Machine Learning in Healthcare

Outline

• Overview of Neural Networks

• History of Machine Learning in Medicine

• Big Data in Medical Applications

• Opportunities/Challenges in Healthcare

• Utility of Machine Learning in Medical Imaging

• Example Applications in Medical Imaging
PART I: Overview of Neural Networks
Artificial Neuron: A mathematical abstraction

Perceptron Model (McCulloch-Pitts)

Soma

Axon (nonlinear) activation function

Synapse

Dendrites Synaptic Weights
Single Neuron: A linear classifier

Dendrites  Synaptic Weights

\textbf{Soma}

\begin{align*}
\sum_{j} \text{weights} \times \text{inputs } x_j &= \text{net input } \text{net}_j \\
\text{transfer function} \quad \varphi(\text{net}_j) &= \alpha_j \\
\text{activation function} \quad \alpha_j &= \theta_j \text{ activation} \\
\end{align*}

\textbf{Axon (nonlinear)}

Output

\textbf{Linear Decision Boundary}

\textbf{Input Space}
Neural Network: Nonlinear mapping

**Single Hidden-Layer Network**

**Inputs**

**Output**

Nonlinear Decision Boundary
Universal Approximation Theorem

- Early 1990s for single hidden-layer networks
- A universal approximator
- Model any continuous nonlinear function (given a sufficient number of neurons)
- No guidance on how to find model parameters…
Why is Deep Learning Hot Today?

**Big Data Availability**
- Facebook: 350 millions images uploaded per day
- Walmart: 2.5 Petabytes of customer data hourly
- YouTube: 100 hours of video uploaded every minute

**New DL Techniques**

**GPU acceleration**

Deep Neural Networks

Raw data
Low-level features
Mid-level features
High-level features

Application components:

- Task objective: e.g. Identify face
- Training data: 10-100M images
- Network architecture: ~10 layers, 1B parameters
- Learning algorithm: ~30 Exaflops, ~30 GPU days

Input

Result
ImageNet Object Recognition Challenge


ImageNet Winners

Before Deep learning After Deep learning

2010 0.282 0.258
2011 0.153
2012 0.112
2013 0.074
2014 0.05
2015 0.036

Human
From Blackbox Models to Dark Magic?

1980s-ERA NEURAL NETWORK

- Input layer
- Hidden layer
- Output layer

- Links carry signals from one node to another, boosting or damping them according to each link's 'weight'.

DEEP LEARNING NEURAL NETWORK

- Multiple hidden layers process hierarchical features
- Identify edges
- Identify combinations of edges
- Identify features

- Identify combinations or features

Input: 'George'

Images:
- Checkerboard
- Textures
- Edges
Task-Specific Priors

**Task:**

```
Cat?
```

**Priors:**

```
Locally-Coded Features
```

![Graph showing Correlation vs. Spatial separation (pixels)](image)
Task-Specific Priors

**Task:**

![Cat image](image)

**Priors:**

- **Spatially Invariant**
  - ![Cat image](image)
  - ![Cat image](image)

- **Scale Invariant**
  - ![Cat image](image)
  - ![Cat image](image)
Ideas: Convolutional Layer

- Filtering with a kernel in small neighborhoods: local processing
- Parameter sharing: translation invariance

\[ y_j = g_j \tanh \left( \sum_i k_{ij} \ast x_i \right) \]

- \( x_i \) is the \( i \)th channel of input
- \( k_{ij} \) is the convolution kernel
- \( g_j \) is a learned scaling factor
- \( y_j \) is the hidden layer
Ideas: Pooling Layer

- Pooling hidden-unit responses: translation invariance
- Analysis at multiple-spatial scales: scale invariance

\[ y_{ijk} = \frac{1}{m^2} \sum_{p,q} x_{i,j+p,k+q} \]

- \( x_{i,j,k} \) is value of the \( i^{th} \) feature map at position \( j,k \)
- \( p \) is vertical index in local neighborhood
- \( q \) is horizontal index in local neighborhood
- \( y_{ijk} \) is pooled and subsampled layer
- \( m \) is the neighborhood height/width
Convolutional Neural Networks

