Deep Learning & Generative AI in Healthcare

Session 01

Course Structure and Objectives

Course Format & Resources

Component	Details
Format	Lecture + Interactive Lab sessions (hands-on coding in Jupyter Notebooks)
Resources	Slides, notebooks, readings on Canvas Code repositories on GitHub
Evaluation	Homework assignments Capstone project (individual)

Prerequisites

✓ Proficiency in **Python** programming

✓ Working knowledge of **PyTorch**

✓ Core ML concepts (supervised/unsupervised, evaluation metrics)

Learning Outcomes

1. Model Development

Design, train, and debug deep-learning models for **medical imaging, drug discovery,** and related tasks

2. Generative AI Application

Apply state-of-the-art generative models to healthcare data:

- GANs (Generative Adversarial Networks)
- VAEs (Variational Autoencoders)
- Diffusion models

3. Safety & Compliance

Assess model **safety, bias,** and **regulatory compliance;** communicate findings to technical and clinical stakeholders

Some Useful Resources

Deep Learning
Foundations and Concepts

Christopher M. Bishop & Hugh Bishop
Springer

Hands-On Machine Learning
with Scikit-Learn, Keras & TensorFlow

Aurélien Géron
O'Reilly

Deep Learning with Python
Second Edition

François Chollet
Manning

Deep Learning for the Life Sciences
Genomics, Microscopy, Drug Discovery & More

Ramsundar, Eastman, Walters & Pande
O'Reilly

Natural Language Processing with Transformers
Building Language Applications with Hugging Face

Tunstall, von Werra & Wolf
O'Reilly

LLMs and Generative AI for Healthcare
The Next Frontier

Kerrie Holley & Marian Mosher
O'Reilly

Hands-On Large Language Models
Language Understanding and Generation

Jay Alammar & Maarten Grootendorst
O'Reilly

Build a Large Language Model
From Scratch

Sebastian Raschka
Manning

Hands-On Generative AI with Transformers and Diffusion Models

Cuenca, Passot & Whitaker
O'Reilly

The Impact of Deep Learning

Unified Framework Advantage

Traditional ML	Deep Learning
Specialized techniques per domain	Single fundamental framework
Domain-specific expertise required	Transferable across domains
Limited scalability	Scales with data and compute

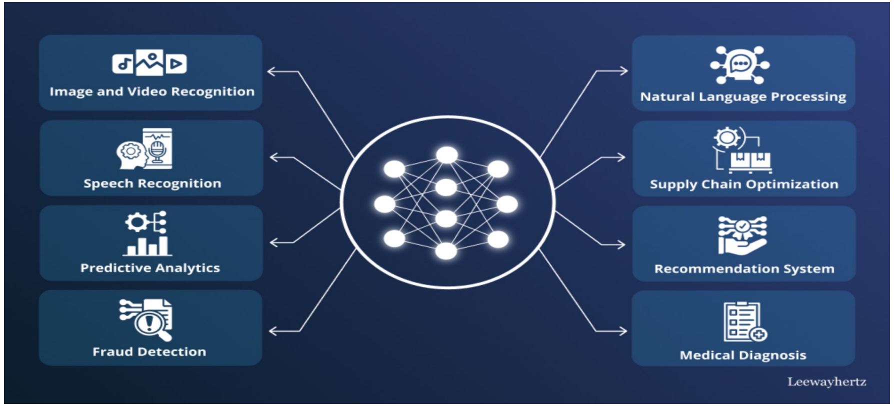
> **Key Insight:** Deep learning's transformational impact extends to virtually every computational domain, including healthcare

Application Breadth

Computer Vision
Image classification, object detection, segmentation

Natural Language
Translation, generation, understanding

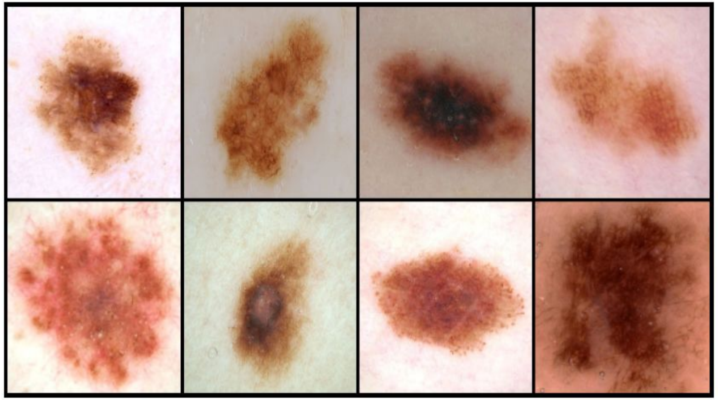
Healthcare
Diagnosis, drug discovery, treatment planning



