Knot too Tight - Knot too Lose

FINAL REPORT

Team Name:
The Knotorious Five - BME 400

Team Members:
Madison Michels (Team Leader)
Lucy Hockerman (Communicator)
Presley Hansen (BWIG)
Madison Michels (Communicator)
Kate Hiller (BSAC)
Sadie Rowe (BPAG)

Client:
Dr. Margene Anderson
Dr. Sara Colopy
Dr. Paul Merkatoris

Advisor:
Dr. Wally Block

December 10th, 2025

Abstract

Proper suture tension is critical for achieving wound closure integrity and preventing tissue damage or knot failure. However, current veterinary training methods rely on subjective evaluation and repeated practice, with no qualitative system to teach students how to achieve optimal suture strength. Existing training tools either lack real-time feedback, are not commercially available, or only measure tension in the first knot throw, leaving a gap in effective suture training devices. To address this, two preliminary concepts were developed and evaluated. The first concept, a force-sensing system using force-sensitive resistors (FSRs) to provide real-time LED feedback, was prototyped but ultimately deemed not viable because it could not reliably translate applied force into knot tightness. The second concept, a visual knot characteristic system, uses a machine learning model to analyze suture knot images and provide real-time feedback on knot security. This design concept was prioritized for its alignment with traditional instructor evaluation methods, minimal interference with suturing technique, and adaptability. A comprehensive knot image database, including both tight and loose knots, was created and used to train five machine learning models across different platforms. Each model was evaluated on the same test set using performance metrics including confusion matrices, accuracy, precision, recall, and F1-scores. The two highest-performing models were statistically compared with a McNemar test, and no statistically significant difference was found between them. Potential sources of errors and system limitations have been identified to guide future model refinement and development of a physical training module capable of delivering real-time model feedback.

Table of Contents

Introduction

Motivation & Impact

Currently, there is no device on the market or in veterinary skills laboratories that allows students to learn and assess proper suture knot tension. Suture knot tying is a skill that is purely learned through "feel" and extensive practice. The security of a knot depends on achieving appropriate tension during the final throw, which ensures sufficient plastic deformation of the suture material. However, students often struggle with either tying the knot too tightly and causing material failure, or too loosely, allowing the knot to unravel after they release the suture material, as shown in Figure 1. Improper suture tension can have significant clinical consequences. Excessive tension may cause tissue damage, while insufficient tension can lead to knot slippage or wound dehiscence. [1]
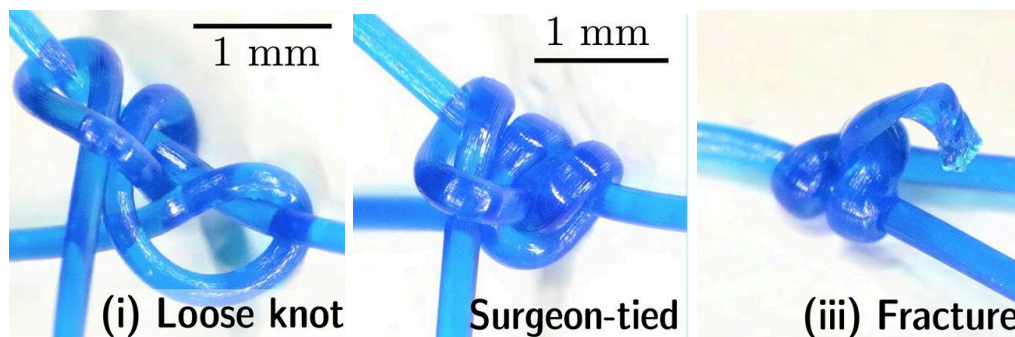

Figure 1. Examples of suture knots tied under different tension conditions, showing a loose knot (left), a correctly tied surgeon's knot (middle), and a fractured knot (right) resulting from excessive tension [2].

After a student completes a knot, the professor evaluates it with a magnifying glass to determine whether the knot was tied correctly. There is no quantitative force value that defines an optimally secure suture knot, as this value depends on the suture material, thread size, pulling speed, and knot style. In the literature, only values of the ultimate tensile strength of some suture materials have been evaluated, with little to no data available on the tension in the final throw of the knot [1, 2, 3].

