Southern Adventist University

Knowledge Exchange

Achieve

Summer 6-7-2023

Machine Learning in Action: Exploring Examples in Multiple Domains

GERMAN HARVEY ALFEREZ
Southern Adventist University, harveya@southern.edu

Follow this and additional works at: https://knowledge.e.southern.edu/achieve

Part of the Computer Sciences Commons

Recommended Citation
ALFEREZ, GERMAN HARVEY, "Machine Learning in Action: Exploring Examples in Multiple Domains" (2023). *Achieve*. 2178.
https://knowledge.e.southern.edu/achieve/2178

This Presentation - Oral - to academic peers, less than or equal to 1 hour is brought to you for free and open access by Knowledge Exchange. It has been accepted for inclusion in Achieve by an authorized administrator of Knowledge Exchange. For more information, please contact jspears@southern.edu.

Machine Learning in Action: Exploring Examples in Multiple Domains

Harvey Alférez, Ph.D.
Southern Adventist University
harveya@southern.edu
www.harveyalferez.com
@harveyalferez

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

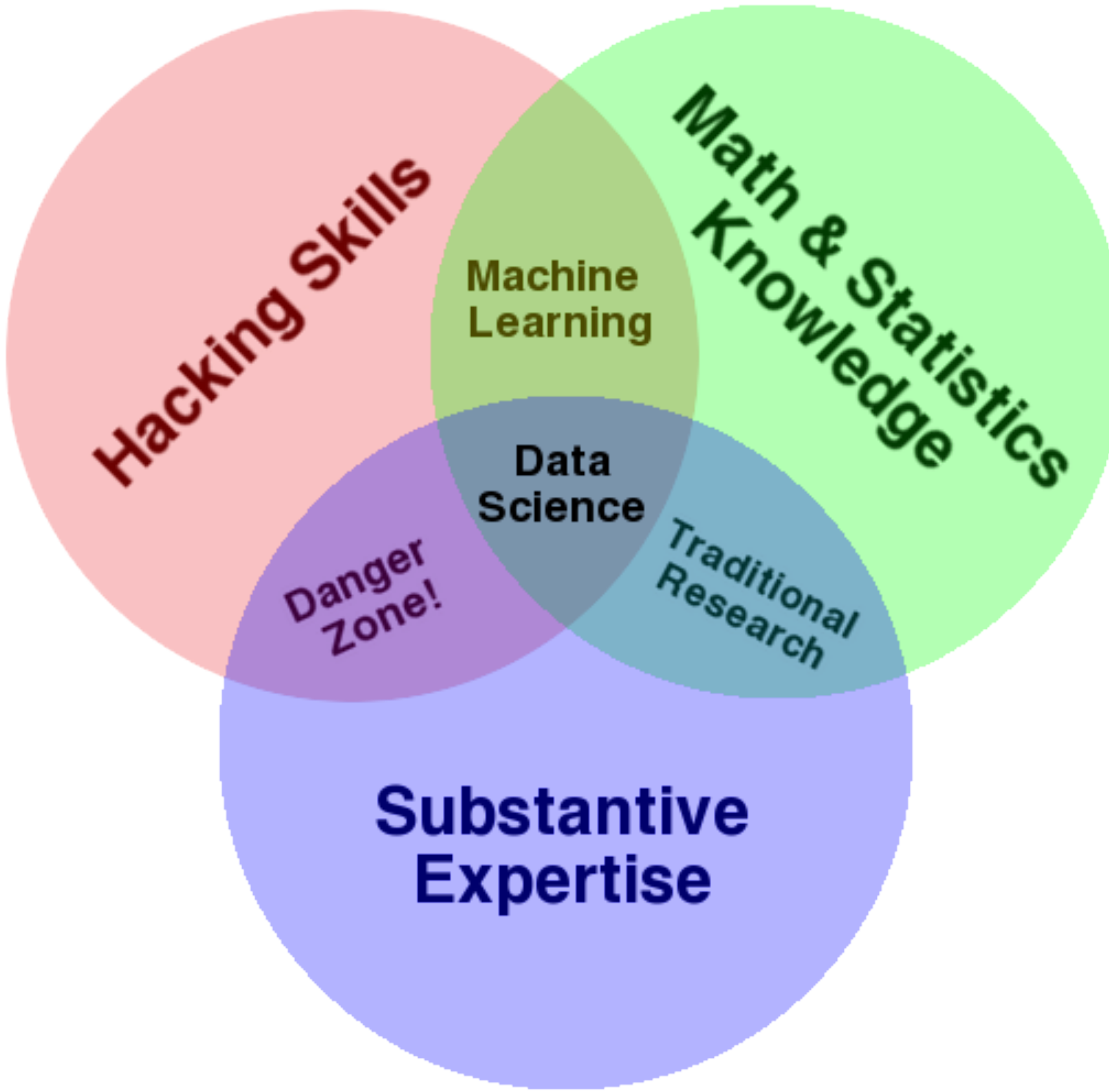
CENTER FOR INNOVATION
AND RESEARCH
IN COMPUTING
Southern Adventist University


HICKMAN
SCIENCE CENTER

- **MSc in Computer Science**
- **MSc in Applied Computer Science**
- **Graduate Certificates:**
  - **Data Analytics**
  - **Network Security**
  - **Web Application Development**

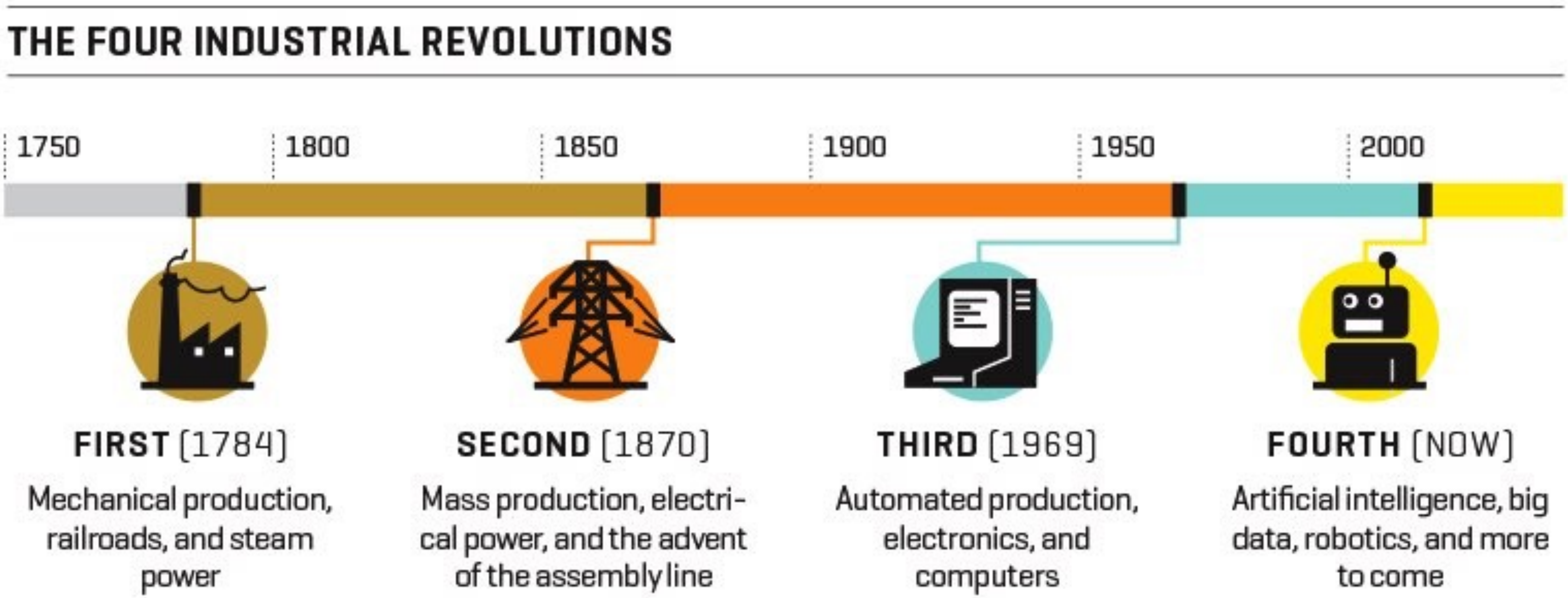
harveya@southern.edu

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

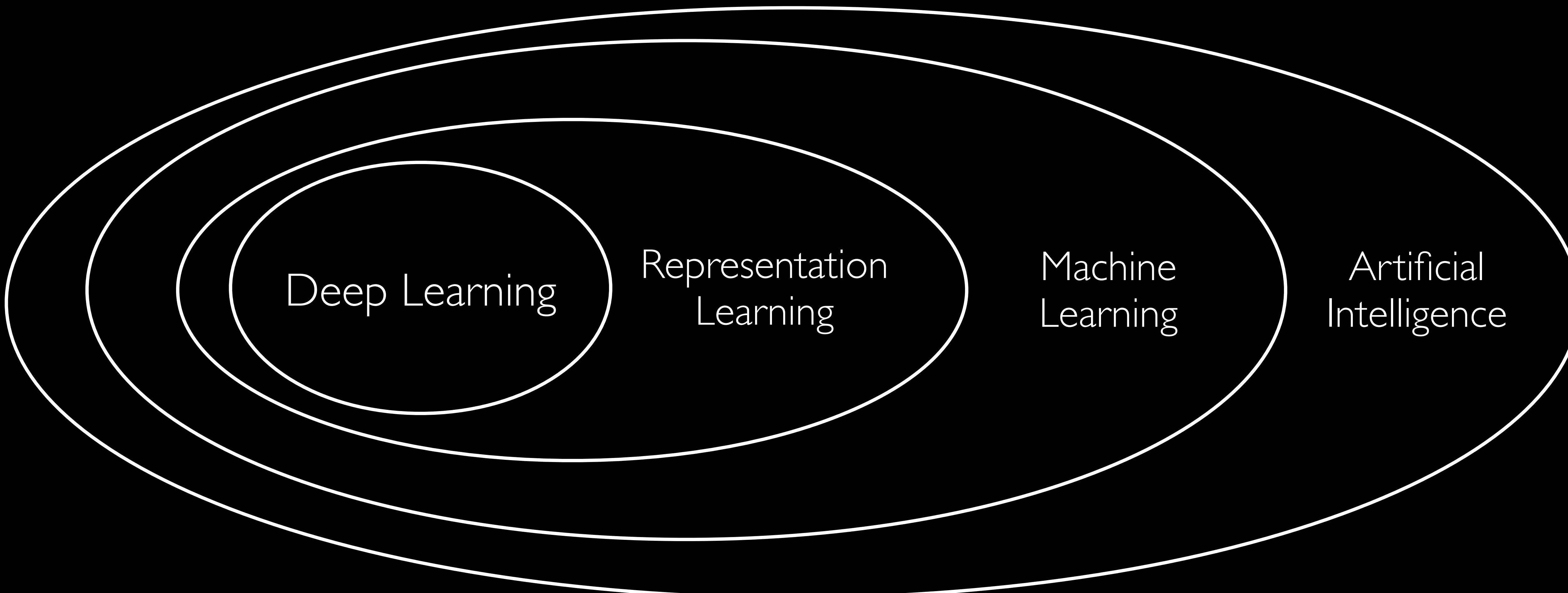


http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University



https://fortune.com/2016/03/08/davos-new-industrial-revolution./



Goodfellow, I., Bengio, Y. y Courville, A. (2016). *Deep learning*. MIT Press.


MACHINE LEARNING

“Learning is any process by which a system improves performance from experience.”

— Herbert Simon

Definition by Tom Mitchell (1998):

Machine Learning is the study of **algorithms** that

improve their performance *P*

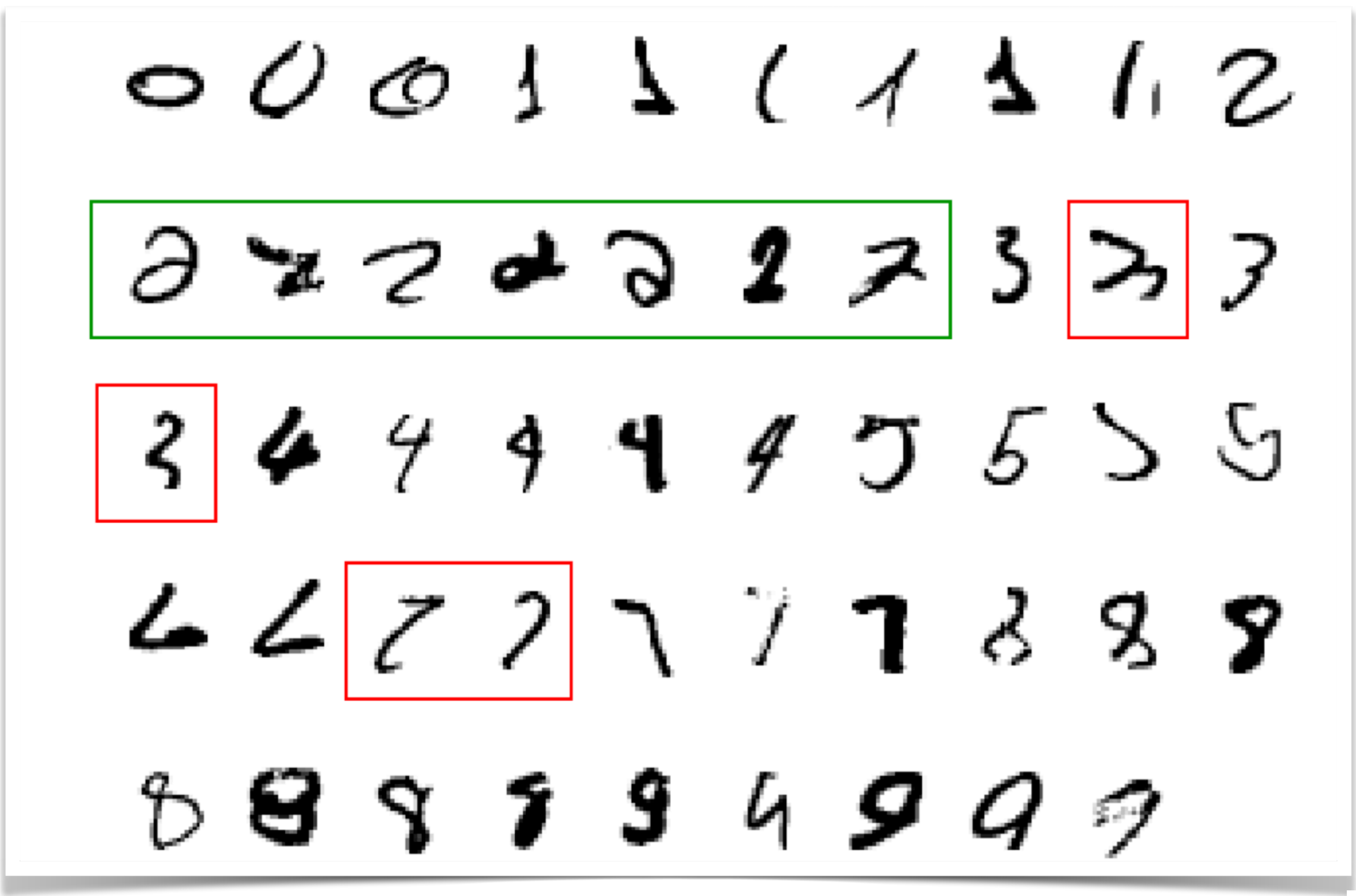
at some task *T*

with experience *E*.

A well-defined **learning task** is given by <*P*, *T*, *E*>.

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

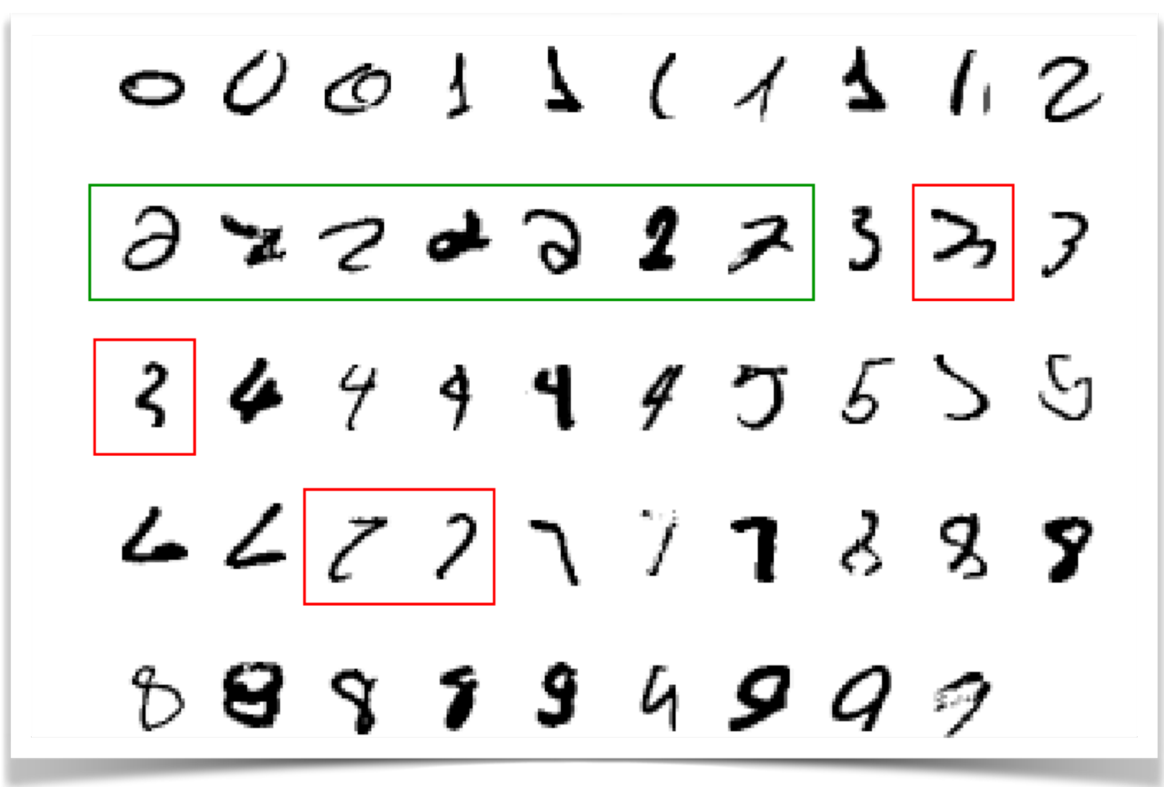
A classic example of a task that requires machine learning: It is very hard to say what makes a 2



CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Improve on task T, with respect to performance metric P, based on experience E

T: Recognizing hand-written numbers
P: Percentage of numbers correctly classified
E: Database of human-labeled images of handwritten numbers



CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

"Machine Learning: Field of study that gives computers the ability to **learn** without being **explicitly programmed**."

— Arthur Samuel (1959)

input → function → output

5 10 → 5 + 10 → 15

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

input → function → output

machine learning

number 1 = 0.00012

number 2 = 0.91000

number 3 = 0.00025

. . .