Medical Diagnosis: Skin Cancer Detection

Clinical Challenge

- **Melanoma:** Most dangerous skin cancer, curable if detected early
- Distinguishing malignant melanomas from benign nevi is extremely challenging
- Impossible to manually code accurate classification algorithms



Deep Learning Solution

Component	Details
Dataset	~129,000 labeled lesion images (biopsy-verified)
Architecture	Deep neural network with ~25 million parameters
Approach	Transfer learning (pre-trained on 1.28M natural images)
Outcome	Accuracy exceeds professional dermatologists

Study: Esteva et al., 2017; Brinker et al., 2019

Performance Achievement

Surpassed Expert Level

Classification accuracy > dermatologists

Key ML Concepts & Transfer Learning Strategy

Training Process

Term	Definition
Training Set	Labeled data used to set parameter values
Weights	Adjustable parameters (~25M for melanoma model)
Learning/Training	Process of determining parameter values from data

Problem Types

Supervised Learning

Network receives correct labels during training

Classification

Output: discrete classes (benign/malignant)

Regression

Output: continuous variables (e.g., chemical yield)

Transfer Learning Workflow

Step 1: Pre-training

1.28M Images

Everyday objects (dogs, buildings, mushrooms)

Learn general properties of natural images

⬇

Step 2: Fine-tuning

129K Images

Lesion images (labeled by biopsy)

Specialize to skin lesion classification

Why Transfer Learning? The 129K lesion dataset is relatively small for deep learning. Pre-training on larger general dataset provides robust feature representations.

Protein Structure Prediction

The Biological Challenge

Proteins: Building blocks of living organisms

- Chains of 22 different amino acid types
- Fold into complex 3D structures
- Shape determines behavior and interactions

Experimental Methods	Limitations
X-ray crystallography Cryogenic electron microscopy Nuclear magnetic resonance	• Time-consuming • Difficult for some proteins • Context-dependent • Requires pure samples

Historical Context

Predicting 3D structure from amino acid sequence: **fundamental open problem for 50 years** with little progress until deep learning

Deep Learning Solution: AlphaFold

Input

Amino acid sequence

Lower cost, higher throughput to determine

⬇ Deep Learning Model ⬇

Output

Predicted 3D protein structure

Critical for understanding biology & drug discovery

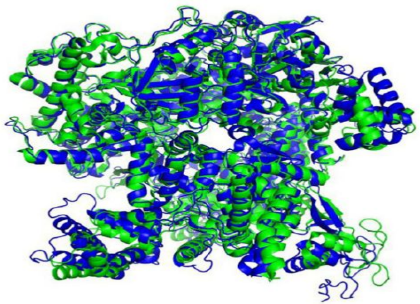


Image Synthesis & Generative AI

Paradigm Shift: Unsupervised Learning

Previous Examples	Image Synthesis
Transform input to output (skin image → classification) (amino acid → 3D structure)	Learn from sample images Generate new similar images (no labels required)
Supervised Learning Labeled training data	**Unsupervised Learning** Unlabeled images only

Generative Models

Definition: Models that generate new outputs differing from training data but sharing the same statistical properties

Unconditional

Generate samples from learned distribution

Conditional

Generate from text prompt (semantics)

Quality

Difficult to distinguish from real data

Generative AI Output Modalities

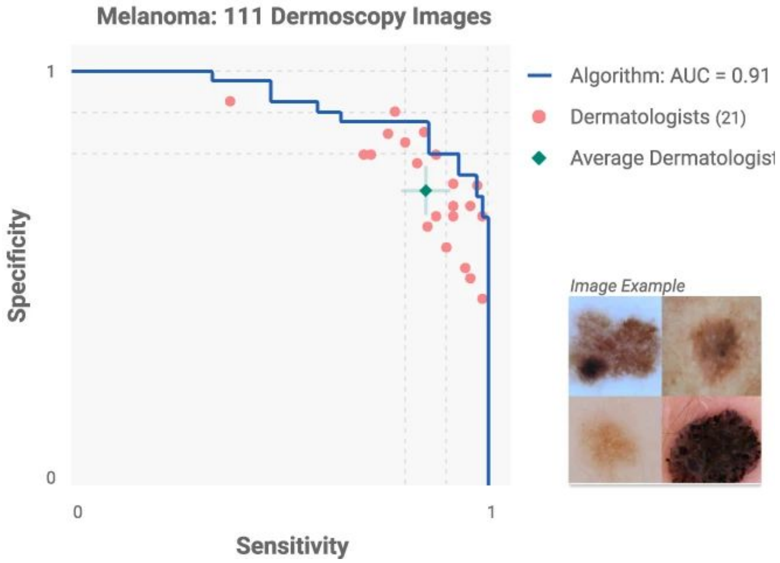
- 📷 Images
- 🎬 Video
- 🔉 Audio
- 📝 Text
- 💊 Drug Molecules
- 🧬 Other Modalities

