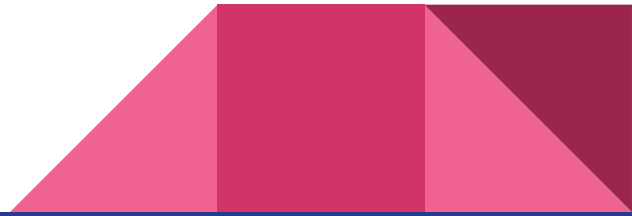
# Differential Privacy

ter.ps/389iweek10

# What is Differential Privacy?

# Healthcare Context

- Anonymize and obfuscate patient data
- Give patients the ability to opt-out of sharing personal information
- Increase privacy without limiting access to data for researchers

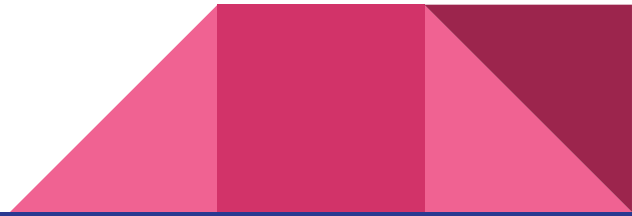


# Regulation: HIPAA

- Patient is in control of the sharing of their personal health information
- Identifiable patient info removed before data is shared for research
  - Treatment records and diagnosis
  - Age, Gender, biological information
- Does not cover all personally identifiable information
  - Results from consumer genetic testing aren’t covered under HIPAA

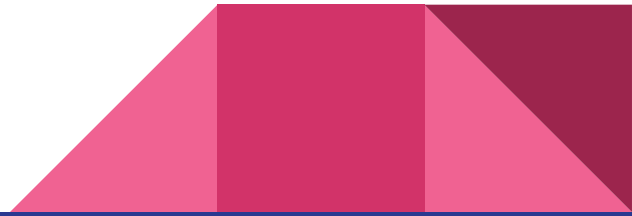
# Current Healthcare Data Landscape

- Healthcare data is controlled by a few large institutions
- Lots of time & money is required to buy datasets and sign BAAs
- Large pharmaceutical companies
- Academic researchers have limited access without industry partnerships



