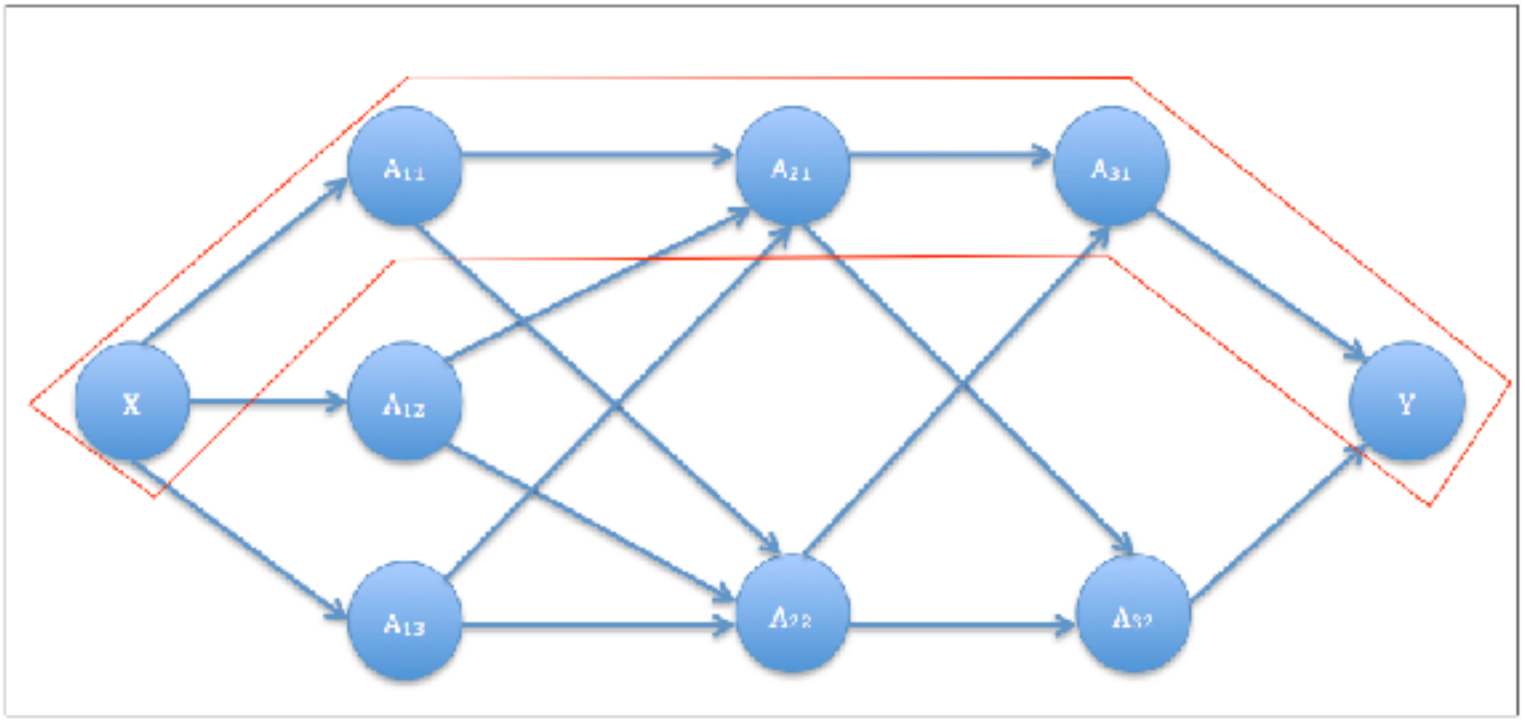
where,
$EC^*_{S_i}$ = Error contribution from step i
$|S_i|$ = Number of algorithms in step i
$E^*_{A_{iz}}$ = Minimum validation error with A_{iz} as the only algorithm in step i
E^* = Global minimum error found in the pipeline.



