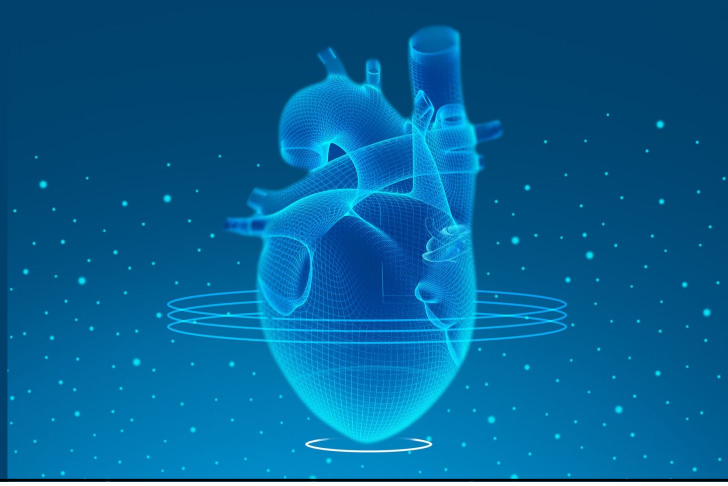
Minneapolis Heart Institute Foundation

GRAND ROUNDS



1

Overhyped or Undervalued? Exploring AI in Medicine

Alex Schmidt, DO and Roopa Patel, DO
Vascular Surgery Fellows

Minneapolis Heart Institute Foundation | GRAND ROUNDS

2

Disclosures

- We have no disclosures or conflicts of interest with the presented material. -

3

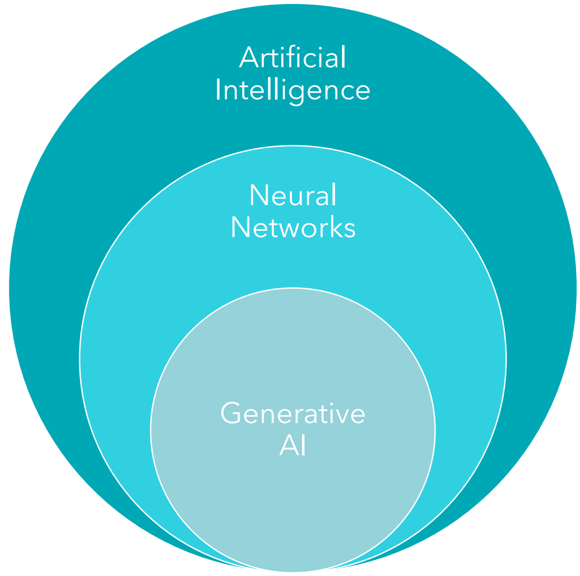
Learning Objectives

- Introduce fundamental concepts needed to understand the use of generative AI in healthcare.
- Evaluate how generative AI tools for clinical documentation and patient communication are improving workflow efficiency.
- Analyze potential AI applications in vascular surgery

4

Key Terms

- **Artificial Intelligence:** Systems that perform tasks thought to require human-like intelligence.
- **Neural Networks:** Multi-layered systems mimicking the human brain structure that can extract features and identify complex patterns from training data.
- **Generative AI (e.g. Large Language Models):** AI that creates new content rather than just analyzing existing data. A specific architecture of deep neural networks trained on massive datasets to produce novel outputs.



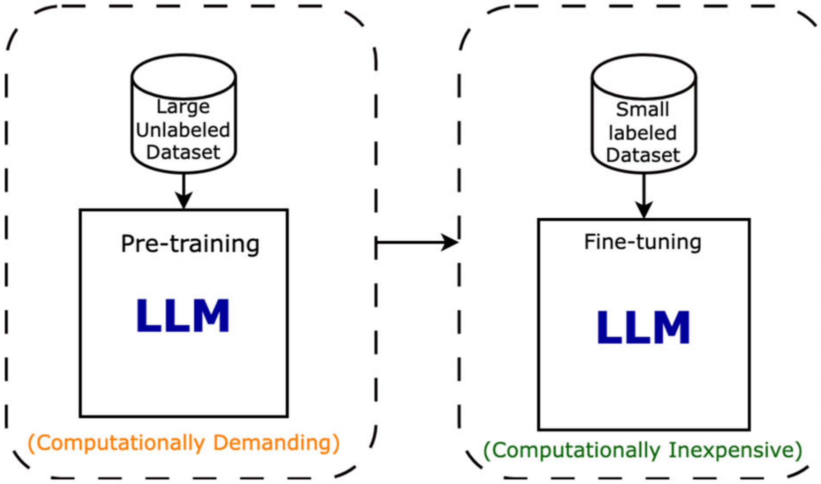
Minneapolis Heart Institute Foundation | GRAND ROUNDS

5

Why has GenAI exploded in the last few years?

