Clinical Training Simulator for Suturing

Team Members: Lucy Hockerman, Presley Hansen, Maddie Michels, Kate Hiller, Sadie Rowe

Clients: Dr. Margene Anderson, Dr. Sara Colopy, Dr. Paul Merkatoris

Advisor: Dr. Walter Block

BME 402, Spring 2026

UW BME Design

College of Engineering
UNIVERSITY OF WISCONSIN–MADISON

Clinical Simulation Training Gap

- There is currently no robust training stimulation used to standardize proper suture tension
- Novice clinical students struggle to judge proper suture tension, often tying knots too loose or too tight during training
- A feedback system is needed to help students learn proper knot tension, improving training efficiency and surgical outcomes

Clinical Context & Impact

Need: System to train the correct tension needed for a secure knot

- Students currently learn by "feel" with instructor feedback after the knot is tied

Too Loose: Suture may unravel, leading to wound reopening and healing complications

Too Tight: Risk of tissue damage and material failure

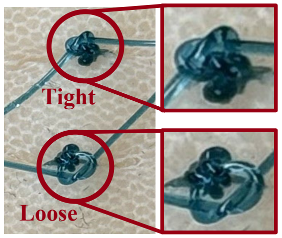

Figure 1: Comparison of loose and tight knots. Loose knots have a visible gap in the top throw.

Impact

- Potential for broad application across a variety of clinical training programs
- Wound opening causes an average of $40,323 in excess hospital charges per patient [1]

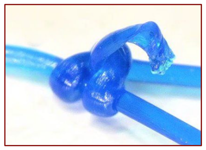

Figure 2: Suture material failure from over-tightening [2].

Design Specifications

Model Requirements

- Classify **Nylon Mono (black) 2-0 sutures**
- Classify knots with an overall **accuracy ≥ 80%** [3]
- Achieve **≥ 80% precision** for "tight" class to minimize false positives [3]
- Attain **≥ 80% recall** for "loose" class to minimize false negatives [3]
- Have a **F1 score of ≥ 0.8** [4]

System Requirements

- Withstand repeated use in training
- Minimally disruptive to the suture process
- Maximum system latency: ≤ 5 seconds

Acknowledgements

- Dr. Walter Block
- Dr. Margene Anderson
- Dr. Sara Colopy
- Dr. Paul Merkatoris
- BME Department
- UW-Madison School of Veterinary Medicine

References

[1] V. K. Shanmugam *et al.*, "Postoperative wound dehiscence: Predictors and associations," *Wound Repair Regeneration*, vol. 23, no. 2, pp. 184–190, Mar. 2015, doi: 10.1111/wrr.12268.

[2] "The ideal surgical knot (according to science)," *Healthcare-in-europe.com*, Jun. 12, 2023. http://healthcare-in-europe.com/en/news/ideal-surgical-knot.html (accessed Oct. 01, 2025).

[3]"Is 80% accuracy good in machine learning?," *Milvus.io*, 2025. https://milvus.io/ai-quick-reference/is-80-accuracy-good-in-machine-learning (accessed Dec. 01, 2025).

[4] "F1 Score in Machine Learning." Accessed: Oct. 07, 2025. [Online]. Available: https://encord.com/blog/f1-score-in-machine-learning/

[5]"Amazon.com: Arducam IMX477 Pi HQ Camera for Raspberry Pi 5, Comes with a 1/2.3" 6mm Focal Length CS Lens, Widely Compatible with Raspberry Pi 4B, 3B+, Zero : Electronics," Amazon.com, 2026. https://www.amazon.com/Arducam-IMX477-Camera-Raspberry-Compatible/dp/B0D95VWCV6?th=1 (accessed Apr. 20, 2026).

Workflow and Testing

Dataset Creation

- Consistent lighting
- Fixed viewpoint
- Cropped to 500x500

Dataset Breakdown

TIGHT	LOOSE
100	100

Figure 3: Arducam IMX477 Pi HQ Camera [5].