AI in Medical Diagnosis

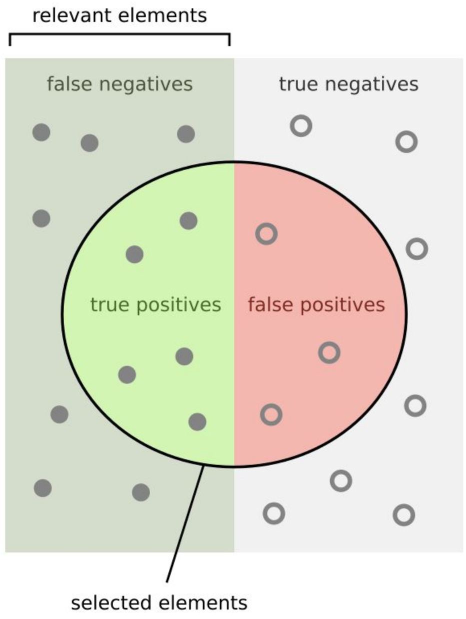
Transformation Potential:

From diagnostics to drug discovery, deep learning is driving innovation.

Melanoma Detection Performance



Classification Metrics



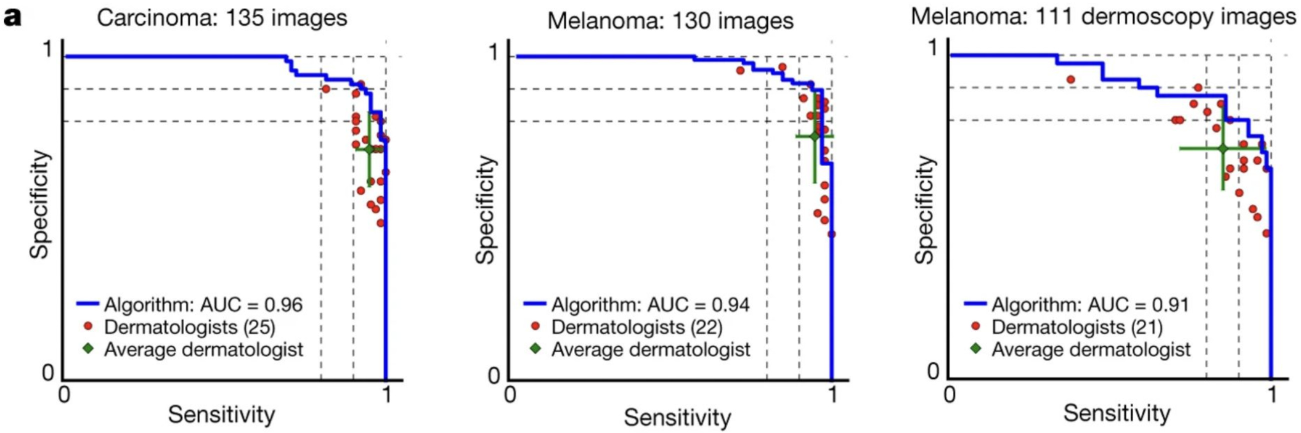
AI in Medical Diagnosis

Growing Need:

Address physician shortages, reduce medical errors, accelerate drug discovery, personalized medicine.

Multi-Cancer Type Performance (Esteva et al., Nature 2017)

Carcinoma	Melanoma	Melanoma
135 images	**130 images**	**111 dermoscopy images**
AUC = 0.96	**AUC = 0.94**	**AUC = 0.91**
vs. 25 Dermatologists	vs. 22 Dermatologists	vs. 21 Dermatologists