6

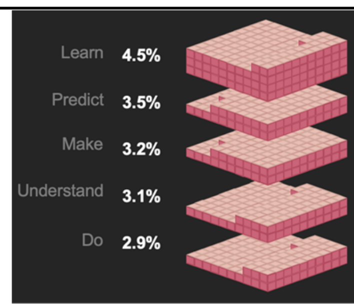
GenAI (LLM) Training



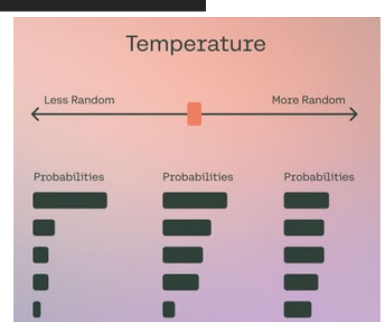
Minneapolis Heart Institute Foundation
GRAND ROUNDS

7

How LLMs Work



The best thing about AI is its ability to,
The best thing about AI is its ability to learn,
The best thing about AI is its ability to learn from,
The best thing about AI is its ability to learn from experience,
The best thing about AI is its ability to learn from experience.,
The best thing about AI is its ability to learn from experience.



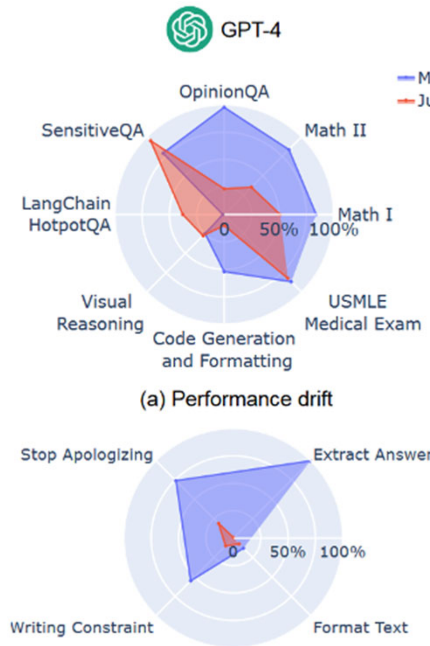
The best thing about AI is its ability to learn. I've always liked the
The best thing about AI is its ability to really come into your world and just
The best thing about AI is its ability to examine human behavior and the way it
The best thing about AI is its ability to do a great job of teaching us

Wolfram, 2023

Minneapolis Heart Institute Foundation
GRAND ROUNDS

8

Model Drift Over Time



(a) Performance drift

(b) Instruction following

Computer Science > Computation and Language

[Submitted on 18 Jul 2023 (v1), last revised 31 Oct 2023 (this version, v3)]

How is ChatGPT's behavior changing over time?

Lingjiao Chen, Matei Zaharia, James Zou

9

Current Clinical Applications Utilizing Generative AI

Gen AI Saves Nurses Time by Drafting Responses to Patient Messages

ABSTRACT

Mayo Clinic uses generative AI to draft responses to patient messages. Initial pilots showed that it saves nurses around 30 seconds per message and drafts more empathetic responses. Mayo Clinic plans to expand access to all LPNs and RNs by mid-2024, which could save 1,500 hours per month.

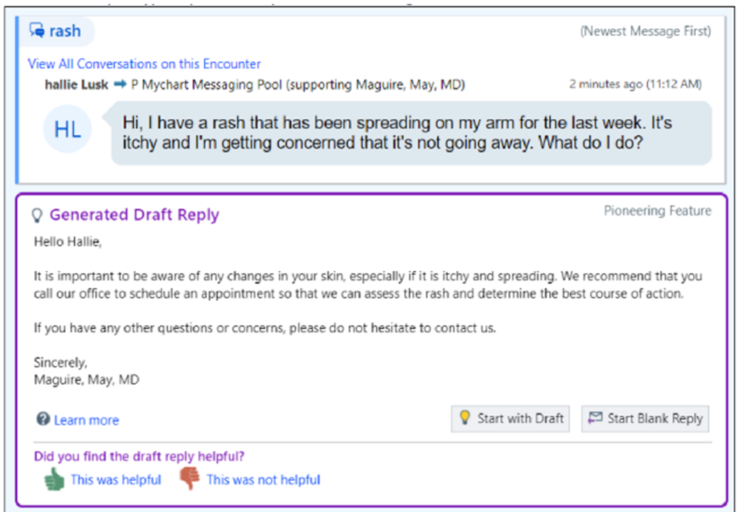
Nuance®
DAX™ Copilot

OpenEvidence®

Minneapolis Heart Institute Foundation® | GRAND ROUNDS

10

Drafting Inbox Message Responses



Minneapolis Heart Institute Foundation | GRAND ROUNDS

11

Original Investigation | Health Informatics

Artificial Intelligence–Generated Draft Replies to Patient Inbox Messages

Patricia Garcia, MD; Stephen P. Ma, MD, PhD; Shreya Shah, MD; Margaret Smith, MBA; Yejin Jeong, BA; Anna Devon-Sand, MPH; Ming Tai-Seale, PhD, MPH; Kevin Takazawa, BBA; Danyelle Clutter, MBA; Kyle Vogt, BA; Carlene Lugtu, MCiM; Matthew Rojo, MS; Steven Lin, MD; Tait Shanafelt, MD; Michael A. Pfeffer, MD; Christopher Sharp, MD

Population: 162 clinicians (141 physicians/APPs, 14 nurses, 8 pharmacists) from Primary Care and Gastroenterology/Hepatology
Intervention: Implementation of GPT-4 LLM to draft replies to patient portal messages
Comparison: Pre-implementation (35 days) vs. Post-implementation (35 days)
Primary Outcome: AI-generated draft reply utilization rate
Secondary Outcomes: Changes in time measures (reply, read, write times), clinician task load and work exhaustion scores, perceptions of utility/quality/time-saving, net promoter score

