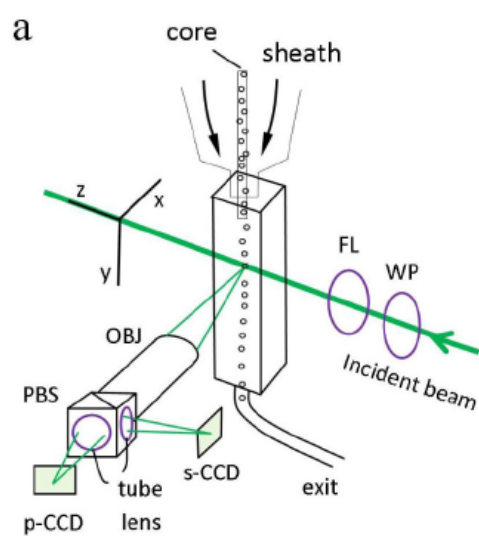# Validating a Deep Learning Framework by Metamorphic Testing
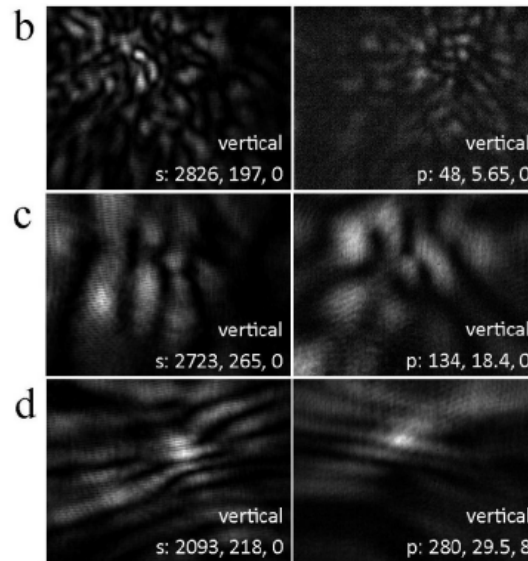
Junhua Ding, Xiaojun Kang, Xin-Hua Hu

East Carolina University

Greenville, NC USA

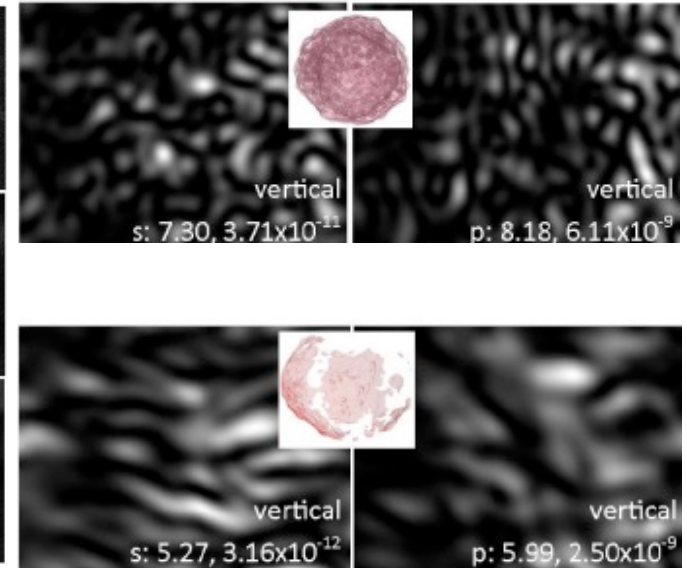# Introduction

- Morphology based cell typing
- Deep learning
  - AlexNet (227 * 227 pixels)
- Support vector machine (SVM)
  - 17 GLCM features

Instrument **cell diffraction images** **cell and fractureddebris**

Ref: J. Ding, Xin-Hua Hu, "A machine learning based framework for quality assurance of massive scale image data"

Validating the Deep Learning Classifier

- Existing Approaches
  - NFCV, confusion matrix
- Proposed approach
  - System level, comparing its accuracy with other classifiers
  - Data set level, testing it with different sizes of data sets
  - Data item level, testing it with different data items

Data set

- Over 6500 diffraction images of three categories of images: normal cells, ghost cell bodies, debris

- 2128 cell images, 2246 ghost cell body images, and 2214 debris images. Size: 640 * 480 pixels, resolutions: 12 bits.

- Pre-labeled manually

Producing deep learning data sets

- Cropping
  - Cut small pieces from original image with different strides in different directions
- Pooling
  - Down-sampling with different different strides.
  - Different pooling functions could be used.
- Produce over 100,000 images for each category, each small image is labeled as its original image. Size: 227 * 227 pixels.

System level testing

- The classification accuracy of the deep learning based classifier is higher than the SVM based classifier on the same data sets.

  - Around 72% vs. over 90%, by 10FCV

Data set level MRs

- MR2: Add 10% of new images into each category of the training data set should not affect the classification accuracy.

- MR3: Duplicate 10\% of images of each category in the training data set should not affect the classification accuracy.

- MR4: Add 10% of images into each category of the validation data set should not affect the classification accuracy.

- MR5: Add 10% of images into each category of the test data set should not affect the classification accuracy.

- MR6: Remove one category of the data from the data set should not affect the classification accuracy of the remaining categories.

- MR7: Add one category of the diffraction images through duplicating one existing category of data in the data set should not affect the classification accuracy.

Data item level MRs

- MR8: The classification accuracy of the classifier trained on the data sets cropped from original images using different stride distances should be almost the same.

- MR9: The classification accuracy of the classifier trained on the data sets cropped from original images using different moving directions should be almost the same.

One more

- Pooling vs. Cropping
  - Pooling > cropping
- Different pooling functions
  - Minimal > maximal > average
- How to evaluate the quality of the generated data?
  - Metamorphic testing

Conclusion

- Metamorphic testing is important for validating machine learning systems.

- Further investigation and experiments will be conducted.