This project aims to develop a data-driven training system to help teach students more efficiently and with fewer repetitions. Also, it would enable students to practice and assess their technique independently, without the need for an instructor. Students use lots of sutures to train, and each stitch costs about $1.73 - $1.83, not only costing the school but also contributing to medical waste. [6] Our system will introduce cost savings as well as minimize environmental impact by reducing the amount of sutures used in training. Additionally, the project has the potential to expand to the greater medical field, having use in applications for medical students.

Existing Suture Training Designs

The ForceTRAP visual force feedback system (Figure 2) is a device referenced in research articles, although limited information is available about the product. The device contains a microcontroller that converts a sensor's output from the model tissue into a calculated reaction force. This force is then translated into LED feedback for the user, where green indicates an in-range force (0.5 N - 1 N), orange indicates a slightly high force (1N - 2N), and red indicates an out-of-range force (>2 N) [7]. It provides feedback on needle insertion, knot-tying force, and alignment of the two tensioned suture threads.
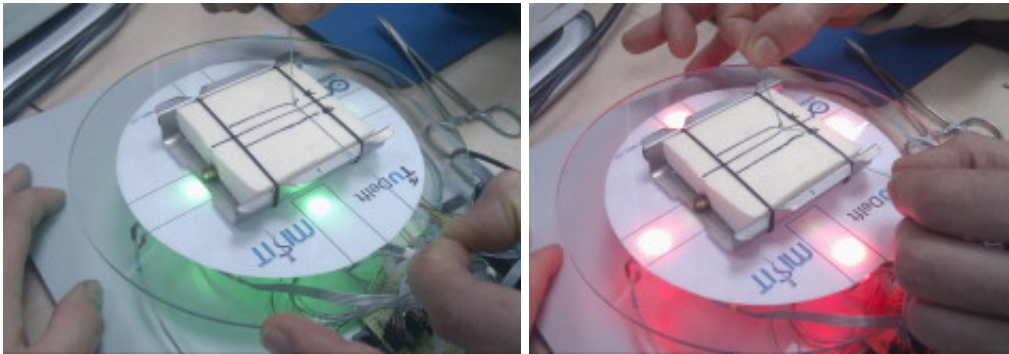
Figure 2. The ForceTrap Training Device is designed to measure and provide feedback on the first throw of a suture knot, which is independent of the final knot tightness [7].
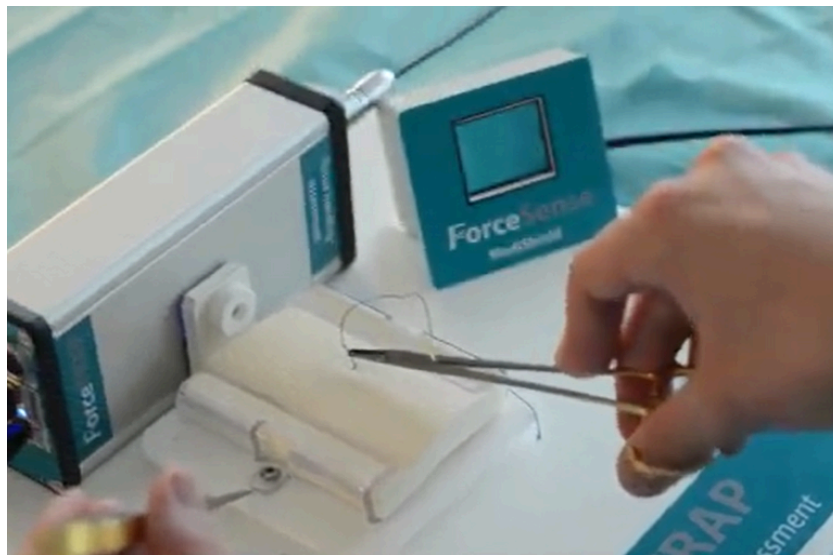
Figure 3. The ForceTRAP Training System integrates a front-mounted camera and embedded sensors within the suture pad to provide real-time feedback on knot tension. A color indicator illuminated to signal to the user when the proper tension has been achieved [8].

There is a standalone ForceTRAP (Figure 3) manufactured by Medishield. However, very limited information is available about this device, as the manufacturer's website no longer exists.