Minneapolis Heart Institute Foundation | GRAND ROUNDS

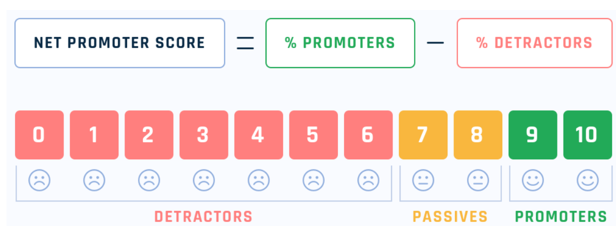
12

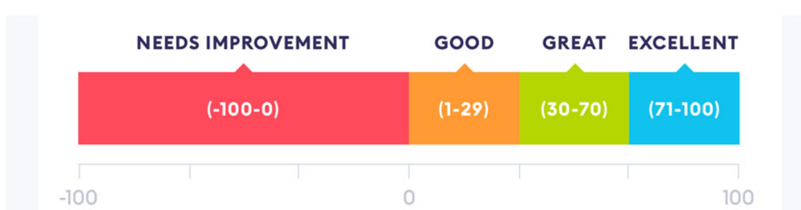
Pre-Measurement Survey	Post-Measurement Survey
Physician Task Load Derivative Index Please reflect on a day that you responded to patient messages in the last 1-2 weeks that is representative of a typical current day responding to patient messages, and move the sliders to indicate your response (Very Low, Low, Medium, High, Very High; 0-100 continuous): • How mentally demanding was the task of responding to patient messages? • How hurried or rushed was the pace of responding to patient messages? • How hard did you have to work to achieve your level of performance responding to patient messages? • How physically demanding was the task of responding to patient messages?	**Physician Task Load Derivative Index** Please reflect on a day that you responded to patient messages in the last 1-2 weeks that is representative of a typical current day responding to patient messages using the AI-generated draft replies tool, and move the sliders to indicate your response (Very Low, Low, Medium, High, Very High; 0-100 continuous): • How mentally demanding was the task of responding to patient messages? • How hurried or rushed was the pace of responding to patient messages? • How hard did you have to work to achieve your level of performance responding to patient messages? • How physically demanding was the task of responding to patient messages?
Burnout and Emotional Exhaustion Score To what degree have you experienced the following? (0-4; Not at all, Very Little, Moderately, A Lot, Extremely) During the past two weeks, I have felt... • A sense of dread when I think about work I have to do. • Physically exhausted at work. • Lacking in enthusiasm at work. • Emotionally exhausted at work.	**Burnout and Emotional Exhaustion Score** To what degree have you experienced the following? (0-4; Not at all, Very Little, Moderately, A Lot, Extremely) During the past two weeks, I have felt... • A sense of dread when I think about work I have to do. • Physically exhausted at work. • Lacking in enthusiasm at work. • Emotionally exhausted at work.

13

Net Promoter Score

- Likelihood to recommend





Minneapolis Heart Institute Foundation | GRAND ROUNDS

14

Response Time Metrics

- Reading Time: No statistically significant change
  - Pre-pilot: 113.0s vs. Pilot: 118.3s
- Writing Time: No statistically significant change
  - Pre-pilot: 162.9s vs. Pilot: 162.0s

Minneapolis Heart Institute Foundation | GRAND ROUNDS

15

Table 2. Draft Utilization per Clinician Stratified by Specialty and Role

Specialty and role	Mean (SD) Reply action count	Reply action count with draft available	Draft used count	Draft utilization rate
Overall	79.3 (95.5)	59.4 (72.6)	8.6 (16.9)	0.203 (0.268)

Table 4. Presurvey and Postsurvey Results

Variable	Score, mean (SD) Presurvey	Postsurvey	*P* value	No. of responses
Physician task load score derivative[a]				
Overall	61.31 (17.23)	47.26 (17.11)	<.001	73
Burnout and work exhaustion score[b]				
Overall	1.95 (0.79)	1.62 (0.68)	<.001	71[c]

16

Table 4. Presurvey and Postsurvey Results (continued)

Variable	Score, mean (SD) Presurvey	Postsurvey	P value	No. of responses
Net promoter score[e]				
Overall	NA	10	NA	73
Primary care	NA	17	NA	42
Physician and APP	NA	13	NA	30
Nurse	NA	−60	NA	5
Clinical pharmacist	NA	71	NA	7
Gastroenterology and hepatology	NA	0	NA	31
Physician and APP	NA	−19	NA	21
Nurse	NA	50	NA	10

Net promoter score is calculated by categorizing likelihood to recommend responses into promoters (score 9-10), passives (score 7-8), and detractors (score 0-6) and then subtracting the percentage of detractors from the percentage of promoters. The net promoter score ranges from −100 to 100, with higher scores indicating higher levels of satisfaction.[20]

17

Application in Surgery?

