LECTURE 19: Machine Learning in Medical Imaging
(A Brief Introduction)

Dr. Ulas Bagci
HEC 221, Center for Research in Computer Vision (CRCV), University of Central Florida (UCF), Orlando, FL 32814.

bagci@ucf.edu or bagci@crcv.ucf.edu
Outline

• Role of Machine Learning (ML) in Radiology
• CAD Systems (ML based detection/diagnosis)
• Image Segmentation using ML
• Image Registration using ML
• Deep Learning and its applications in radiology applications
Motivation

Radiologists need to interpret high volume of images and as the number of images increases, radiologists’ workload increases as well.
Motivation

Radiologists need to interpret high volume of images and as the number of images increases, radiologists’ workload increases as well.

The increasing number and complexity of the images threatens to overwhelm radiologists’ capacities to interpret them.
Motivation

Radiologists need to interpret high volume of images and as the number of images increases, radiologists’ workload increases as well.

The increasing number and complexity of the images threatens to overwhelm radiologists’ capacities to interpret them.

Automated and intelligent image analysis and understanding are becoming an essential part or procedure, such as image segmentation, registration, and computer-aided diagnosis and detection.
Motivation

Radiologists need to interpret high volume of images and as the number of images increases, radiologists’ workload increases as well.

The increasing number and complexity of the images threatens to overwhelm radiologists’ capacities to interpret them.

Automated and intelligent image analysis and understanding are becoming an essential part or procedure, such as image segmentation, registration, and computer-aided diagnosis and detection.

Machine learning algorithms underpin the algorithms and software that make computer-aided diagnosis/prognosis/treatment possible.
Typical goal of machine learning

**input**
- images/video
- audio
- text

**output**
- Label: “Motorcycle”
- Suggest tags
- Image search
- Speech recognition
- Music classification
- Speaker identification
- Web search
- Anti-spam
- Machine translation
Typical goal of machine learning

**Input**
- Images/video
- Audio
- Text

**Output**
- Label: “Motorcycle”
  - Suggest tags
  - Image search
  - ...
- Speech recognition
  - Music classification
  - Speaker identification
  - ...
- Web search
  - Anti-spam
  - Machine translation
  - ...

Feature engineering: most time consuming!
Our goal in object classification
Why is this hard?

You see this:

But the camera sees this:

<table>
<thead>
<tr>
<th>194</th>
<th>210</th>
<th>201</th>
<th>212</th>
<th>199</th>
<th>213</th>
<th>215</th>
<th>195</th>
<th>178</th>
<th>158</th>
<th>182</th>
<th>209</th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td>189</td>
<td>190</td>
<td>221</td>
<td>209</td>
<td>205</td>
<td>191</td>
<td>167</td>
<td>147</td>
<td>115</td>
<td>129</td>
<td>163</td>
</tr>
<tr>
<td>114</td>
<td>126</td>
<td>140</td>
<td>188</td>
<td>176</td>
<td>165</td>
<td>152</td>
<td>140</td>
<td>170</td>
<td>106</td>
<td>78</td>
<td>88</td>
</tr>
<tr>
<td>87</td>
<td>103</td>
<td>115</td>
<td>154</td>
<td>143</td>
<td>142</td>
<td>149</td>
<td>153</td>
<td>173</td>
<td>101</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>102</td>
<td>112</td>
<td>106</td>
<td>131</td>
<td>122</td>
<td>138</td>
<td>152</td>
<td>147</td>
<td>128</td>
<td>84</td>
<td>58</td>
<td>66</td>
</tr>
<tr>
<td>94</td>
<td>95</td>
<td>79</td>
<td>104</td>
<td>105</td>
<td>124</td>
<td>129</td>
<td>113</td>
<td>107</td>
<td>87</td>
<td>69</td>
<td>67</td>
</tr>
<tr>
<td>68</td>
<td>71</td>
<td>69</td>
<td>98</td>
<td>89</td>
<td>92</td>
<td>98</td>
<td>95</td>
<td>89</td>
<td>88</td>
<td>76</td>
<td>67</td>
</tr>
<tr>
<td>41</td>
<td>56</td>
<td>68</td>
<td>99</td>
<td>63</td>
<td>45</td>
<td>60</td>
<td>82</td>
<td>58</td>
<td>76</td>
<td>75</td>
<td>65</td>
</tr>
<tr>
<td>20</td>
<td>43</td>
<td>69</td>
<td>75</td>
<td>56</td>
<td>41</td>
<td>51</td>
<td>73</td>
<td>55</td>
<td>70</td>
<td>63</td>
<td>44</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>57</td>
<td>69</td>
<td>75</td>
<td>75</td>
<td>73</td>
<td>74</td>
<td>53</td>
<td>68</td>
<td>59</td>
<td>37</td>
</tr>
<tr>
<td>72</td>
<td>59</td>
<td>53</td>
<td>66</td>
<td>84</td>
<td>92</td>
<td>84</td>
<td>74</td>
<td>57</td>
<td>72</td>
<td>63</td>
<td>42</td>
</tr>
<tr>
<td>67</td>
<td>61</td>
<td>58</td>
<td>65</td>
<td>75</td>
<td>78</td>
<td>76</td>
<td>73</td>
<td>59</td>
<td>75</td>
<td>69</td>
<td>50</td>
</tr>
</tbody>
</table>
Pixel-based representation