Therefore, this device is currently not on the market. Additionally, it only measures the force in the suture thread of the first throw of the knot, which is independent of final suture knot tightness, as the last two to three throws determine a suture knot's tension.

Another measurement system used to measure force in surgical sutures is the Hook In Force (HIF) sensor. It utilizes a sensor and a magnet to measure the displacement of the thread once it is fed into the system. The purple line in Figure 4 represents the thread within the HIF sensor. This system can measure displacement in the range of 1 mm to 3 mm, converting measured displacement into a corresponding force output. The HIF sensor is referenced only in research articles and is not currently available on the market. Additionally, it is also not integrated into a system to measure the displacement of the final throw of a suture knot [9].
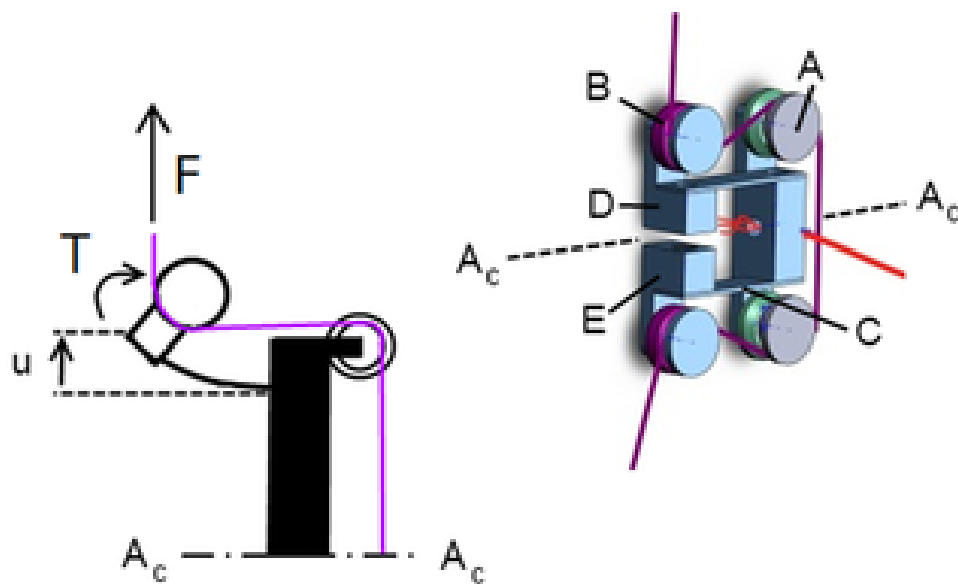
Figure 4. Illustrations and 3D models of the Hook in Force (HIF) Sensor published in literature. This device is not integrated into a training system for suture tension evaluation [9].

Lastly, the wheel sensor (Figure 5) uses a Hall sensor that detects magnetic field changes as a magnet moves when the thread is pulled through the system. A greater displacement of the magnet corresponds to a greater pulling force. The Hall sensor outputs a voltage signal, which is then processed by a microcontroller that converts it into pulling force values and translates these into LED feedback. A green LED indicates a safe working range for the pulling force, while a red LED is a warning that the pulling force is too high. The wheel sensor is only found in research literature and is not currently on the market. It is also not integrated into a training system for students to use.

Figure 5. The wheel sensor prototype using a Hall sensor and magnet to detect suture pulling force. This design remains a research concept published in literature, and not integrated into a training system [10].