- Direct evidence for LLM summarization in surgery is somewhat limited.
- In the prior study, GI (procedural specialty) showed worse physician/APP perceived value (NPS -20).
- While no direct harm was identified, the low utilization in similar fields suggests cautious exploration is warranted before broader surgical implementation.
- **Overall: Reasonable to try but close provider quality monitoring needed**

18

Ambient Scribes

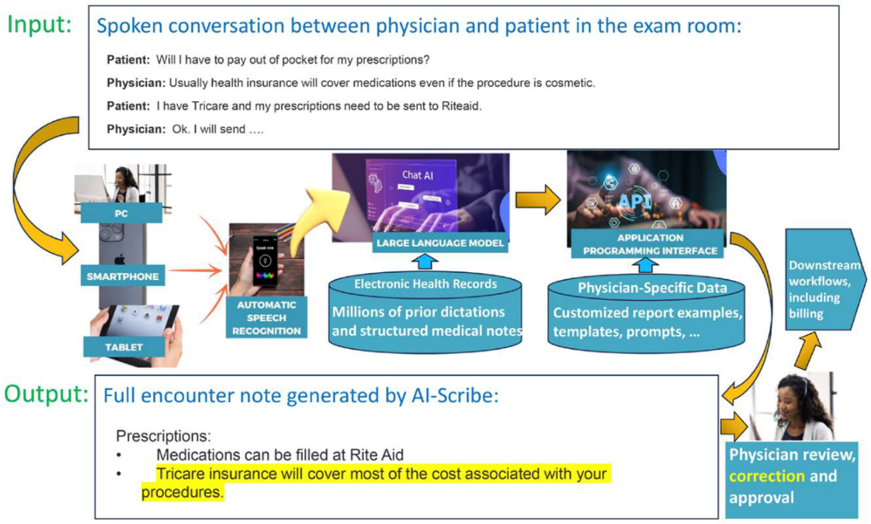


Fig. 1. Diagram of how AI scribe works.

Mess, Sarah A., et al. "Artificial Intelligence scribe and large language model technology in healthcare documentation: Advantages, limitations, and recommendations." *Plastic and Reconstructive Surgery - Global Open*, vol. 13, no. 1, Jan. 2025, https://doi.org/10.1097/gox.0000000000006450.

19

The association between use of ambient voice technology documentation during primary care patient encounters, documentation burden, and provider burnout

Lance M. Owens[1], Joshua J. Wilda[2], Peter Y. Hahn[3], Tracy Koehler[3], Jeffrey J Fletcher[3,*]

Population: 110 primary care providers
Intervention: UMHW first in country to implement DAX at a primary care clinic
Comparison: High (>60% of encounters) vs low DAX users (<60%); pre-implementation period documentation metrics
Primary Outcomes: Provider burnout (Oldenburg Burnout Inventory, OLBI) and documentation time per patient encounter
Secondary Outcomes: Time spent documenting outside scheduled hours; provider contribution to notes; documentation length

Minneapolis Heart Institute Foundation | GRAND ROUNDS

20

Characteristic	Value
Years in practice (mean ± SD)	13.8
Full time equivalent (%)	71%
IT DAX™ use (% of visits)	
- 0-20%	49%
- 21%-40%	11%
- 41%-60%	12%
- 61%-80%	15%
- >80% of encounters	13%
Provider Type	
- MD, DO (IM, Family)	50%
- APP	50%

Oldenburg Burnout Inventory scores (mean ± SD)	
- OLBI-D Disengagement score [range 8-32]	17.9 ± 3.6
- OLBI- E Exhaustion score [range 8-32]	19.7 ± 4.1
- Full burnout inventory score [range 16-64]	37.6 ± 7

		strongly agree	agree	disagree	strongly disagree
1.	I always find new and interesting aspects in my work (D)	1	2	3	4
2.	There are days when I feel tired before I arrive at work (E.R.)	1	2	3	4
3.	It happens more and more often that I talk about my work in a negative way (D.R)	1	2	3	4
4.	After work, I tend to need more time than in the past in order to relax and feel better (E.R)	1	2	3	4
5.	I can tolerate the pressure of my work very well (E)	1	2	3	4
6.	Lately, I tend to think less at work and do my job almost mechanically (D.R)	1	2	3	4

21

Measure	High DAX™ use	Low DAX™ use	Difference (95% CI)
OLBI-D	16.3 (4.2)	18.4 (3.2)	-2.1 (-3.8 to -0.5)
OLBI-E	19.0 (5.5)	20.0 (3.4)	-1.0 (-2.9 to 0.9)
OLBI-T	35.4 (9.2)	38.4 (5.8)	-3.1 (-6.3 to 0.8)

Measure	Pre-DAX	Post-DAX	Difference
Average patient visits per provider	293.6	287.6	6.0 (-0.2 to 12.2)
Scheduled workdays per month	22.2	22.3	-0.1 (-0.4 to 0.23)
Average documentation time per note	5.9 min	4.1 min	-1.8 min (-2.2 to -1.4), 29%
Time documenting outside scheduled hours	-	-	-4.0 min (-7.2 to -1.0), 12%
Percent contribution to note by provider	-	-	-33% (-41.8% to -24.2%)
Documentation length (characters)	-	-	+542 (283.9 to 800.1)

