



# Comparing Machine Learning Methods for Hyperspectral Image and Text Classification



Daria Garkavtseva<sup>1</sup>, Calarina Muslimani<sup>2</sup>, Ekaterina Rapinchuk<sup>3</sup>

University of Massachusetts, Amherst<sup>1</sup>, University of Maryland, College Park<sup>2</sup>, Michigan State University<sup>3</sup>

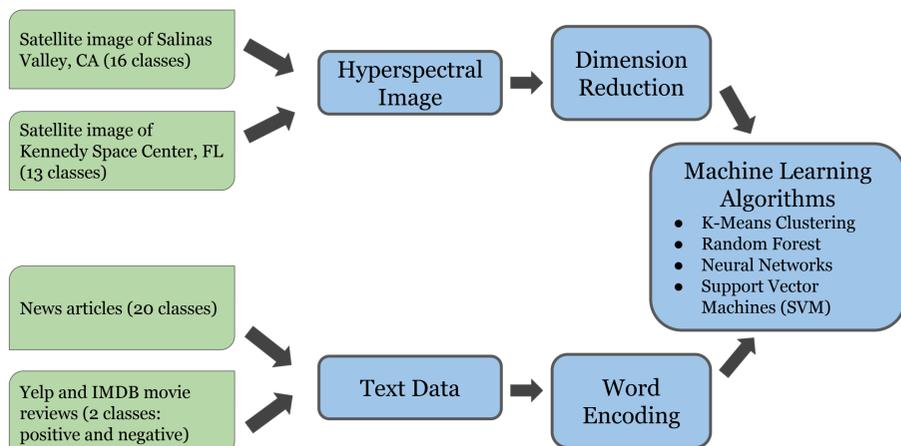
## Background

Hyperspectral images have **hundreds of channels** that span across the **electromagnetic spectrum**, unlike regular color images that only have red, green, and blue color channels. They are able to provide more information than regular images, which is useful in tasks such as object and mineral detection.

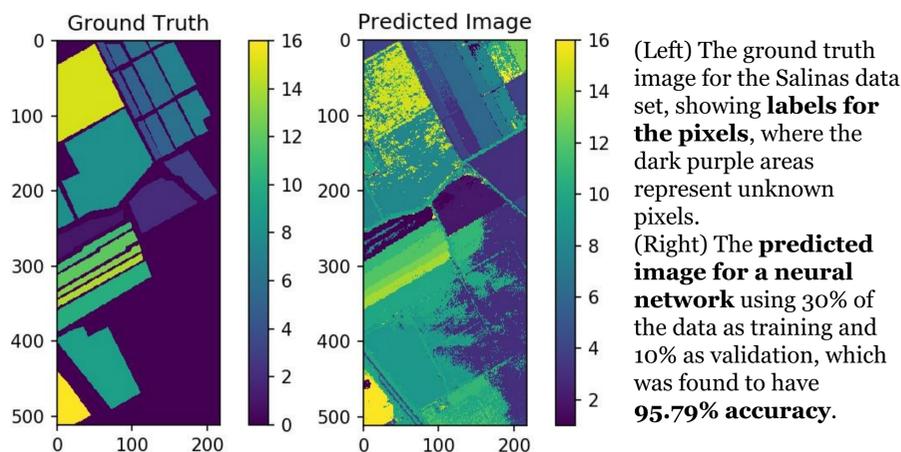
## Objectives

- Apply and compare the performance of **machine learning algorithms** for **hyperspectral image segmentation**.
- Compare the performance of these algorithms for **hyperspectral images vs text classification**.