Input

Raw image

Motorbikes
“Non”-Motorbikes

Learning algorithm
Pixel-based representation

Input

Raw image

Motorbikes

“Non”-Motorbikes

Learning algorithm
Pixel-based representation

Input

Raw image

Motorbikes
“Non”-Motorbikes

Learning algorithm
What we want

Input

- handlebars
- wheel

Raw image

+ Motorbikes
- “Non”-Motorbikes

Feature representation

E.g., Does it have Handlebars? Wheels?

Learning algorithm

Features

- Wheels
- Handlebars

pixel 2

pixel 1
Some feature representations

SIFT

Spin image

HoG

RIFT

Textons

GLOH
Some feature representations

Coming up with features is often difficult, time-consuming, and requires expert knowledge.
Feature Engineering?

Feature Learning?
Representation Learning

• Is there some way to extract **meaningful features** from data in a *supervised* or *unsupervised* manner?
Representation Learning

• Is there some way to extract meaningful features from data in a supervised or unsupervised manner?

• Biologically inspired systems
  – to make the computer more robust, intelligent, and learn, …
Representation Learning

• Is there some way to extract **meaningful features** from data in a *supervised* or *unsupervised* manner?

• Biologically inspired systems
  – to make the computer more robust, intelligent, and learn, …
  – Model our systems after **the brain**!
Representation Learning

• Is there some way to extract meaningful features from data in a supervised or unsupervised manner?

• Biologically inspired systems
  – to make the computer more robust, intelligent, and learn, …
  – Model our systems after the brain!
    • Brain interprets imprecise information from the senses at an incredibly rapid rate!
    • It discerns a whisper in a noisy room, a face in a dimly lit alley, and hidden agenda in a political statement.
Recap: Types of Learning

• **Supervised (inductive) learning**
  – Training data includes desired outputs

• **Unsupervised learning**
  – Training data does not include desired outputs

• **Semi-supervised learning**
  – Training data includes a few desired outputs

• **Reinforcement learning**
  – Rewards from sequence of actions
Applications of ML in Radiology

- Image Segmentation
- Image Registration
- CAD (computer aided detection/diagnosis)
- Brain function or activity analysis
- Neurology disease diagnosis
- Text analysis of radiology reports
- …
Computerized Detection/Classification Methods

- Rule-Based
  - Training
  - Feature Selection
  - Thresholding
- kNN
  - Training
  - Feature Labeling
  - Estimate Class Label
- ANN
  - Training
  - Feature Learning
  - Estimate Class Label
- Decision Trees
  - Training
  - Feature Learning
  - Predict Target Variables
- Bayesian
  - Statistical Model/Training
  - Feature Labeling
  - Estimate Class Label
- LDA
  - Estimation of class Mean/SD
  - Feature Selection
  - Prediction by Decision Boundary
- SVM
  - Training
  - Feature Labeling
  - Prediction by Decision Boundary
Supervised Learning

In supervised learning the predictive model represents the assumed relationship between input variables in $x$ and output variable $y$. 
Linear Models and Regression

• Linear models assume that there is a linear relationship between the input of the model and the output of the model.

• Perhaps it is the simplest method for classification and regression.

