Expanding the Horizons of Delirium Prediction by Leveraging Artificial Intelligence
Presenters: Heidi Lindroth, PhD, RN, FAAN and Anirban Bhattacharyya, MD, MS, MPH

Time	Section
02:12	**Introduction of Heidi Lindroth and Anirban Bhattacharyya**
03:37	**Expanding the Horizons of Delirium Prediction (Heidi Lindroth)**
04:11	**Altered States on ICU** • Martyn Stones and Janice Sharp's book on hallucinations from people in the ICU • Spoke of a specific example from her experience with someone who was experiencing severe delirium and hallucinations • Thinks AI can help physicians by being a decision support tool
08:02	**Current Delirium Prediction** • Machine learning is a fundamental process of AI, algorithm based, and there are different types (supervised, unsupervised, and reinforcement) • AI→ Machine Learning→ Deep Learning Venn diagram

08:35	**Machine Learning: Different Ways of Learning**
	• Supervised
	  o Labeled data (we know the variables going in)
	  o Transparent, explainable
	  o Human informed and guided
	  o Types/examples:
	    ▪ Logistic & Linear regression, decision-trees (random forest), support vector machine, neural networks, gradient descent, etc.)
	  o About 95% of existing delirium prediction models
	• Unsupervised
	  o Unlabeled data (we do not know variables)
	  o Find hidden structures or patterns within data
	  o Considered "Black Box" (so, this is not as trusted)
	  o Types/examples:
	    ▪ Clustering, more advanced neural nets
	    ▪ Convolutional Neural Nets, Recurrent Neural nets, transformers, etc.
	  o 5% future potential to unlock hidden potential?
	• Reinforcement Learning
	  o Decision making
	  o Repeated trial & error, learns through rewards
	  o Learn series of actions (similar to Bayesian)
	  o Human in the loop
	  o Has been used to understand clinician decision making (inverse)
	  o 0%: need to explore how to apply in delirium prediction

10:35	**Data Types**
	• Structured:
	  o Defined, searchable
	  o EHR flowsheet data
	  o Relational Database/Datamart
	  o This is most often used because easy to use
	• Unstructured:
	  o Data types vary
	  o Text, images, waveform
	  o Data Lakes
	  o Not easy to use

11:27	**<u>Model Development & Testing</u>**
	• Typical Schema: Train→ Validate→ Test
	• Train:
	  ○ Develop model and select features (i.e. variables)
	• Validate
	  ○ Evaluate developed model and improve (hyperparameter tuning—iterative process)
	• Test
	  ○ New data (model has not seen) and report performance

12:13	**<u>Evaluating Performance of ML Models</u>** • Look for these 5 statistics: Area under the receiver operator curve (AUROC), Accuracy, Precision, Recall, F1 Score • Need to find a balance between precision and recall because the more thorough your model, the less precise it will be • These all run from 0 to 1 and a higher score means better performance
13:25	**<u>How is Supervised ML different from regular Logistic Regression?</u>** • Terminology: variables vs. features/labels • Intent: Examining relationships vs. finding the best performing model • Problem Being Solve: Practice-based or Research-based? Exploratory, hypothesis generating?
15:47	**<u>What is the Same?</u>** • Inform Model (our assumptions continue to inform modeling) • Hidden Potential (thinking outside of our assumptions, allow assistance in identifying patterns in data that we cannot see because of our assumptions)

17:47	**<u>Where are we Now?</u>**
	• Search terms "Delirium" AND "Machine Learning" (PubMed search diagram)
	• 3 systematic reviews done already
	  ○ Postoperative (PMID 39395856): random forest most frequently used, pooled AUC 0.792, Ensemble models perform better (AUC 0.805)
	  ○ All Adult Settings (PMID 35922015): pooled performance AUROC: 0.89
	  ○ All Settings (PMID 34373042): random forest, AUROC 0.79-0.91