Error contribution from algorithms using the *agnostic* methodology

$$\boxed{EC^*_{A_{ij}} = \frac{1}{|\theta_{ij}|}\sum_{z=1}^{|\theta_{ij}|} {E^z_{A_{ij}}}^* - E^*_{A^p_{ij}},}$$

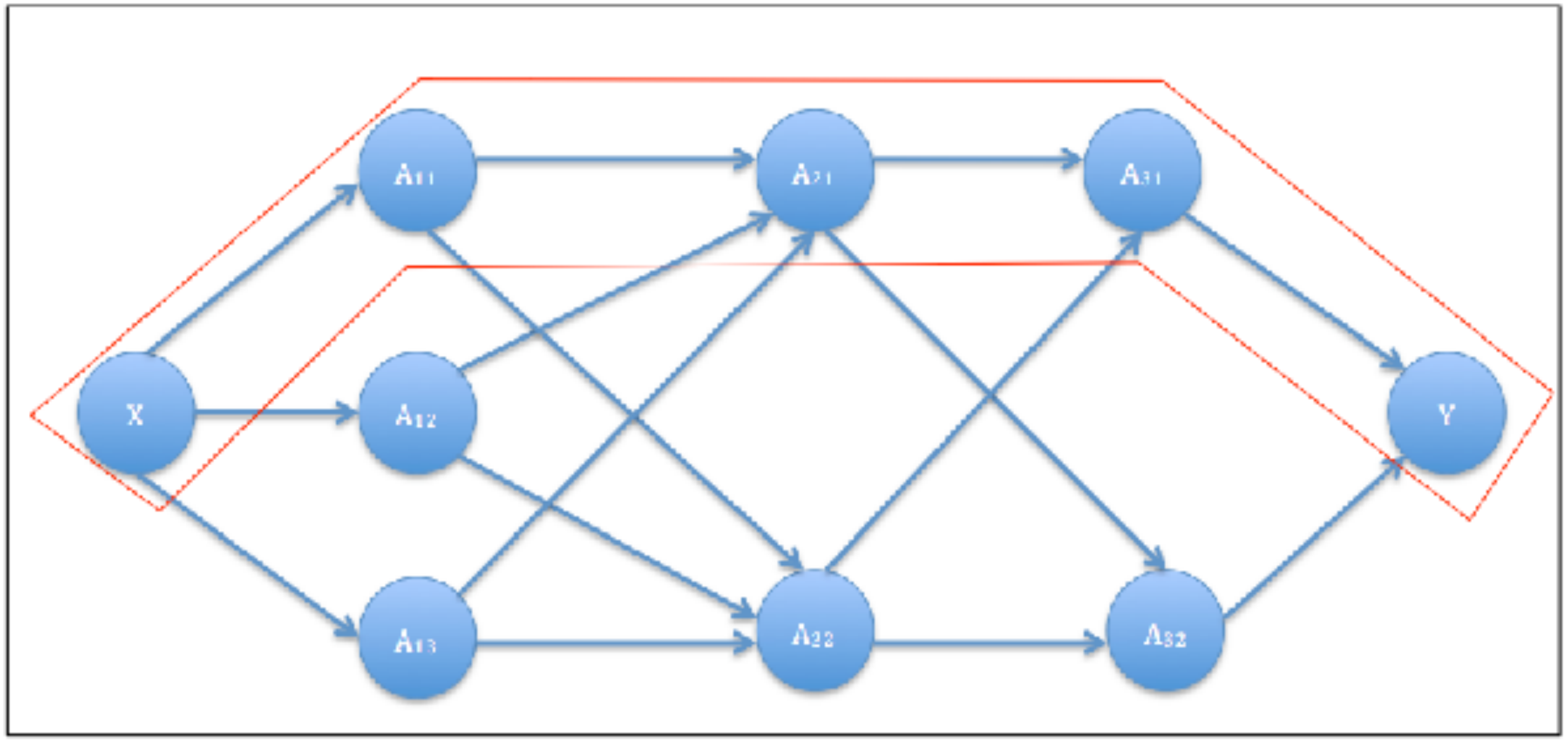
where,
$EC^*_{A_{ij}}$ = Error contribution from algorithm A_{ij},
$|\theta_{ij}|$ = Number of hyperparametric configurations of A_{ij},
${E^z_{A_{ij}}}^*$ = Minimum error obtained with the z-th configuration of θ_{ij},
$E^*_{A^p_{ij}}$ = Minimum error found over the path p that consists of algorithm A_{ij}