• It has been widely used in computer-aided classification.
**Ex: LDA (Linear Discriminant Analysis)**

\[ J(w) = \frac{w^T S_B w}{w^T S_w w} \]

\( S_B = (m_1 - m_2)(m_1 + m_2)^T \) is called the “between” scatter matrix (\( m_i \) is the mean of samples from class \( i, i \in \{1, 2\} \)), and \( S_w = S_1 + S_2 \) is called the “within” scatter matrix \( (Si=\sum_{x \in Di}(x-m_i)(x-m_i)^T, D_i \) is the collection of samples from class \( i, i \in \{1, 2\} \)).

\[ f(x) = w^T x + w_0 \]
Ex. Artificial Neural Networks

- Artificial neural networks (ANNs) are techniques that were inspired by the brain and the way it learns and processes information.
- Neural networks are composed of nodes and interconnections. Nodes usually have limited computation power. They simulate neurons by behaving like a switch, just as neurons will be activated only when sufficient neurotransmitter has accumulated.
Ex. Learning with Kernels

- By applying traditional supervised and unsupervised learning methods in the feature space, kernel methods provide powerful tools for data analysis and have been found to be successful in a number of real applications.

SVM
Support vector machines
(a) The Fisher LD fails to separate two classes because training example D adversely influences decision boundary T.

(b) The SVM defines the decision boundary using only points A, B, and C, called support vectors, and is not influenced at all by point D.
Ex: CAD for finding micro-calcifications in mammogram region

Mammogram Region (a)  SVM Output (b)  Detected Lesion Positions (c)
Ex. CAD for Pulmonary Abnormalities
Ex. CAD for Pulmonary Abnormalities
Ex. CAD for Pulmonary Abnormalities
Ex. Tree-in-Bud (TIB) Detection
TIB Appearance

- Thickened bronchial structures
- Locally surrounded by clusters of 2-3 mm micro-nodules

1. Homogeneity (small)
2. Gradient (high)
3. A few mm in length
Scale and Rotation Invariant Features

What is known about TIB?

- Small (scale is known)
- Gradient is high (variation is known)
- Micro-nodule (dots) and thickened vessels should be observed

Design a local scale filter to select candidate regions

Derive a suitable shape features
TIB Overview

Image Acquisition → Segmentation → Candidate Selection

Pattern Detection ← SVM Classification ← Feature Extraction
Local Scale Concept

At any voxel v in a scene,

- **b-scale**: largest homogeneous ball centered at v.
- **t-scale**: largest homogeneous ellipsoid centered at v.
- **g-scale**: largest connected homogeneous region containing v.
Local Scale Map

brain PD slice

ball scale

tensor scale
generalized scale
Computation of Ball-Scale

A hyperball $B_{k,\nu}$ of radius $k \geq 0$ with center at $c \in C$

$$B_{k,\nu} = \left( e \in C \mid \sqrt{\left( \sum_{i=1}^{n} \frac{\nu_i^2(c_i - e_i)^2}{\min_j [\nu_j^2]} \right)} \right)$$

For a hyperball $B_{k,\nu}$, we define the fraction of object $FO_{k,\nu}$ as

$$FO_{k,\nu} = \frac{\sum_{e \in B_{k,\nu}(c) - B_{k-1,\nu}(c)} W_\phi(\|f(c) - f(e)\|)}{B_{k,\nu}(c) - B_{k-1,\nu}(c)}$$
b-scale object scale estimation algorithm

Input: \( c \) is pixel in a scene \( C \), \( ts \): homogeneity threshold, \( FO \) is object-fraction based on regional homogeneity

Output: \( r(c) \): b-scale value for pixel \( c \)

Begin

\[
\begin{align*}
& \text{Set } k = 1 \\
& \text{While } FO(c) > ts \text{ do} \\
& \quad \text{Set } k \text{ to } k + 1 \\
& \text{EndWhile} \\
& \text{Set } r(c) \text{ to } k - 1 \\
End
\end{align*}
\]
b-scale object scale estimation algorithm

Input: c is pixel in a scene C, ts: homogeneity threshold, FO is object-fraction based on regional homogeneity

Output: r(c): b-scale value for pixel c

Begin
  Set k=1
  While FO(c) > ts do
    Set k to k+1
  EndWhile
  Set r(c) to k-1
End
b-scale object scale estimation algorithm

Input: \( c \) is pixel in a scene \( C \), \( ts \): homogeneity threshold, \( FO \) is object-fraction based on regional homogeneity