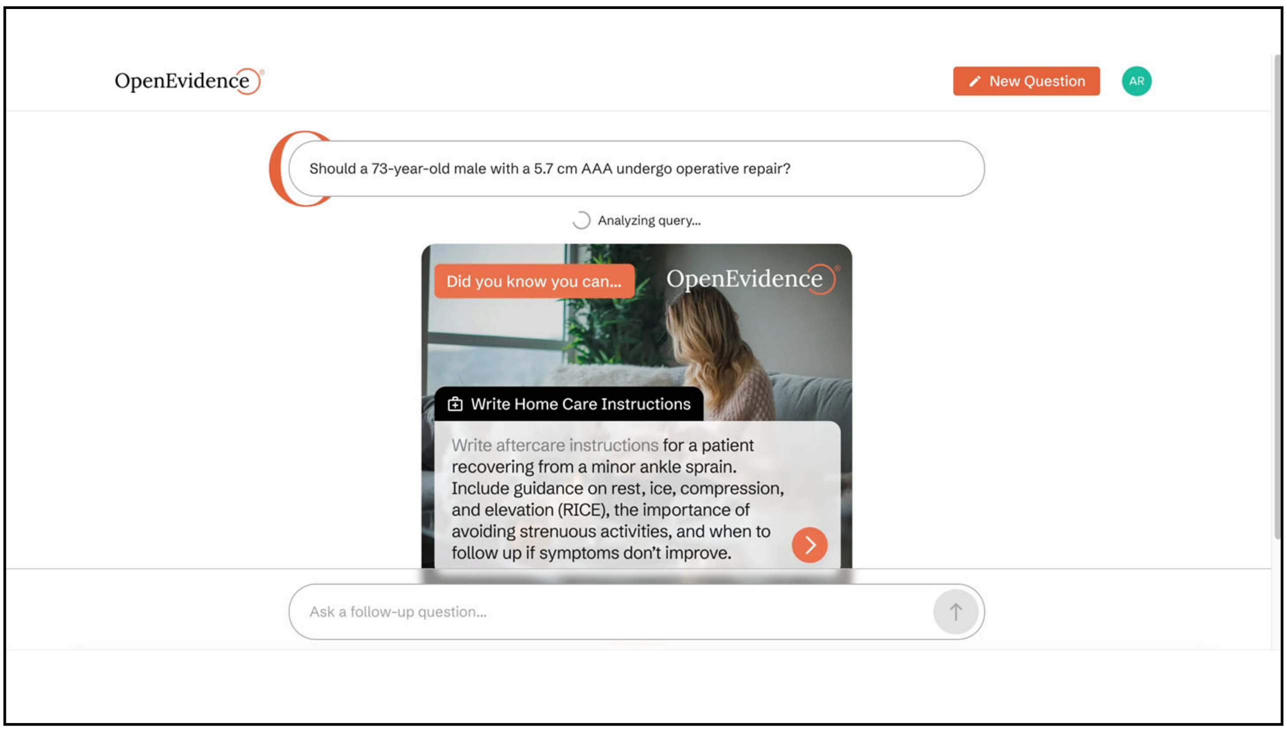
22

Application in Surgery?

- Evidence is currently unclear on the direct impact of ambient AI on physician burnout
- Study suggested providers may spend less time to produce longer (?more detailed / helpful) notes
- **Overall: Given the potential to reduce documentation time and create richer notes, implementation appears reasonable, though further research on burnout is needed.**

Minneapolis Heart Institute Foundation | GRAND ROUNDS

23



24

Applications in Vascular Surgery

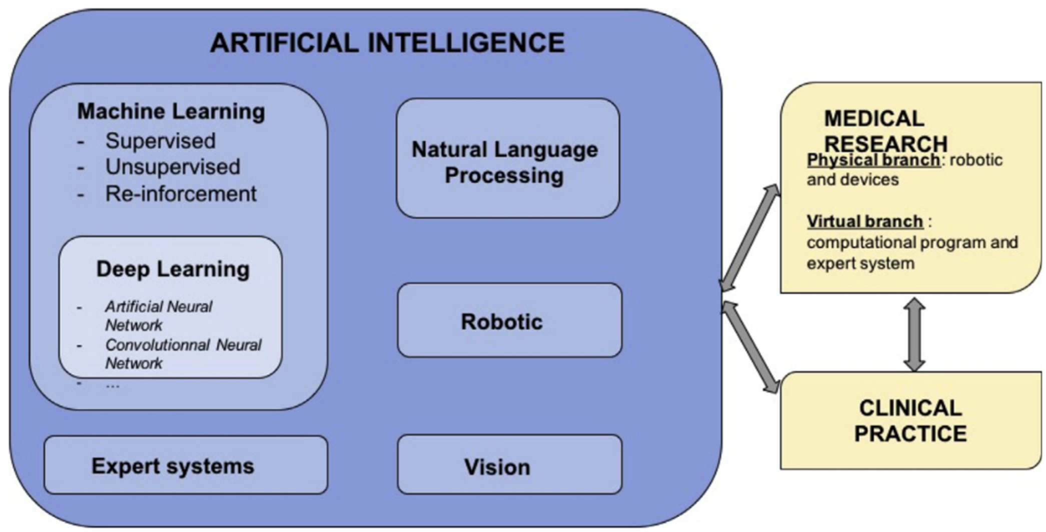
25

Why Vascular Surgery is Ideal for AI Integration

- **Image-Rich Specialty:** The endovascular revolution has transformed vascular surgery into a technology-driven field with abundant imaging data perfect for ML-based analysis
- **Objective Disease Definitions:** Conditions like AAA and PAD have precise diagnostic criteria (≥3 cm, ABI <0.9), enabling automated diagnosis with minimal human input
- **High-Stakes Outcomes:** Procedures on complex patients with multiple comorbidities benefit from AI's ability to analyze past cases and predict surgical outcomes