Input Image → Feature maps → Pooling → Feature maps → Pooling → Fully Connected → Fully Connected → Output

Categories:
- Dog (0.1)
- Cat (0.4)
- Deer (0.94)
- Lion (0.2)
PART II: History of ML in Medicine
1970’s: MYCIN expert system

- 1970’s (Stanford): MYCIN expert system for identifying bacteria causing severe infections
- Proposed a good therapy in ~69% of cases. Better than infectious disease experts

**FIGURE 33-1** Short sample dialogue. The physician’s inputs appear in capital letters after the double asterisks.

**Dialogue interface**

- I am ready
- **THIS IS A 26 YEAR OLD MALE PATIENT**
- My understanding is:
  - The age of the patient is 26
  - The sex of the patient is male
- **FIVE DAYS AGO, HE HAD RESPIRATORY-TRACT SYMPTOMS**
- What is his name?
- **JO**
- My understanding is:
  - The name of the patient is Jo
  - Respiratory-tract is one of the symptoms that the patient had
- **A COUPLE OF DAYS BEFORE THE ADMISSION, HE HAD A MALAISE**
- Please give me the date of admission
- **MARCH 12, 1979**
- My understanding is:
  - The patient was admitted at the hospital 3 days ago
  - Malaise is one of the symptoms that the patient had 5 days ago
QMR Models

1980’s: INTERNIST-1/QMR model

- 1980’s (Univ. of Pittsburgh): INTERNIST-1/Quick Medical Reference
- Diagnosis for internal medicine

Probabilistic model relating:
- 570 binary disease variables
- 4,075 binary symptom variables
- 45,470 directed edges

Elicited from doctors:
- 15 person-years of work

Led to advances in ML & AI
(Bayesian networks, approximate inference)

Problems:
1. Clinicians entered symptoms manually
2. Difficult to maintain, difficult to generalize
1980’s: automating medical discovery

Dr. Robert Blum, "Discovery, Confirmation and Incorporation of Causal Relationships from a Large Time-Oriented Clinical Data Base: The RX Project." Dept. of Computer Science, Stanford. 1981

TABLE 1
Hypothetical Time-Oriented Record for One Patient

<table>
<thead>
<tr>
<th>Visit number</th>
<th>Date</th>
<th>Knee pain</th>
<th>Fatigue</th>
<th>Temperature</th>
<th>Diagnosis</th>
<th>White blood count</th>
<th>Creatinine clearance</th>
<th>Blood urea nitrogen</th>
<th>Prednisone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>January 17, 79</td>
<td>Severe</td>
<td>Moderate</td>
<td>38.5</td>
<td>Systemic lupus</td>
<td>3500</td>
<td>45</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>June 23, 79</td>
<td>Mild</td>
<td>—</td>
<td>37.5</td>
<td>—</td>
<td>4700</td>
<td>—</td>
<td>33</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>July 1, 79</td>
<td>Mild</td>
<td>—</td>
<td>36.9</td>
<td>—</td>
<td>4300</td>
<td>—</td>
<td>—</td>
<td>20</td>
</tr>
</tbody>
</table>

Discovers that prednisone elevates cholesterol (Annals of Internal Medicine, ‘86)
Neural Networks in Medicine

1990’s: neural networks in medicine

- Neural networks with clinical data took off in 1990, with 88 new studies that year
- Small number of features (inputs)
- Data often collected by chart review

Problems: 1. Did not fit well into clinical workflow
           2. Poor generalization to new places
## Disease Diagnosis