Output: \( r(c) \): b-scale value for pixel \( c \)

Begin

Set \( k = 1 \)

While \( FO(c) > ts \) do

Set \( k \) to \( k + 1 \)

EndWhile

Set \( r(c) \) to \( k - 1 \)

End
Candidate TIB Patterns
Willmore Energy/Flow

- **Canham-Helfrich model**: deformation of cell membranes is uniquely determined by its shape. In classical Gaussian (K) or mean curvature (H) based energy models are not taking into account topological changes!! However, in CH model, arbitrary topology changes are allowed.

\[ S = \int_{\Sigma} \alpha + \beta(H)^2 - \gamma K dA. \]

\[ S_w = \int_{\Sigma} (H^2 - K) dA = \int_{\Sigma} |H|^2 dA - \int_{\partial \Sigma} |K| ds, \]

Mobius invariance (inversion and affine invariant)
Mean Curvature
\[ H = \left( \kappa_1 + \kappa_2 \right)/2 \]

Gaussian Curvature
\[ K = \kappa_1 \kappa_2 \]
Other Shape Features

- Let $k_1$ and $k_2$ be eigenvalues of the local Hessian matrix for any given local patch where $k_1 \geq k_2$

  Mean curvature ($H = (k_1 + k_2) / 2$)
  Gaussian curvature ($K = k_1 k_2$)
  Shape Index ($SI$): $(2/\pi) \cdot \arctan((k_1+k_2)/(k_1-k_2))$
  Elongation: $(k_1/k_2)$
  Shear: $(k_1-k_2)^2/4$
  Compactness: $1/\sqrt{k_1 k_2}$
  Distortion: $k_1-k_2$

- Overall, we have 8 features extracted from a patch (if there is at least one b-scale pattern in the patch)
Comparison to Other Methods

Table 1: Accuracy ($A_z$) of the CAD system with given feature sets.

<table>
<thead>
<tr>
<th>Features</th>
<th>Dimension</th>
<th>Patch Size</th>
<th># of patches (TIB)</th>
<th># of patches (Normal)</th>
<th>Area under ROC curve: $A_z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape &amp; GLCM</td>
<td>$8+18=26$</td>
<td>17x17</td>
<td>14144</td>
<td>12032</td>
<td>0.8991</td>
</tr>
<tr>
<td>Shape &amp; GLCM</td>
<td>$8+18=26$</td>
<td>13x13</td>
<td>24184</td>
<td>20572</td>
<td>0.9038</td>
</tr>
<tr>
<td>Shape &amp; GLCM</td>
<td>$8+18=26$</td>
<td>9x9</td>
<td>50456</td>
<td>42924</td>
<td>0.9096</td>
</tr>
<tr>
<td>Shape</td>
<td>8</td>
<td>17x17</td>
<td>14144</td>
<td>12032</td>
<td>0.7941</td>
</tr>
<tr>
<td>Shape</td>
<td>8</td>
<td>13x13</td>
<td>24184</td>
<td>20572</td>
<td>0.7742</td>
</tr>
<tr>
<td>Shape</td>
<td>8</td>
<td>9x9</td>
<td>50456</td>
<td>42924</td>
<td>0.7450</td>
</tr>
<tr>
<td>Steer. Wavelets &amp; Shape</td>
<td>$6x17x17+8=1742$</td>
<td>17x17</td>
<td>14144</td>
<td>12032</td>
<td>0.7846</td>
</tr>
<tr>
<td>Steer. Wavelets &amp; Shape</td>
<td>$6x13x13+8=1022$</td>
<td>13x13</td>
<td>24184</td>
<td>20572</td>
<td>0.7692</td>
</tr>
<tr>
<td>Steer. Wavelets &amp; Shape</td>
<td>$6x9x9+8=494$</td>
<td>9x9</td>
<td>50456</td>
<td>42924</td>
<td>0.7908</td>
</tr>
<tr>
<td>Steer. Wavelets</td>
<td>$6x17x17=1734$</td>
<td>17x17</td>
<td>14144</td>
<td>12032</td>
<td>0.7571</td>
</tr>
<tr>
<td>Steer. Wavelets</td>
<td>$6x13x13=1014$</td>
<td>13x13</td>
<td>24184</td>
<td>20572</td>
<td>0.7298</td>
</tr>
<tr>
<td>Steer. Wavelets</td>
<td>$6x9x9=486$</td>
<td>9x9</td>
<td>50456</td>
<td>42924</td>
<td>0.7410</td>
</tr>
<tr>
<td>GLCM</td>
<td>18</td>
<td>17x17</td>
<td>14144</td>
<td>12032</td>
<td>0.7163</td>
</tr>
<tr>
<td>GLCM</td>
<td>18</td>
<td>13x13</td>
<td>24184</td>
<td>20572</td>
<td>0.7068</td>
</tr>
<tr>
<td>GLCM</td>
<td>18</td>
<td>9x9</td>
<td>50456</td>
<td>42924</td>
<td>0.6810</td>
</tr>
</tbody>
</table>
Deep Learning