Problem Statement

In veterinary training, mastering the skill of applying appropriate suture tension is essential for successful wound closure and patient recovery. However, novice practitioners often struggle to judge the correct amount of force needed, leading to either insufficient tension or excessiv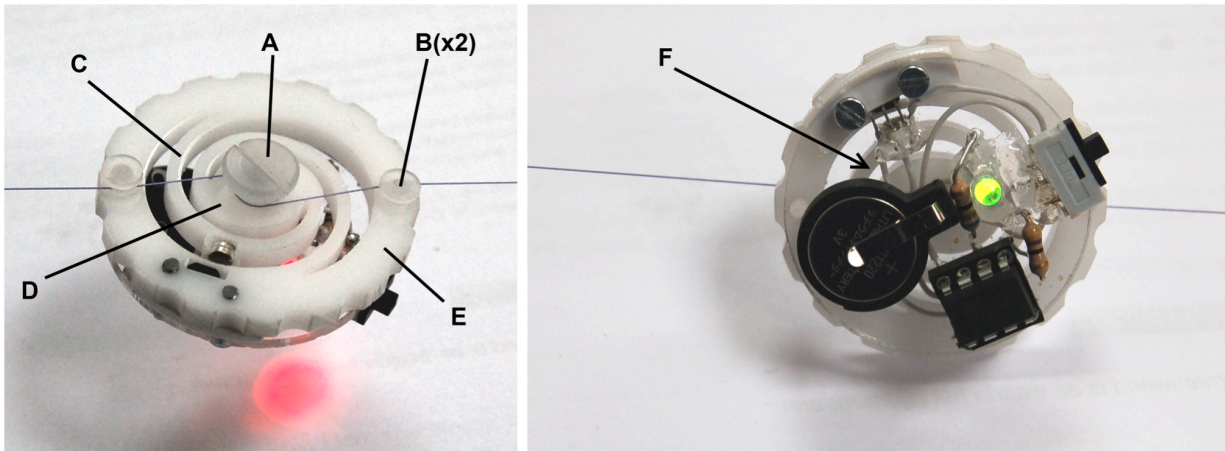e tension, which can cause plastic deformation of the suture material or tissue damage. Currently, the evaluation of suture technique relies heavily upon subjective instructor feedback, lacking objective, real-time metrics to guide learners. This gap hinders consistent skill development and increases the risk of procedural errors. There is a critical need for a real-time suture tension measurement and feedback system to help students learn to apply optimal tension, prevent material or tissue compromise, and improve surgical outcomes through data-driven training.

Background

Variability in Suturing

The team was tasked with creating a data-driven, real-time feedback system to teach students proper suture knot tension. A knot is secure when the suture material plastically deforms. Many variables affect knot security, including suture material, knot type, and material size.

A "throw" is the term used to describe the action of crossing the ends of the suture to form a loop and then wrapping one end of the suture around the other [11]. See Figure 6 for the first throw of a suture knot. The client requested that the team focus on square knots (Figure 6) when developing the measurement system, with plans to expand the measurement system to accommodate additional suture knot types.

FIG 1. SIMPLE SUTURE TECHNIQUE

a. Insert the needle at 90° angle to skin to evert wound edge

b. The suture should exit at 90° with each side a mirror image of the other. This is the 'bite'

c. Pass a double loop around the needle holder

d. Clamp the free end of the suture

e. Pull it through to create the first knot or throw

f. Create a single loop in the opposite direction to the first knot and tighten. Single 'throws' can secure if

Figure 6. The first throw of an interrupted square knot suture, demonstrating the initial loop formation and hand position required to establish proper tension and alignment of the suture strands [11].
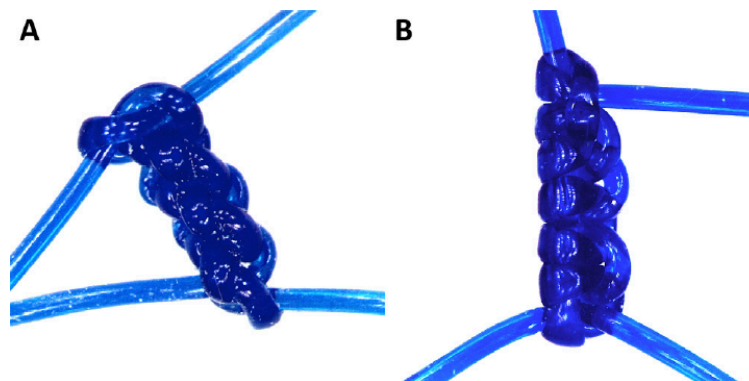
Figure 7. There are many knot types physicians use, including square knots (left) and slip knots (right). This project focuses on evaluating square knot tension, with integration of other knot types in the future [12].

