Data Science, ML, and AI for Medical Students

Hour 1: Linking AutoAnalyzer with Colab Notebooks for Clinical Insight

David Liebovitz, MD

Associate Vice Chair for Clinical Informatics, Department of Medicine

Data Science | Machine Learning | Artificial Intelligence

Digital Health & Data Science Training Series

Course Topics

- Core Concepts in Digital Health and Data Science
- AI within CVD
- **Data Exploration and Visualization ALC ← Today**
- Machine Learning and AI Models ALC
- Video-based Gait Analysis using AI
- Telemedicine in Psychiatry
- Wearables and Reproductive Health
- LLMs in Healthcare
- Ethical Considerations for AI in Healthcare

Key Competencies

1. **Healthcare Delivery Science:** Evaluate digital health tools impact
2. **Health Data Ecosystem:** EHRs, PACS, device data, omics
3. **Health IT Regulatory:** Mandates, interoperability, compliance
4. **Data Science Methods:** ML/AI for diagnosis, precision medicine
5. **Clinical Decision Support:** CDS system principles
6. **Modeling Applications:** Apply predictive models to clinical problems
7. **Bias, Ethics, Health Equity:** Algorithm bias, ethical implications
8. **Sociotechnical Context:** Implementation science principles

⭐ Why Digital Health & AI Skills Are Essential

🏥 Current Healthcare Reality

📊 Data-Driven Medicine

EHRs generate 2.3 exabytes of healthcare data annually. Your patients' care increasingly relies on algorithms for diagnosis, treatment planning, and risk stratification.

🤖 AI Integration Accelerating

FDA has approved 500+ AI medical devices. By 2030, AI will be standard in radiology, pathology, drug discovery, and clinical decision support.

🔬 Precision Medicine

Genomics, proteomics, and personalized treatment protocols require computational skills to interpret and apply effectively.

🚀 Your Future Career

⚕️ Clinical Practice

You'll need to interpret AI recommendations, understand model limitations, identify bias, and communicate uncertainty to patients and colleagues.

⚠️ Critical Thinking

As shown in today's data quality example, you must critically evaluate the data and algorithms that inform your medical decisions.

🌐 Leadership & Innovation

Whether in research, administration, or clinical practice, you'll shape how AI is developed, validated, and implemented in healthcare.

💡 The Bottom Line

Digital health literacy isn't optional—it's as fundamental to modern medicine as anatomy or physiology. Today's session gives you practical tools to understand, evaluate, and use data-driven insights in your future practice.

🎯 Model Performance: AUROC vs Calibration

🧠 Neurological Example: Migraine Prediction

Two AI Models for Migraine Risk *(Hypothetical Example)*

Model A (Good AUROC, Poor Calibration):
- AUROC: 0.87 (excellent discrimination)
- Says "40% migraine risk" for patients with 15% actual risk
- **Problem:** Over-prescribing preventive medications

Model B (Good AUROC, Good Calibration):
- AUROC: 0.85 (still excellent discrimination)
- Says "40% migraine risk" for patients with 38% actual risk
- **Benefit:** Accurate shared decision-making

Patient Conversation Impact

Patient: "Doctor, the AI says I have 40% migraine risk. What does that mean?"

With good calibration: "Out of 100 patients like you, about 40 will develop migraines in the next year."

📊 Understanding Model Metrics

AUROC (Discrimination)

"Can the model rank patients from low to high risk?"

• 0.5 = random guessing • 0.8+ = good • 0.9+ = excellent

Calibration

"Are the predicted probabilities accurate?"

If model predicts 30% risk, do ~30% of patients actually have the outcome?

Population Variation

Performance can vary dramatically across age, sex, ethnicity, and comorbidities

Always test subgroups separately!

💡 Clinical Takeaway

Both metrics matter: **AUROC for identifying high-risk patients, calibration for making treatment decisions.** Test performance across all relevant patient populations.

📊 Statistical Reasoning: Beyond P-Values

🧠 Neurological Example: Coffee and Stroke Risk

Study Findings *(Hypothetical Example)*

Observational study results:
• Heavy coffee drinkers: 15% stroke rate
• Non-coffee drinkers: 10% stroke rate
• $p < 0.001$ (highly significant!)
• **Conclusion: Coffee causes strokes?**

The Hidden Confounders

Heavy coffee drinkers also had:
• Higher smoking rates (60% vs 10%)
• More sedentary jobs (desk work)
• Higher stress levels
• Poor sleep habits
The association disappeared after controlling for these factors!