Error contribution from hyper parameters using the *agnostic* methodology

$$\boxed{EC^*_{\theta_{ijk}} = \frac{1}{|\theta_{ijk}|}\sum_{z=1}^{|\theta_{ijk}|} E^{z}_{\theta_{ijk}}{}^{*} - E^*_{A^p_{ij}}}$$

where,
$EC^*_{\theta_{ijk}}$ = Error contribution of hyperparameter θ_{ijk},
$|\theta_{ijk}|$ = Number of configurations of θ_{ijk},
$E^{z}_{\theta_{ijk}}{}^{*}$ = Minimum error obtained with the z-th configuration of θ_{ijk}
$E^*_{A^p_{ij}}$ = Minimum error found over the path p that consists of algorithm A_{ij}



Datasets and pipeline used in this work

Datasets

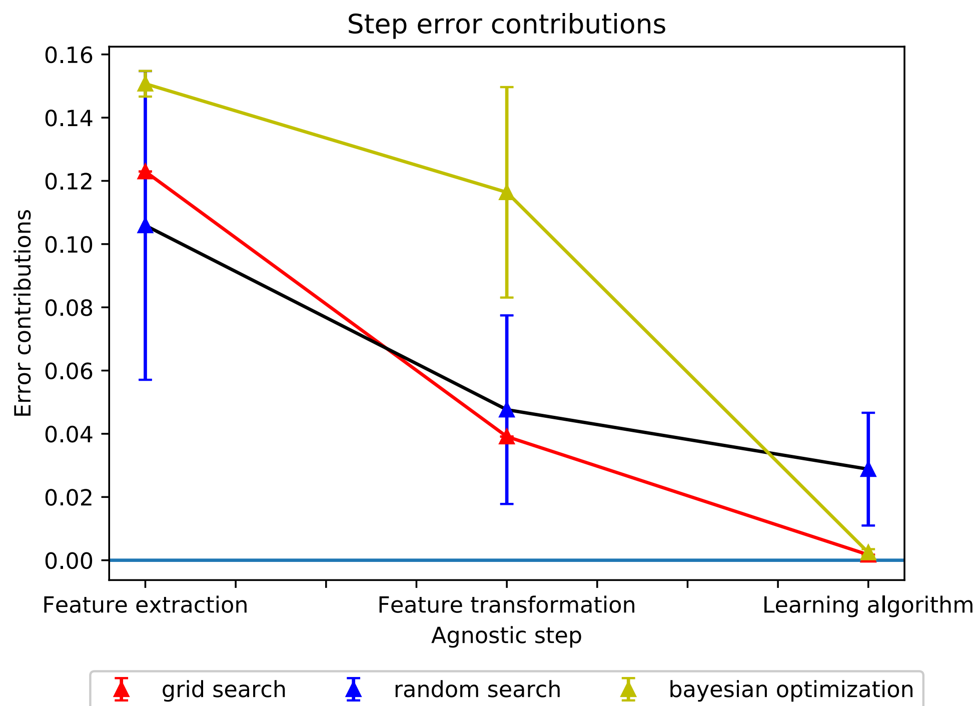
Dataset (notation)	Distribution of classes
Breast cancer (*breast*)	*benign*: 151, *in-situ*: 93, *invasive*: 202
Brain cancer (*brain*)	*glioma*: 16, *healthy*: 210, *inflammation*: 107
Material science 1 (*matsc1*)	*dendrites*: 441, *non-dendrites*: 132
Material science 2 (*matsc2*)	*transverse*: 393,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