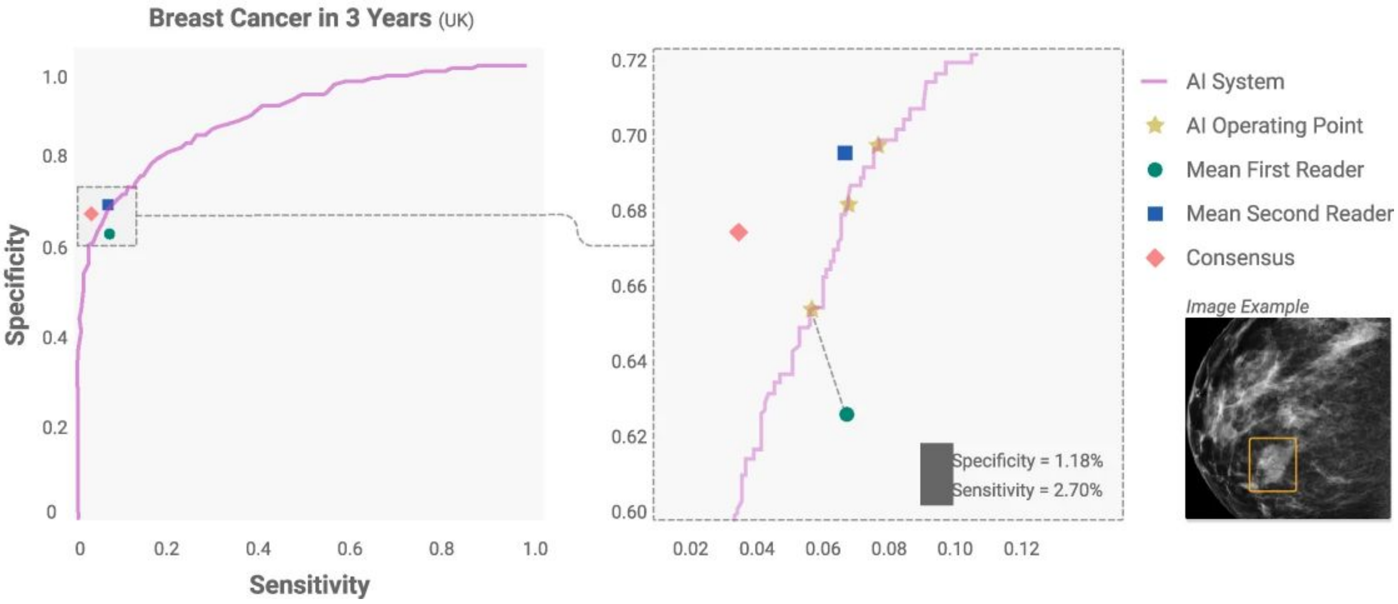
AI in Medical Diagnosis

Transformation Potential:

From diagnostics to drug discovery, deep learning is driving innovation.

Breast Cancer Detection (UK Study)

AI vs Human Readers Comparison



AI in Medicine: Drug Discovery & Molecular Dynamics

Geometric Learning

Input: Amino acid (AA) sequence

⬇

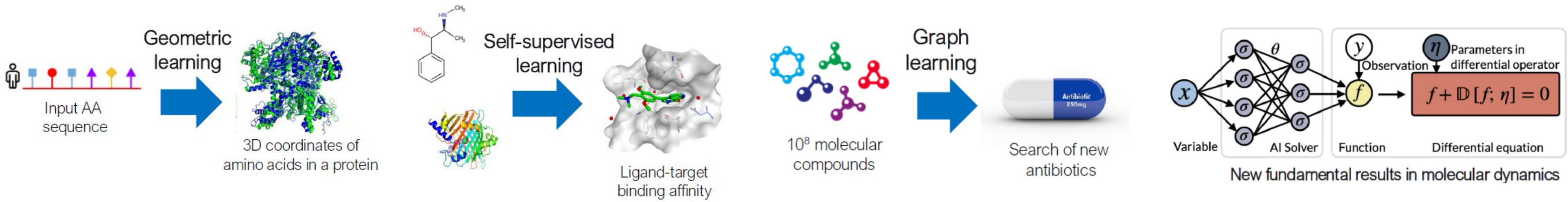
Output: 3D coordinates of amino acids in protein

Graph Learning

Scale: 10^8 molecular compounds

⬇

Application: Search for new antibiotics



Self-Supervised Learning

Application: Ligand-target binding affinity prediction

AI for Molecular Dynamics

Breakthrough: New fundamental results in molecular dynamics

Variable → AI Solver → Function → Differential equation

Parameters embedded in differential operators

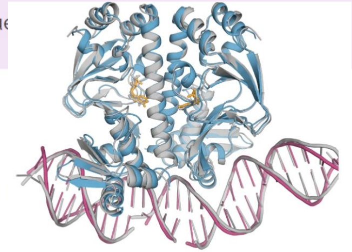
AI in Drug Discovery: AlphaFold 3

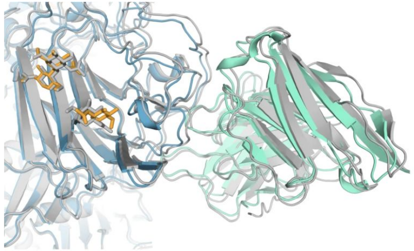
Growing Need: Address physician shortages, reduce medical errors, accelerate drug discovery, personalized medicine.