📈 Statistical Reasoning Principles

- **Association ≠ Causation:** Strong statistical associations can be misleading without considering confounders
- 📊 **P-value Limitations:** Statistical significance ≠ clinical significance. A tiny effect can be "significant" with large sample sizes
- **Effect Sizes Matter:** A 50% reduction from 0.02% to 0.01% risk is statistically dramatic but clinically minimal
- 🎯 **Confidence Intervals:** Wide CIs suggest uncertainty, regardless of statistical significance

💡 Clinical Takeaway

Always ask: **"Is this association clinically meaningful?"** and **"What else could explain this relationship?"** before making treatment decisions.

🎯 For Your Practice

Focus on effect sizes, confidence intervals, and biological plausibility—not just p-values. Consider confounders and alternative explanations.

📈 The Evolving Story of Statins: A Data Science Case Study

📅 Timeline of Discovery

- **1987** — **First Statin Approved:** Lovastatin introduced for cholesterol reduction
- **1995** — **WOSCOPS Trial Published:** Demonstrated cardiovascular benefits
- **2001** — **Unexpected Finding:** WOSCOPS post-hoc analysis revealed diabetes association
- **2008+** — **Pattern Confirmed:** Multiple trials confirmed increased diabetes risk

🔗 Neurologic Connection

Dual Purpose: Statins prevent both coronary artery disease (CAD) and stroke through cholesterol reduction and plaque stabilization, making them essential in neurology practice.

⚠️ The Unexpected Association

WOSCOPS (2001): The First Signal

- Originally showed 30% reduction in diabetes cases
- 5,974 participants, 139 developed diabetes
- Used non-standardized criteria for diabetes diagnosis
- **Initially appeared protective**

Subsequent Studies: Pattern Reversal

- JUPITER trial: Higher diabetes incidence with rosuvastatin
- Meta-analyses show 9-12% increased risk
- Risk greater with high-intensity statins
- **Consistent pattern across multiple studies**

🔍 Key Clinical Insight

Benefits still outweigh risks: statins prevent ~5 cardiovascular events for every new diabetes case

⚡ Why This Matters for Your Future Practice

In our rapidly evolving healthcare environment, **unexpected associations like this are discovered continuously.** The tools and analytical approaches you're learning today—data exploration, statistical testing, and machine learning—enable much earlier detection of these patterns, potentially improving patient safety and outcomes.

🎯 Session Goals

🔬 Start with code-free analysis using AutoAnalyzer for immediate insights

📈 Explore a real dataset (stroke risk) through interactive visualizations

💻 Dive deeper with Colab notebooks (don't worry - AI will help!)

🚀 Build predictive models and prepare for future sessions

🎓 **Learning Approach: Start simple → Add complexity → AI-assisted coding**

Part 1: Code-Free Analysis with AutoAnalyzer

Getting Started

Access URLs:
• autoanalyze.azurewebsites.net
• autoanalyze-beta.azurewebsites.net

⚠️ **Work in small groups** - server capacity limited
🔒 **Public data only** - hosted on Northwestern's Azure for privacy

AutoAnalyzer Interface

```
Demo 5 (stroke)
✓ Data validation passed
390 rows, 15 columns, 0.09 MB
Data Exploration   Machine Learning
```

Future Option

Available on GitHub for local deployment
→ Consider informatics elective for local use with sensitive data

Our Analysis Plan

1. Clean & Filter
Prepare dataset for analysis

2. Explore & Visualize
Charts, plots, patterns

3. Predict Stroke
Build ML model

💻 Part 2: Deeper Dive with Colab Notebooks

🤖 Don't Fear the Code! AI is Your Assistant

When we move to code, remember: **AI tools like ChatGPT, Claude, and GitHub Copilot** can help you understand, debug, and write code. You're not expected to memorize syntax!