<table>
<thead>
<tr>
<th>Table 1</th>
<th>25 Neural Network Studies in Medical Decision Making*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject</strong></td>
<td><strong>No. of Examples</strong></td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Breast cancer&lt;sup&gt;4&lt;/sup&gt;</td>
<td>57</td>
</tr>
<tr>
<td>Vasculitis&lt;sup&gt;5&lt;/sup&gt;</td>
<td>404</td>
</tr>
<tr>
<td>Myocardial infarction&lt;sup&gt;6&lt;/sup&gt;</td>
<td>351</td>
</tr>
<tr>
<td>Myocardial infarction&lt;sup&gt;8&lt;/sup&gt;</td>
<td>356</td>
</tr>
<tr>
<td>Low back pain&lt;sup&gt;11&lt;/sup&gt;</td>
<td>100</td>
</tr>
<tr>
<td>Cancer outcome&lt;sup&gt;13&lt;/sup&gt;</td>
<td>5,169</td>
</tr>
<tr>
<td>Psychiatric length of stay&lt;sup&gt;17&lt;/sup&gt;</td>
<td>957</td>
</tr>
<tr>
<td>Intensive care outcome&lt;sup&gt;23&lt;/sup&gt;</td>
<td>284</td>
</tr>
<tr>
<td>Skin tumor&lt;sup&gt;21&lt;/sup&gt;</td>
<td>150</td>
</tr>
<tr>
<td>Evoked potentials&lt;sup&gt;35&lt;/sup&gt;</td>
<td>100</td>
</tr>
<tr>
<td>Head injury&lt;sup&gt;57&lt;/sup&gt;</td>
<td>500</td>
</tr>
<tr>
<td>Psychiatric outcome&lt;sup&gt;54&lt;/sup&gt;</td>
<td>289</td>
</tr>
<tr>
<td>Tumor classification&lt;sup&gt;55&lt;/sup&gt;</td>
<td>53</td>
</tr>
<tr>
<td>Dementia&lt;sup&gt;57&lt;/sup&gt;</td>
<td>75</td>
</tr>
<tr>
<td>Pulmonary embolism&lt;sup&gt;59&lt;/sup&gt;</td>
<td>607</td>
</tr>
<tr>
<td>Heart disease&lt;sup&gt;62&lt;/sup&gt;</td>
<td>460</td>
</tr>
<tr>
<td>Thyroid function&lt;sup&gt;62&lt;/sup&gt;</td>
<td>3,600</td>
</tr>
<tr>
<td>Breast cancer&lt;sup&gt;62&lt;/sup&gt;</td>
<td>350</td>
</tr>
<tr>
<td>Diabetes&lt;sup&gt;62&lt;/sup&gt;</td>
<td>384</td>
</tr>
<tr>
<td>Myocardial infarction&lt;sup&gt;63&lt;/sup&gt;</td>
<td>2,856</td>
</tr>
<tr>
<td>Hepatitis&lt;sup&gt;65&lt;/sup&gt;</td>
<td>39</td>
</tr>
<tr>
<td>Psychiatric admission&lt;sup&gt;76&lt;/sup&gt;</td>
<td>319</td>
</tr>
<tr>
<td>Cardiac length of stay&lt;sup&gt;83&lt;/sup&gt;</td>
<td>713</td>
</tr>
<tr>
<td>Anti-cancer agents&lt;sup&gt;83&lt;/sup&gt;</td>
<td>127</td>
</tr>
<tr>
<td>Ovarian cancer&lt;sup&gt;91&lt;/sup&gt;</td>
<td>75</td>
</tr>
<tr>
<td><strong>MEDIAN VALUE</strong></td>
<td>350</td>
</tr>
</tbody>
</table>

---

*For reference citations, see the reference list.

†P = prior probability of most prevalent category.

‡D = ratio of training examples to weights per output.