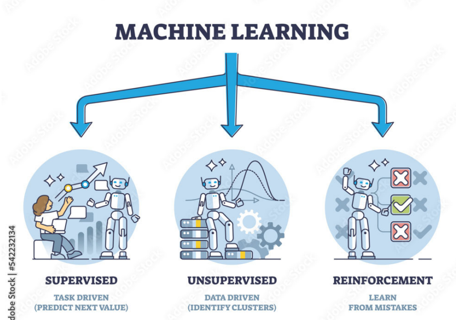
Fischer et al, 2023

26



27

Machine Learning



- Supervised
  - Uses set of labeled data, independent and dependent variables are known
  - Algorithm learns how to map input to the corresponding outcomes- knows outcomes tried to link with known characteristics
- Unsupervised
  - Explores patterns in independent variables without prior knowledge of outcomes
  - Can uncover unknown relationships
- Reinforcement
  - Algorithm makes decisions and performs task which are couple with rewards or penalties
  - Learns patterns with trial and error to train the algorithm

Minneapolis Heart Institute Foundation | GRAND ROUNDS

28

Deep Learning

Artificial Neural Network: feed forward

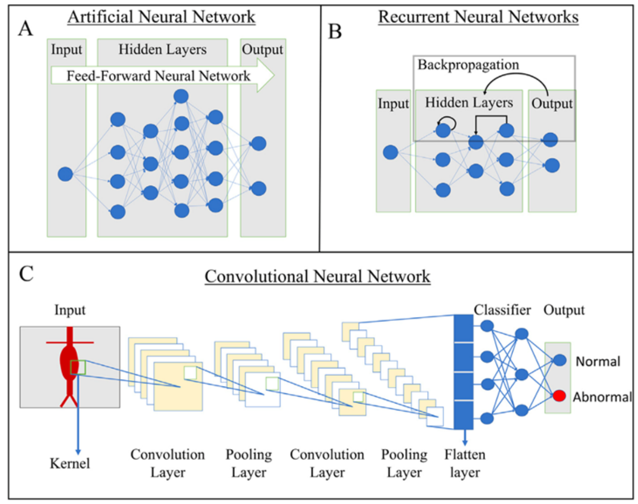
- input→ output

Recurrent Neural Network: feed back connections

- process sequential data, integration of data from previous steps

Convolutional Neural Network: process with grid-like structure developed feature map

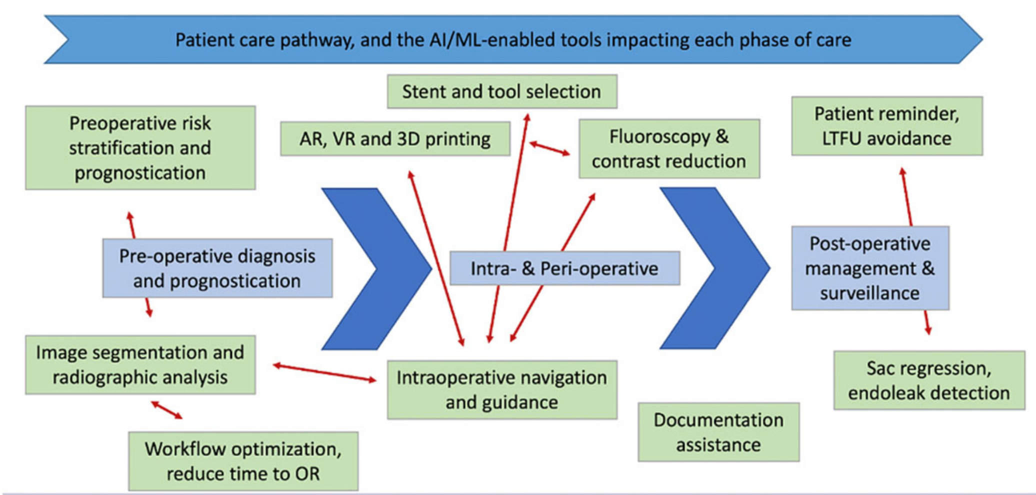
- kernel to create image segmentation



Minneapolis Heart Institute Foundation | GRAND ROUNDS

29

AI Enabled Applications in Aortic Surgery



Minneapolis Heart Institute Foundation | GRAND ROUNDS

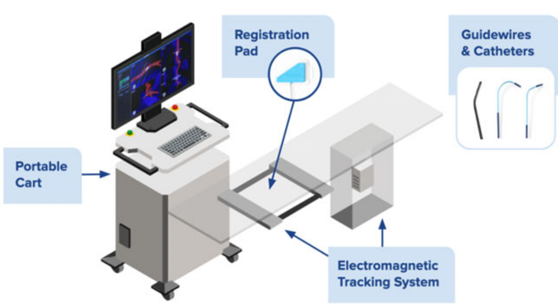
30

AI Enabled Applications in Aortic Surgery

- Early applications: radiologic image evolution, diagnosis, risk stratification
  - Diagnosis: image segmentation: Viz.ai's Vascular Suite – using imaging as well as data to search for vascular disease
    - VIZ.ai algorithm detected aortic dissection with a sensitivity 94.2% and specificity of 97.3%, still needs a human adjunctor and cannot work independent of human evaluation
      - Flags imaging-with high risk pathology and moves imaging to the top of radiology work-list
  - Risk Predication: evaluated aneurysm sac expansion rates, surveil the aneurysm sac after endovascular repair
    - PRAEVAorta Nurea- fully automated analysis of AAA sac measurement before and after endovascular repair – plan to use this data to predict how AAA will evolve based on geometry and blood flow characteristics
  - Surgical Planning: pre-operative and intraoperative- goal of improving operating efficiency and decrease radiation exposure
    - Intraoperative positioning system- anatomic mapping and modeling algorithm
    - Philips' Fiber optic real shape- overlay system for tracking endovascular tools

