Artificial Intelligence for Automated Evaluation of High-Fidelity Simulation

Asad Siddiqui MD, M. Ed, FRCPC
The Hospital for Sick Children
University of Toronto

Acknowledgements

Dr. Tobias Everett
Prof. Frank Rudzicz
Zhuojie Zhou
Chuer Chu

Disclaimer

- No financial conflicts of interest
- Not a mathematician or computer scientist
Objectives

- Simulation-based evaluation
- Need for AI/ML in simulation
- AI/ML in healthcare & medical education
- Development of AI/ML algorithms for high-fidelity simulation

The problem

- Medical education → CBME
  - Greater emphasis is placed on evaluation, feedback and assessment
  - Problem: Higher frequency of evaluations required → resource intensive
- High-fidelity simulation
  - Used as summative and evaluation tool
  - Availability of experts limiting factor
  - 27% perceived the lack of trained faculty as a barrier

Need an innovative solution to lack of human resources available in high-fidelity simulation
MEPA

- International simulation course
- Standardized operating room crisis situations
- 10 YEAR REVIEW PAPER

Pediatric Anesthesia

Global Rating Scale is reliable in evaluation of MEPA

The Managing Emergencies in Pediatric Anesthesia global rating scale is a reliable tool for simulation-based assessment in pediatric anesthesia crisis management.

MEPA GRS can differentiate between different grades of experience
- Robust tool for evaluation – highly validated
CanNASC

- National evaluative simulation curriculum
- Evaluative programs require:
  - Consistent/Reliable
  - Quality assurance
- Increasingly difficult to find expertise with consistency

Automation of evaluation in simulation would have many benefits:

- Decreased human experts needed
- Quality improvement for evaluative simulations
- Increased access to simulations with evaluative feedback
- Lateral translation to assessment of real cases in OR

Artificial Intelligence & Machine Learning could provide a way for automation of simulation evaluations
Machine Learning in Medicine

"I think that if you work as a radiologist you are like Willy L. Coover in the cartoon. You've already over the edge of the cliff, but you can't yet look down. There's no ground under your feet. It's all asymptote.

Deep learning is going to do better than radiologists. It might be any years."

Machine Learning in Medicine

- ML benefits beyond image diagnostics
- ML can recognize complex physiological patterns
Machine Learning in Medicine

**Will artificial intelligence solve the human resource crisis in healthcare?**

- MCC short answer questions
- Over 90% accuracy on six short answer questions

**Machine Learning in Medical Education**

- Tool, hand and eye motion tracking have been used in surgical assessment of trainees
- Include computer vision, deep learning techniques for analysis
- Simulated and OR setting
Machine Learning in Medical Education

- Model venous anastomoses by experts and novices
- Hand motions tracked
- ML delineated experts and novices at accuracy of 83%

Artificial Intelligence for Automated Evaluation of High-Fidelity Simulation

Project

To create an AI computer-generated algorithm for scoring anesthesiology trainees in a high-fidelity simulation scenario (anaphylaxis)

80 anaphylaxis videos

Previously rated by expert raters on GRS 1-6
Methods

Development of Automated Evaluation Algorithm

- Phase 1 – Transcription
  - Time-Frame Tagging of Key Words
- Phase 2 – Vital Signs Extraction
- Phase 3 – Movement Analytics
- Phase 4 – Audio Analytics
- Phase 5 – Creation of model to incorporate algorithms from phase 1-4

---

Phase 1 - Transcription

- Pre-trained automated speech recognition model for transcription
- Low accuracy due to audio quality
- Manual transcription
  - Approx. 70% rate of audio-text transcribed
  - Key words + Time Stamp

---

Phase 2 – Vital Signs Extraction

- 12-layer CNN for multi-digit number classification
- Images within defined box coordinates extracted
- Defined vital signs per second
- 10% of vital signs extraction manually checked
- 99% accuracy of extraction
Phase 3 – Movement Analytics

- Extraction class to divide video into second marks
- Holistic movement dynamics
- Coarse arm motion

Phase 4 – Audio Analytics

- Audio files separated from video
- Affective Computing
- OpenSMILE (feature extraction tool) to obtain emotion features

Final Model

- Core part of model is bidirectional transformer encoder
- Preprocessed data – four parts – put in model for each second of simulation
- Output: pass (GRS 4-6) vs fail (GRS 1-3)
- Target computed with true value and loss-function used to updated and backward the parameters in the model
Final Model

- Small sample size
- Cross-validation to fully use data
- 5-folder cross used
Solution to small sample size to maximize the dataset

Results

- Overall accuracy of best model: 71%

Accuracy: 71%
Results

1. Transcription
   - Vital Signs: Pass
   - Movement Analytics: Fail
   - Audio Analytics: Fail
   - Accuracy: 55.8%

2. Transcription
   - Vital Signs: Fail
   - Movement Analytics: Fail
   - Audio Analytics: Fail
   - Accuracy: 61.5%

3. Transcription
   - Vital Signs: Fail
   - Movement Analytics: Fail
   - Audio Analytics: Fail
   - Accuracy: 67.3%
Results

**Accuracy:** 65.4%

Importance

- Test
- Vital Signs
- Audio Analytics
- Movement Analytics

Results Details

- Agreement expert raters pass/fail in data set was 90%
- Successful transcription of approx. 70% of audio to text
- Transcriptions from poor audio
Future Directions

Expand video data set and expand training data

Lateral application to anaphylaxis in different simulations

Performance prediction in real clinical scenarios

Summary

Evaluative simulations increasing in frequency

Problem: Lack of human resources

ML/AI prevalent in healthcare and emerging in medical education

ML/AI algorithm for automated evaluation in simulation anaphylaxis

75% accuracy

Acknowledgements

Dr. Tobias Everett
Prof. Frank Rudzicz
Zhuojie Zhou
Chuer Zhu