Surgeons learn numerous suture patterns. The client demonstrated simple interrupted and simple continuous suture techniques (Figure 8). To simplify the task of designing this system, the team will focus on measuring tension in simple interrupted sutures.
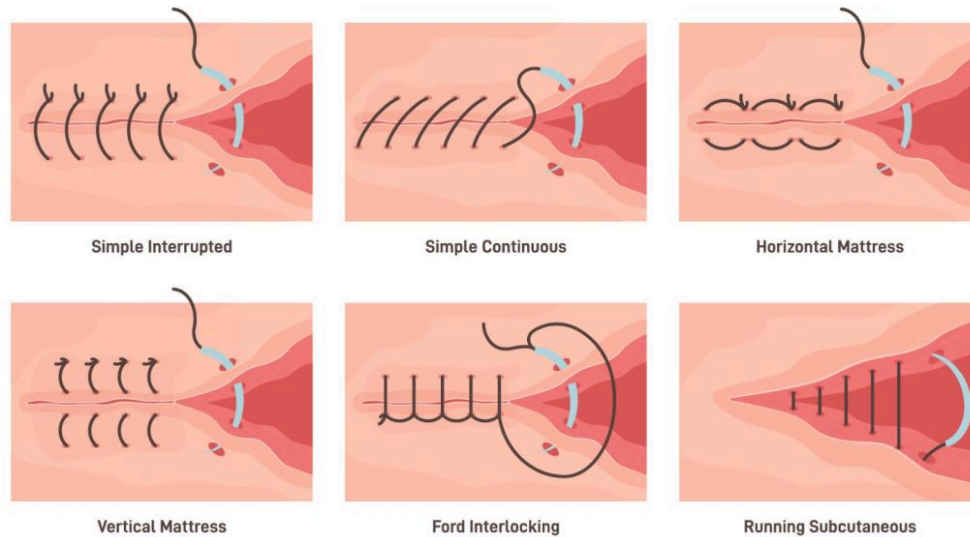
Figure 8. Basic suture patterns, showing common techniques physicians use for wound closure. Each pattern serves a different purpose depending on the tissue type, desired tension distribution, and wound stability requirements. The proposed training system focuses on evaluating simple interrupted stitches [13].

The client primarily uses Maxon, Biosyn, and PDS for their sutures. Maxon is an absorbable monofilament made from polyglyconate (glycolide and trimethylene carbonate) [14]. Biosyn is another absorbable monofilament made from Glycomer 631 [15]. Polydioxanone or PDS is a synthetic absorbable monofilament homopolymer of poly(p-dioxanone) [16]. PDS generally exhibits higher maximum tensile strength and lower stiffness [5]. Biosyn absorbs much more quickly and has better flexibility. Maxon has similar material properties of absorption and strength to PDS, but Maxon exhibits greater memory [17]. Since these sutures all have different material properties, the tension required for a secure knot will vary by material.

Lastly, there are various suture thread diameters standardized by the United States Pharmacopeia (USP) scale (Figure 9) [18]. Larger numbers followed by a zero indicate thinner suture material, while larger numbers without zeros indicate thicker sutures. Suture size affects knot security, as thicker suture sizes are more difficult to plastically deform, while thinner sutures are more prone to material failure during tightening. The team will develop a training system allowing for the evaluation of various suture colors, materials, and thicknesses.

Figure 9. Suture diameter size chart, showing the relationship between suture diameter and common name used in the field [19].

Client Information

The client, Dr. Margene Anderson, from the Veterinary School of Medicine, oversees educational advancement and faculty development. Dr. Anderson serves as the primary point of contact for the project. Dr. Paul Merkatoris, a large animal surgeon, and Dr. Sara Colopy, a small animal general surgeon, also support this project. They teach veterinary students and provide the team with clinical insight into the suturing process. Together, they provide feedback throughout the design process to ensure that the device is feasible, functional, and meets the project's needs. Additionally, the clients have taught the team proper suture techniques to deepen the team's understanding of the project needs and strengthen design ideas.

Product Design Specifications

The machine learning model design requirements focus on the ability to classify a variety of suture sizes, colors, and materials while accurately identifying knot tension. The model must achieve an overall accuracy ≥ 80% and a precision of ≥ 80% for the "tight" class to minimize false positives, as identifying a loose knot as a tight knot was determined to be the worst case scenario. The model shall also have a recall of ≥ 80% for the "loose" class to minimize false negatives. Additionally, the model shall have an F1score of 0.8 or greater to ensure reliable and balanced performance across all classifications. [20]