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Machine learning is not statistics

	Statistics	Machine Learning
Approach	Hypothesis-driven development	Creating systems that learn from data
Goal	Relationships between variables	Optimization; prediction accuracy
Assumptions	Some knowledge about population is usually required	None
Data Complexity	Usually applied to low-dimensional data	Usually applied to high-dimensional data; ML learns from data

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Types of machine learning

Supervised (inductive) learning

– Given: training data + desired outputs (labels)

Unsupervised learning

– Given: training data (without desired outputs)

Semi-supervised learning

– Given: training data + a few desired outputs

Reinforcement learning

– Rewards from sequence of actions

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Types of machine learning

Supervised (inductive) learning

– Given: training data + desired outputs (labels)

Unsupervised learning

– Given: training data (without desired outputs)

Semi-supervised learning

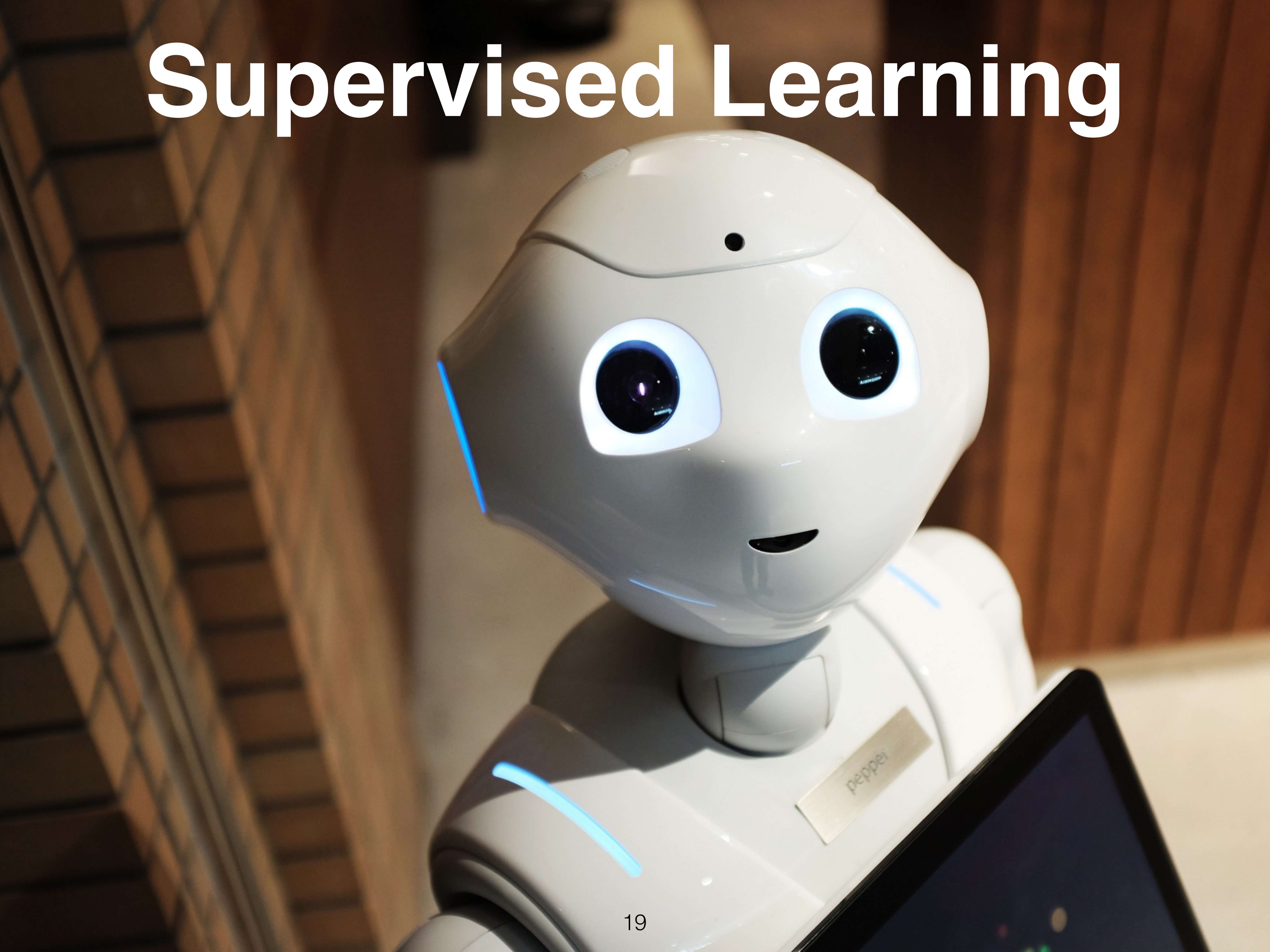
– Given: training data + a few desired outputs

Reinforcement learning

– Rewards from sequence of actions

CENTER FOR INNOVATION
AND RESEARCH
IN COMPUTING
Southern Adventist University

Supervised Learning



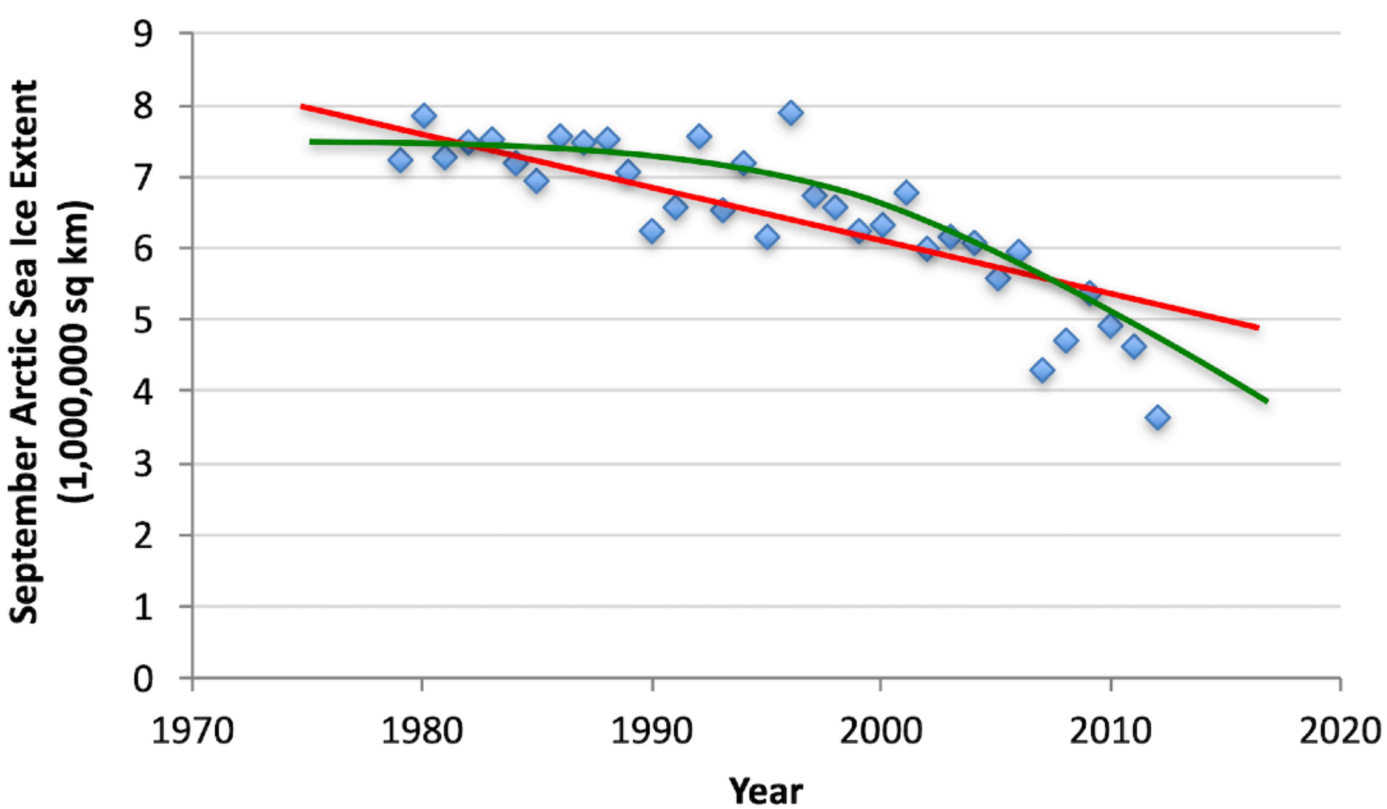
If you're trying to predict a target value, then you need to look into **supervised learning**.

– Given: training data + desired outputs (labels)

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Supervised learning: regression

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
    - y is **real-value** == regression

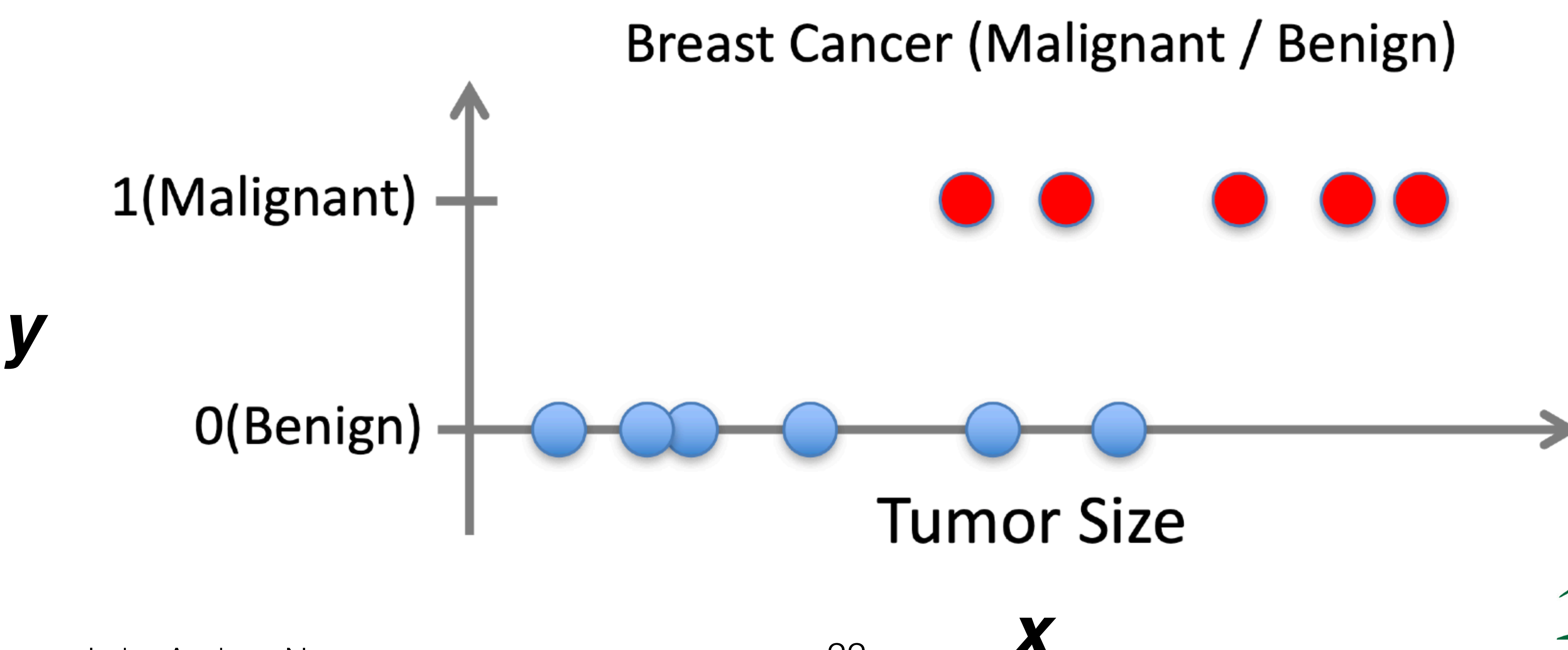


Data from G. Witt. Journal of Statistics Education, Volume 21, Number 1 (2013)

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Supervised learning: classification

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is **categorical** == classification

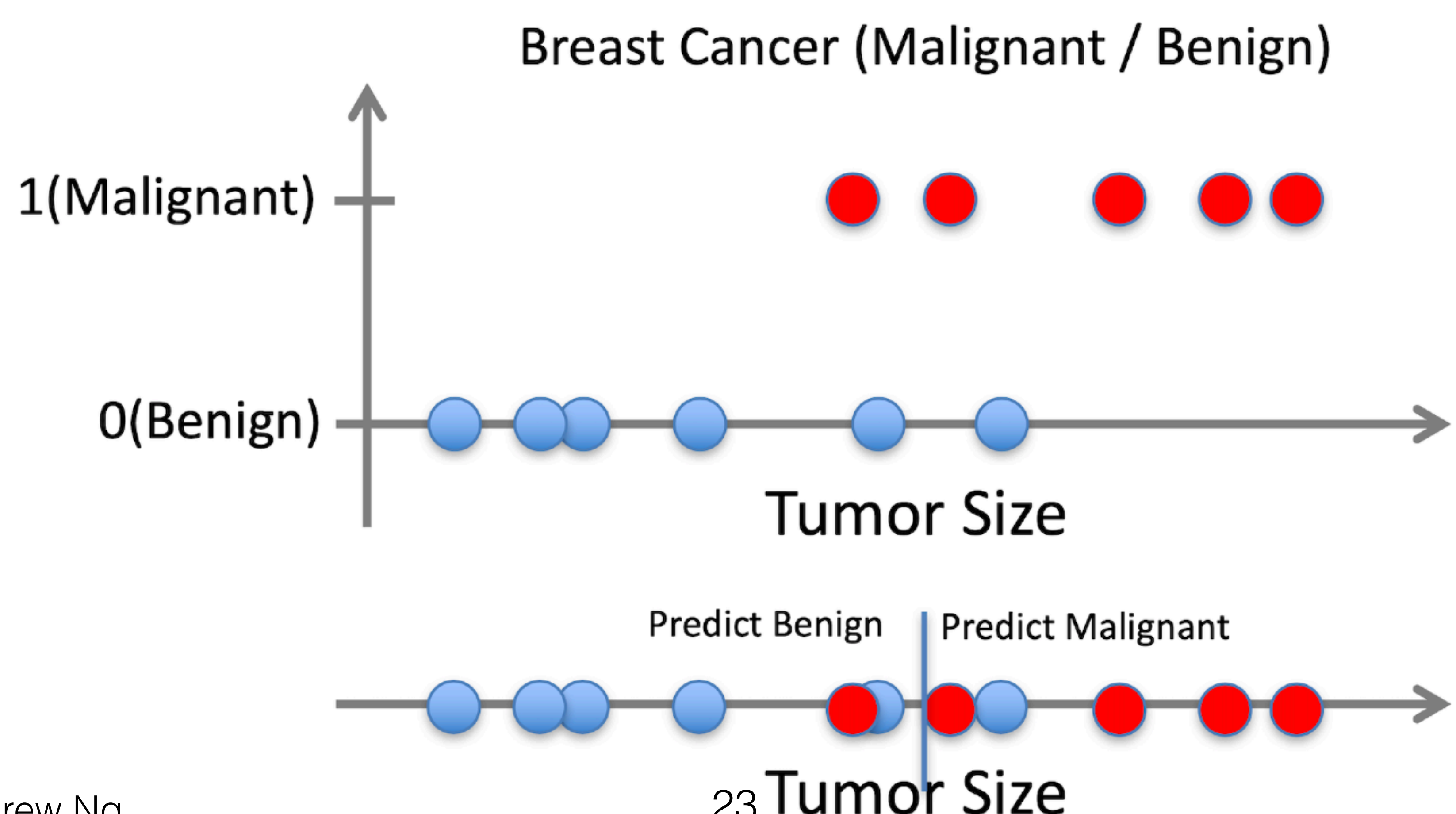


Based on example by Andrew Ng

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Supervised learning: classification

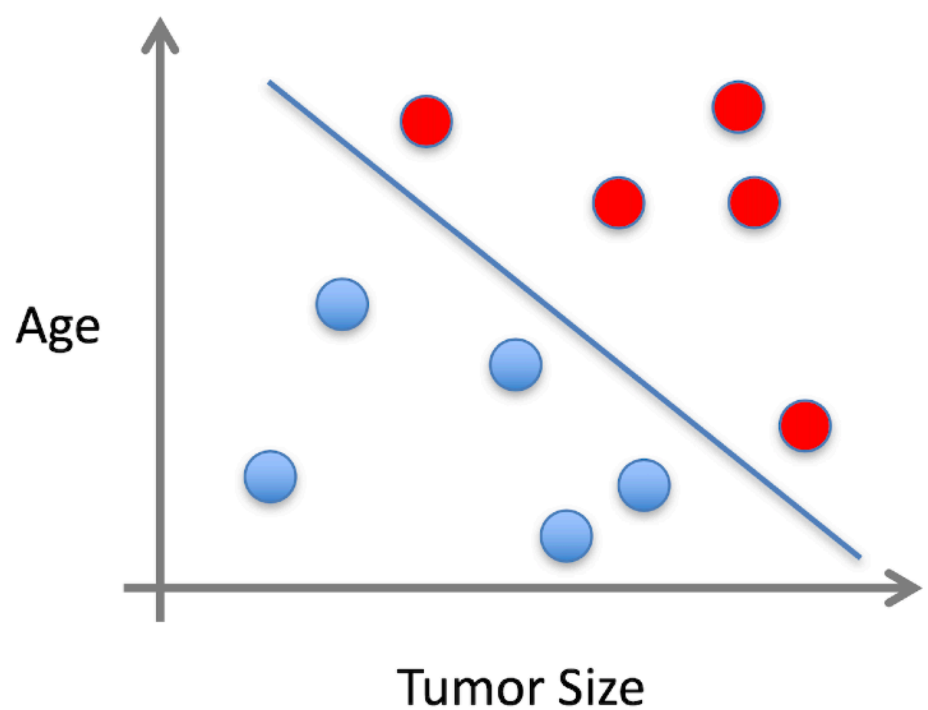
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is **categorical** == classification



