UT Southwestern Medical Center
Department of Surgery

Developing Artificial Intelligence Models for Medical Student Suturing and Knot-Tying Video-Based Assessment and Coaching

Madhuri B. Nagaraj MDMS, Babak Namazi PhD, Ganesh Sankaranarayanan PhD, Daniel J. Scott MD
Department of Surgery, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas TX 75390 USA

BACKGROUND

• Due to the social distancing implications of COVID-19, an at-home suturing and knot-tying curriculum with video-based assessment was developed for second-year medical students (MS2).

• Deliberate practice (distributed learning, directed feedback) has been shown to impact skill acquisition.

•Assessment and directed feedback is required for skill acquisition, but this process can be time-consuming and expensive.

• The need for feedback and frequent assessments is an enduring concern for both remote and in-person training.

AIM: We aimed to use artificial intelligence (AI) to develop a deep learning model that would (1) distinguish pass-fail performances to develop an assessment tool and (2) identify the reason for task failure as the basis for future formative feedback tools.

METHODS

• MS2s trained to proficiency on a simple interrupted knot on a penrose drain with instrument tying technique

• Proficiency was defined as performing the task < 2min with no critical errors (touching the needle, improper instrument hold, improper knot-tying technique).

• Videos were graded as pass-fail

• Comparison was made between manual grading by a senior surgical resident and AI model grading.

Manual Assessment

Input: Expert video review

Output: Pass-Fail

AI Assessment

Input: Extracted frames of each video

Output: Probability of Pass-Fail

TECHNOLOGY & MODEL

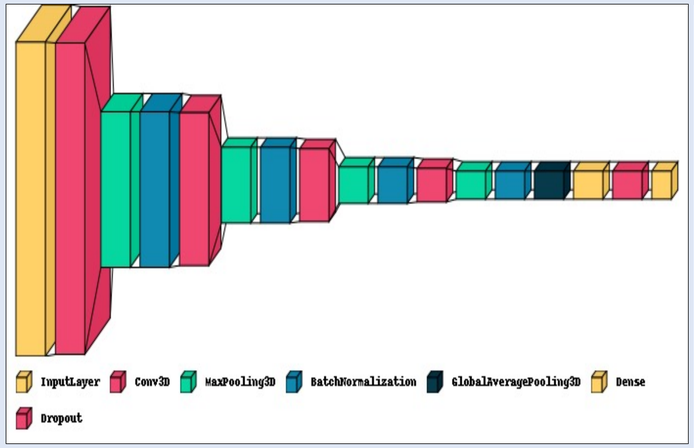


Fig 1. The block diagram of the 3D CNN model

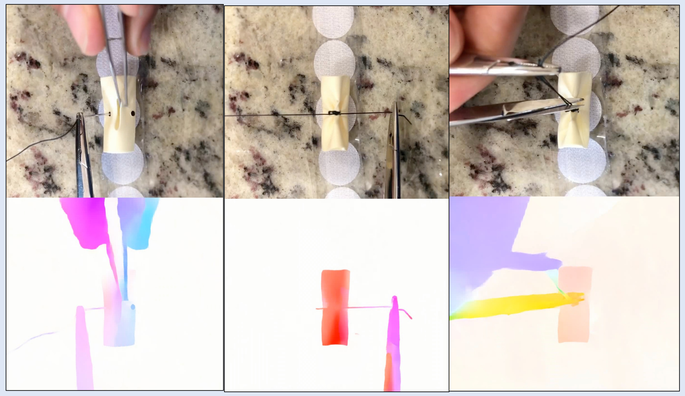


Fig 2. Visualization of the optical flows (not used for training)

•A 3D Convolutional Neural Network (CNN) was developed and trained.

•The model architecture consists of 5 layers of 3D convolution, max pooling and normalization, followed by a fully connected layer for classification.

• All extracted frames were resized to 224x224 pixels, and random cropping, flipping and translations were used as data augmentation.

• The model was trained with cross-entropy loss for 100 epochs using SGD optimizer.

• For each video, 80 frames were used for training and testing. Shorter videos were padded with blank images.

•Approximately 80% of the videos were used for training and 20% for validation.

•The movements of the objects were extracted by visualizing the optical flow features (not used for training the model)

Definitions

• Convolutional Neural Networks (CNN) – A class of deep neural networks most often applied to visual imagery (i.e. image classification, object recognition). - In 3D CNN, kernels slide in all three dimensions to extract both the spatial and temporal features that are essential for classifying a video.

• F-1 score – Metric for AI model accuracy using precision and recall. Ranges 0% (poor) to 100% (perfect).

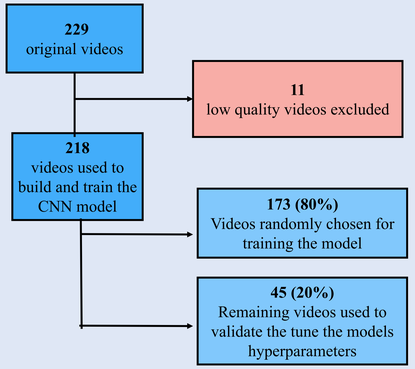
• Optical flow: the pattern of apparent motion of objects in a visual scene

MANUAL RESULTS

• 229 videos were manually reviewed (150 pass, 79 fail).

• Critical error distribution: 15 knot-tying errors, 47 instrument-holding errors and 17 both errors.

PRELIMINARY MODEL RESULTS



Current Model Metrics

• Accuracy: 75.55%
• F-1 score of 52.17%.

CONCLUSIONS & FUTURE DIRECTIONS

•Computer vision and deep learning techniques can instead be reliably employed to perform automated surgical video analysis.

• The lower F-1 score is likely due to the high variations in the video quality and camera parameters, and high-class imbalance.

•Next steps:

- Incorporating motion features like optical flow (figure 2).
- Train the model to perform automated task segmentation to identify the cause of failure and provide step-by-step guidance and feedback.
- Augmenting the dataset by collecting more standardized videos

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the University of Texas Southwestern Simulation Center and the medical students who participated.