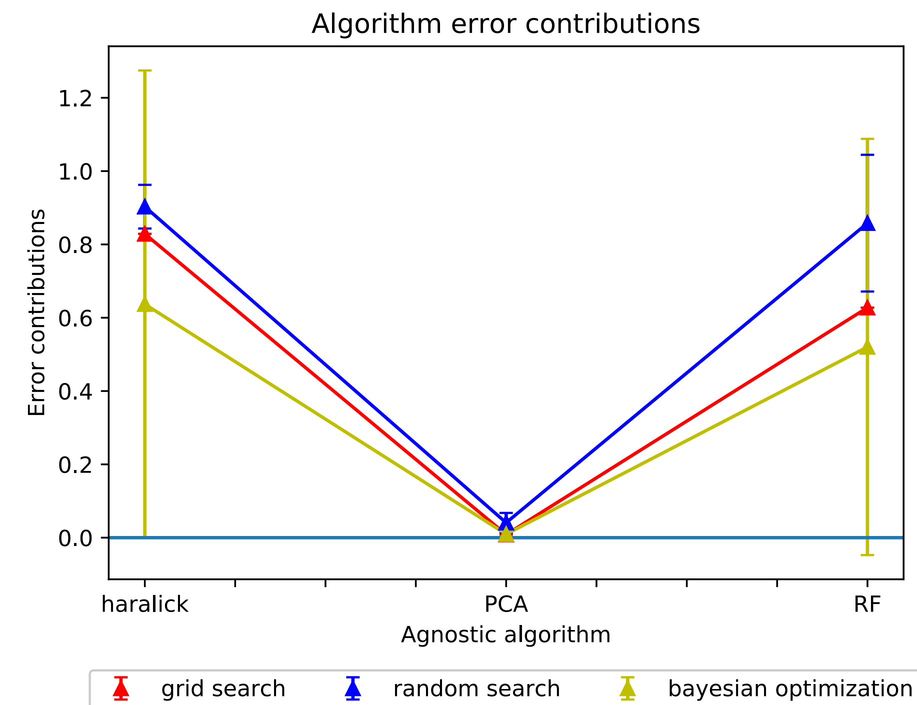
Pipeline

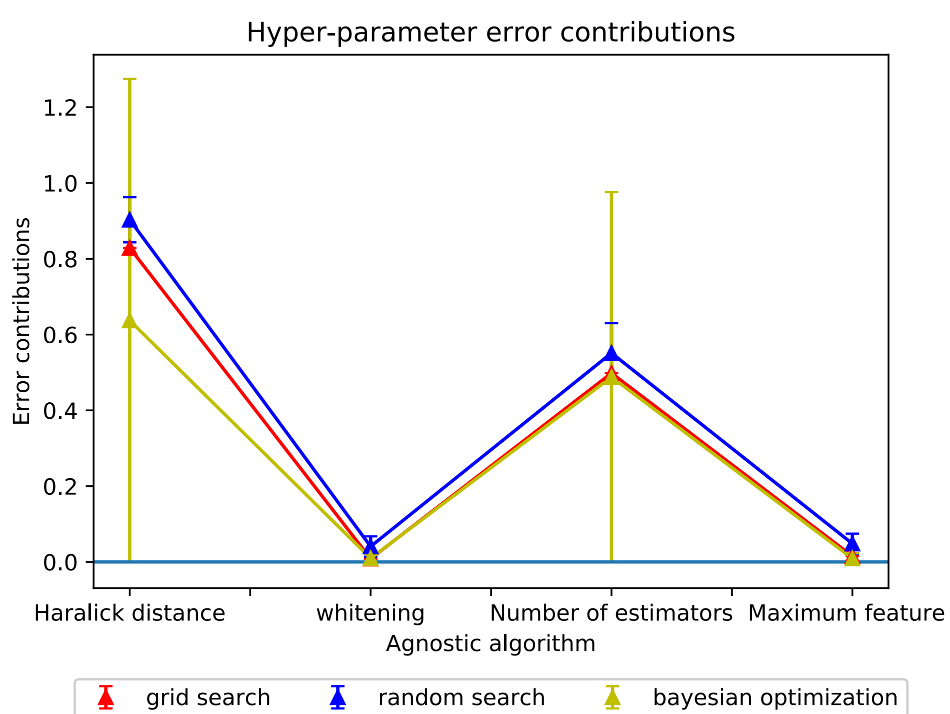
Step	$A_{ij}(\theta_{ij})$	Definition
Feature extraction	$A_{11}(\theta_{11})$	Haralick texture features (*distance*)
	$A_{12}(\theta_{12})$	Pre-trained CNN trained on ImageNet database with VGG16 network
	$A_{13}(\theta_{13})$	Pre-trained CNN trained on ImageNet database with Inception network
Feature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