Based on example by Andrew Ng

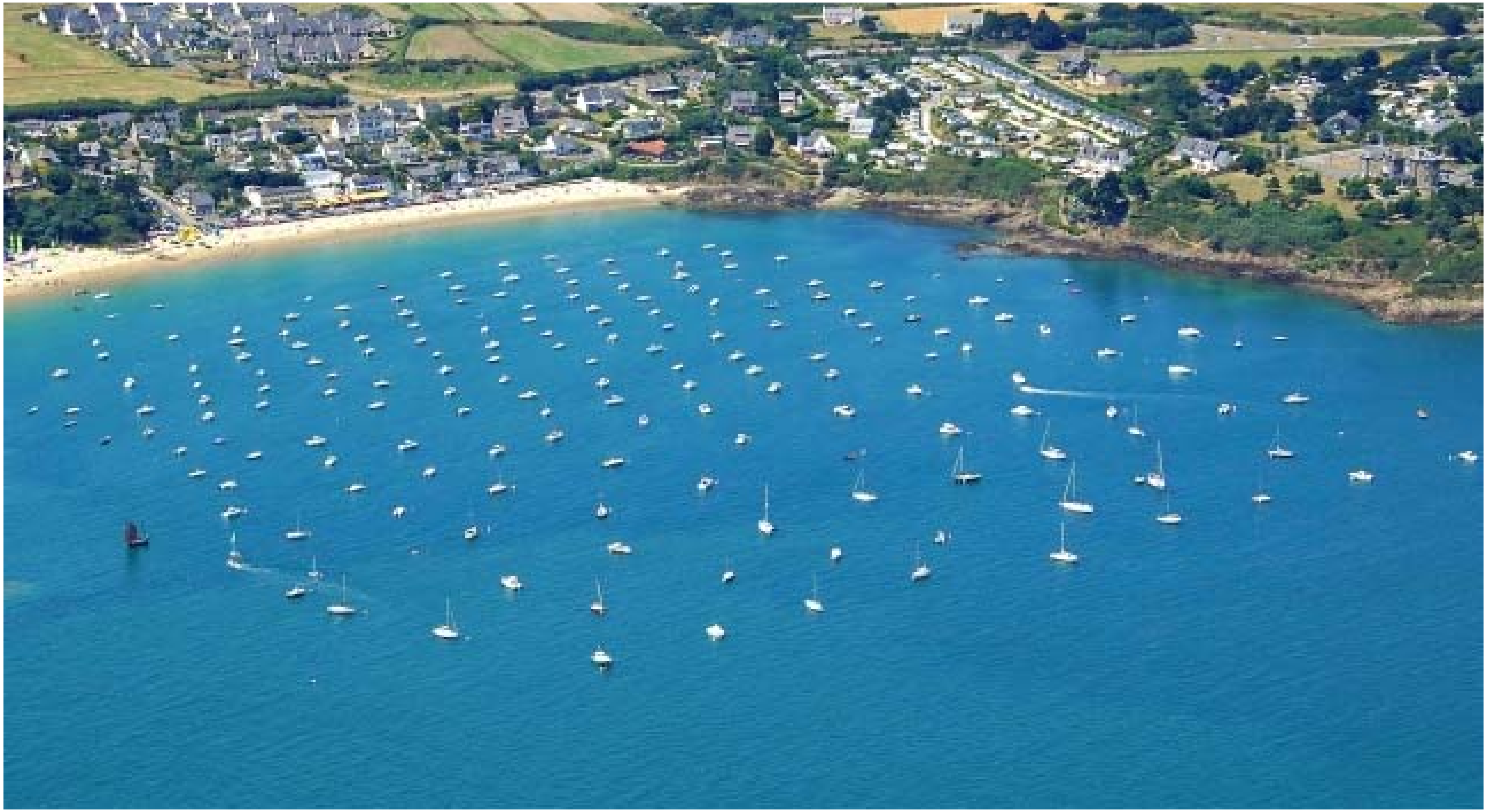
Supervised learning

- x can be multi-dimensional
    - Each dimension corresponds to an attribute



- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

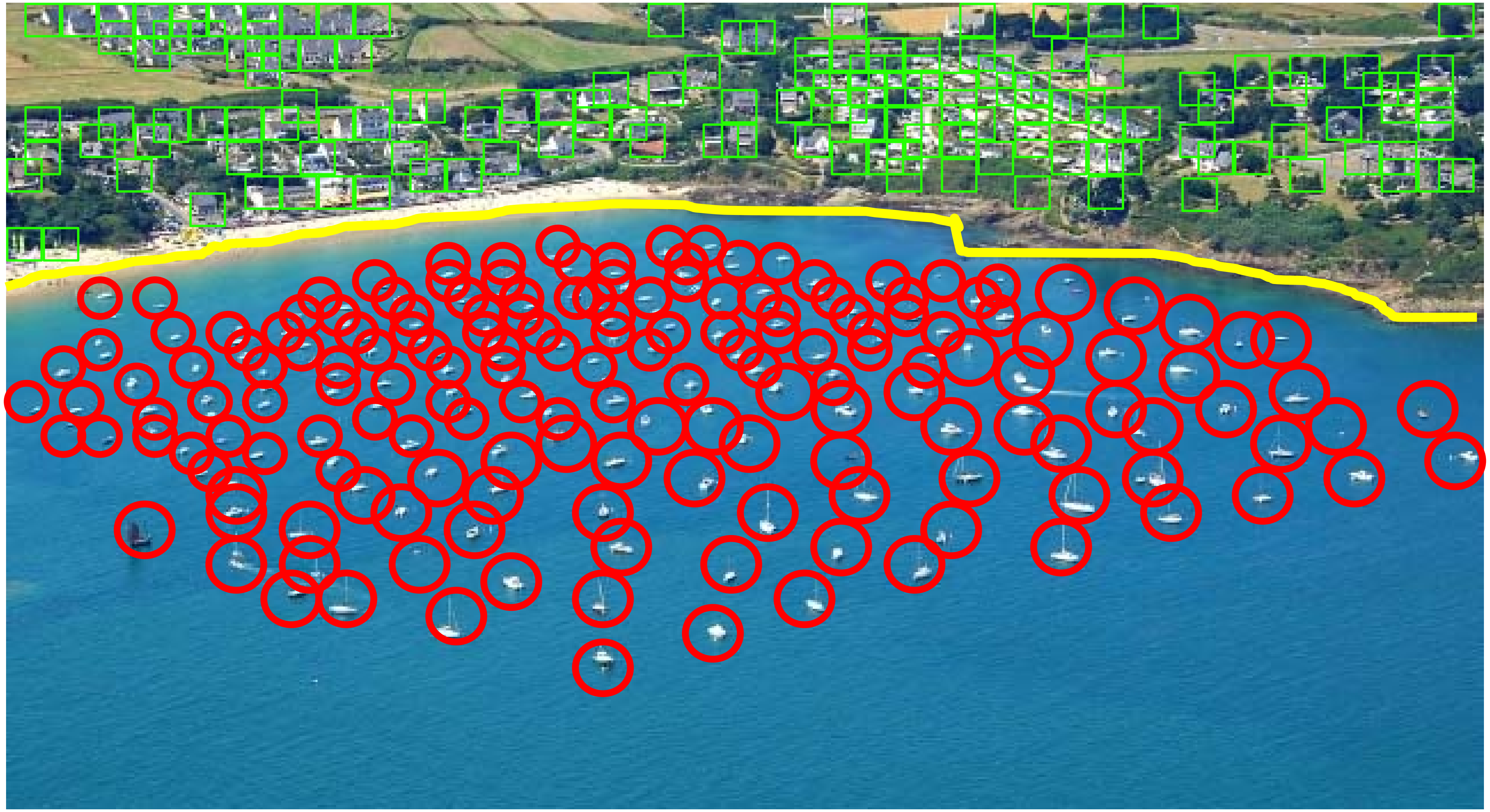
Basic principles of classification



- Want to classify objects as boats and houses.

CENTER FOR INNOVATION
AND RESEARCH
IN COMPUTING
Southern Adventist University

Basic principles of classification

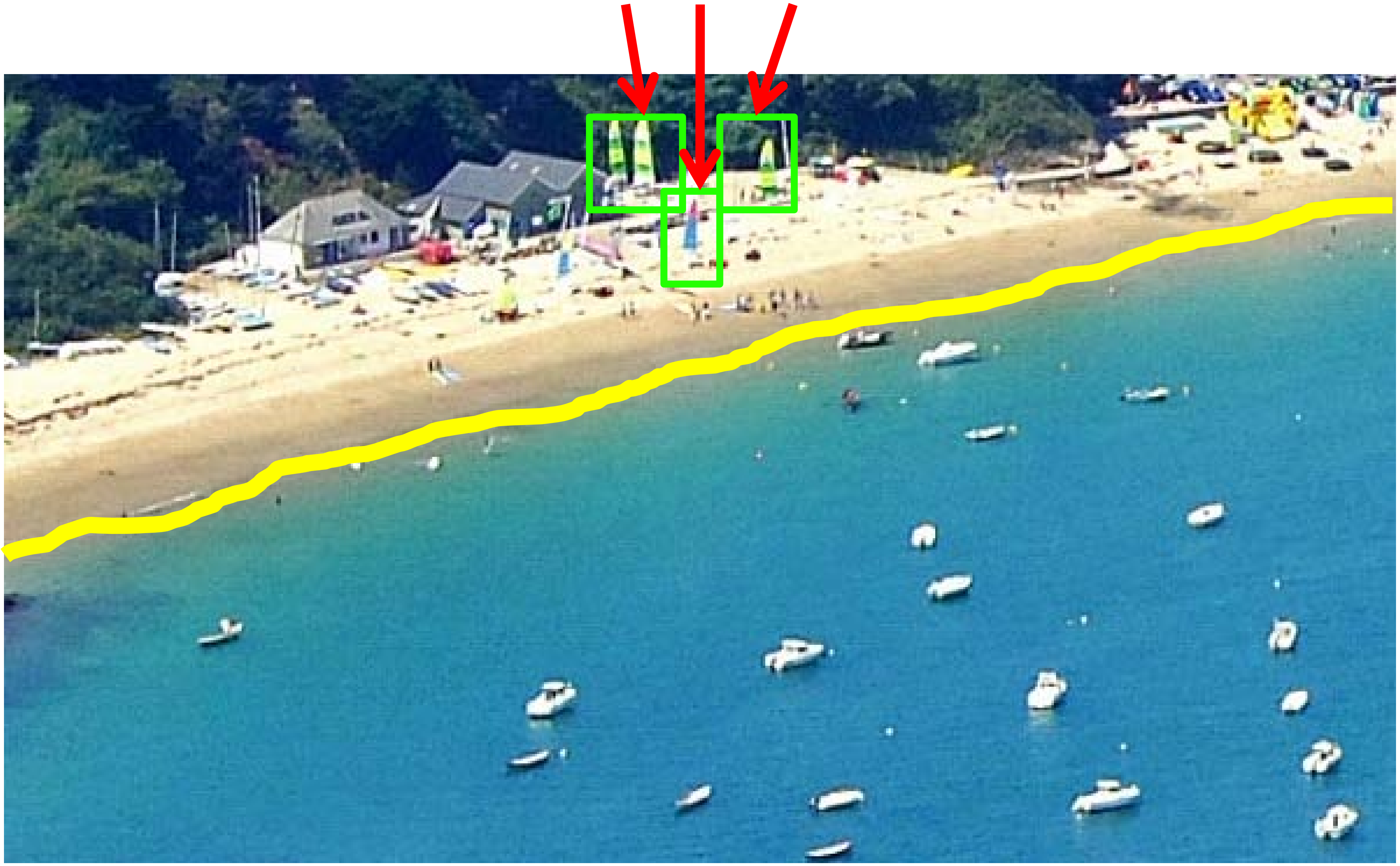


- All objects before the coast line are boats and all objects after the coast line are houses.
- Coast line serves as a *decision surface* that separates two classes.

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

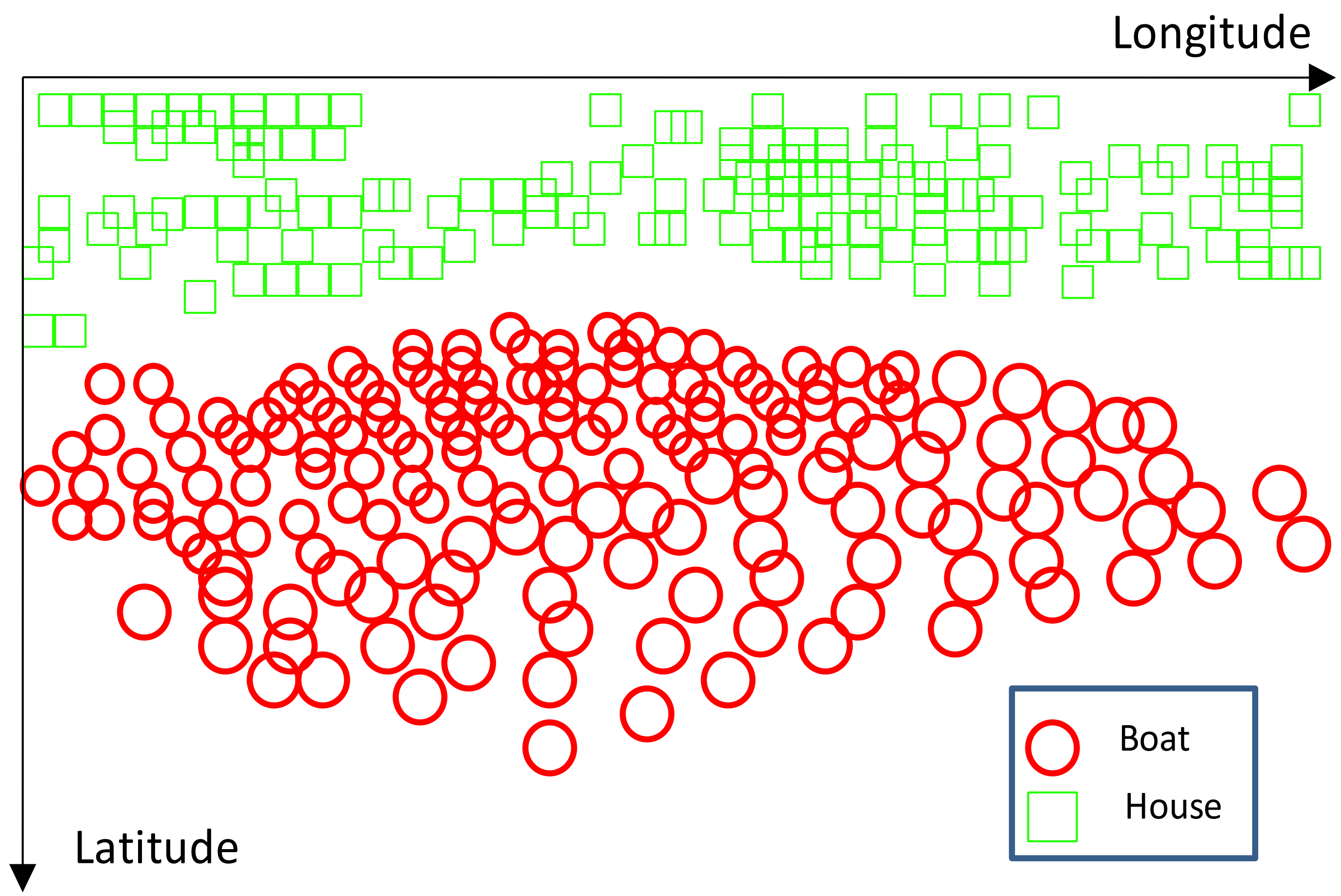
Basic principles of classification

These boats will be misclassified as houses



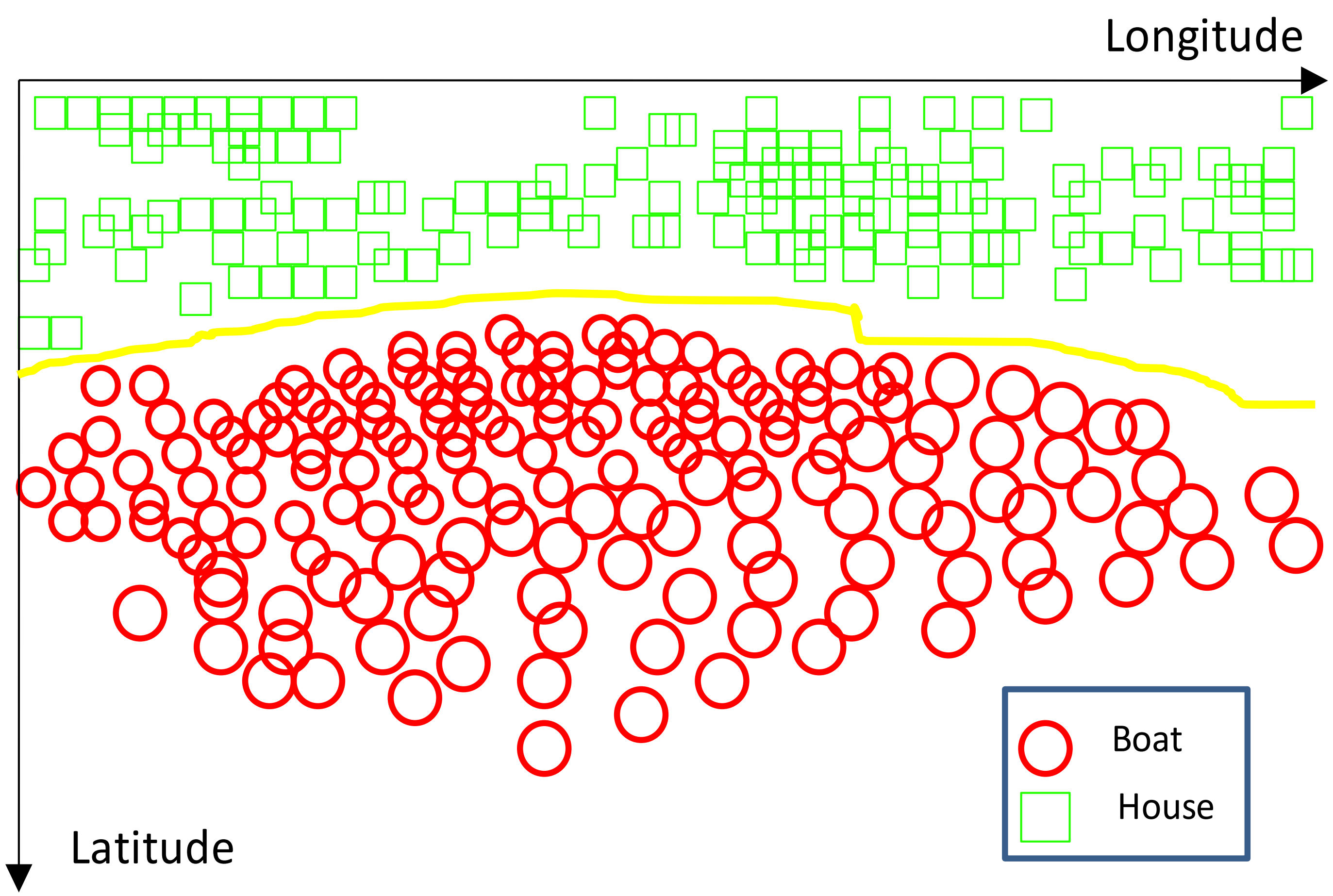
CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Basic principles of classification



- The methods that build classification models (i.e., "*classification algorithms*") operate very similarly to the previous example.
- First all objects are represented geometrically.

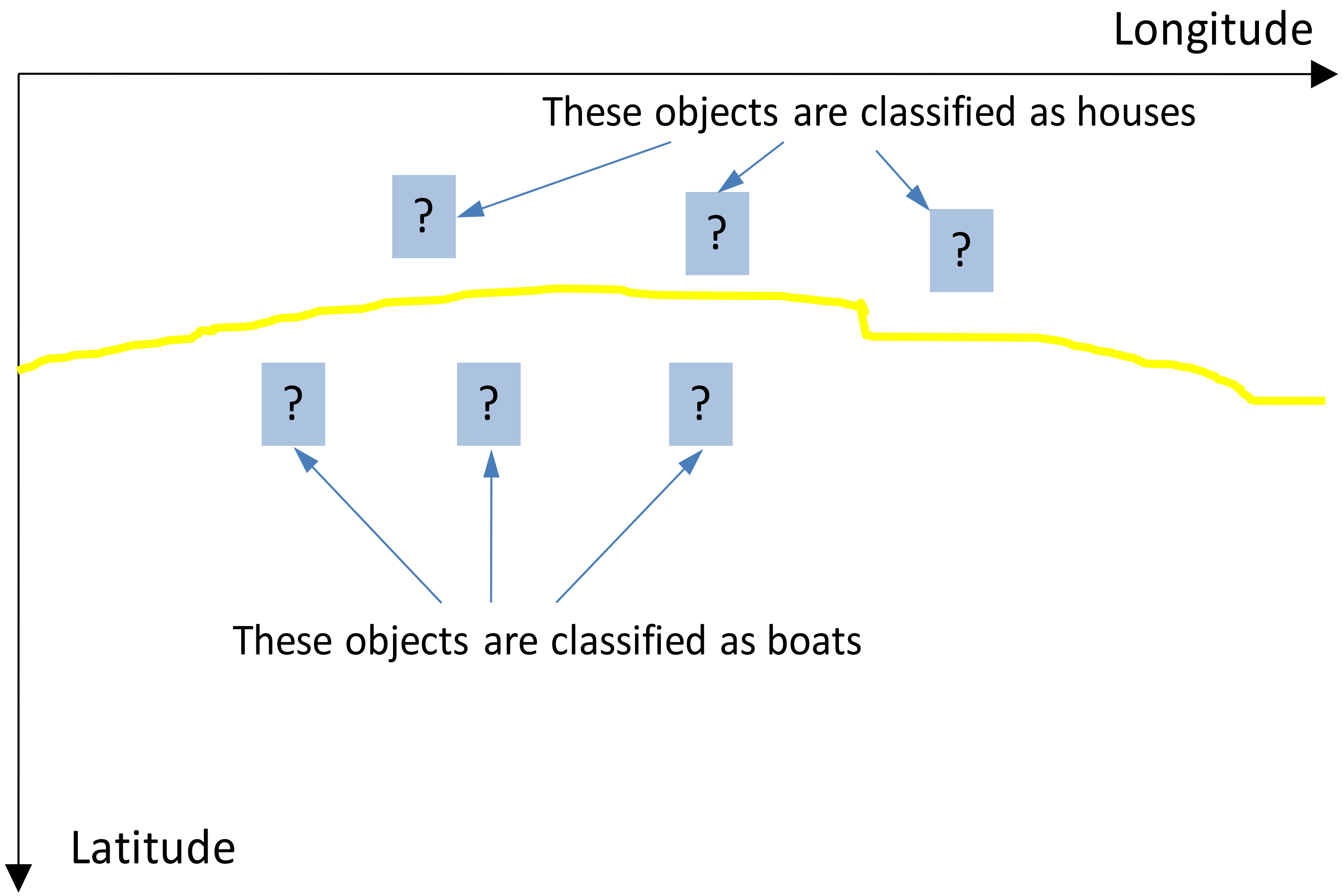
Basic principles of classification



Then the algorithm seeks to find a decision surface that separates classes of objects

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Basic principles of classification