§A single integer in the accuracy column denotes percentage overall classification rate and a single real number between 0 and 1 indicates the AUROCC value. Neural = accuracy of neural net. Other = accuracy of best other method.
PART III: Big Data in Medical Applications
Sources of Medical Data

- Inpatient health monitoring
- Medical imaging
- Examination
- Patient-generated health data
- Laboratory results
Sources of Medical Data

Diversity of digital health data

- Lab tests
- Proteomics
- Imaging
- Social media
- Phone
- Vital signs
- Devices
- Genomics
Adoption of Electronic Health Records (EHR) has increased 9x since 2008

[Henry et al., ONC Data Brief, May 2016]
Standard Classes of Data

- **PUBLIC**: Data that may be freely disclosed to the public
- **INTERNAL ONLY**: Internal data not meant for public disclosure
- **CONFIDENTIAL**: Sensitive data that if compromised could negatively affect operations
- **RESTRICTED**: Highly sensitive corporate data that if compromised could put the organization financial or legal risk

Medical Data
Large, Public Databases are Emerging

**THE PRECISION MEDICINE INITIATIVE**

- Baseline health exam
- Clinical data derived from electronic health records (EHRs)
- Healthcare claims
- Laboratory data

**UK Biobank**

UK Biobank is a national and international health resource with unparalleled resources to improve the prevention, diagnosis, and treatment of a wide range of serious diseases, such as diabetes, arthritis, osteoporosis, eye disorders, depression, and forms of dementia, and to provide health information, which does not always identify them, to the researchers. To make the most of the research featured on this website, please read the background materials before registering. To our part, without you, none of the research featured on this website would be possible.

**THE PRECISION MEDICINE INITIATIVE**

- Baseline health exam
- Clinical data derived from electronic health records (EHRs)
- Healthcare claims
- Laboratory data

**MIMIC**

De-identified health data from ~40K critical care patients

- Demographics
- Vital signs
- Laboratory tests
- Medications
- Vital signs
- Laboratory tests
- Medications
- Notes

**UK BIOBANK MAKES INFECTION AND HEALTH DATA AVAILABLE TO TACKLE COVID-19**

- Baseline health exam
- Clinical data derived from electronic health records (EHRs)
- Healthcare claims
- Laboratory data
PART IV: Opportunities/Challenges in Healthcare
Example: Emergency Departments

Emergency Department:
• Limited resources
• Time sensitive
• Critical decisions
Example: Emergency Departments

Triage Information (Free text)

MD comments (free text)

Specialist consults

Physician documentation

Lab results (Continuous valued)

Repeated vital signs (continuous values) Measured every 30 s

Disposition

Data in Emergency Department (ED) Collaboration with Steven Horng, MD

Electronic records for over 300,000 ED visits
How can Machine Learning Help?

• Triggering clinical pathways
• Context-specific displays
• Risk stratification
• Improving clinical documentation

Pathways have been shown to reduce in-hospital complications without increasing costs [Rotter et al 2010]
Automatic Protocol Selection

- Triggering clinical pathways
- Context-specific displays
- Risk stratification
- Improving clinical documentation

Our task: Determine whether a patient has or is suspected to have cellulitis
Disease-specific Recommender Systems

- Triggering clinical pathways
- Context-specific displays
- Risk stratification
- Improving clinical documentation

**Our task:**
Determine whether patient complained of chest pain, or is a psych patient

*Automatically place* specialized order sets on patient displays
Minimizing Risk

- Triggering clinical pathways
- Context-specific displays
- **Risk stratification**
- Improving clinical documentation

Ex 1: Likelihood of mortality or admission to ICU

Ex 2: Early detection of severe sepsis
# Real-time Disease Prediction

<table>
<thead>
<tr>
<th><strong>History</strong></th>
<th><strong>Acute</strong></th>
<th>Deep vein thrombosis</th>
<th>Laceration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcoholism</td>
<td>Abdominal pain</td>
<td>Employee exposure</td>
<td>Motor vehicle accident</td>
</tr>
<tr>
<td>Anticoagulated</td>
<td>Allergic reaction</td>
<td>Epistaxis</td>
<td>Pancreatitis</td>
</tr>
<tr>
<td>Asthma/COPD</td>
<td>Ankle fracture</td>
<td>Gastroenteritis</td>
<td>Pneumonia</td>
</tr>
<tr>
<td>Cancer</td>
<td>Back pain</td>
<td>Gastrointestinal bleed</td>
<td>Psych</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>Bicycle accident</td>
<td>Geriatric fall</td>
<td>Obstruction</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Cardiac etiology</td>
<td>Headache</td>
<td>Septic shock</td>
</tr>
<tr>
<td>HIV+</td>
<td>Cellulitis</td>
<td>Hematuria</td>
<td>Sexual assault</td>
</tr>
<tr>
<td>Immunosuppressed</td>
<td>Chest pain</td>
<td>Intracerebral hemorrhage</td>
<td>Suicidal ideation</td>
</tr>
<tr>
<td>Liver malfunction</td>
<td>Cholecystitis</td>
<td>Infection</td>
<td>Syncope</td>
</tr>
<tr>
<td></td>
<td>Cerebrovascular accident</td>
<td>Kidney stone</td>
<td>Urinary tract infection</td>
</tr>
</tbody>
</table>
Improving Clinical Documentation