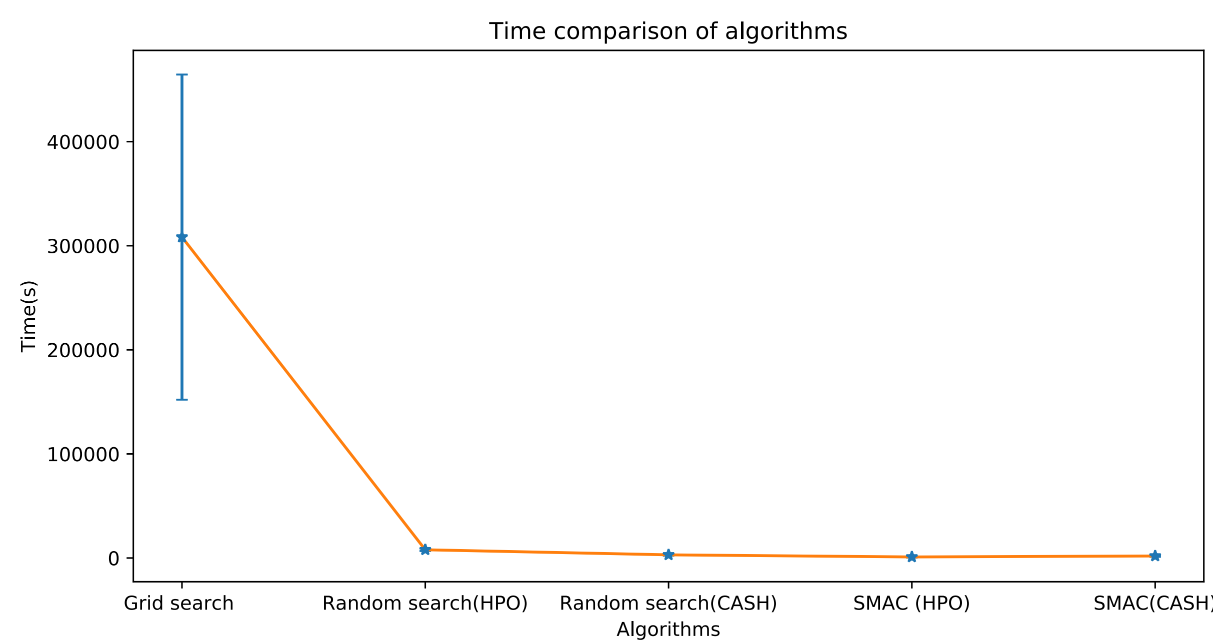
Comparisons of error contributions from components of the pipeline averaged over the datasets

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









Discussion

- We propose a method (known as the *agnostic* methodology) 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.

Outline

- Problem definition and motivation
- Our contributions
    1. Reduction of classification error
    2. Quantification of the contribution of components in a learning pipeline
- Conclusion and future work
- References

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 :
  - Locally by modifying the individual components of the pipeline.
  - Globally modifying the components of the pipeline as a whole.
- The contributions of the components in image classification pipelines 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 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.

Thank you

Outline

- Problem definition and motivation
- Our contributions
  1. Reduction of classification error
  2. Quantification of the contribution of components in a learning pipeline
- Conclusion and future work
- References

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. 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.

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 convo- lutional 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.