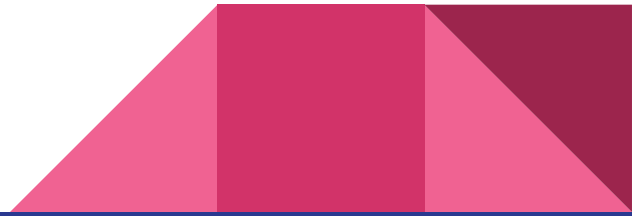
# Privacy and Healthcare ML / Research

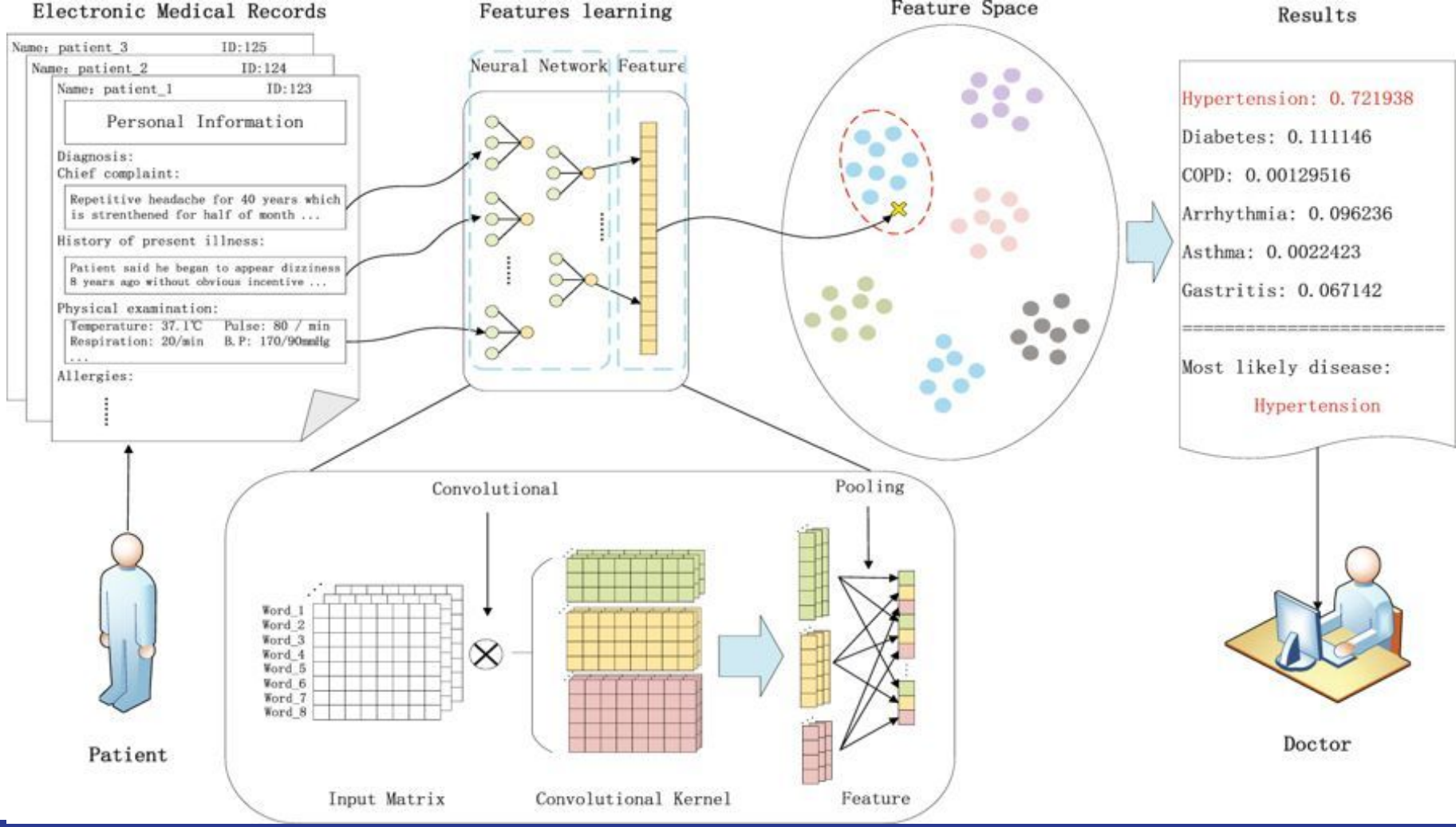
- Hot topic in research right now
- Deep Learning models, Neural Networks
- Generate fake data for researchers to use
- Lead to cheaper datasets
- Ethical issues:
  - Potential problems with accuracy