Table 1. Performance of the different negation detection algorithms on 200 test sentences.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NegEx</td>
<td>0.699</td>
<td>0.875</td>
<td>0.777</td>
</tr>
<tr>
<td>Added rules</td>
<td>0.833</td>
<td>0.982</td>
<td>0.901</td>
</tr>
<tr>
<td>Perceptron</td>
<td>0.901</td>
<td>0.925</td>
<td>0.913</td>
</tr>
</tbody>
</table>

Table 2. Performance of the linear SVMs on chief complaint prediction, without and with negation detection. The Best-n accuracy measures how often the list of n most likely predicted labels actually contained all of the true chief complaints, and DCG stands for the Discounted Cumulative Gain, which measures the quality of the whole ranking.

<table>
<thead>
<tr>
<th>SVM Configuration</th>
<th>Best-5</th>
<th>Best-10</th>
<th>DCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negation detection none</td>
<td>0.496</td>
<td>0.615</td>
<td>0.381</td>
</tr>
<tr>
<td>Perceptron</td>
<td>0.511</td>
<td>0.620</td>
<td>0.393</td>
</tr>
<tr>
<td>SVM</td>
<td>0.753</td>
<td>0.819</td>
<td>0.601</td>
</tr>
<tr>
<td>SVM</td>
<td>0.757</td>
<td>0.825</td>
<td>0.613</td>
</tr>
</tbody>
</table>

Figure 1. Screenshots of the system now running at BIDMC hospital on note: 69 y/o M patient with severe intermittent RUQ pain. Began soon after eating bucket of ice cream and cupcake. Also is a heavy drinker. Left: the system correctly proposes both ‘RUQ abdominal pain’ and ‘Allergic reaction’ as possible chief complaints. Right: If the nurse does not see the label they want, they can start typing and see a list of suggested auto-completes. Again, the four most likely labels describe ‘RUQ abdominal pain’ and ‘Allergic reaction’.

References
Improving Clinical Documentation

Percentage of *standardized* chief complaints (per week)
At a Broader Time Scale...

**Demographic data:**
- Age/gender
- Socioeconomic status, lifestyle
- Company code

**Medical Claims:**
- ICD9 diagnosis code
- CPT code (procedure)
- Specialty
- Location of service
- Date of Service

**Medications:**
- NDC code (drug name)
- Days of supply
- Quantity
- Service Provider ID
- Date of fill

**Lab Tests:**
- LOINC code (urine or blood test name)
- Results (actual values)
- Lab ID
- Range high/low
- Date
• Find markers of disease stage and progression, statistics of what to expect when
  – What is the “typical trajectory” of a female diagnosed with Sjögren’s syndrome at the age of 19?

• Estimate a patient’s future disease progression
  – When will a specific individual with smoldering multiple myeloma (a rare blood cancer) transition to full-blown multiple myeloma?
  – Which second-line diabetes treatment should we give to a patient?
Personalized Prescriptions

Me

???

Drug A
or
Drug B

Patient 1

Drug A

Drug C

Patient 2

Drug B
From Data Generation to Decision Making
Many Challenges Unique to Medicine

• Life or death decisions
  – Need robust algorithms
  – Checks and balances built into ML deployment
  – (Also arises in other applications of AI such as autonomous driving)
  – Need fair and accountable algorithms

• Many questions are about unsupervised learning
  – Discovering disease subtypes, or answering question such as “characterize the types of people that are highly likely to be readmitted to the hospital”?