AlphaFold 3 Capabilities

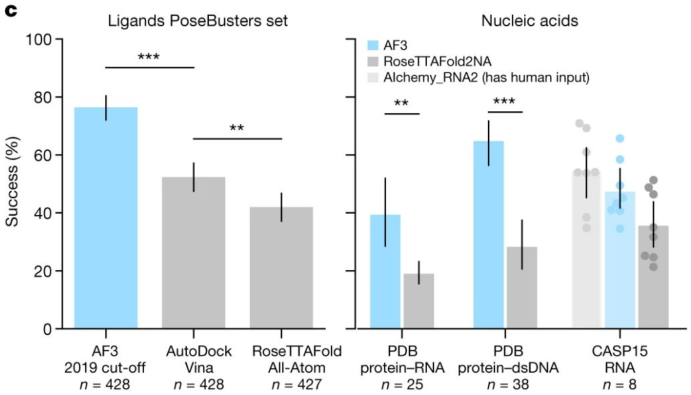
Updated diffusion-based architecture capable of predicting the joint structure of complexes including:

- Proteins
- Nucleic acids
- Small molecules
- Ions
- Modified residue

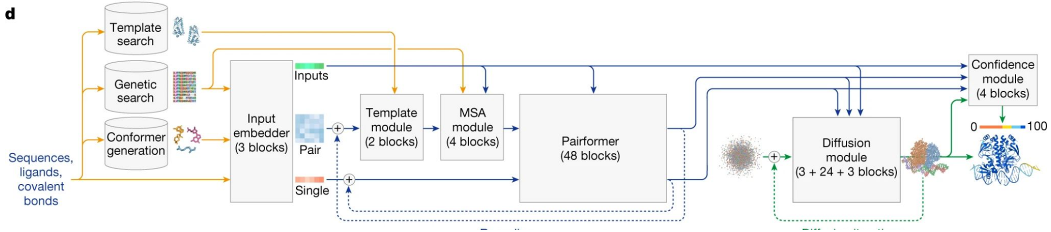




Performance Across Prediction Tasks



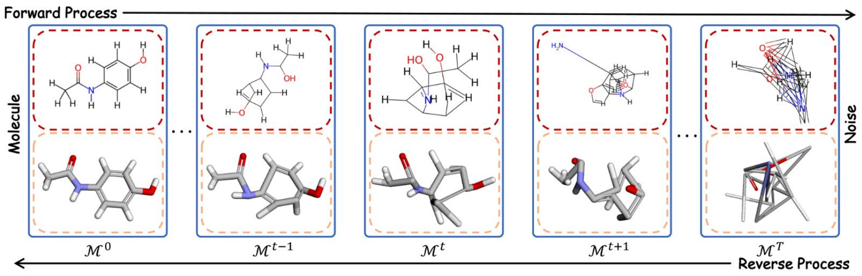
Architecture Workflow



AI in Drug Discovery: Equivariant Diffusion

Equivariant Diffusion for Small Molecule Generation

Forward & Reverse Process



Mathematical Formulations:

- **DDPM:** Denoising Diffusion Probabilistic Models
- **SMLD:** Score Matching with Langevin Dynamics
- **SDE:** Stochastic Differential Equations

Conditional Generation Tasks

Optional Conditions:

- Property (e.g., non-toxic, specific solubility)
- Target (protein binding site)
- Fragment (molecular scaffold)
- Composition (e.g., C8H9NO2)

Task 1: De Novo Generation

Generate novel molecules from scratch based on desired properties

Task 2: Molecular Optimization

Iteratively improve existing molecules for specific targets

Task 3: Conformer Generation

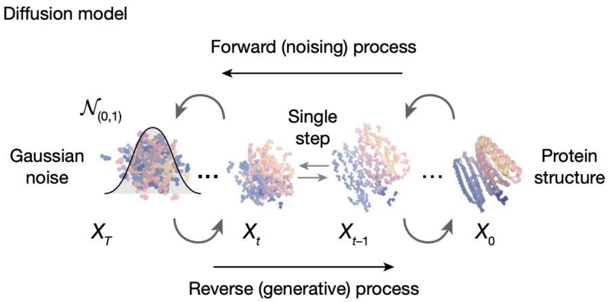
Generate 3D conformations of molecules

AI in Drug Discovery: RFdiffusion for Protein Design

De novo design of protein structure and function with RFdiffusion

Nature 620, 1089–1100 (2023) | **2024 Nobel Prize in Chemistry**

Diffusion Model for Proteins



Symmetric Extensions:

- Symmetric noise application
- Symmetric oligomer generation
- Maintains symmetry constraints during diffusion

Protein Design Applications

Binding Target
Design proteins that bind specific targets

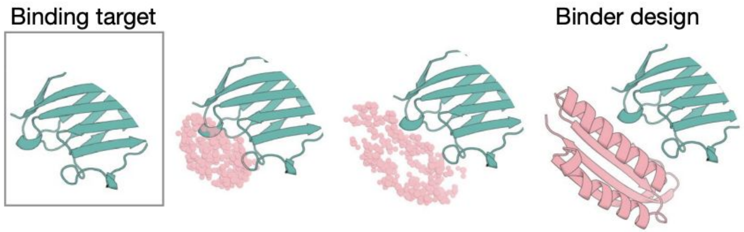
Binder Design
Progressive binder scaffolding

Functional Motif
Incorporate functional elements

Motif Scaffolding
Build scaffolds around motifs

Symmetric Motif
Design symmetric functional units

Symmetric Scaffolding
Build symmetric protein assemblies



AI in Healthcare: All-Purpose Prediction Engines

Health system-scale language models are all-purpose predictions engines

Clinical Tasks (Physician)

In-hospital mortality prediction
How likely is the patient to die in the hospital before discharge?