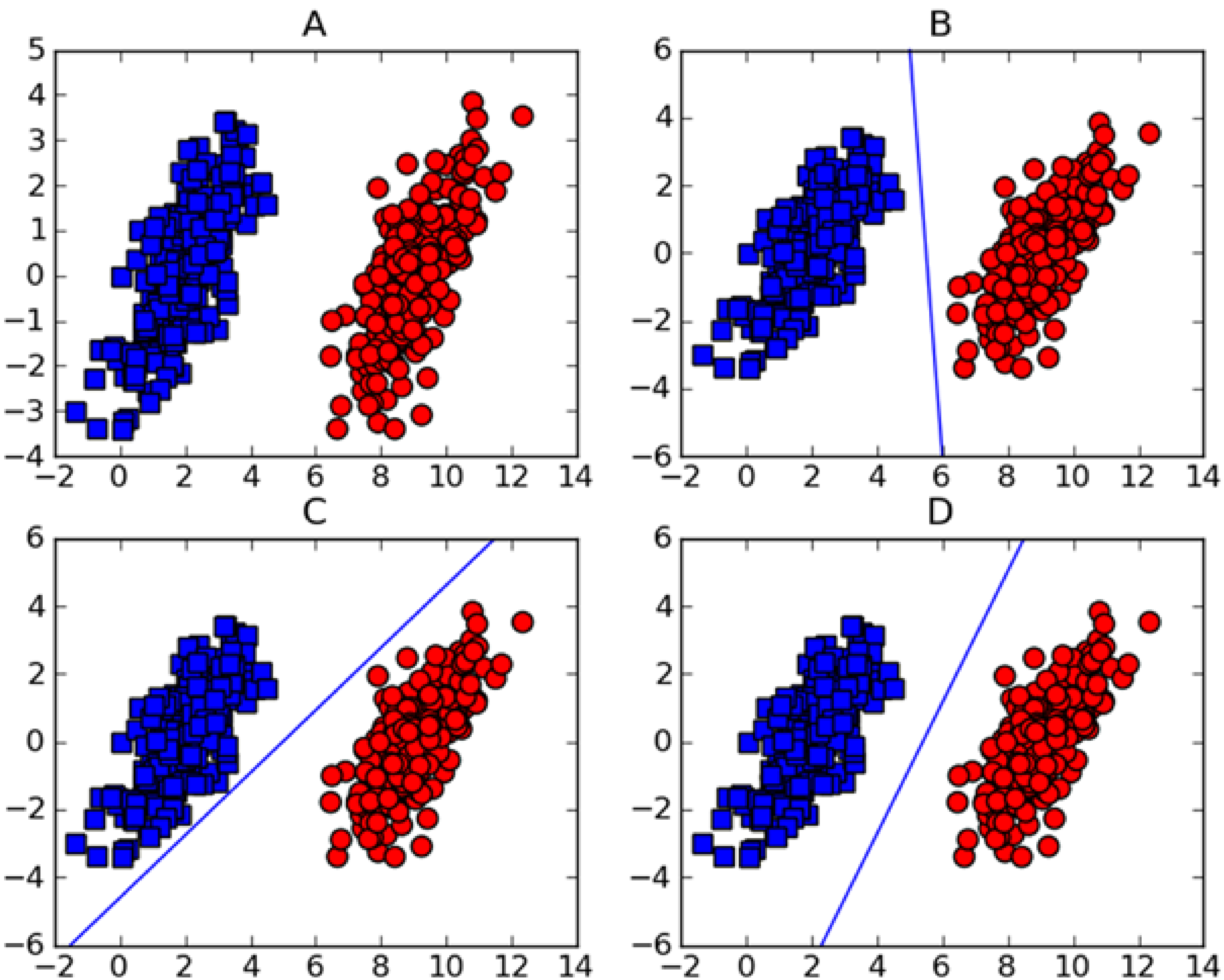


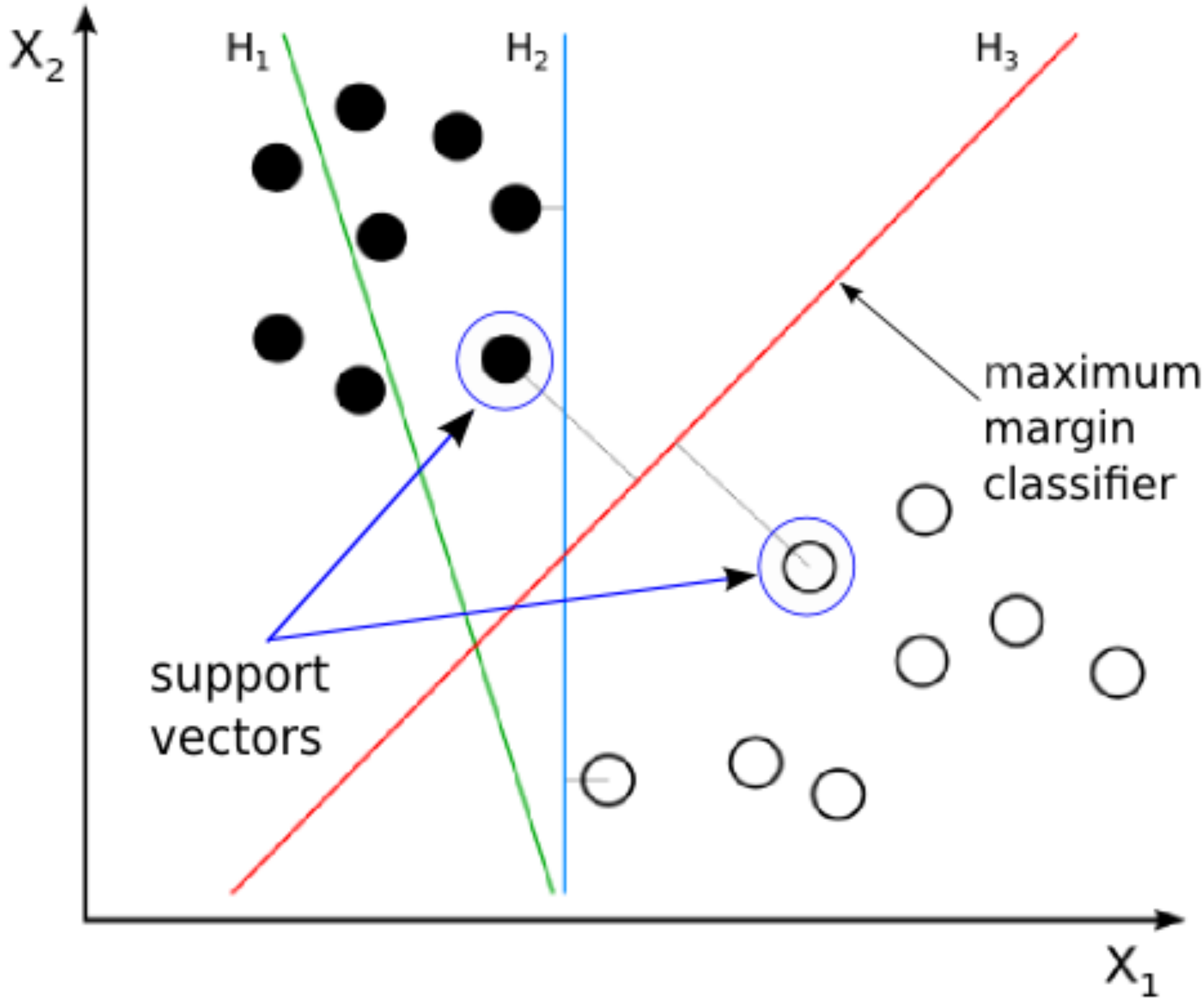
Unseen (new) objects are classified as “boats” if they fall below the decision surface and as “houses” if the fall above it

SVM

- **Support vector machines** are considered by some people to be the **best stock classifier**.
  - Support vector machines make **good decisions** for **data points** that are **outside** the **training set**.

CENTER FOR INNOVATION
AND RESEARCH
IN COMPUTING
Southern Adventist University

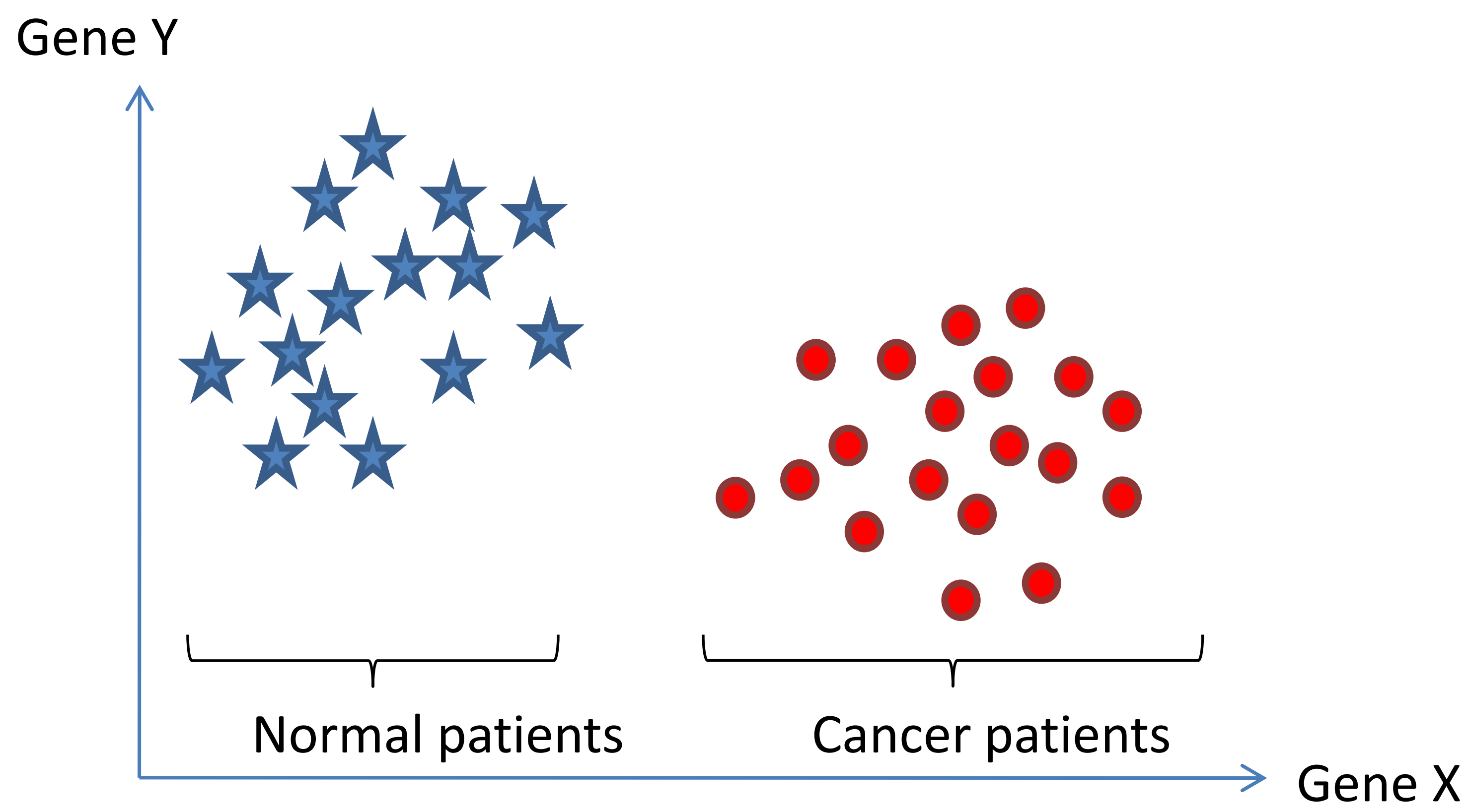

A
B
C
D
5 4 3 2 1 0 −1 −2 −3 −4
6 4 2 0 −2 −4 −6
−2 0 2 4 6 8 10 12 14



The point closest to the separating hyperplane and make sure this is as far away from the separating line as possible.

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

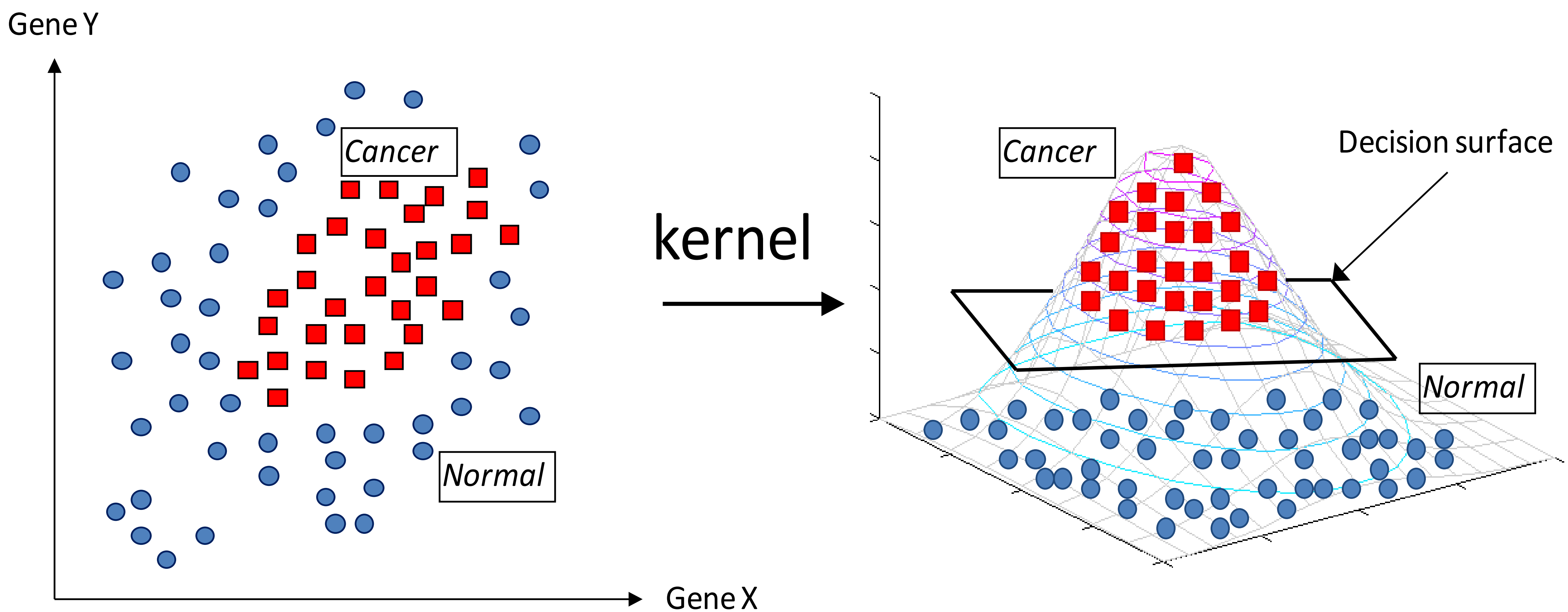
Main ideas of SVMs



- Consider example dataset described by 2 genes, gene X and gene Y
- Represent patients geometrically (by "vectors")

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Main ideas of SVMs

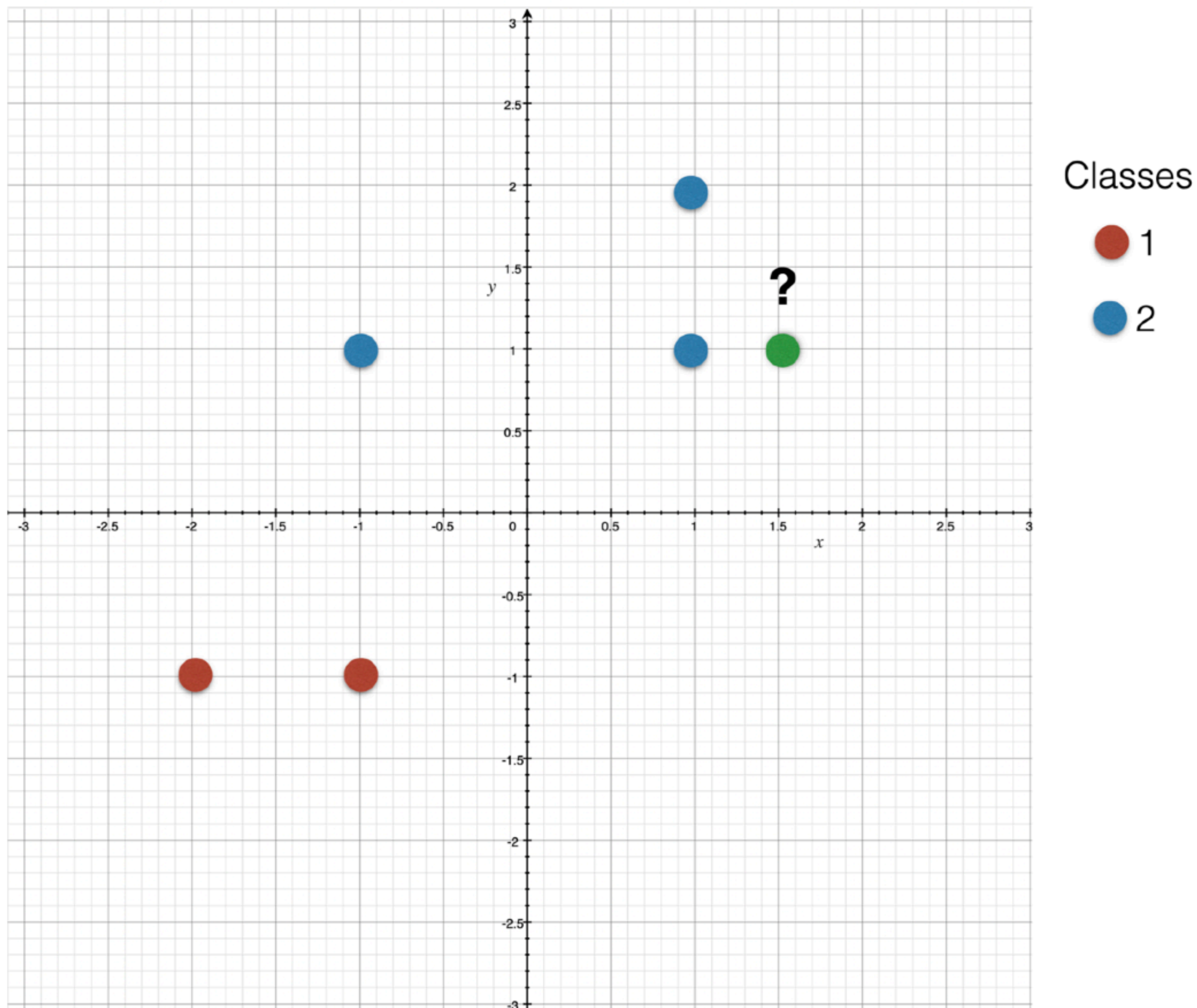


- If such linear decision surface does not exist, the data is mapped into a much higher dimensional space ("feature space") where the separating decision surface is found;
- The feature space is constructed via very clever mathematical projection ("kernel trick").