• Many of the questions we want to answer are causal
  – Naïve use of supervised machine learning is insufficient
Problems with “Data”

- Often very little labeled data (e.g., for clinical NLP)
  - Motivates semi-supervised learning algorithms
- Sometimes small numbers of samples (e.g., a rare disease)
  - Learn as much as possible from other data (e.g. healthy patients)
  - Model the problem carefully
- Lots of missing data, varying time intervals, censored labels
Problems with Clinical Integration

• Difficulty of de-identifying data
  – Need for data sharing agreements and sensitivity

• Difficulty of deploying ML
  – Commercial electronic health record software is difficult to modify
  – Data is often in silos; everyone recognizes need for interoperability, but slow progress
  – Careful testing and iteration is needed
PART V: Utility of ML in Medical Imaging
Machine Learning for Diagnosis

### Top-10 Diseases

- **Neoplasms**
- **Nervous**
- **Cardiovascular**
- **Urogenital**
- **Pregnancy**
- **Digestive**
- **Respiratory**
- **Skin**
- **Endocrine**
- **Nutritional**

### Top-10 Techniques

- Support Vector Machine, 42%
- Neural Network, 31%
- Others, 11%
- Linear Regression, 3%
- Random Forest, 1%
- Discriminant Analysis, 8%
- Logistic Regression, 2%
- Naïve Bayes, 2%
- Nearest Neighbor, 3%
- Decision Tree, 2%
Deep Learning on the Rise

Number of DL Studies

DL Studies Based on Data Type

- Diagnostic Imaging
- Electrodiagnosis
- Genetic diagnosis
- Clinical Laboratory
- Mass Screening
- Others
Medical images are high-dimensional (volumetric and temporal)
Medical images are mostly interpreted by radiologists (manual labor)
Humans are quite poor in seeing fine-grained patterns in static images
Similarities medical–natural images (closely tied to computer vision)
Imaging Morphology and Function

Anatomical

Functional

Diffusion
Modern Imaging Modalities

X-ray

(a)

Nuclear Medicine

(b)

Ultrasound

(c)

MRI

(d)
Ultrasound

- Uses sound pressure waves
- We cannot hear these sounds: > 20 kHz
- Typically 2-18 MHz
Ultrasound

(a) Ultrasound machine

(b) Ultrasound image
Ultrasound

- Higher resolution with increasing frequency
- Cannot image too deep if the frequency is high
- Lots of application areas: cardiology, urology, obstetrics,...

Doppler Ultrasound
X-ray

- Uses X-ray photons
- Photons at VERY high frequency: \(~10^{16}-10^{19}\) Hz
- Different tissues attenuate photons differently \(\rightarrow\) contrast
- Very good at detecting bone structure

- Projection images
Chest X-ray
Projection Through the Body
Computerized Tomography

- Extends X-ray imaging to 3D format
- Rotate X-ray source and detectors all together
Computerized Tomography
Nuclear Medicine

- A radionuclide is injected into the blood stream
- Goes to cancer locations before cancer cells are hyperactive
- Emitted Gamma rays are detected for imaging
Positron Emission Tomography

- Resolution is not good, but very sensitive to cancer
MRI
Magnetic Resonance Imaging (MRI)

- FLEXIBLE CONTRAST
- Arbitrary geometries
- Non-invasive, non-ionizing imaging
- Relatively slow imaging
Medical Imaging Pipeline

Serial medical imaging pipeline

Current

Acquisition ➔ Reconstruction ➔ Analysis ➔ Diagnosis

Future

End-to-end integrated medical imaging pipeline

Acquisition ➔ Diagnosis
• Radiologists need to interpret an excessively large number of images
• Their capacity to correctly interpret images is overwhelmed
• Automated image analysis systems are needed for error reduction
• Machine learning underpins the algorithms for such systems
PART V: Example Applications in Medical Imaging
Examples: Detecting Micro-calcifications