18:47	**<u>How Could AI Improve Delirium Prediction?</u>** • Reduce noise in data • Dynamic modeling (incorporate real-time information, adapt to changing circumstances) • Incorporate various data types (structured, unstructured wearables, environmental, genetic, lifestyle, ambient sensing, etc.) • Improve screening for studies (more precise, efficient, and dynamic screening for eligibility) • What else? (passive digital markers, detection, prognosis, ???)
20:14	**<u>Previous Work: Static Models & Dynamic Models</u>** • In static models we look at a specific time period for what data is going to predict the outcome of delirium and then once we have that data, we have a pretty large time frame of when that delirium might occur • In dynamic models, we're able to change and shorten that lead time. ○ The ROC improved as the time got shorter, which makes sense that we're able to predict delirium the closer to the event actually happening
21:38	**<u>What Could We Do?</u>** • Are we predicting the risk of delirium? Low, medium, high risk? • Are we predicting the presence of delirium? Diagnostic focused • Are we predicting the prognosis? The likely course and outcomes for this individual? • Are we predicting treatment response? • Are we predicting disease progression?

23:14	**<u>Considerations for the Lifecycle of an Algorithm</u>** • Once the algorithm is done, what does it need next? • Participation, research, skills, concern, knowledge (diagram)
25:44	**<u>Expanding Horizons: Developing and Implementing AI Models for Delirium Prediction in Critical Care (Anirban Bhattacharyya)</u>**
26:06	**<u>Learning Objectives</u>** • Understand the principles behind AI model development for delirium prediction • Identify challenges like transparency, fairness, and bias in AI • Explore the lifecycle of an AL algorithm and its clinical implementation • Envision the future of AI in delirium care, including multimodal approaches • Apply insights to improve AI development and integration in healthcare
27:40	**<u>Model Building</u>** • Delirium prediction to use as a screening tool • 16546 patients • Continuous prediction using sliding observation windows
28:58	**<u>Model Building- Outcome</u>** • Typically diagnosed using CAM-ICU (administered by nurses and done once or twice a shift) o Distribution of CAM-ICU (overwhelming majority had delirium in first week of admission) • We analyzed timing and frequency to plan prediction
29:47	**<u>Model Building- Features</u>** • Have to use features that are commonly available • So, in the ICU that means: vital signs, routinely done labs • In their model also included additional labs and medications, and this improved the performance of their model (however the added lab is not done that frequently in the ICU, and if it is done there is suspicion that something else is going on) • So, they went with a model with all vital signs and routinely done labs even though it sacrifices some of the performance
32:38	**<u>Model Building- Missingness</u>** • Because the features are routinely collected, in the eICU there was very little missingness • Used 1-hour bins for vital signs • 6-hour bins for labs
33:29	**<u>Model Building- Observation and Prediction Windows</u>** • Used 3 different approaches to machine learning (logistic regression, random forest, and LSTM) • Ran the analysis on different observation and prediction windows (the AUC graphs you can see on the slide) • LSTM models performed the best • Found that if you observed for 12 hours and then predicted 12 hours ahead, there was pretty good performance • Observing for 12 and 24 hours was better than 48 for predicting, meaning that whatever causes delirium is happening in the near short term and not very far off
37:05	**<u>Modeling Building- Optimization</u>** • Did feature ranking for ventilation, heart rate, age, WBC, SOFA score, and Vasopressor dose • Chose to have a more sensitive test
38:34	**<u>Model Building- Challenges Ahead</u>** • Fairness metrics, transparency, explainability→causality, real time performance, deal with discrepancies
42:11	**<u>Implementation: Research to Practice</u>** • Standards, best practices, and operational tools

	• Quality assurance • Transparency and accountability • Risk management for AI models in health care • AI Lifecycle: o Planning & Design→ Data Collection & Management→ Model Building & Tuning→ Verification & Validation→ Model Deployment→ Operation & Monitoring→ Real-World Performance Evaluations
43:49	**Reporting of Delirium Prediction Algorithms** • Less than 10% of the models actually show the specific purpose of why they were building it or how it would be used • Do a good job at verifying and validating the model (in about 40% of the studies) • Concerns in model interpretability and operations & monitoring. Very few reports on that
45:15	**Future of AI Models for Delirium** • Model evolution (different models for different outcomes, personalized care, proactive & preventive) • Multimodal AI (Structured Data, Computer Vision, Knowledge base, Tie data to context) • Trust building and Integration (Data visualization tools, Action items integrating with A2F bundle, Clinician training)
47:35	**What Do We Do Next?** • Think about your role • Advocacy for AI integration, data sharing • Standardize AI building and implementation • Multidisciplinary teams and settings to address challenges
48:35	**Questions and Answers**