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

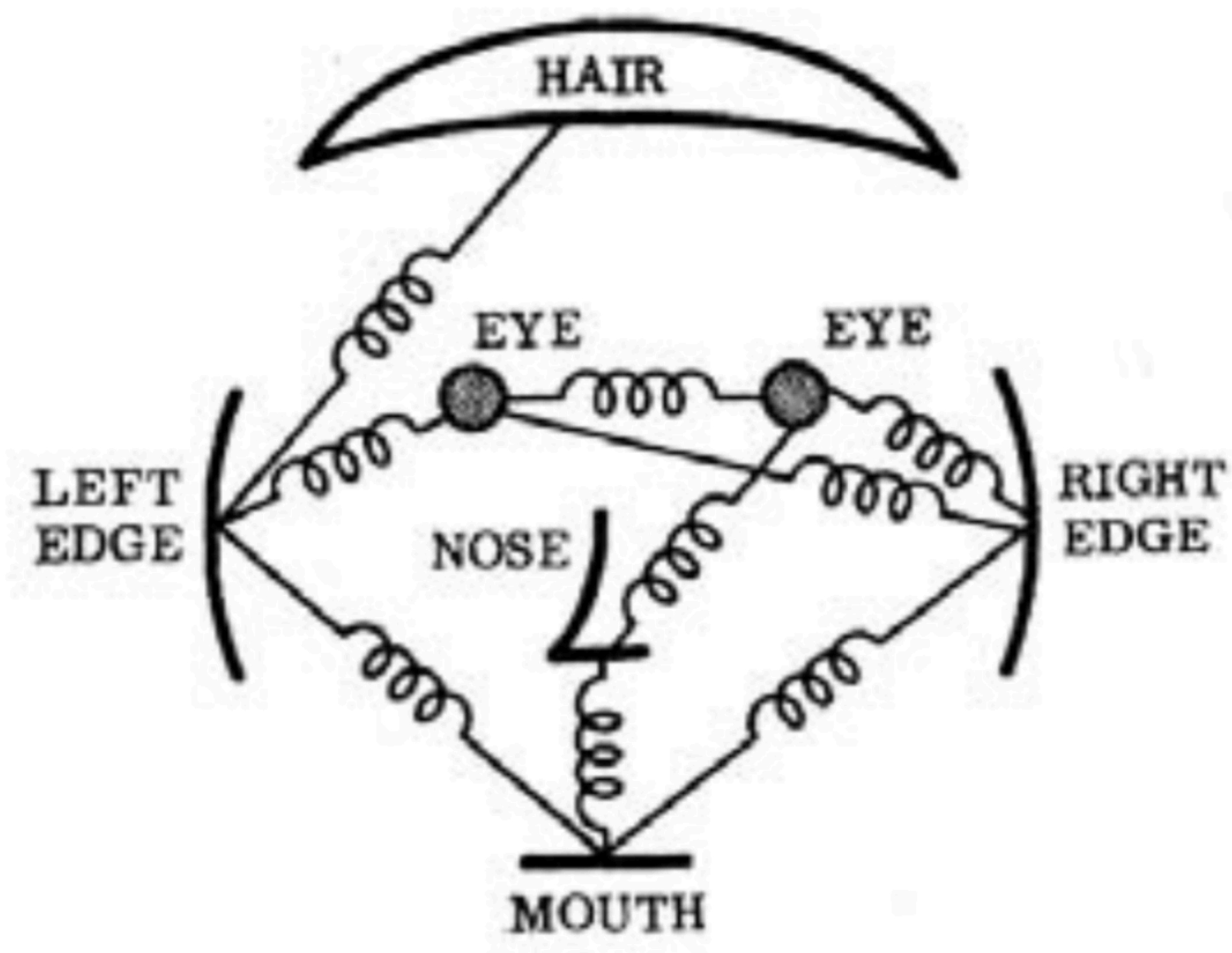
```python
import numpy as np
features = np.array([[-1, -1], [-2, -1], [1, 2], [-1, 1], [1, 1]])
labels = np.array([1, 1, 2, 2, 2])
from sklearn.svm import SVC
clf = SVC()
clf.fit(features, labels)
print(clf.predict([[1.5, 1]]))
```

```
[2]
```



Supervised Learning through

Deep Learning

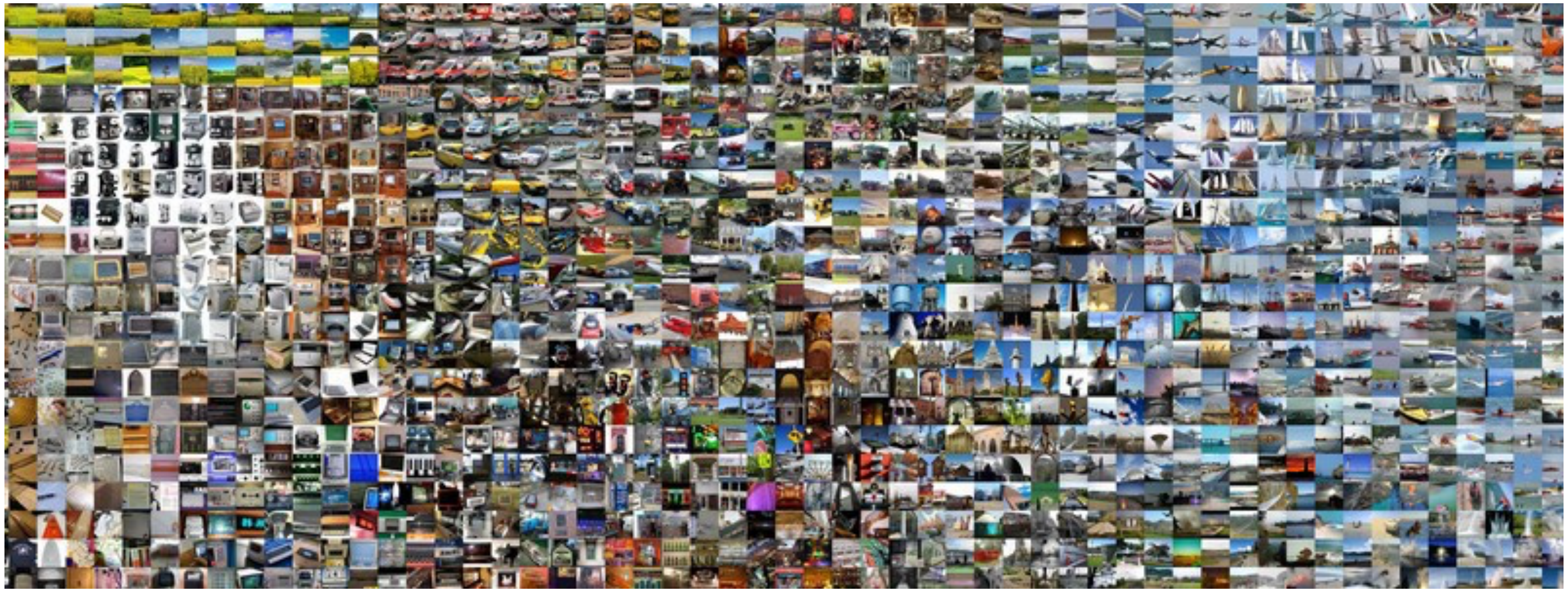


Pictorial structures (Fischler and Elschlager, 1973)

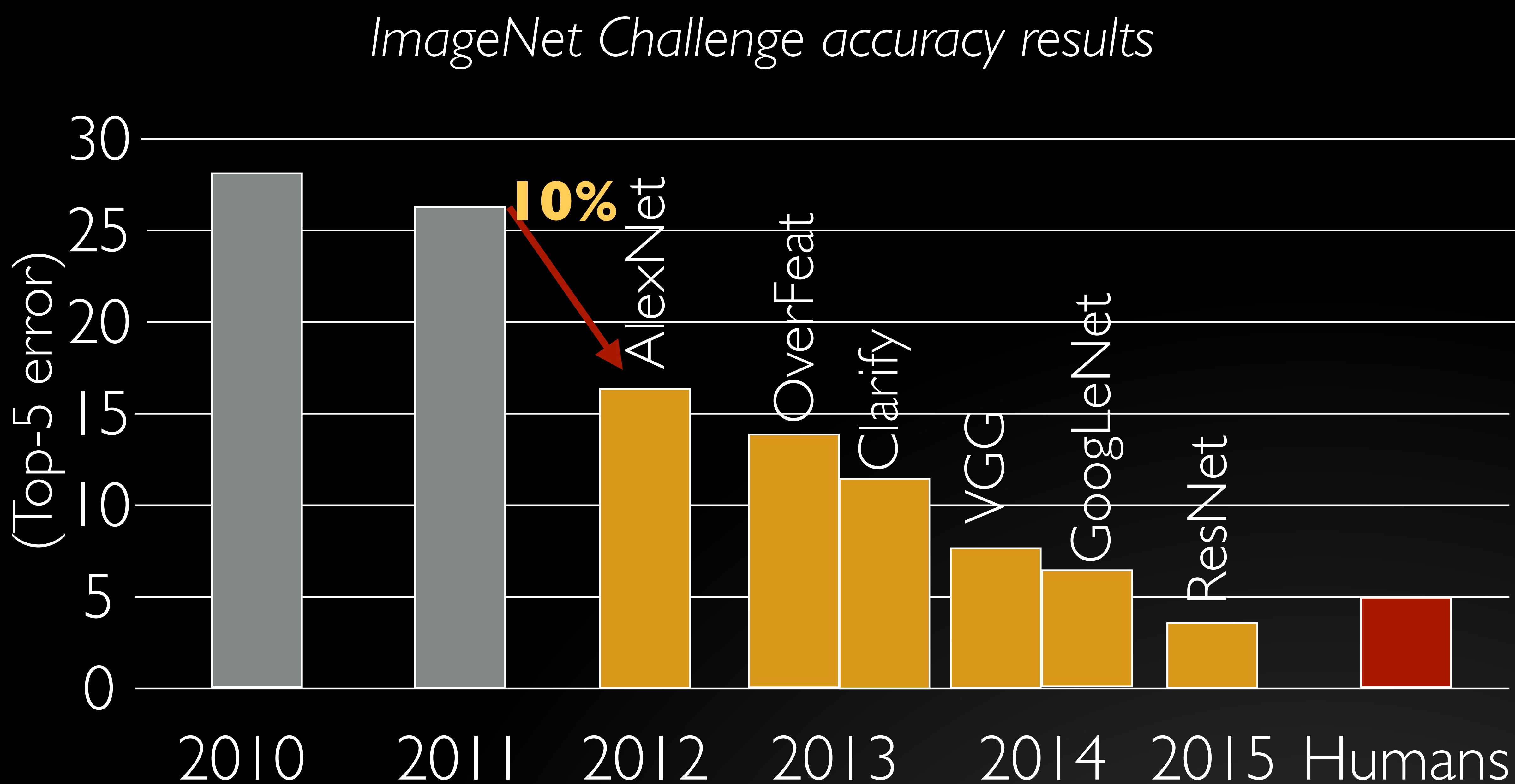
CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Why is Everyone So Excited about Deep Learning?

- Computer vision
  - ImageNet (1 million labeled images)
    - Yearly contest: train a classifier with the smallest possible error rate
    - Since 2012 - deep learning has been the winner!



CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University



Sze, V., Chen, Y.-H. y Yang, T.-J. (2017). Efficient processing of deep neural networks: a tutorial survey. *Proceedings of the IEEE*, 105. https://doi.org/10.1109/JPROC.2017.2761740

Deep Learning

- Deep learning = layering
- **Goal:** build complex, hierarchical representations from simple building blocks.
- Traditional artificial neural networks = few layers
- Deep artificial neural networks = several layers (6 - 8, similar to the visual cortex)

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

input → function → output

fundus of the eye picture

machine learning

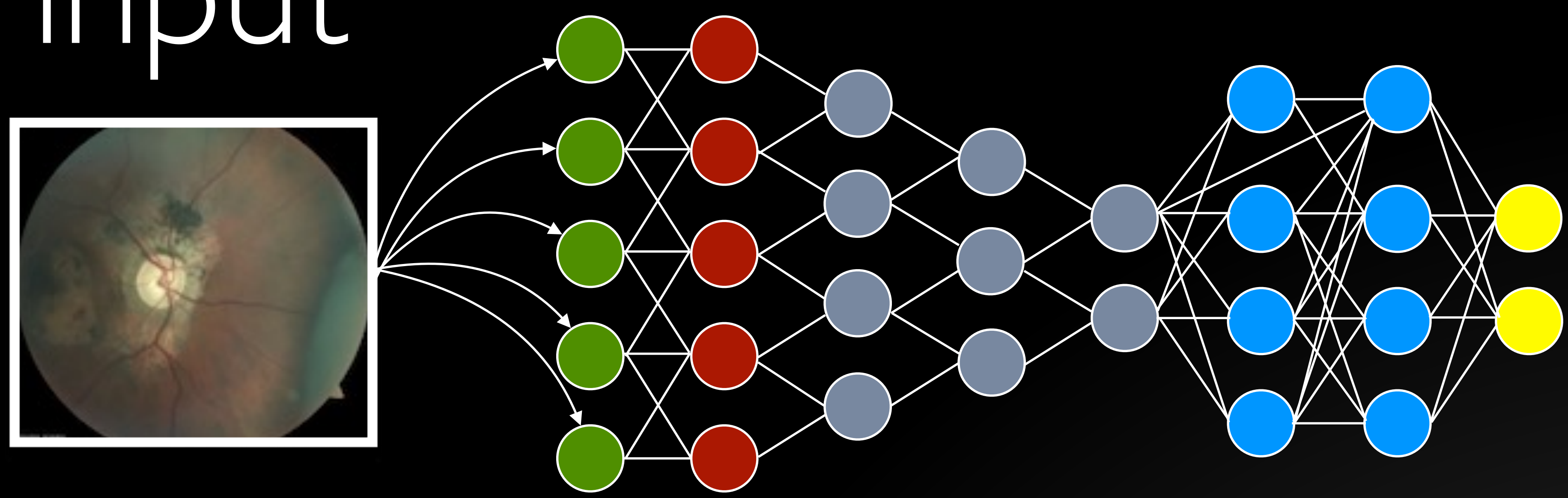
probability of classes

glaucoma = 0.99624

no glaucoma = 0.00869

input

output



glaucoma = 0.99624

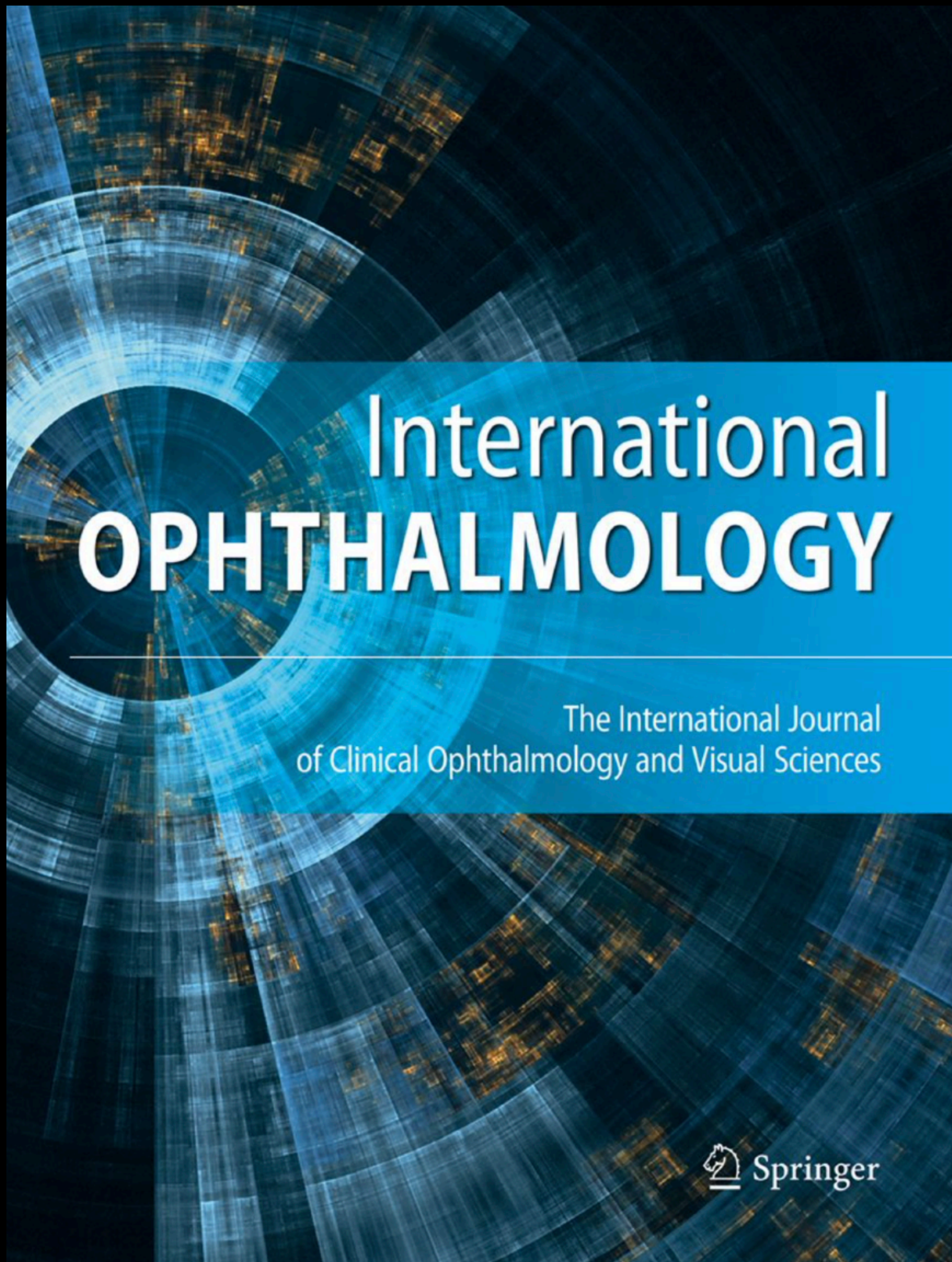
no glaucoma = 0.00869

fundus of the eye picture

Imágenes de pacientes **sin glaucoma**	Clasificación automática: **no glaucoma**
	0.99762
	0.99444
	0.9948

Imágenes de pacientes **con glaucoma**	Clasificación automática: **glaucoma**
	0.99754
	0.99989
	0.99624

Espinoza, M., Alférez, G.H. y Castillo, J. (2018). Prediction of glaucoma through convolutional neural networks. En H. R. Arabnia, L. Deligiannidis, G. Jandieri, A. M. G., F. G. Tinetti y Q.-N. Tran (Eds.), *International Conference on Health Informatics and Medical Systems: The 2018 WorldComp International Conference Proceedings* (p. 90-95). CSREA Press.



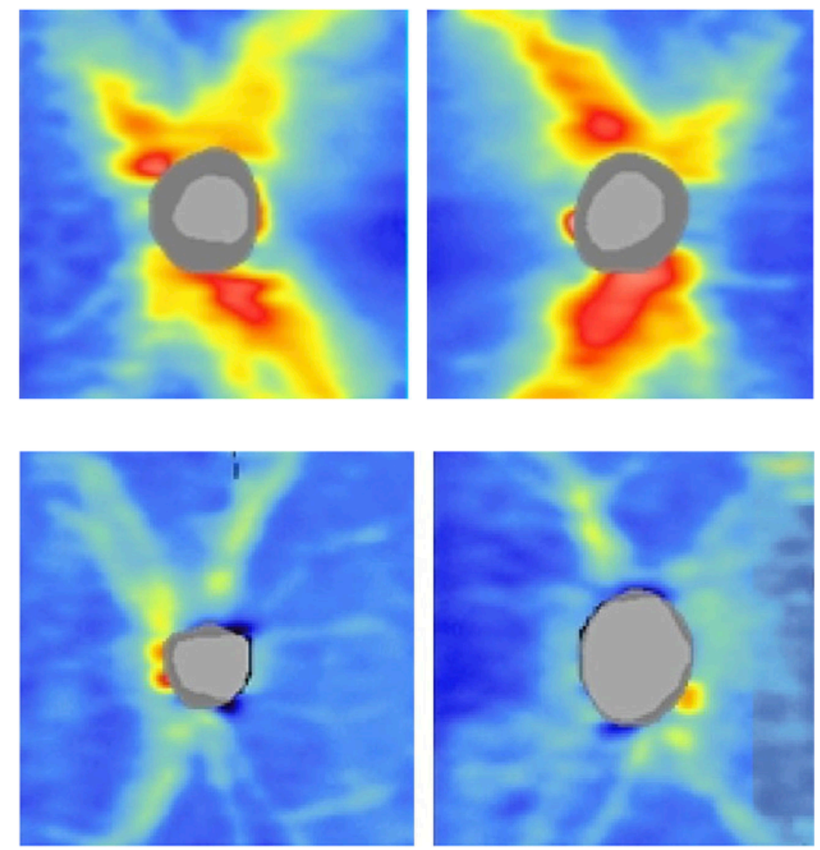


Fig. 3 Non-glaucomatous left and right eyes, respectively

Fig. 4 Glaucomatous left and right eyes, respectively

Olivas, L.G., Alférez, G.H., & Castillo, J. (2021). Glaucoma detection in Latino population through OCT's RNFL thickness map using transfer learning. International Ophthalmology. https://doi.org/10.1007/s10792-021-01931-w

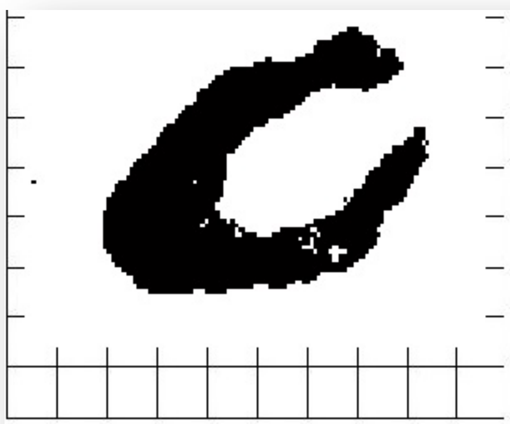
“Misdiagnosis in the United States is disconcertingly common. A review of three very large studies concluded that there are about 12 million significant misdiagnoses a year.”

— Eric Topol. “Deep Medicine.”