📓 Notebook 1: Data Exploration

- Load and inspect the stroke dataset
- Identify variable types and missingness
- Visualize distributions (histogram, bar chart)
- Form clinical questions from observed patterns

📓 Notebook 2: Statistical Analysis

- Compare groups using t-tests and chi-square
- Calculate effect sizes: risk difference and ratio
- Interpret p-values, confidence intervals
- Recognize limits of observational associations

📓 Notebook 3: Prediction Model

- Split data into train/test for generalization
- Build logistic regression with preprocessing
- Evaluate discrimination (AUROC) and calibration
- Discuss model limitations: leakage, bias, fairness

🔗 AutoAnalyzer → Colab Bridge

Use AutoAnalyzer insights to guide your code-level learning in Colab notebooks

Data Quality & Integrity: A Growing Concern

Recent Findings: Federal Data Manipulation

Freilich J, Kesselheim AS. The Lancet (2025)
Study of 232 U.S. public health datasets revealed concerning patterns

49% of Datasets Altered

114 of 232 datasets changed without public logging

93% Changed Gender → Sex

Most common alteration: terminology changes

Only 13% Logged

Vast majority of changes unrecorded

Clinical Implications

- **Research Validity:** Unlogged changes make studies irreproducible
- **Variable Meaning:** 'Sex' ≠ 'Gender' - different biological vs. social constructs
- **Clinical Decisions:** Policy changes based on manipulated data
- **Public Trust:** Transparency essential for scientific credibility

For Your Practice

Always verify data provenance
• Check data source documentation
• Look for version control and change logs
• Understand variable definitions
• Document your own data processing steps

Remember: Good data science starts with trustworthy, well-documented data

Source: Freilich J, Kesselheim AS. Data manipulation within the US Federal Government. *The Lancet*. 2025;406(10500):227-8. doi: 10.1016/S0140-6736(25)01249-8.

⚖️ Subgroup Analysis: Why Fairness Matters

🔍 Case Study: Obermeyer et al., Science (2019)

Analysis of a commercial algorithm used by health systems to identify patients for care management programs

🎯 The Algorithm's Goal

Identify patients who would benefit from extra care management

📊 What It Actually Did

Used healthcare costs as a proxy for health needs

⚠️ The Bias

Selected less sick white patients over sicker Black patients

🔍 Root Cause Analysis

1. **Flawed Proxy:** Healthcare costs ≠ health needs
2. **Access Disparities:** Black patients had less access to expensive care
3. **Historical Bias:** Algorithm learned from biased historical data
4. **No Subgroup Testing:** Algorithm wasn't tested across racial groups

✅ The Solution

Mandatory subgroup analysis revealed the bias.
When researchers tested performance by race, they found:

- Black patients were significantly sicker at same risk scores
- Algorithm needed recalibration for equitable care

🎯 **Key Takeaway: Always test AI models across different patient populations before deployment**

🎯 Calibration vs Discrimination: Understanding Model Performance

📊 Discrimination (AUROC)

"Can the model separate high-risk from low-risk patients?"

- AUROC = 0.5: Random guessing
- AUROC = 0.7: Acceptable
- AUROC = 0.8: Good
- AUROC = 0.9: Excellent

⚖️ Calibration

"Are the predicted probabilities accurate?"

If model says "30% stroke risk," do 30 out of 100 similar patients actually have strokes?

🏥 Clinical Example: Stroke Risk

Model A: Good discrimination, poor calibration

- AUROC: 0.85 (Excellent at ranking patients)
- But predicts 20% risk for patients with 5% actual risk
- **Problem:** Over-treatment, unnecessary procedures

Model B: Good discrimination, good calibration

- AUROC: 0.82 (Still excellent at ranking)
- Predicts 20% risk for patients with 19% actual risk
- **Benefit:** Accurate shared decision-making

💡 Clinical Impact

Patient conversation:
"Doctor, you said I have a 20% stroke risk. What does that mean for me?"

With good calibration: "Out of 100 patients like you, about 20 will have a stroke in the next year."

🎯 **Both matter: Discrimination for identifying high-risk patients, calibration for making treatment decisions**

Essential Facts — What to Remember

Garbage in → garbage out: prioritize data quality

Association ≠ causation; confounding is common

Calibration is as important as discrimination

Always check subgroup performance for fairness

Document and share methods for reproducibility

AI tools are your coding companions, not replacements

🔧 AutoAnalyzer Integration

📤 **Upload Dataset**
Load your data
→
🔍 **Explore & Analyze**
Run basic statistics
→
🤖 **Build Models**
Create predictions
→
💡 **Clinical Insight**
Interpret results

🚀 Next Steps — Future Sessions

📝 **NLP for EHR Text**
Process clinical narratives

🧠 **Large Language Models**
Hands-on with LLMs

⚕️ **Clinical Decision Support**
AI integration in practice

🛡️ **Ethics & Privacy**