- Mammogram Region (a)
- SVM Output (b)
- Detected Lesion Positions (c)
Examples: Detecting Pulmonary Abnormalities
Examples: Detecting Pulmonary Abnormalities
Examples: Segmentation of Ventricles

- **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)
**Topics:** deep belief network

- The idea of pre-training came from work on deep belief networks (DBNs)
  - it is a generative model that mixes undirected and directed connections between variables
  - top 2 layers' distribution \( p(h^{(2)}, h^{(3)}) \) is an RBM
  - other layers form a Bayesian network:
    - the conditional distributions of a layers given the one above it are
      \[
      p(h_j^{(1)} = 1|h^{(2)}) = \text{sigm}(b^{(1)} + W^{(2)\top} h^{(2)}) \\
p(x_i = 1|h^{(1)}) = \text{sigm}(b^{(0)} + W^{(1)\top} h^{(1)})
      \]
    - this is referred to as a sigmoid belief network (SBN)
  - a DBN **is not** a feed-forward network
Examples: Segmentation of Ventricles

DEEP BELIEF NETWORK

Topics: deep belief network

• This is where the RBM stacking procedure comes from
  ‣ idea: improve prior on last layer by adding another hidden layer
  ‣ how do we train these additional layers?

\[
p(x, h^{(1)}) = p(x|h^{(1)}) \sum p(h^{(1)}, h^{(2)})
\]

\[
p(x) = \sum_{h^{(1)}} p(x, h^{(1)})
\]

\[
p(h^{(1)}, h^{(2)}) = p(h^{(1)}|h^{(2)}) \sum_h p(h^{(2)}, h^{(3)})
\]
Examples: Segmentation of Ventricles

- 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)

<table>
<thead>
<tr>
<th>Training image</th>
<th>Negatives</th>
<th>Positives</th>
<th>Perpendicular lines</th>
<th>Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Rigid training patches</td>
<td></td>
<td></td>
<td>b) Non-rigid training profiles</td>
<td></td>
</tr>
</tbody>
</table>

(a) Layer 1  
(b) Layer 2  
(c) Layer 3  
(d) Layer 4
Examples: Segmentation of Hippocampus

- **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
Examples: Segmentation of Hippocampus

Hand-Crafted Features

- Limited discriminative power

Extracting patches from a 7T MR image

Responses of Haar filters for the image patches in a-c
Examples: Segmentation of Hippocampus

Hierarchical Feature Extraction

Stacked two-layer convolutional ISA (Independent Subspace Analysis)
Examples: Segmentation of Hippocampus

Qualitative Evaluations

Ground Truth  Haar + Texture Features  Hierarchical Features
Determine accurate correspondences between images
Examples: Image Registration

![Diagram](Image)

- Encoder
- Decoder
- Representations

Morphological signatures for image registration

Input image patches
Examples: Image Registration

Difficulty #2: How to deal with high dimensional training data?
Solution: Use the convolutional RBM in each layer

- $N_w$ - $N_v$
Examples: Tissue-Specific Segmentation

- Image Volume
- 2D Image Input
- Encoder Network
- Decoder Network
- Convolution Layer
- BN Layer
- ReLU Layer
- Pooling Layer
- Upsampling Layer
- Softmax Layer
- 2D Label Output
- Segmentation
Examples: Tissue-Specific Segmentation

Segmentation of tumorous tissues:

- Active cells
- Necrotic core
- Edema
- Background

Multi-channel 3D MRI input data

Ex. Tissue Specific Segmentation of Brain Tumors
Examples: Tissue-Specific Segmentation
Examples: Tissue-Specific Segmentation

- Location
- Shape
- Intensity
- Texture

Patient A
Patient B
Patient C
Patient D
Patient E

Challenge: variability of input data
Examples: Predicting Survival from Histopathology