31

AI Enabled Applications in Aortic Surgery



Intraoperative positioning system

Philips Fiber Optic RealShape

32

AI Enabled Applications in Aortic Surgery

Future: eventual development may lead to entire vascular surgical care pathways where the diagnosis, surgical planning, and surgical therapy are directly assisted or facilitated by AI/ML suite enabled tools.

Minneapolis Heart Institute Foundation | GRAND ROUNDS

33

Diagnostic Performance of a Deep Learning-Powered Application for Aortic Dissection Triage Prioritization and Classification

Vladimir Laletin [1,*], Angela Ayobi [1], Peter D. Chang [2,3], Daniel S. Chow [2,3], Jennifer E. Soun [2,3], Jacqueline C. Junn [4], Marlene Scudeler [1], Sarah Quenet [1], Maxime Tassy [1], Christophe Avare [1], Mar Roca-Sogorb [1] and Yasmina Chaibi [1]

- Retrospective study from January 2019 to December 2022 that evaluated the diagnostic performance of a deep learning based application for detecting, classifying and highlighting suspected aortic dissections from CTA
- 200 US and European cities, reading 1300 CTA
- DL tools for identifying AD- expedite the process for radiologists, reducing time for clinical decision making
- Goal was to evaluate potential for faster triage in the clinical setting

Minneapolis Heart Institute Foundation | GRAND ROUNDS

34

AI Enabled Applications in Aortic Surgery

- DL CNN algorithm for image segmentation of the aorta
- Localization of the AD focusing on the visible intimal flap
- 2 radiologist read all the studies first, then the imaging was run through the CINA-CHEST applications and processed
- Categorized into Type A or B – processing times were recorded

Table 1. CINA-CHEST (AD) inclusion and exclusion criteria.

The Inclusion Criteria for CINA-CHEST (AD)
Chest or thoraco-abdominal CTA scans
Age $\geq$ 18 y/o
Matrix size $\geq 512 \times 512$ (rectangular matrix accepted)
Axial acquisition only
Slice thickness $\leq$ 3 mm with no gap between successive slices
Radiation dose parameters: 60 kVp to 160 kVp
Reconstruction diameter above 200 mm
Density threshold in the aorta $\geq$ 140 HU
Soft tissue reconstruction kernel
Field of view including the aortic arch and thoracic aorta
The Exclusion Criteria for CINA-CHEST (AD)
Parameters not compatible with acquisition protocol
Thoracic aorta out of the field of view
Significant motion artefacts (uninterpretable images)
Significant streak artefacts (uninterpretable images)
Significant noise (uninterpretable images)
Bad bolus timing (uninterpretable images)

Minneapolis Heart Institute Foundation | GRAND ROUNDS

35

AI Enabled Applications in Aortic Surgery

- Diagnostic accuracy of the device and its capacity for detection of types of aortic dissection assessed
- DL based application identified 129 or the 137 patients with AD - 94.2% sensitive, 1134 of the 1166 patients without AD, found to be 97.3% specific
- Time to identification 27.9 seconds
- Type A dissections were accurately identified in 100% of patients
- Type B dissections were accurately identified in 89.2% of patients

Table 4. Reasons for AD misdiagnoses by CINA-CHEST (AD).

Main Reasons for False Negatives (n = 8)	Main Reasons for False Positives (n = 32)
Intramural hematoma (IMH) (4)	Inadequate contrast opacification (13)
Penetrating atherosclerotic ulcer (PAU) (2)	Motion artefacts (10)
Acquisition artefacts (2)	Instances of pathology mimicking dissection (7)
	Interference from stent grafts (2)

Minneapolis Heart Institute Foundation | GRAND ROUNDS

36

Carotid Image Segmentation

- Early applications
  - Edge based segmentation: not effective in images with blurred boundaries
  - Region based segmentation: can cause over segmentations
  - Threshold based segmentation: can't deal with gray-level images with unimodal distribution
- Preprocessing: higher needs for US due to lower quality and contrast between tissues when compared to CT or MRI
  - DL methods: normalization, denoising and image enhancement

Carotid Image Segmentation

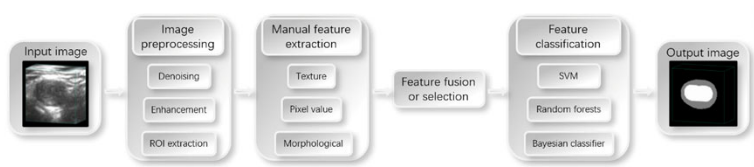


FIGURE 6. Traditional machine learning segmentation process.