1

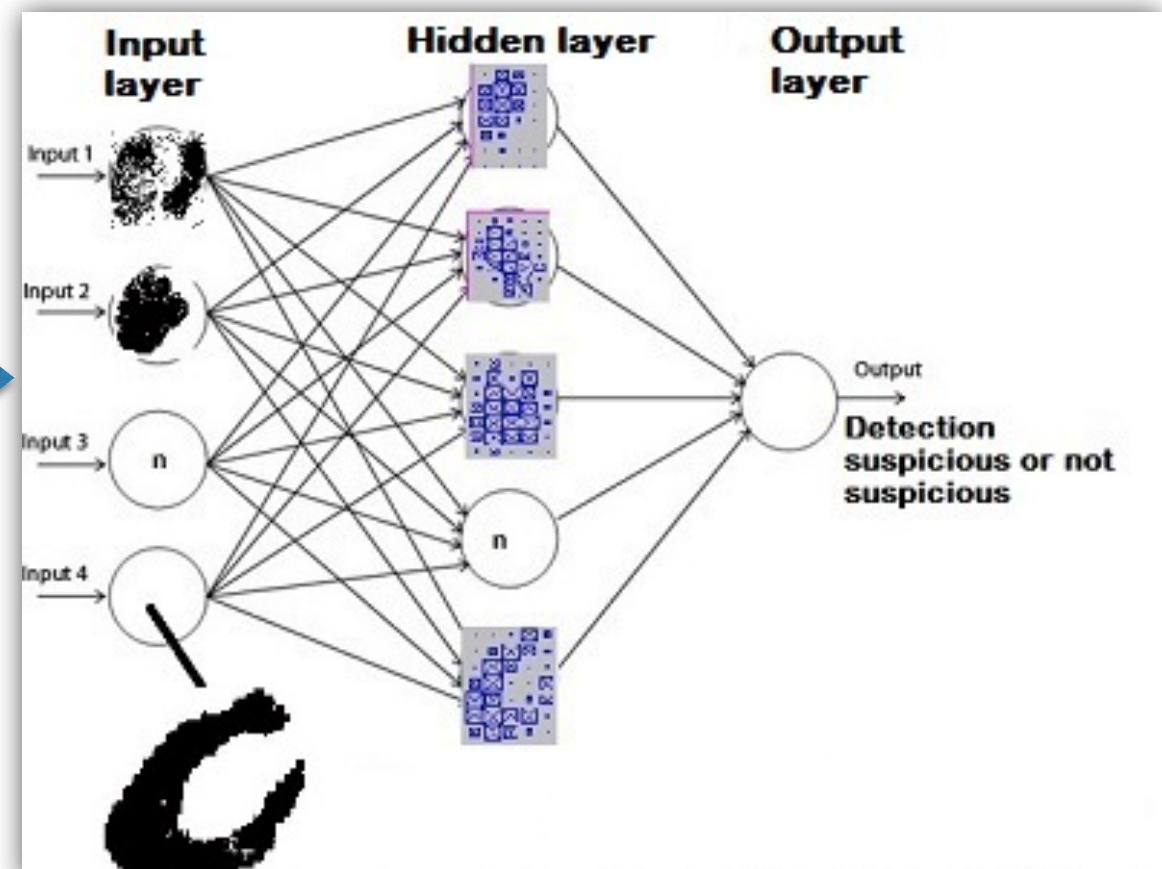
2

3



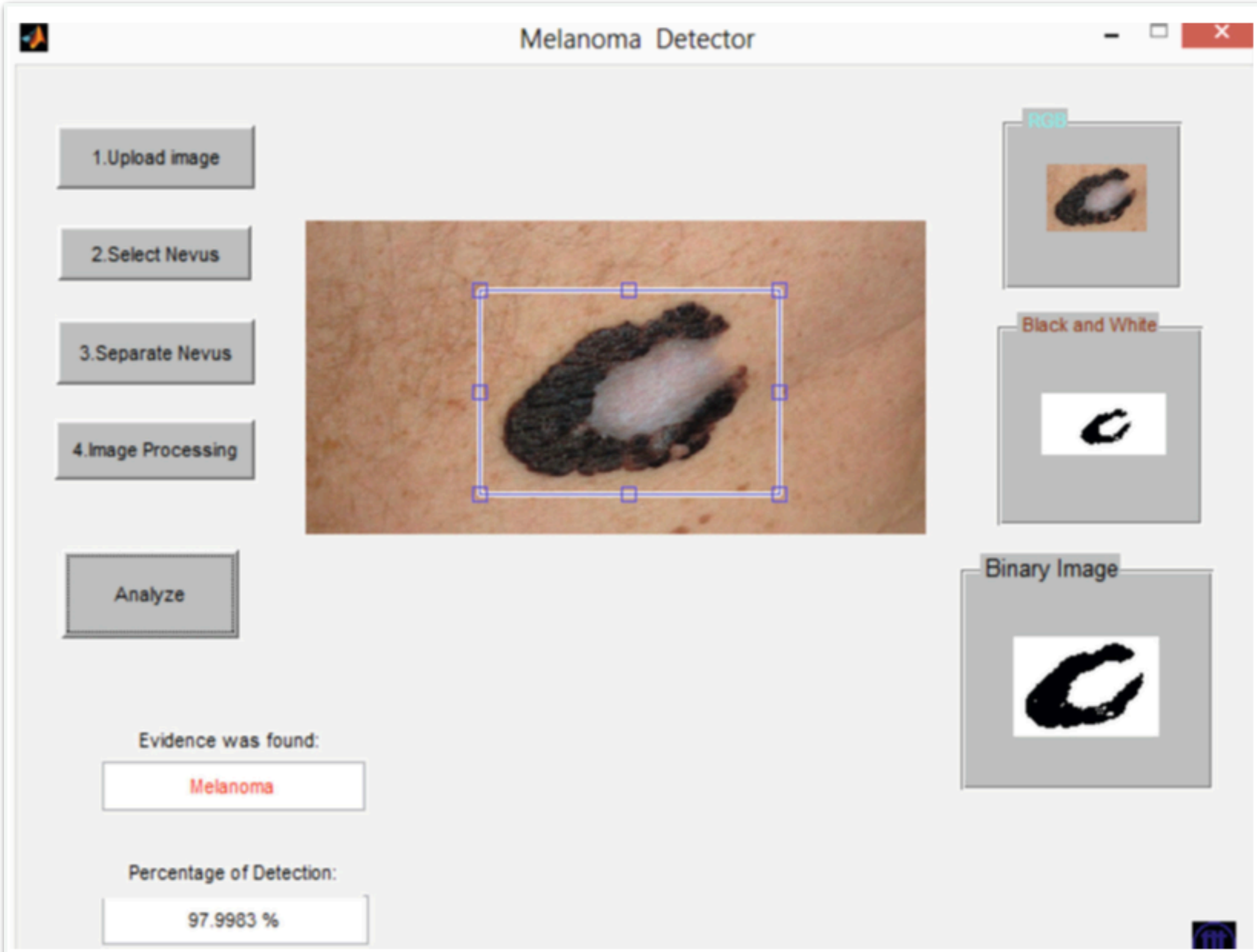
0	0	0	1	1	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0
0	0	1	1	1	1	1	0	0	0
0	1	1	0	0	1	1	1	0	0
0	1	1	1	1	1	1	1	1	0
0	1	0	0	0	0	0	1	1	0
0	1	0	0	0	0	0	0	1	1
0	1	0	0	0	0	0	0	1	1

4



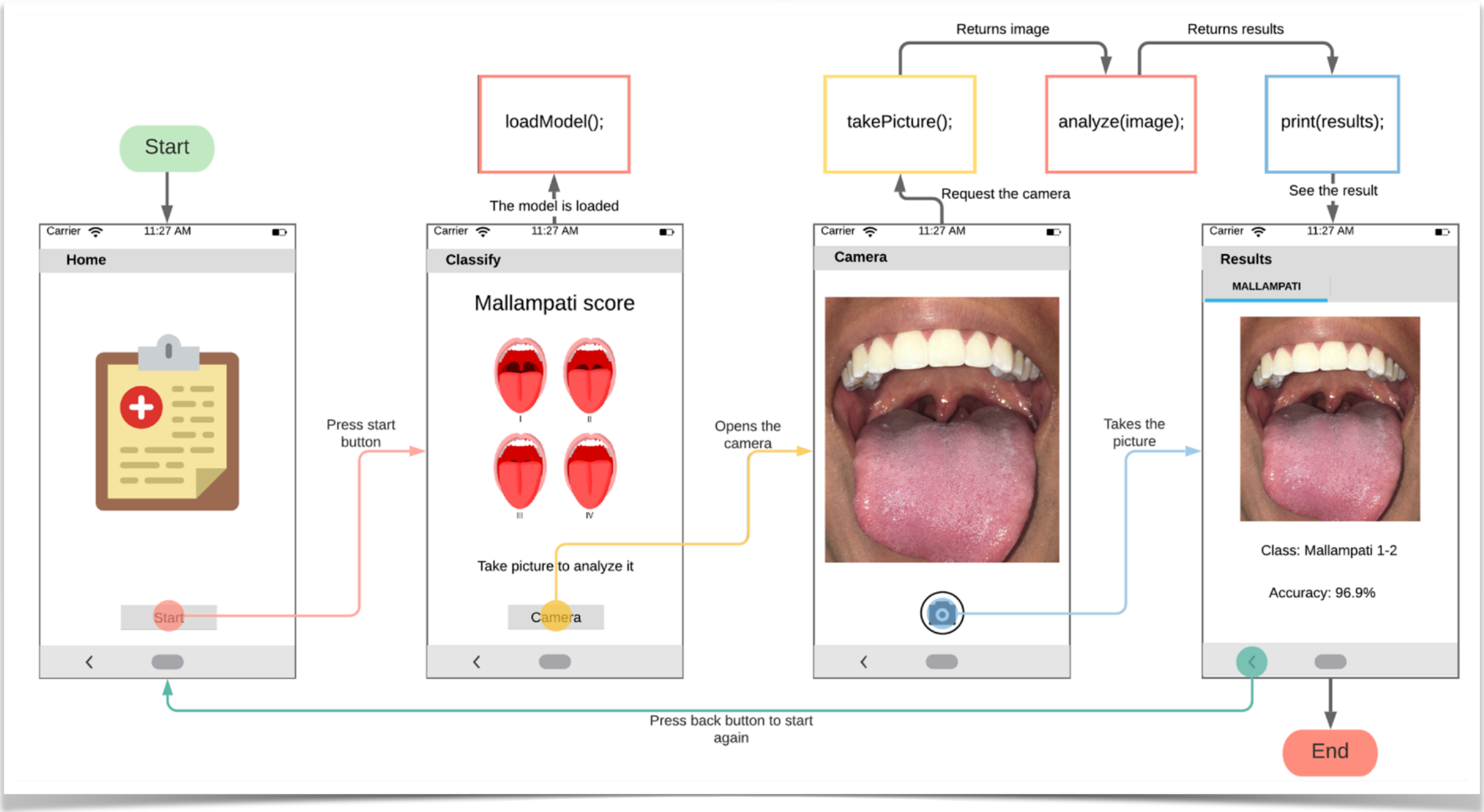
Marín, C., Alférez, G.H. Córdova, J. y González, V. (2015). Detection of melanoma through image recognition and artificial neural networks. En D. A. Jaffray (Ed.), *World congress on medical physics and biomedical engineering* (pp. 832–835). Springer.

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University



Marín, C., Alférez, G.H. Córdova, J. y González, V. (2015). Detection of melanoma through image recognition and artificial neural networks. En D. A. Jaffray (Ed.), *World congress on medical physics and biomedical engineering* (pp. 832–835). Springer.

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University



Aguilar, K, Alférez, G.H. y Aguilar, C. (2020). Detection of difficult airway using deep learning. *Machine Vision and Applications*, 31. https://doi.org/10.1007/ s00138-019-01055-3

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Sylnice, J. & Alférez, G.H. (2018). Dynamic Evolution of Simulated Autonomous Cars in the Open World Through Tactics. *Proceedings of the Future Technologies Conference 2018 (FTC 2018)*, Vancouver, Canada.

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

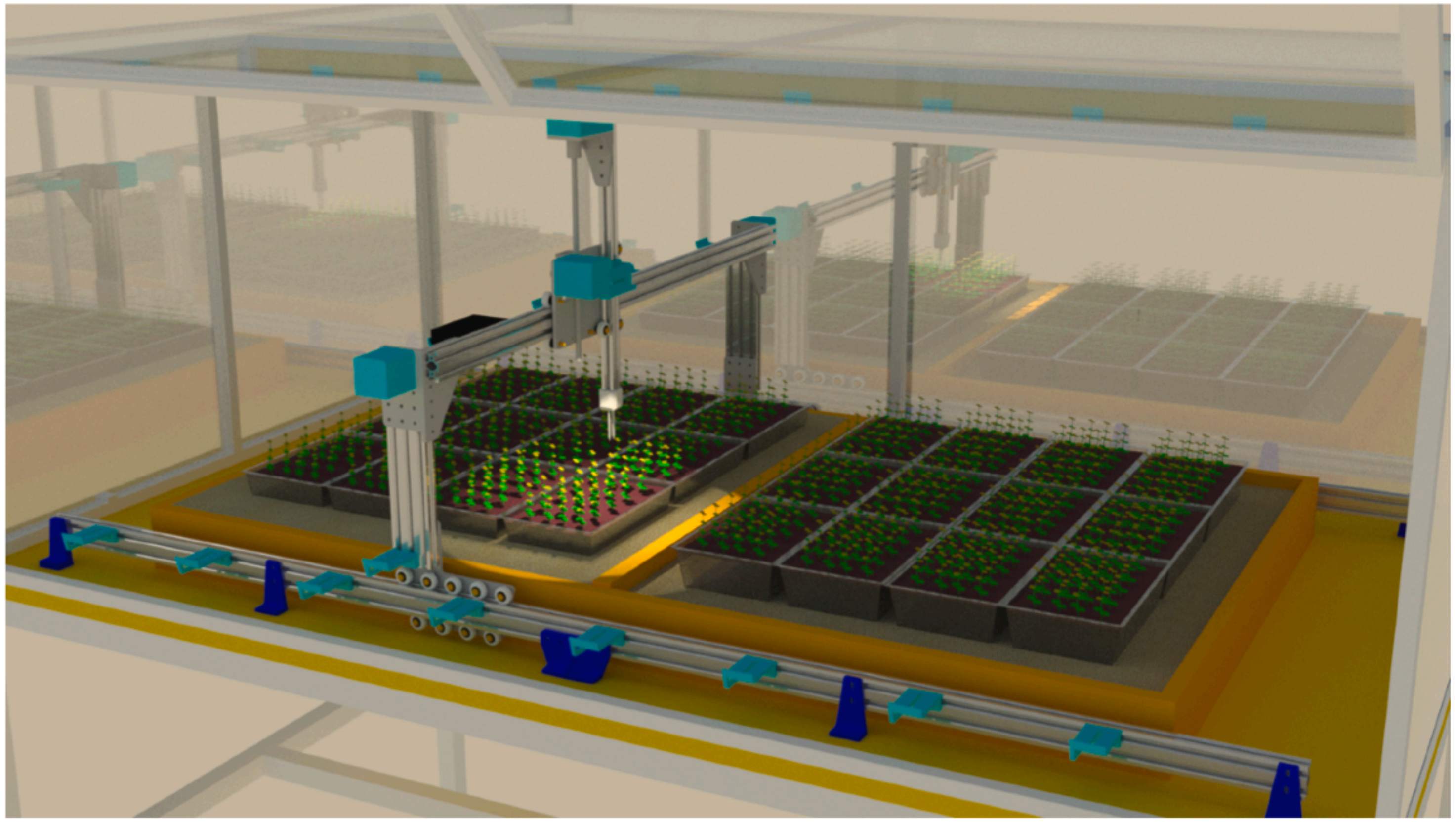


Fig. 1. CNC robotic system.

Fig. 2. Tray with living Legacy blueberry plants.

Fig. 3. Tray without living Legacy blueberry plants.

Fig. 4. No tray.

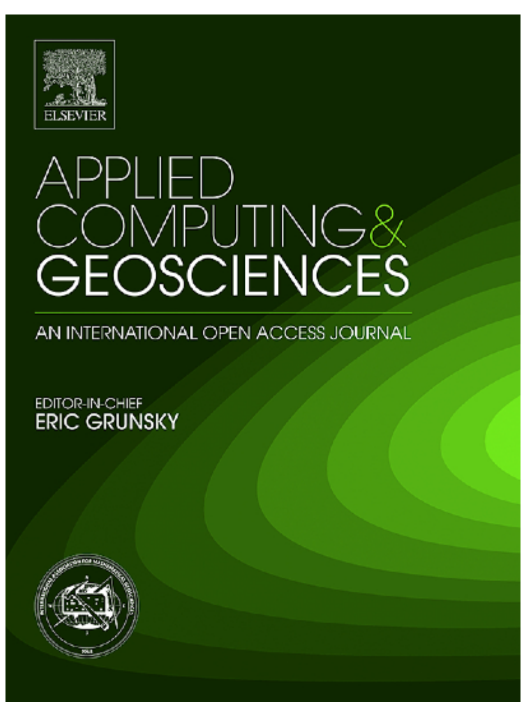
Quiroz, I. A. y Alférez, G. H. (2020). Image recognition of Legacy blueberries in a Chilean smart farm through deep learning. Computers and Electronics in Agriculture, 168, 105044. https://doi.org/10.1016/ j.compag.2019.105044

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Automatic classification of plutonic rocks with deep learning

Germán H. Alférez, Elías L. Vázquez, Ana María Martínez Ardila, Benjamin Clausen

Applied Computing and Geosciences - Elsevier
https://doi.org/10.1016/j.acags.2021.100061



GEOSCIENCE RESEARCH INSTITUTE

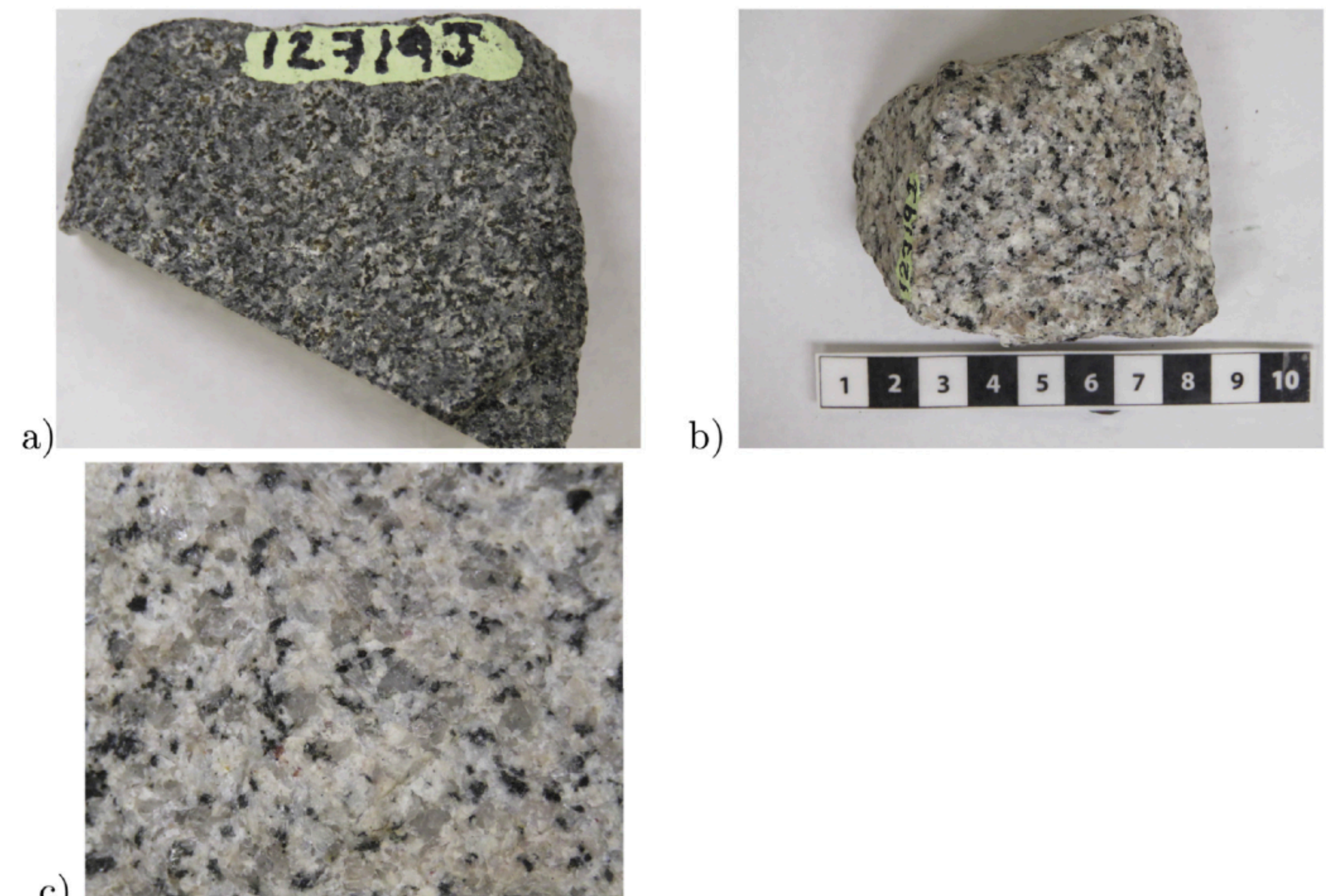


Fig. 2. a) Diorite image taken under white light; b) granite image with a ruler; and c) a cropped granite picture.