- Deep mapping and representation
- Won the 1st place in many competitions.
- Industrial applications (Google, IBM, Microsoft, Baidu, Facebook, Samsung, Yahoo, Intel, Apple, Nuance, BBN, ...)

![Neural Network Diagram]
Deep Learning in Medical Imaging

- Deeper representations $\rightarrow$ abstractions $\rightarrow$ disentangling
- Manifolds are expanded and flattened
Ex. Automatic LV Segmentation from US

- **Goals**
  - Automated functional analysis of the heart
  - Improve workflow, reduce user variability

- **Challenges**
  - Low signal-to-noise ratio, edge dropout, shadows
  - Training set (machine learning methods need lots of annotated images)
• Coarse to fine search strategy (3 scales)
• ROI detected from sampling initial distribution (fewer initial points, compared to grid search)
• Gradient-based search in fine stages (less computation than grid-based search)
Ex. Hippocampus Segmentation Using 7T MRI

- **Importance**
  The volume of the hippocampus is an important trait for early diagnosis of neurological diseases (e.g., Alzheimer’s disease)

- **Challenges**
  - The hippocampus is small ($\approx 35 \times 15 \times 7 \text{mm}^3$)
  - The hippocampus is surrounded by complex structures
  - Low imaging resolution ($\approx 1 \times 1 \times 1 \text{mm}^3$) of 1.5T or 3T MRI scanners
Challenges in 7T

- The characteristics of 7T MR images
  - Much richer structural information
  - Severe intensity inhomogeneity problem
  - Less partial volume effect

Comparison between 1.5T and 7T MR images
Hand-Crafted Features

- Limited discriminative power

Extracting patches from a 7T MR image

Responses of Haar filters for the image patches in a-c
Deep Learning Features

- Multi-atlas segmentation
- Classification using high-level shape information via auto-context models (ACM)
- Hierarchical feature representation via unsupervised deep learning

Basis filters generated by unsupervised deep learning

Context feature
Hierarchical Feature Extraction

Stacked two-layer convolutional ISA (Independent Subspace Analysis)
Multi-Atlas Segmentation

Training Stage

Alignment training images in each atlas space 1...N

2-layer ISA

image patches

learned features

Classifier sequence 1

Atlas space 1

2-layer ISA

Classifier sequence 2

Atlas space 2

... 

2-layer ISA

Classifier sequence N

Atlas space N

Testing Stage

2-layer ISA

Adaptively weighted fusion

classification maps 1...N

probability map

Level set

segmentation result

Subject image space
Qualitative Evaluations

Ground Truth

Haar + Texture Features

Hierarchical Features
Ex. Registration of Brain MR Images

Determine accurate correspondences between images

Individual

Model
Deep Learning for Image Registration

Input image patches

Morphological signatures for image registration

Encoder

Decoder

Representations

y'(1)  

y'(2)  

x

z
Deep Learning for Image Registration

**Difficulty #1:** How to determine the number of hidden nodes and the number of layers?

**Solution:** Use affinity propagation to roughly estimate the number of hidden units.
Deep Learning for Image Registration

**Difficulty #2:** How to deal with high dimensional training data?

**Solution:** Use the convolutional RBM in each layer.
Deep Learning for Image Registration

**Difficulty #3:** How to deal with large number of training samples?

**Solution:** Select key points across training images to focus on distinctive image regions.
Summary

• ML algorithms are key to detection and diagnosis problem
• Deep Learning is getting a lot of interests due to its strong features for CAD tasks
• Image analysis tasks can be enhanced via ML algorithms
Slide Credits and References

• Wang and Summers, Medical Image Analysis, 2012.
• Kim, M., MLMI, MICCAI 2013.