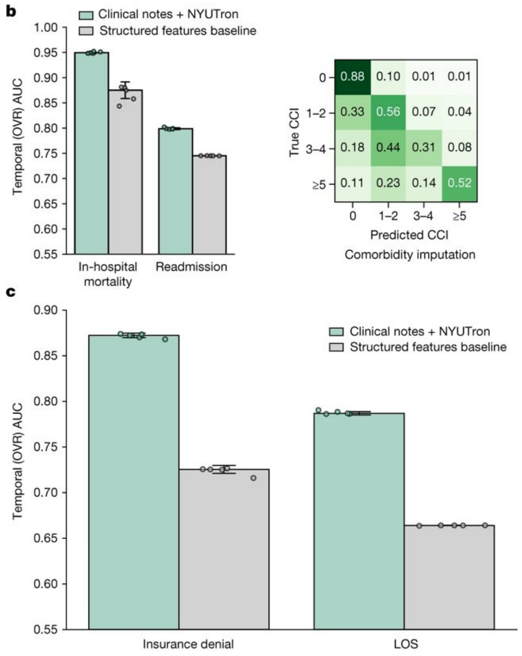
Binned comorbidity index imputation
Without structured ICDS, how sick/chronically ill is the patient?

30-day all-cause readmission prediction
How likely is the patient to come back within 30 days of discharge?

Operational Tasks (Admin)

Binned LOS prediction
How long will the patient stay in the hospital?

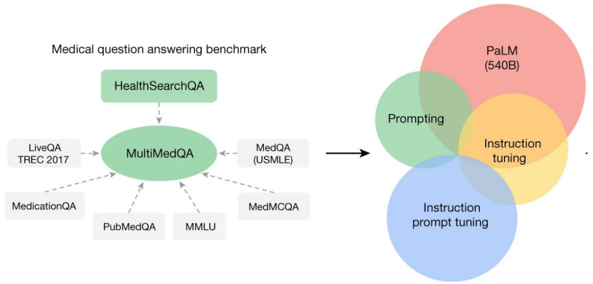
Performance: NYUTron System



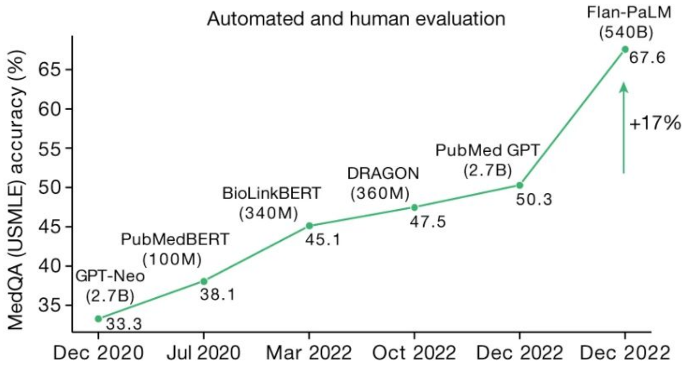
AI in Healthcare: Clinical Knowledge Encoding

Large language models encode clinical knowledge

Medical Question Answering Benchmark



Performance Evolution



Training Approaches

Prompting

Instruction tuning

Instruction prompt tuning

PaLM (540B)

Example Question

Q: How long does it take for newborn jaundice to go away?

Med-PaLM Answer: Newborn jaundice is caused by buildup of bilirubin. It is common and typically harmless, but can be a sign of a more serious condition. The jaundice typically goes away on its own within a few weeks...

Result: Med-PaLM performs encouragingly on consumer medical question answering

What Makes Biomedical Data Different?

Data Availability:

Electronic Health Records (EHRs), medical imaging repositories, genomics data.

Multi-Source Data Integration

Diverse Data Sources:

- Low Activity measurements
- SNP Array (genetic data)
- Proteomics
- Histology (tissue samples)
- Patient History
- Drug Use records

Healthcare Data Ecosystem

Data Flow Architecture:

- Molecular data (genomics, proteomics)
- Imaging (radiology, pathology)
- Patient History (demographics, encounters)
- ↓ Aggregated in EHR systems
- ↓ Stored in Cloud infrastructure
- ↓ Accessed by Enterprise & Research
- ↓ Used by Physicians for care delivery

What Makes Biomedical Data Different?