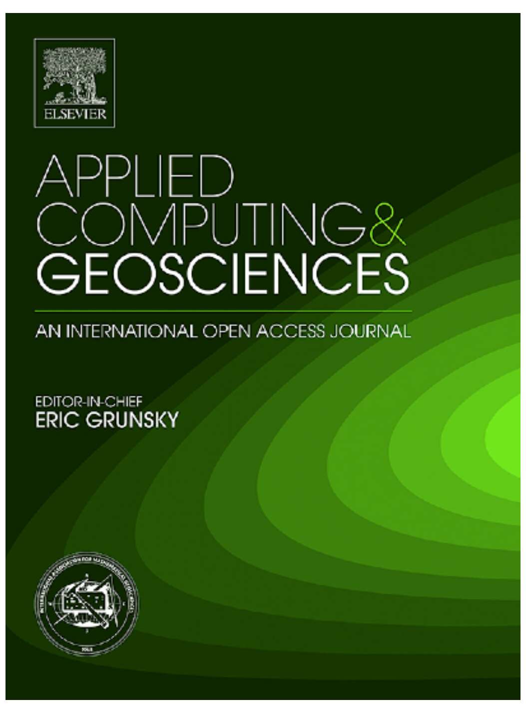
Automatic classification of plutonic rocks with deep learning

Germán H. Alférez, Elías L. Vázquez, Ana María Martínez Ardila, Benjamin Clausen

Applied Computing and Geosciences - Elsevier
https://doi.org/10.1016/j.acags.2021.100061

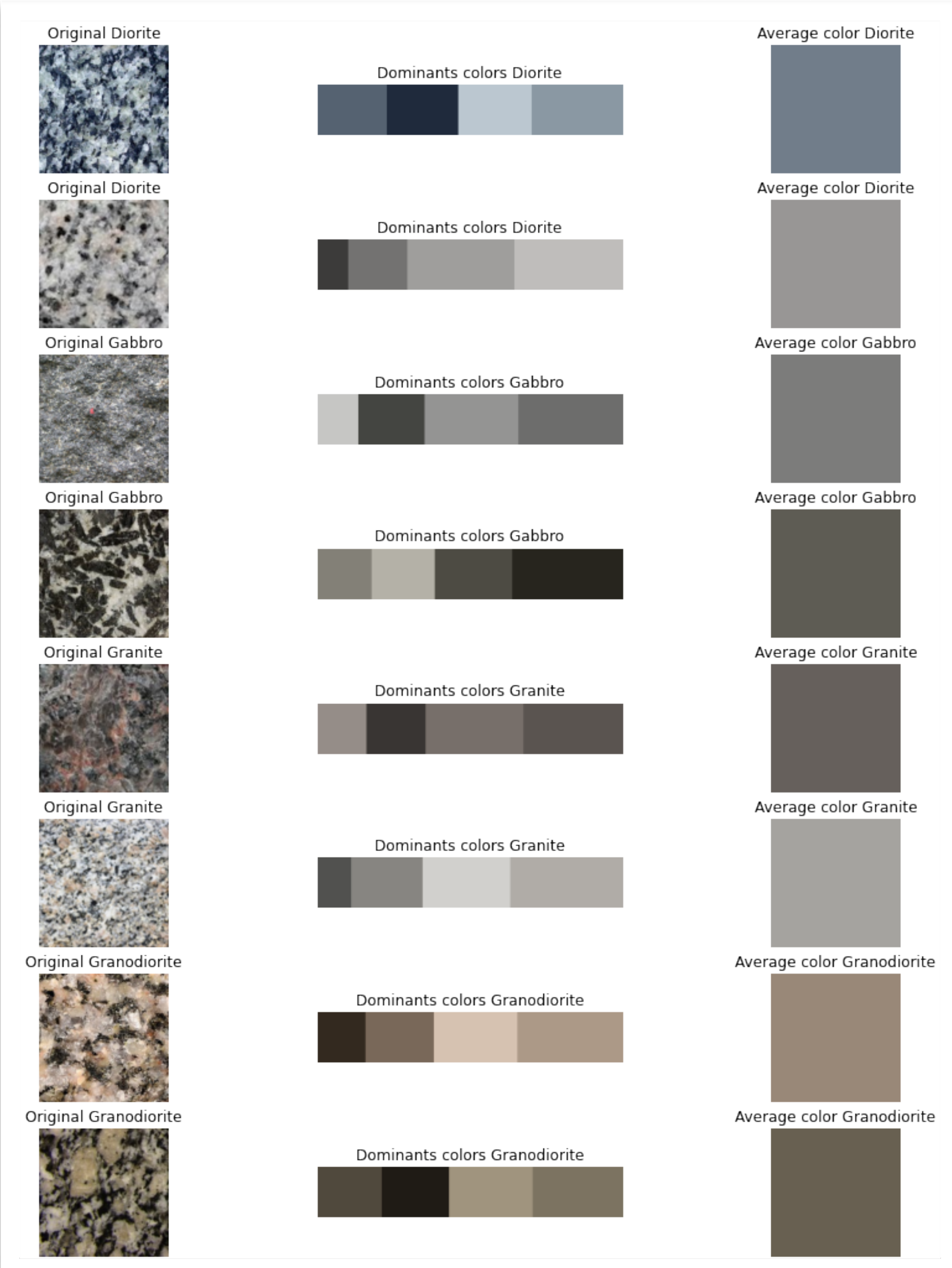
Rock Classes	Precision	Recall	F_1 Score
gabbro	0.98	0.94	0.96
diorite	1.00	0.94	0.97
granodiorite	0.85	1.00	0.92
granite	1.00	0.94	0.97
Average	***0.96***	***0.95***	***0.95***

Table 5: Validation results per class for the gabbro, diorite, granodiorite, and granite combi
nation with an accuracy of 0.95.



GEOSCIENCE
RESEARCH INSTITUTE

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

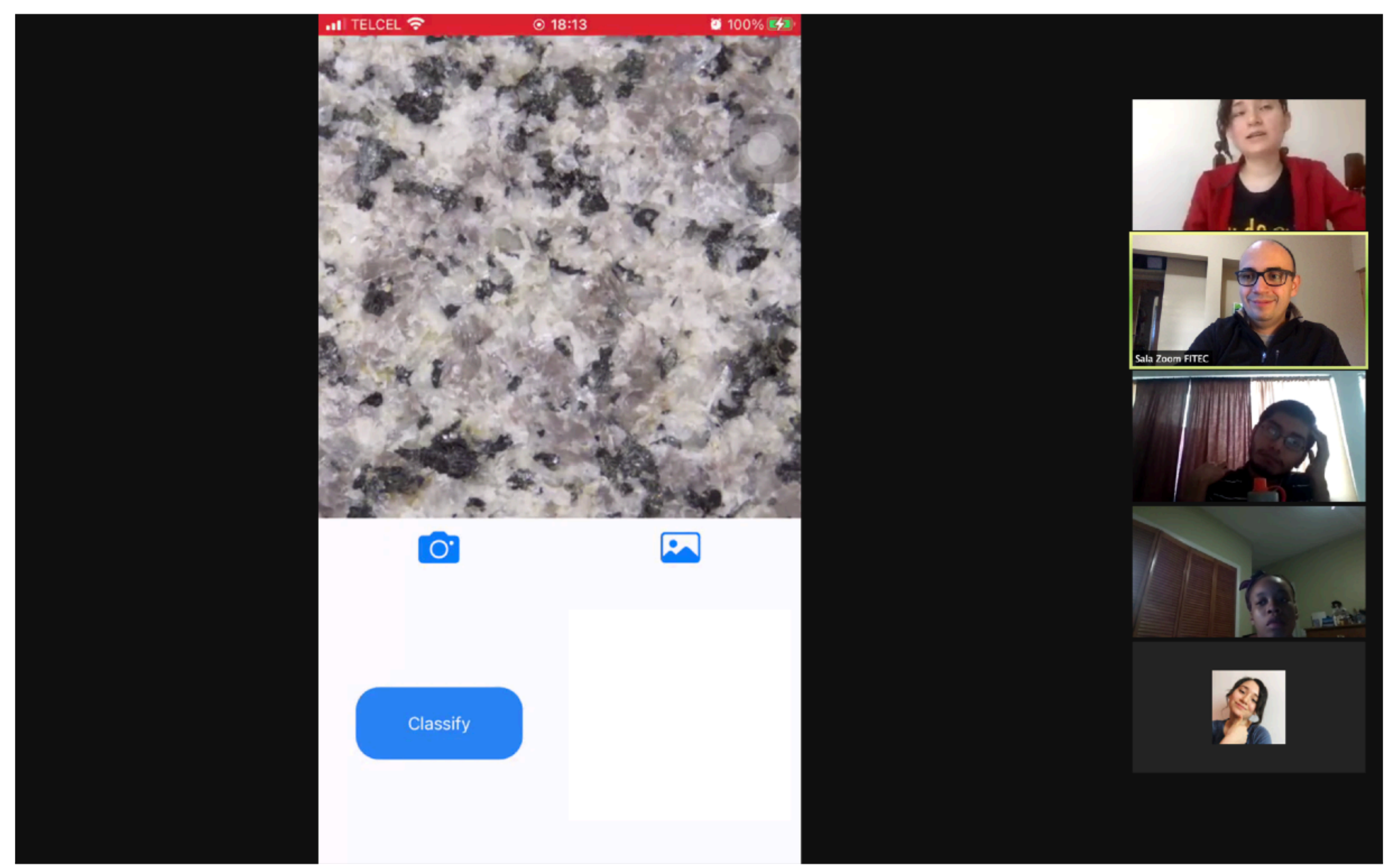


Automatic Classification of Plutonic Rocks with Machine Learning Applied to Extracted Shades and Colors on iOS Devices

Alférez, G.H., Hernández Serrano, S., Martínez Ardila, A.M., & Clausen, B.L. (2021). Automatic Classification of Plutonic Rocks with Machine Learning Applied to Extracted Shades and Colors on iOS Devices.

THE GEOLOGICAL SOCIETY OF AMERICA

FTC 2021

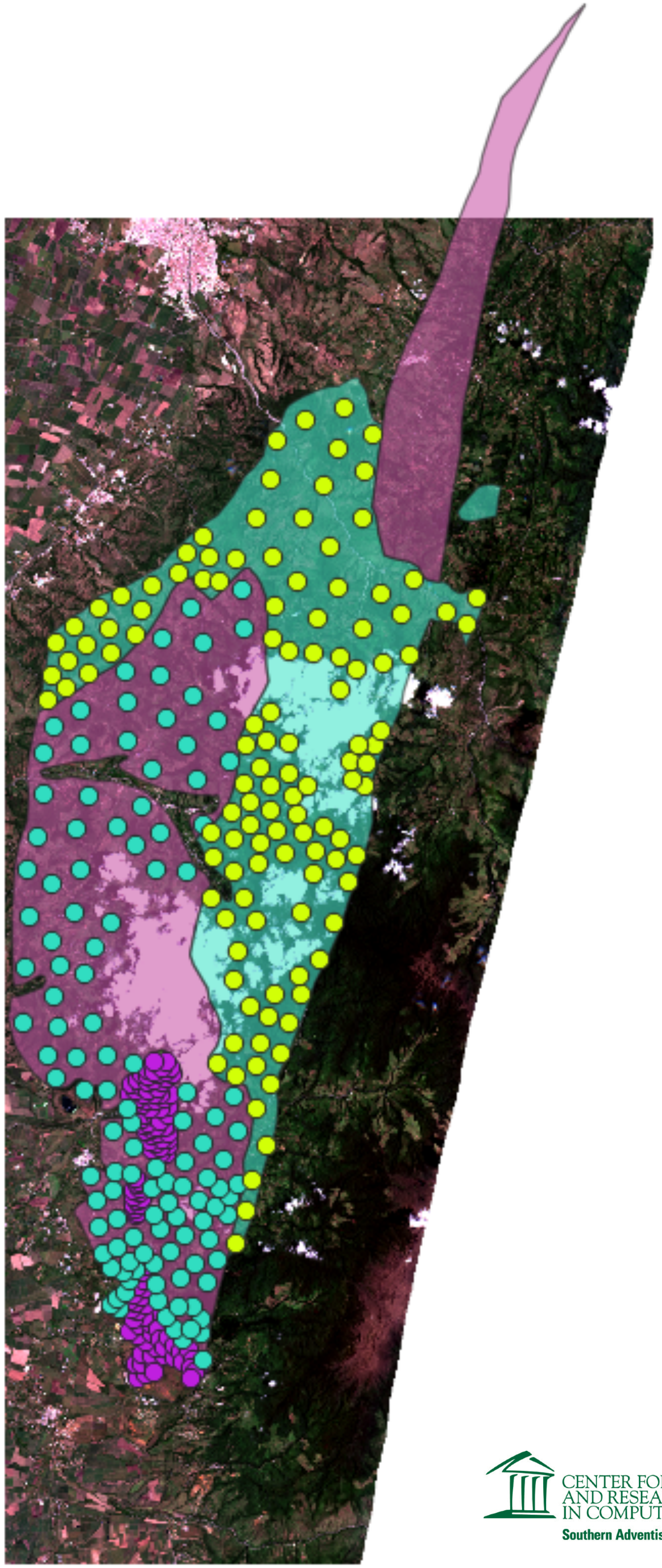


GEOSCIENCE RESEARCH INSTITUTE

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

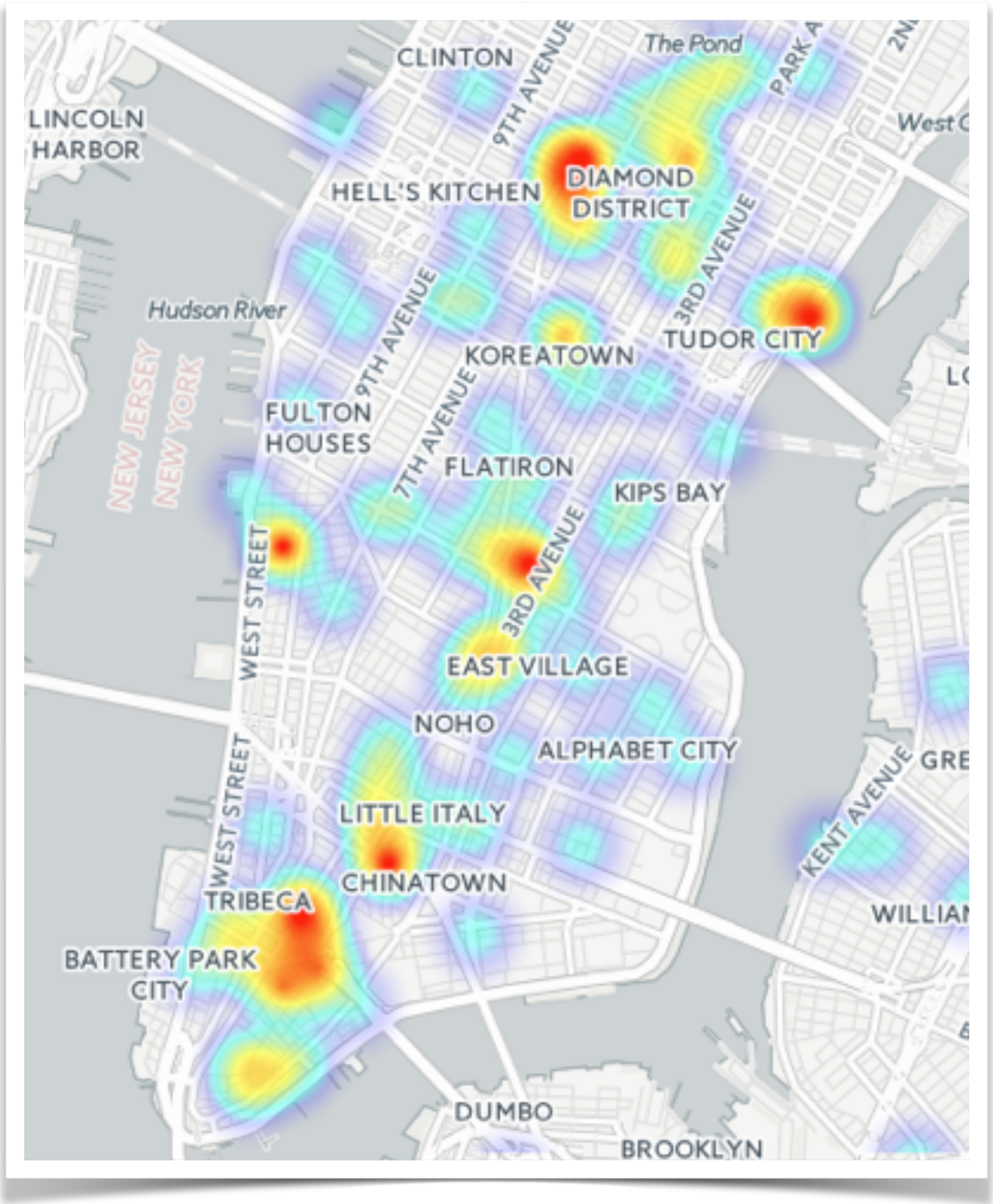
Automatic Classification of Felsic, Mafic, and Ultramafic Rocks in Satellite Images from Palmira and La Victoria, Colombia

GEOSCIENCE
RESEARCH INSTITUTE



Bosquez, S., Alférez, G.H., Martínez Ardila, A.M., & Clausen B.L. (2022). Automatic Classification of Felsic, Mafic, and Ultramafic Rocks in Satellite Images from Palmira and La Victoria, Colombia. *Proceedings of the Computing Conference 2022*, London, England.