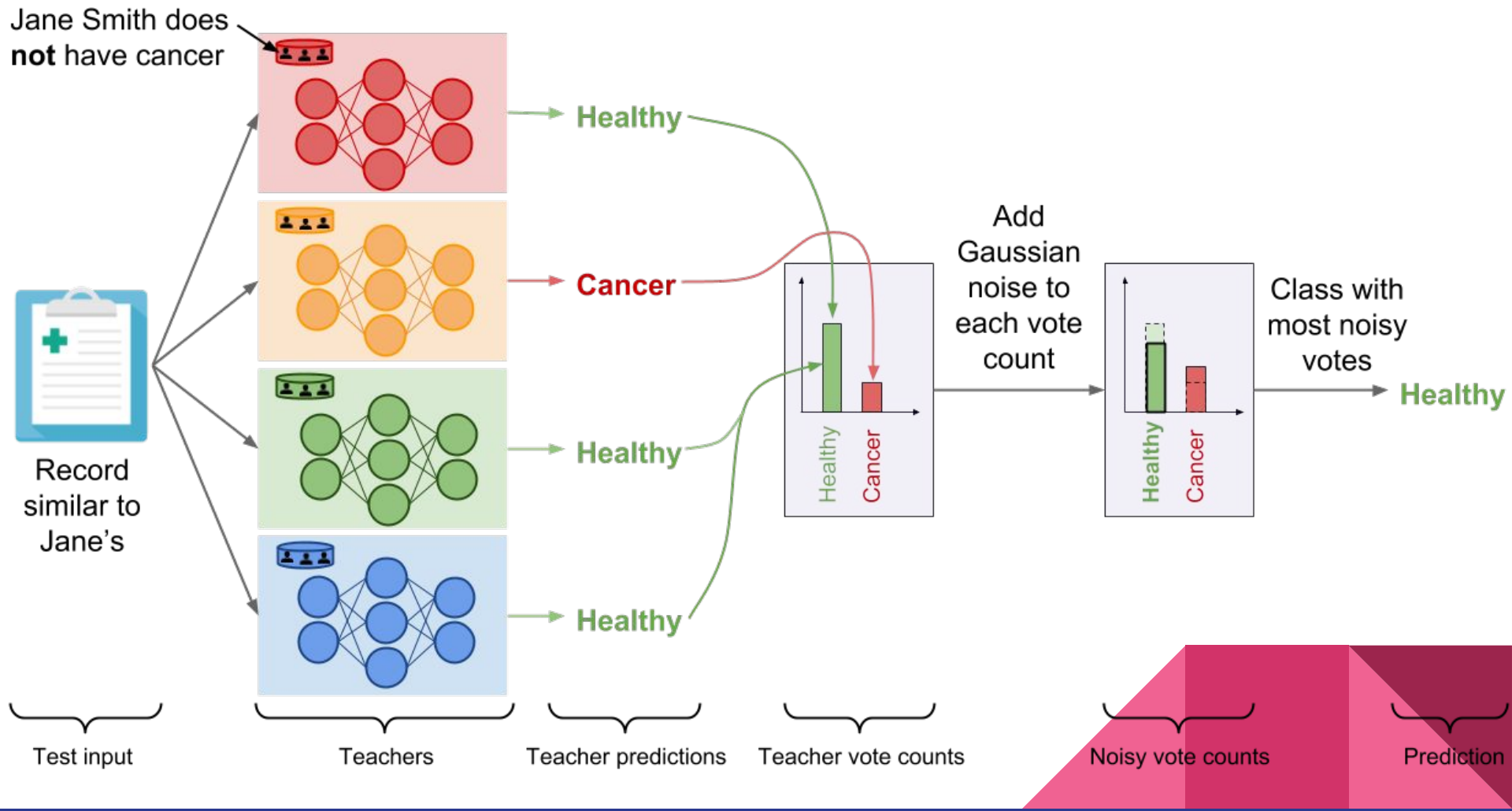


# Deep Learning & Neural Networks

- Exploratory models that generate weights for features
- Models generate data that mimics real patient data
- https://www.nature.com/articles/s41598-018-24389-w
- http://www.cleverhans.io/privacy/2018/04/29/privacy-and-machine-learning.html