A. Whole-slide imaging
   - Resection / biopsy
   - Whole-slide image

B. Region of interest selection
   - Web viewer
   - Regions of interest

C. Survival Convolutional Neural Network (SCNN)
   - Prediction error (negative log-likelihood)
   - Convolutional layers
   - Fully connected layers
   - Cox model
   - Patient survival

High power field (20X objective)

Convolution
Pooling
Rectifier
Examples: Classifying Retinal Disease
Examples: Classifying Retinal Disease
Examples: Denoising/Dealiasing Images

\[ \hat{x} = \min_{x} \left\| F_{u} x - y_{u} \right\|_{2}^{2} + \left\| C(x_{u}) - x \right\|^{2} \]

Data consistency  Consistency with network
Model-based Deep Learning: Cascaded CNNs
Examples: Denoising/Dealiasing images

Deep CNN

Training

Undersampled

Fully-sampled

Testing

Undersampled

Reconstruction

Forward pass
Model parameters = Θ_{img}

Backward pass
Θ_{img-ft}(updated) = Θ_{img} + ΔΘ_{img}(update)

Conv1 → DC → ⋯ → Conv5 → DC
Medical Data Are Scarce

ZF

# of training samples = 250

Neural Network

Reconstruction
Transfer Learning

Forward pass
Model parameters = Θ

Neural Network

Backward pass
Θ_{IMG}(updated) = Θ + ΔΘ_{IMG}(update)

ZF

Ground truth
Fine Tuning

No. of samples for fine-tuning = 0

SSIM: 0.933
PSNR: 30.72

SSIM: 0.959
PSNR: 33.57

ImageNet-trained
4000 samples

T1-trained
4000 samples
### Examples: Denoising/Dealiasing images

<table>
<thead>
<tr>
<th></th>
<th>CS</th>
<th>ImageNet-trained</th>
<th>T1-trained</th>
<th>T2-trained</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM</td>
<td>0.928</td>
<td>0.956</td>
<td>0.958</td>
<td>0.956</td>
<td>1.00</td>
</tr>
<tr>
<td>PSNR</td>
<td>30.79</td>
<td>33.29</td>
<td>33.60</td>
<td>33.39</td>
<td>Inf</td>
</tr>
</tbody>
</table>

![Images showing the denoising/dealiasing process](image-url)
Examples: Synthesizing Missing Images

![Diagram showing the training and testing phases of synthesizing missing images.](image-url)
What is a Generative Model?

- Density estimation
- Sample generation

Training examples → Model samples
Generative Adversarial Network (GAN)

$D(x)$ tries to be near 1

$x$ sampled from data

Differentiable function $D$

$D$ tries to make $D(G(z))$ near 0,
$G$ tries to make $D(G(z))$ near 1

$x$ sampled from model

Differentiable function $G$

Input noise $z$
Examples: Synthesizing Missing Images

<table>
<thead>
<tr>
<th>pGAN</th>
<th>Multimodal</th>
<th>Replica</th>
<th>Reference</th>
<th>Source</th>
</tr>
</thead>
</table>

Error: 0 - 0.6
Examples: Synthesizing Missing Images

<table>
<thead>
<tr>
<th>pGAN</th>
<th>Multimodal</th>
<th>Replica</th>
<th>Reference</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Error

0 - 0.6
Future Outlook:

**DL Tasks:**
- Image Reconstruction
- Image Synthesis
- Data Acquisition
- Image Segmentation
- Tumor/Lesion Detection
- Disease Grading
- Monitoring
- Treatment Planning
Acknowledgment

Slide Credits

• Book: Principles of Neural Science
• Book: Medical Imaging Signals and Systems
• Lecture: Pascal Vincent
• Lecture: David Sontag
• Lecture: Ulas Bagci
• Lecture: Emine U Saritas
• Paper: Jarret, CVPR 2009
• Paper: Rueckert, arXiv 2019
• Paper: Waldrop, PNAS 2019
• Paper: Krizhevsky, NIPS 2012
• Paper: Jiang, SVN 2017
• Site: asimovinstitute.org
• Site: vinodsblog.com
• Site: doi.org/10.1016/j.media.2017.07.005