Image Preprocessing

Image Preprocessing	K-Fold Cross Validation Avg. Accuracy (K=6)
Binary Images (A)	53.00%
Contrasted Images (B)	61.57%
Normal Images (C)	73.93%

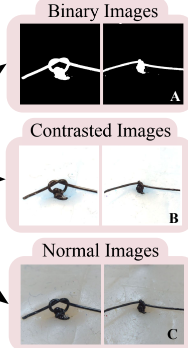

Figure 4: (A) Binarization enhances edge detection. (B) Contrast adjustment reduces background information. (C) Original, unedited images.

ResNet50 Convolutional Neural Network

Table 1: Confusion matrix for ResNet performance testing.

Predicted Label \ True Label	Loose	Tight
Tight	0.10	1.00
Loose	0.90	0.00

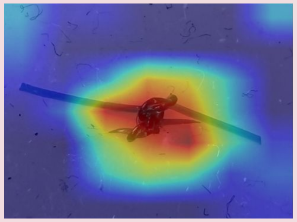

Figure 5: Heatmap is showing the model is properly focusing on knot morphology.

F1: 0.95 **Precision:** 90% **Recall:** 100% **Accuracy:** 95%

Model uploaded to Pi

Camera Preview → Image Captured → Raspberry Pi → Knot Classification (TIGHT / LOOSE)

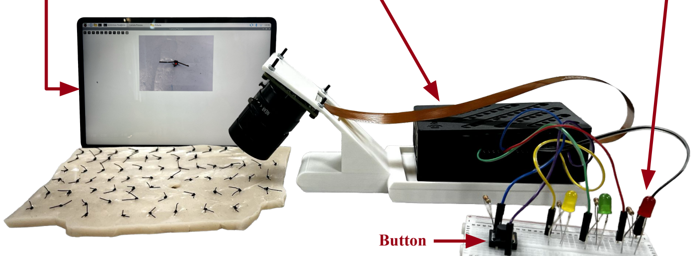

Figure 6: Overall system setup showing HQ camera, Raspberry Pi, and breadboard hardware. User input via button initiates a camera preview. Second click initiates image capture that is passed through the model (housed on the Pi) and model output indicated by red and green lights.

Testing & Results

System Performance

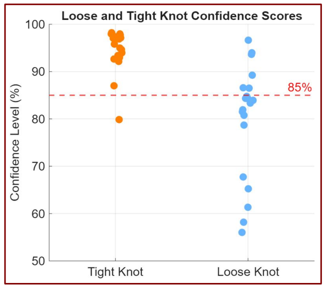

Figure 7: Distribution of model confidence scores for predicting knot class, with an 85% temporary confidence decision threshold used for classifying tight knots.

Accuracy: 82.5% **Precision:** 76%

Recall: 95% **F1:** 0.84

	True Loose	True Tight
Predicted Tight	6 (FP)	19 (TP)
Predicted Loose	14 (TN)	1 (FN)

Table 2. Confusion matrix for model performance with threshold implemented. A true positive (TP) is defined as a tight knot correctly classified as tight (n = 40). Table shows accuracy of model is greater than the 80% goal when threshold is implemented.

Latency Testing

Overall Mean Latency: 3.93 seconds

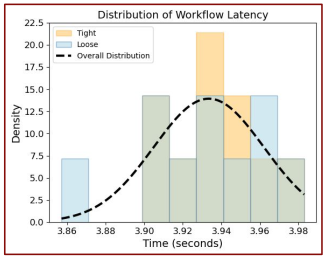

Figure 8: Standard distribution of latency (seconds) from second button press to LED feedback for tight and loose knot classification.

Discussion

Key Insights

- Confidence threshold improves model reliability by filtering low-confidence predictions that are prone to error
- Pre-processing analysis:

	Binarization	Contrast
Pros	Clearly defines edge boundaries Negates lighting inconsistencies	Reduces background noise while preserving structure and depth
Cons	Oversimplifies knot features Reduces depth cues	Exaggerates shadows Overexposes suture material

Limitations

- **Discrepancy between evaluation methods:** ResNet-based vs manual testing
  - Noise from pi integration may cause model prediction inaccuracy
  - Model trained/tested on limited dataset → learned from narrow distribution
- **Rigid imaging setup:** non-adjustable camera stand → limited suture placement

Future Work

- Improve model accuracy
  - Expand and diversify dataset
  - Hyperparameter tuning to optimize training settings
- Future testing to refine confidence threshold
- Improve hardware durability
- Develop an adjustable camera stand