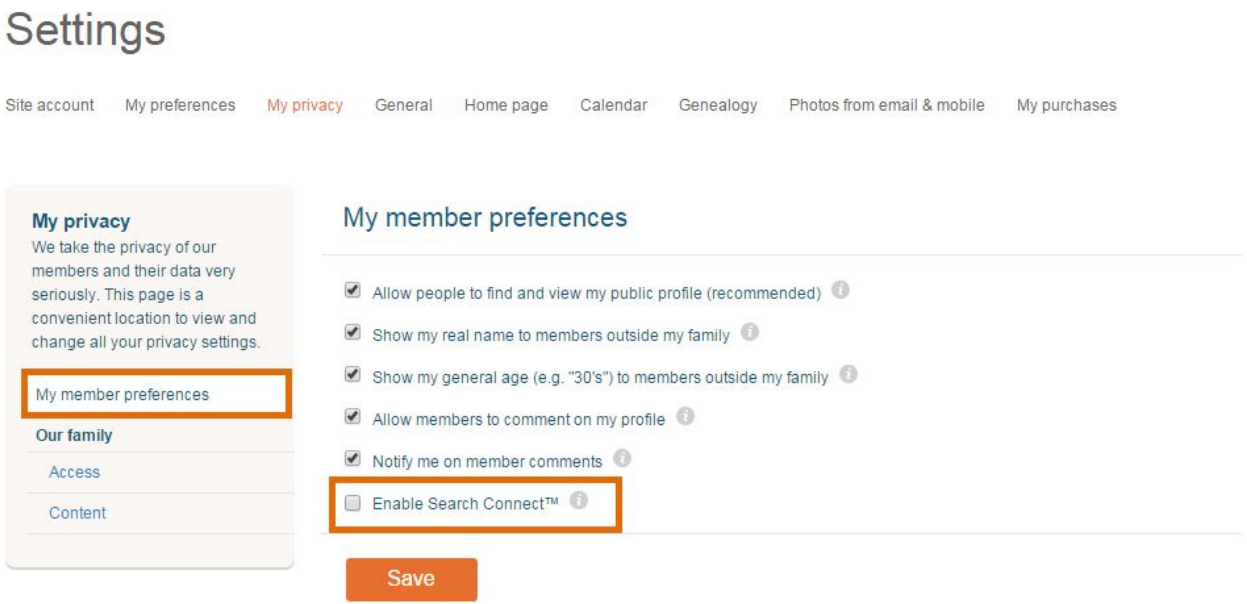

Electronic Medical Records
Features learning
Feature Space
Results
Name: patient_3 ID:125
Name: patient_2 ID:124
Name: patient_1 ID:123
Personal Information
Diagnosis:
Chief complaint:
Repetitive headache for 40 years which is strenthened for half of month ...
History of present illness:
Patient said he began to appear dizziness 8 years ago without obvious incentive ...
Physical examination:
Temperature: 37.1℃ Pulse: 80 / min
Respiration: 20/min B.P: 170/90mmHg
...
Allergies:
Neural Network Feature
Hypertension: 0.721938
Diabetes: 0.111146
COPD: 0.00129516
Arrhythmia: 0.096236
Asthma: 0.0022423
Gastritis: 0.067142
Most likely disease:
Hypertension
Convolutional
Pooling
Word_1
Word_2
Word_3
Word_4
Word_5
Word_6
Word_7
Word_8
Input Matrix
Convolutional Kernel
Feature
Patient
Doctor


Jane Smith does **not** have cancer
Healthy
Cancer
Healthy
Healthy
Record similar to Jane's
Healthy
Cancer
Add Gaussian noise to each vote count
Healthy
Cancer
Class with most noisy votes
Healthy
Test input
Teachers
Teacher predictions
Teacher vote counts
Noisy vote counts
Prediction

# Privacy & DNA Testing

- DNA is currently not regulated under HIPAA
  - Major players: 23 and Me, Helix, Ancestry, My Heritage
  - Companies can sell data to insurance providers, government
- DNA breaches are harder to fix
- https://www.theverge.com/2018/8/1/17638680/genetic-data-privacy-consumer-rights-guidelines-23andme-ancestry

# My Heritage Privacy Settings



# Blockchain?

- Potential to give patients more power
  - Patients can restrict access to data without exposing themselves
  - Get paid for sharing information
  - Smart contracts to aggregate large quantities of anonymized data
- More research needs to be done before widespread adoption
- Regulatory challenges
- https://www.forbes.com/sites/forbestechcouncil/2018/04/13/blockchain-in-health-care-the-good-the-bad-and-the-ugly/#5ab202296278

# Readings

- https://pdfs.semanticscholar.org/65a5/37c9cd327c2925676f59ddffa01cf4afbe51.pdf
- https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf
- https://www.broadinstitute.org/count-me-in
- https://www.theverge.com/2018/8/1/17638680/genetic-data-privacy-consumer-rights-guidelines-23andme-ancestry
- https://www.theverge.com/2018/6/6/17435166/myheritage-dna-breach-genetic-privacy-bioethics