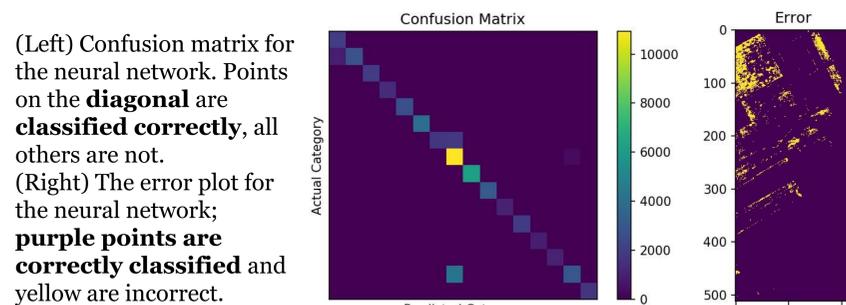
## Methods



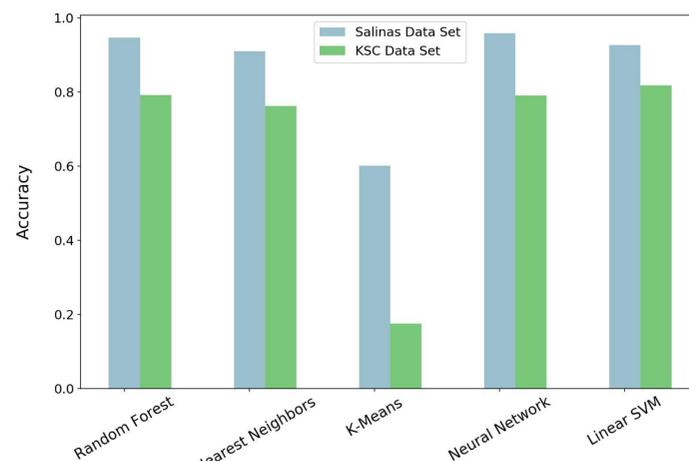
## Results: Hyperspectral Images



## Results: Hyperspectral Images

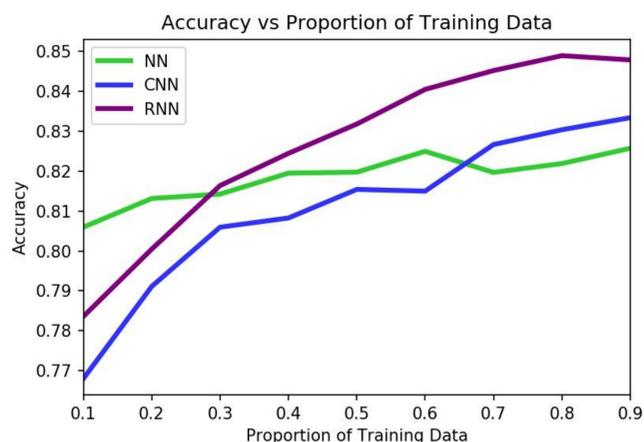


Accuracy of Machine Learning Methods for Hyperspectral Images



(Above) A comparison of the accuracies of machine learning methods for 2 data sets: Salinas Valley and Kennedy Space Center (KSC). **Supervised learning algorithms have consistently high accuracies while K-Means, an unsupervised learning algorithm, does significantly worse.**

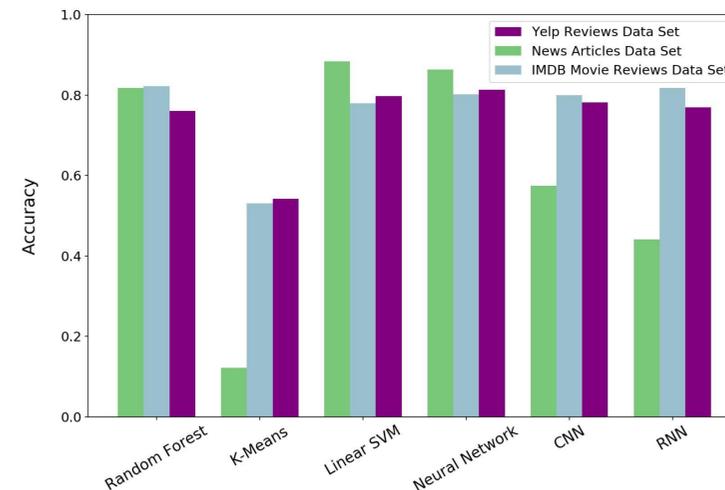
## Results: Text



(Above) A graph comparing the performance of neural networks (NN), convolutional neural networks (CNN), and recurrent neural networks (RNN) for **varying amounts of training data**. The performance of **RNNs and CNNs is affected more by amount of training data** than NNs.

## Results: Text

Accuracy of Machine Learning Methods for Text Data



(Above) A **comparison of the accuracies** of machine learning methods for 3 data sets using 60% training and 10% validation data. **CNNs and RNNs, which consider order of the words more accurately classify based on meaning** (Yelp and IMDB data sets), while **the other algorithms, which consider only word frequency better classify based on topic** (News articles data set).

## Conclusions

- **Supervised methods consistently performed better** than unsupervised methods for all data.
- These algorithms generally perform **better** for **hyperspectral image segmentation** than **text classification**.
- **Text data required more training data** than **hyperspectral images** to achieve similar accuracies (in this case, 60% vs 30% training data).
- The Salinas data set performed **more than 15% better** than the KSC data set for most methods.
- Methods that consider **order** of the words more accurately classify by **meaning** whereas algorithms that consider word **frequency** more accurately classify by **topic**.

## References and Acknowledgements

We acknowledge the support of the MSU ACRES REU program and the NSF through grant OAC1560168.

- [1]Chollet, Francois, et al. *Keras*, 2015. <https://keras.io>.
- [2]"Hyperspectral Remote Sensing Scenes." *Hyperspectral Remote Sensing Scenes - Grupo De Inteligencia Computacional (GIC)*, 2014.
- [3]Kotzias, Dimitrios, et al. "From Group to Individual Labels using Deep Features." *KDD*, 2015.
- [4]Pedregosa, Fabian, et al. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [5]"Practical Text Classification With Python and Keras." *Real Python*, 23 Oct. 2018. [realpython.com/python-keras-text-classification/](https://realpython.com/python-keras-text-classification/).