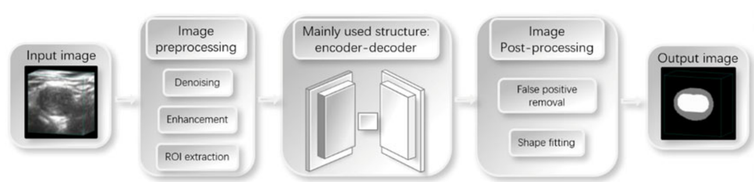


FIGURE 7. Deep learning segmentation process.

Carotid Image Segmentation

- Carotid Intima-media thickness (CIMT): difficult to assess, segmentation allows for CIMT measurements
  - CIMT >1.18mm has showed an increased incidence of stroke
  - AI DL neural networks applied to imaging converting 3D→2D imaging for processing
  - Ultimately using pixels to obtain the boundaries to calculate CIMT
- Plaque Segmentation: evaluate the plaque surface appearance, plaque density, total plaque area and volume, goals to eventually predict risk of plaque rupture

39

30-Day Risk Score for Mortality and Stroke in Patients with Carotid Artery Stenosis Using Artificial Intelligence Based Carotid Plaque Morphology

Rohini J. Patel, Daniel Willie-Permor, Austin Fan, Sina Zarrintan, and Mahmoud B. Malas, San Diego, California

- Retrospective study from 2010 to 2021 evaluating head and neck CTA in patients with carotid artery stenosis.
- Using AI plaque morphology software taking 2D images and to determined morphology in a 3D model
  - Software produces its own set of variable to describe the plaque
    - Calcium volume, intraplaque hemorrhage, lipid rich necrotic core, matrix
- 372 patients evaluated and stratified into symptomatic (36%) or asymptomatic (64%)
- No significant difference between the type of plaque variables in the two groups.

40

30-Day Risk Score for Mortality and Stroke in Patients with Carotid Artery Stenosis Using Artificial Intelligence Based Carotid Plaque Morphology

Rohini J. Patel, Daniel Willie-Permor, Austin Fan, Sina Zarrintan, and Mahmoud B. Malas, San Diego, California

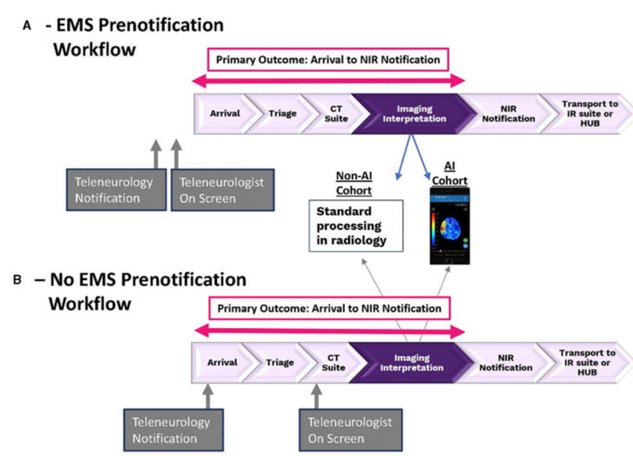
- Model A: clinical variables only
- Model B: software plaque morphology variables only
- Model C: clinical and software variables- best predictor
- Significant risk factors
  - Clinical: Age, sex, history of stroke or TIA, BMI, Hyperlipidemia, COPD
  - Plaque morphology: matrix volume, perivascular adipose tissue, lipid rich necrotic core volume
  - Stronger RF: age >80, COPD, previous stroke, matrix volume
- Risk calculator demonstrated the ability to individualize 30d risk prior to intervention

Minneapolis Heart Institute Foundation | GRAND ROUNDS

41

VALIDATE study

- 76 AI facilities, 90 non-AI facilities 2021-2022
- Door to NIR time was 39.5 minutes in AI groups and 89.5 minutes in non-AI groups
- Door to NIR in non-TC 33 minutes (AI) versus 97 minutes (non-AI)
- Door to NIR in TC 34 minutes (AI) versus 78 minutes (non-AI)



Minneapolis Heart Institute Foundation | GRAND ROUNDS

42

AI in Surgical Education

- Personalized training: evaluate a surgical trainee on an individual basis to create personalize learning experiences based on deficiencies and improve specific skills
- Simulation based training: Virtual reality provides another form of surgical simulation to mimic real scenarios for trainee practice purposes
- Predictive Modeling: AI software to help with evaluation and identify potential surgical complications with case preparation
- Augmented reality: provide surgeons with real-time information during surgery with head mounted devices that map from pre-operative imaging

Minneapolis Heart Institute Foundation | GRAND ROUNDS

43

Barriers to use of AI in medicine

- Risks and barriers : limited availability of large and high-quality datasets – general lack of standardization
  - Datasets may not be representative of all racial, ethnic, sex, and other demographic differences
- Black Box Effect: end users do not understand how decisions are made by an algorithm within the hidden layer
  - Explainable AI is being developed and enforced to help with transparency
- Who is the responsible party for negative outcomes?
- Is patient data protected in large data sets?

Minneapolis Heart Institute Foundation | GRAND ROUNDS

44

AI: Overhyped or Undervalued?

- Evidence is still emerging but generative AI tools in the clinic have potential to improve efficiency and reduce clinician burnout
- Promising adjuncts to help improve patient care and time to treatment
- Helps to guide individualized care based on risk calculators
- A way to progress surgical education