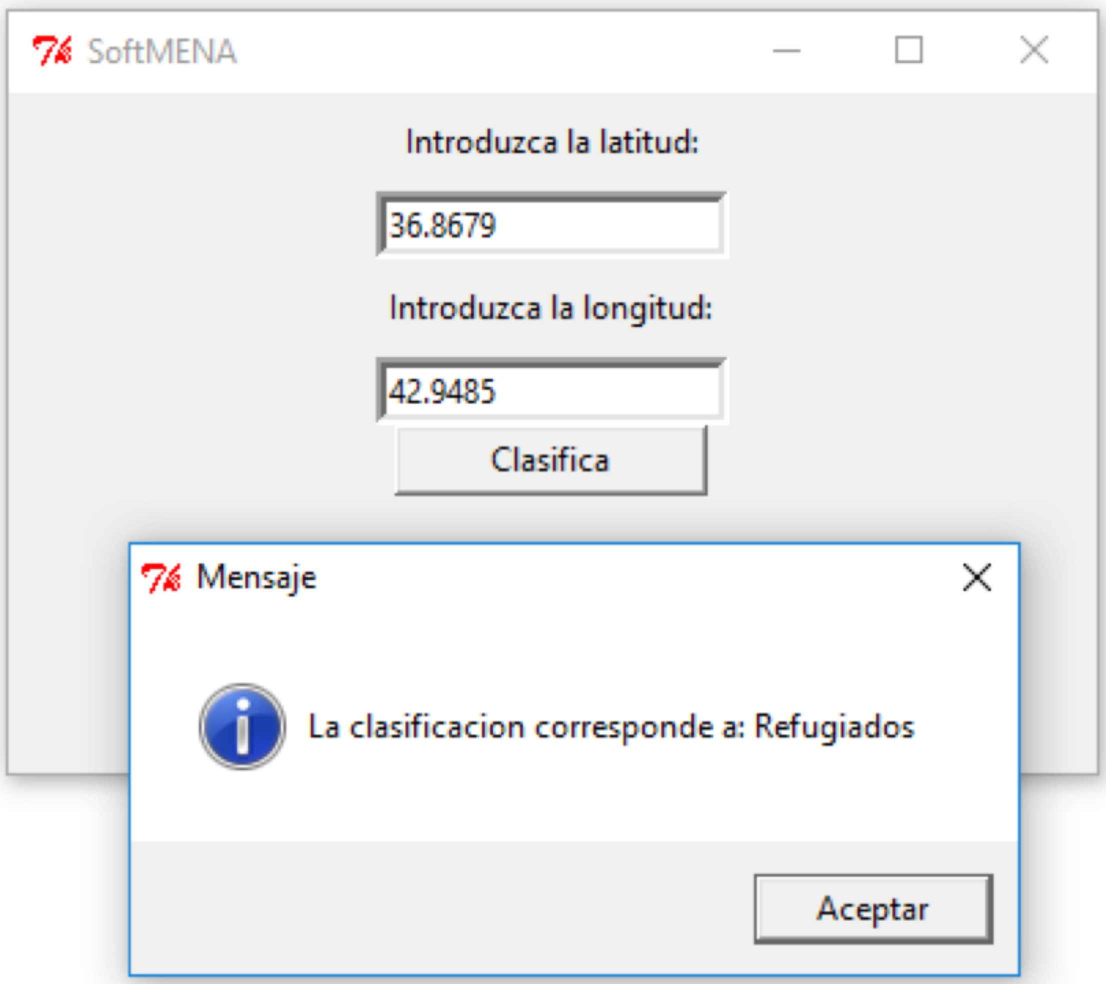
CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

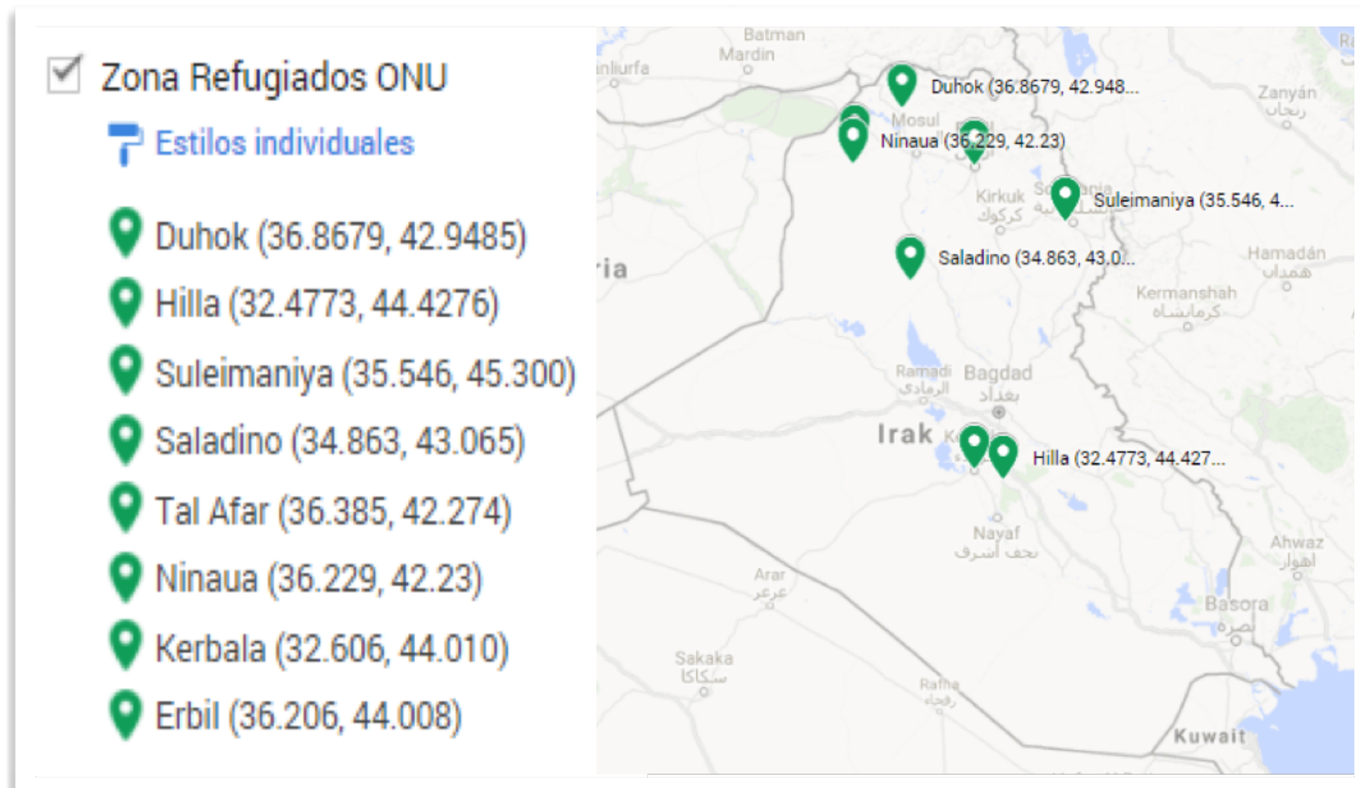


Areas with negative tweets in Manhattan

Alférez, G.H. (2016). Tweeting in New York City - Data Science Can Teach Us to Sympathize. Adventist Review, 193(2), 47-49

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University





Refugee zones (adapted from UNHigh Commissioner for Refugees, 2017)

Classification Duhok, Iraq

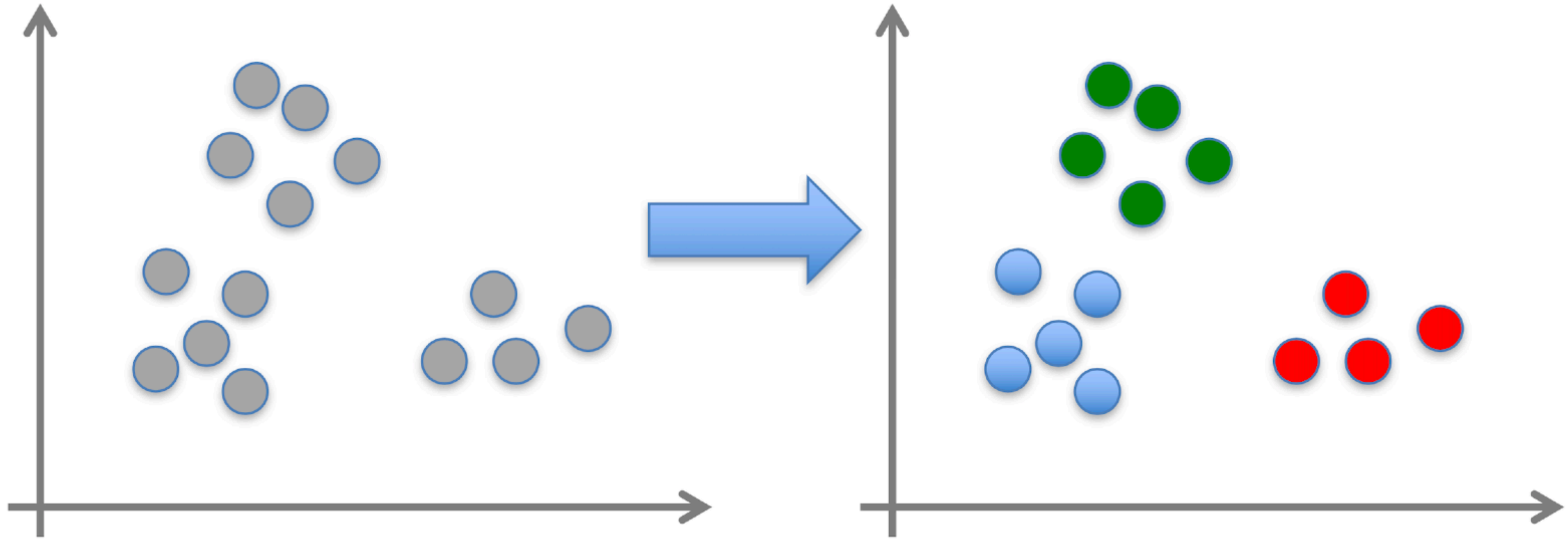
González, M. & Alférez, G.H. (2020). Application of Data Science to Discover Violence-Related Issues in Iraq. arXiv. https://arxiv.org/abs/2006.07980

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Unsupervised Learning

- If you're not trying to predict a target value, then you need to look into **unsupervised learning**.
    - E.g., **clustering**.

- Given x_1, x_2,…, x_n (without labels)
  - Output hidden structure behind the x's



CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Interpreting the Geochemistry of Southern California Granitic Rocks using Machine Learning

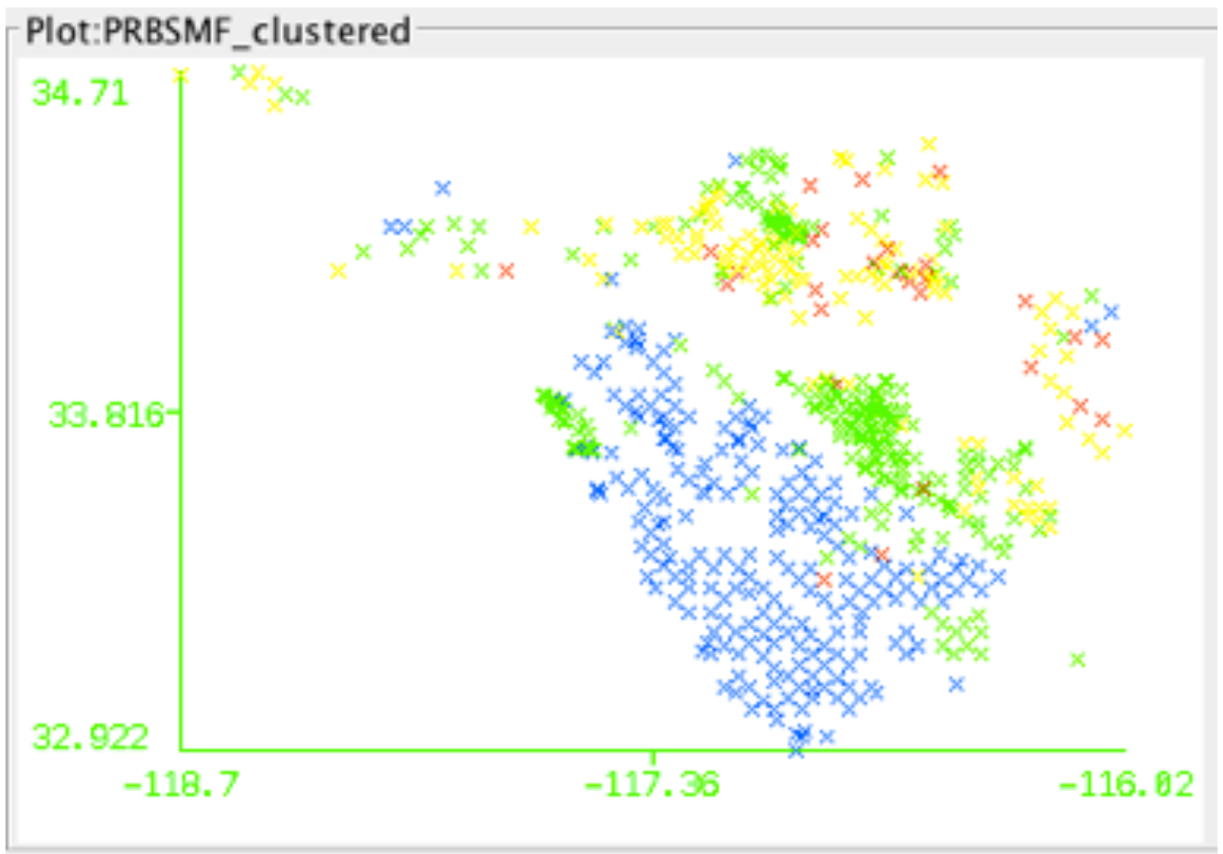


Figure 4. Cluster assignment visualization for Sri.
Cluster 0 is in yellow, **Cluster 1 is in green**, **Cluster 2 is in red**, and **Cluster 3 is in blue**

Sri (Initial $^{87}Sr/^{86}Sr$ ratios) Analysis

Table 2. WEKA results for Sr_i

Cluster #	Number of samples	Isotope ratio
0	135	0.7091
1	358	0.7068
2	31	0.7126
3	243	0.7042

Alférez, G. H., Rodríguez, J., Clausen, B. y Pompe, L. (2015). Interpreting the geochemistry of Southern California granitic rocks using machine learning. En D. de la Fuente, R. Dziegiel, E. B. Kozerenko, P. M. La-Monica, R. A. Liuzzi, J. A. Olivas y T. Waskiewicz (Eds.), Proceedings of the 2015 International Conference on Artificial Intelligence (ICAI) (pp. 592–598). CSREA Press. http://worldcomp-proceedings.com/ proc/p2015/ICA2786.pdf

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

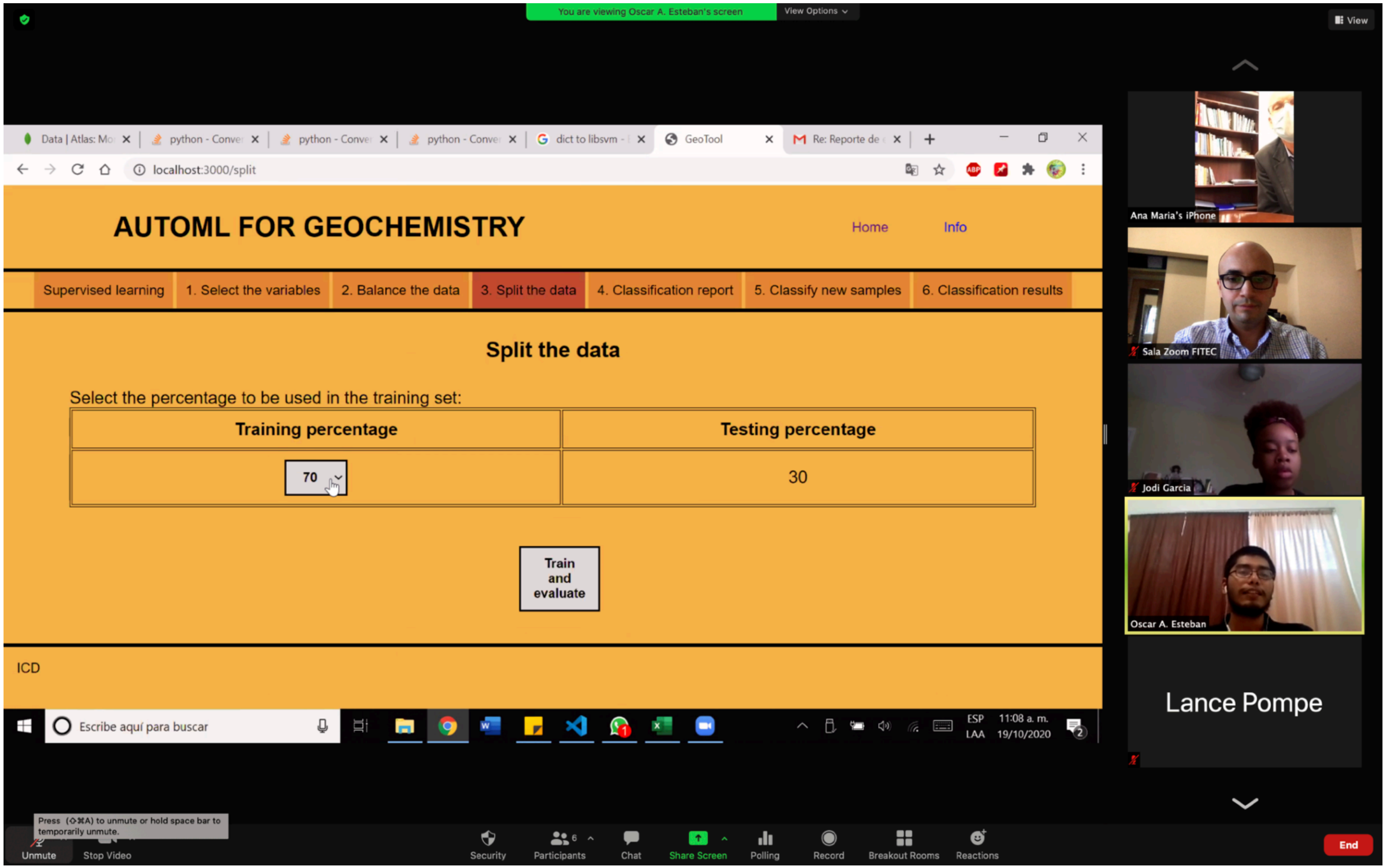
Machine learning in practice

- Understand domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learn and evaluate models
- Interpret results
- Consolidate and deploy discovered knowledge

Based on a slide by Pedro Domingos

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Automated machine learning pipeline for geochemical analysis



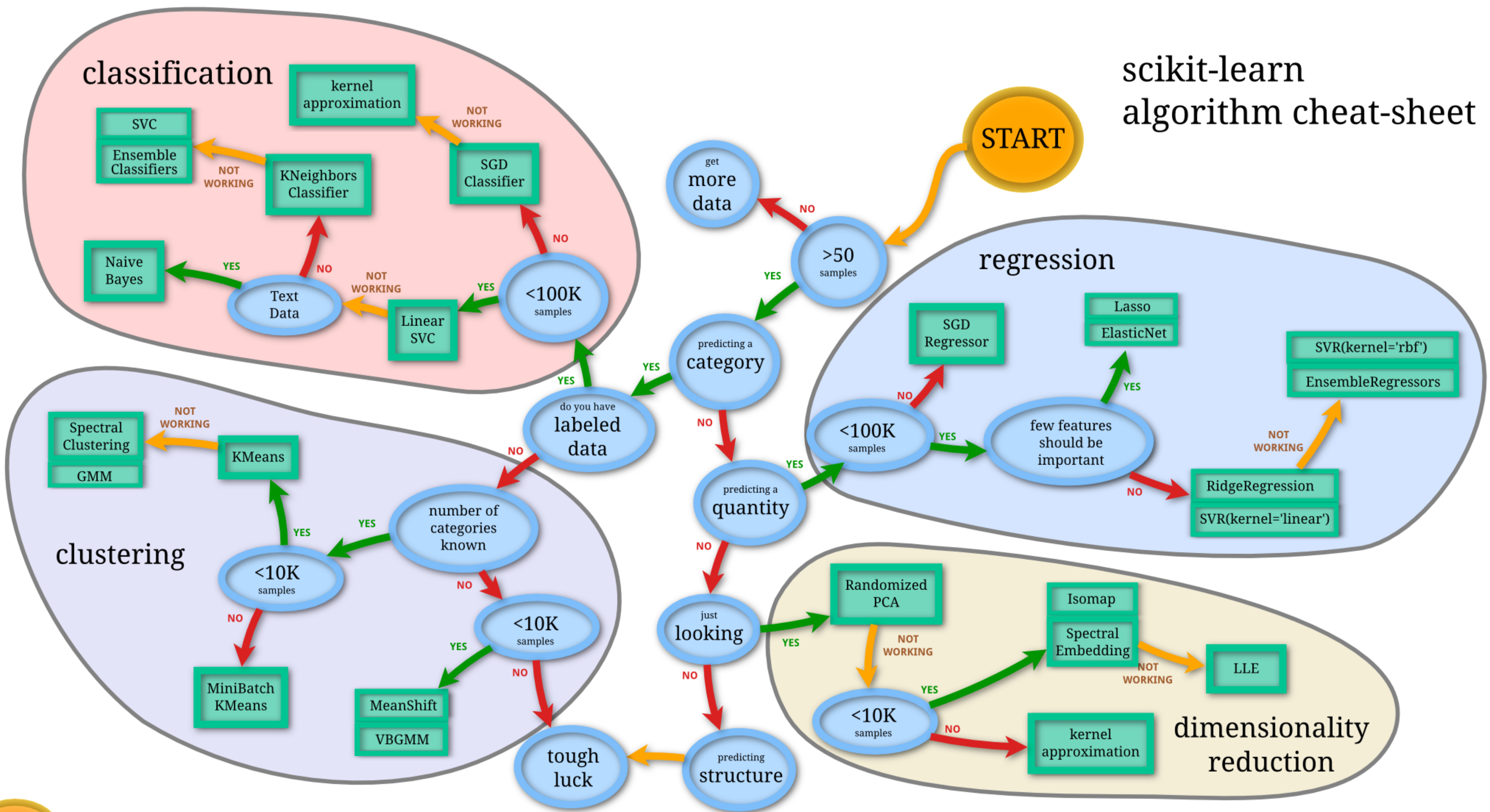
THE GEOLOGICAL SOCIETY OF AMERICA

GEOSCIENCE RESEARCH INSTITUTE

Alférez, G.H., Esteban, O.A., Clausen, B.L., & Martínez Ardila, A.M. (2022). Automated machine learning pipeline for geochemical analysis. Earth Science Informatics. https://doi.org/10.1007/s12145-022-00821-8

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Technologies



CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Why TensorFlow?

- Developed by Google
- Python API
- Portability: deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API
- Large community

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

Companies Using TensorFlow

- Google
- OpenAI
- DeepMind
- Snapchat
- Uber
- Airbus
- eBay
- Dropbox
- A bunch of startups

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University

The possibilities are endless!

Machine Learning in Action: Exploring Examples in Multiple Domains

Harvey Alférez, Ph.D.
Southern Adventist University
harveya@southern.edu
www.harveyalferez.com
@harveyalferez

CENTER FOR INNOVATION AND RESEARCH IN COMPUTING
Southern Adventist University