Complexity of Healthcare Data:

Structured (EHRs, lab results), unstructured (clinical notes), image data (X-ray, MRI), multimodal data, etc.

Beyond Accuracy: Critical Requirements

Accuracy alone is not sufficient

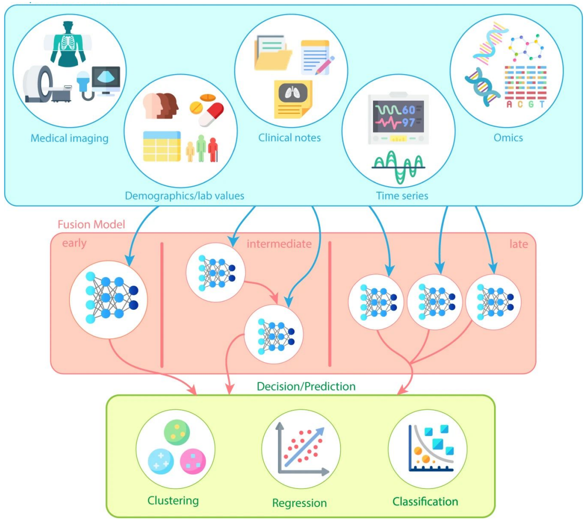
Explainability
Model decisions must be interpretable by clinicians

Non-discriminatory predictions
Fair across demographics and populations

Privacy-preserving
Protect sensitive patient information (HIPAA compliance)

Causal
Understand causal relationships, not just correlations

Multimodal Fusion Pipeline



What Makes Biomedical Data Different?

Key Challenges in Biomedical Data

Little labeled data - narrow generalization
Annotation requires expert knowledge and is time-consuming

Lots of missing data, varying time intervals, censored labels
Incomplete records, irregular measurements, lost to follow-up

Difficult to correct biases and inequities
Historical biases embedded in training data

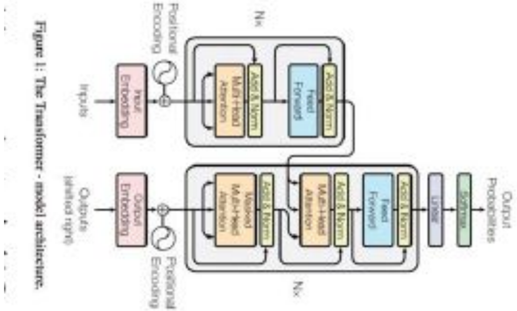
Motivates semi-supervised and self-supervised learning
Leverage large amounts of unlabeled data to improve model performance

Motivates generative AI for rare diseases
Synthesize data for conditions with limited samples

Inspiration from Computer Vision

Russakovsky et al. '14

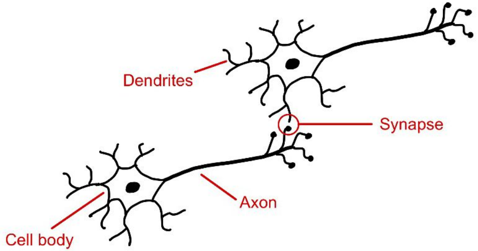
Transfer Learning Architectures



Key Insight: Pre-training on large general datasets, then fine-tuning on specific biomedical tasks

A Brief History of Machine Learning

Biological Inspiration



Human Brain:

- ~90 billion neurons
- Each neuron: several thousand synapses
- Total: ~100 trillion (10^{14}) synapses
- Synaptic strengths change → learning and memory

From Biology to Math

Artificial Neural Networks:

Mathematical models capturing neural properties through simple operations (McCulloch & Pitts, 1943)

Neuron Components & Function

Component	Function
Dendrites	Receive input signals
Cell body	Processes signals
Axon	Transmits electrical impulses
Synapses	Junctions connecting neurons Release neurotransmitters

Neural Firing Mechanism

Excitatory synapses: Stimulate subsequent neuron firing

Inhibitory synapses: Reduce likelihood of firing

Threshold behavior: Neuron fires only with sufficient stimulation

Key Insight: Changes in synaptic strengths represent the mechanism for learning from experience

Neural Network Evolution: Three Eras

Mathematical Formulation

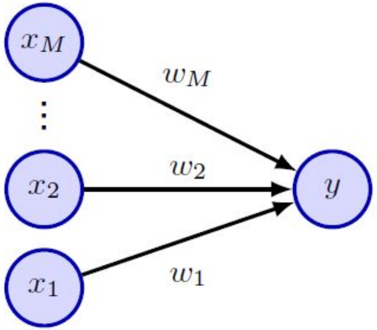
Basic Neuron Model:

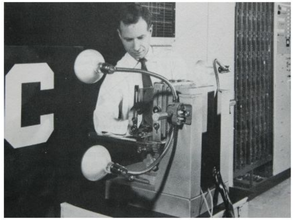
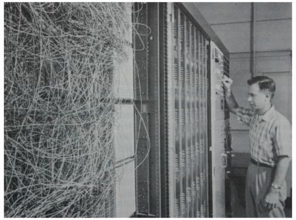
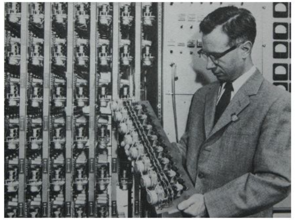
$a = \Sigma\ w_i x_i$ (pre-activation)

$y = f(a)$ (activation)

- **$x_1,\ldots,x_m$:** inputs from other neurons
- **$w_1,\ldots,w_m$:** weights (synaptic strengths)
- **f(·):** activation function (nonlinear)

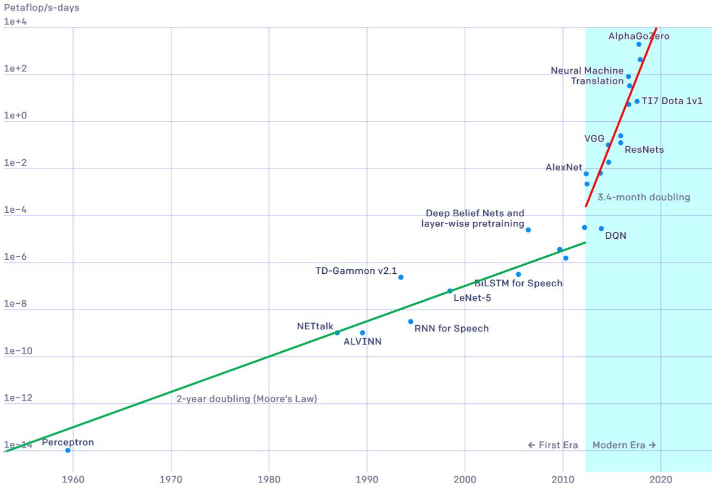
Era	Key Characteristics	Limitations
1. Single-Layer (1960s)	• Perceptron (Rosenblatt, 1962) • Step function activation • Guaranteed convergence • Analog hardware implementation	• Limited capabilities • Minsky & Papert (1969) formal proofs • Dampened enthusiasm (1970s-80s)
2. Backpropagation (mid-1980s)	• Multi-layer networks trainable • Continuous differentiable activations • Gradient-based optimization • Stochastic gradient descent	• Only final 2 layers learned useful features • Hand-crafted pre-processing needed • Few applications beyond CNNs • Reaching limits by 2000
3. Deep Networks (2010s-present)	• Many layers trainable effectively • Massive scale (trillions of parameters) • GPU acceleration • Representation learning • Foundation models	• Computational cost • Data requirements • Hyperparameter tuning complexity



The Deep Learning Revolution

Computational Scaling



Period	Scaling Pattern
1960-2012	2-year doubling (Moore's Law)
2012-present	3.4-month doubling (10× per year!)

Parameter Scale Evolution:
1980s: Hundreds to thousands
→ Millions → Billions → Trillions (10^{12})

Key Enabling Technologies

GPUs (Graphics Processing Units)

Massive parallelism enables layer-wise computation. Training now uses thousands of GPUs linked by high-speed interconnections

Residual Connections (He et al., 2015a)

Address vanishing gradients, enable training of networks with hundreds of layers

Automatic Differentiation

Backpropagation code generated automatically from forward propagation specification. Enables rapid architecture experimentation

Open Source Ecosystem

Researchers build on others' work, accelerating progress

Machine Learning Fundamentals: Curve Fitting Example

Synthetic Data Setup

Component	Description
Input variable	*x* (continuous, real axis)
Target variable	*t* (continuous, real axis)
Training set	*N* observations: $\{x_1,...,x_n\} \rightarrow \{t_1,...,t_n\}$
Goal	Predict *t* for new value of *x*

Key Concept: Generalization
Ability to make accurate predictions on previously unseen inputs

Data Generation Process

Example: N = 10 data points

- Input: x uniformly spaced in [0,1]
- Target: t = sin(2πx) + Gaussian noise
- Captures real-world property: underlying regularity corrupted by random noise

Solution: Minimize E(**w**) by finding optimal coefficients $\mathbf{w}^*$. Since E is quadratic in **w**, has unique closed-form solution.

Linear Model: Polynomial Fitting

Polynomial Function:

$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_mx^M$$

- *M* = order of polynomial
- **w** = coefficient vector $\{w_0,...,w_m\}$
- Linear in coefficients **w** (despite nonlinear in x)

Error Function

Sum of Squares Error:

$$E(\mathbf{w}) = ½ \Sigma \{y(x_n, \mathbf{w}) - t_n\}^2$$

Measures misfit between predictions and training data