The overall system requirements emphasize durability, responsiveness, and seamless integration into the suturing process. The system must withstand repeated use in a training environment and provide real-time feedback on knot tension with a latency of less than or equal to 3 seconds. If a quantitative measuring system is used, the tensioning device must withstand forces up to 30 N, with a working range of 0-20 N and accuracy of ±0.5 N [10], [21]. The feedback must be immediate and intuitive to the user through visual or sound indicators to guide the user. It must survive over 500 training cycles without mechanical or electrical failure while

adhering to electrical safety standards outlined in IEC 62368 and IEC 61010-1 [22], [23]. Also, the design shall not interfere with the user's ability to perform suturing techniques or increase suturing time by more than 15 seconds. The system shall be compatible with multiple suture sizes, including 2-0, 3-0, and 4-0 diameters. Lastly, the total device cost shall remain within the client's budget of $250, with the possibility of additional funding as the project progresses. Additional details on the Product Design Specifications can be found in Appendix A.

Preliminary Designs

Visual Knot Characterization

Currently, trainees' knots are evaluated visually by an instructor to determine whether the suture has achieved sufficient tension to induce plastic deformation and ensure the final throw is adequately secured. During this evaluation, instructors look for gaps between the final throw and the base of the knot, as well as signs of unraveling both of which indicate insufficient tightening.

The proposed solution aims to automate this assessment using an overhead camera combined with a machine learning model to analyze the knot directly. Rather than relying on instructors to provide feedback after the procedure, this system would optically detect loosening in real time, enabling students to immediately recognize whether they have applied proper tension. This reduces the delay between performance and feedback, making it easier for trainees to associate knot security with the technique and force they applied.
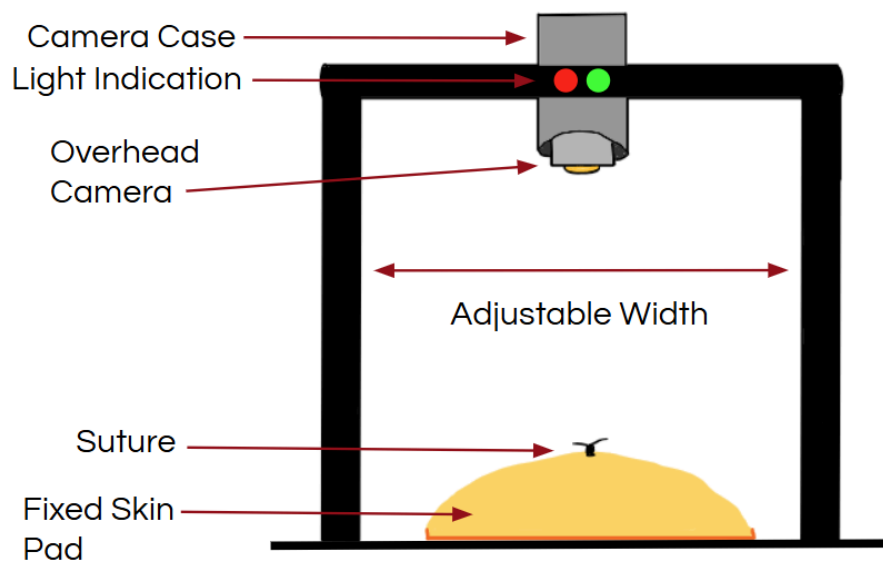

Figure 10. Optical Tension Design.

The proposed design is shown above in Figure 10. The training module would consist of a sturdy frame with an adjustable width to accommodate different users, and a fixed,

interchangeable skin pad to maintain consistent imaging conditions. An overhead camera is housed in a removable case attached to the frame and is connected to a red/green light for real-time feedback.

This design concept does not utilize direct measurement of the tension in the suture, but rather would be trained to evaluate knots the same way instructors do currently. Machine learning (ML) is a subset of artificial intelligence (AI) that is focused on enabling machines to learn from provided data to improve their performance without being explicitly programmed. For this design concept, a machine learning model could be trained on a large dataset of knot images, such as the one pictured in Figure 11, to learn visual features that distinguish secure knots from insecure ones. The process would involve extensive image training and exploration of the most effective methods for visual recognition. Potential features include pixel coloration differences within the gap between the final throw and the base of the knot (as shown with the red arrow in the bottom right corner of Figure 11), variations in knot diameter between tight and loose knots, or other extractable characteristics relevant to assessing knot quality.
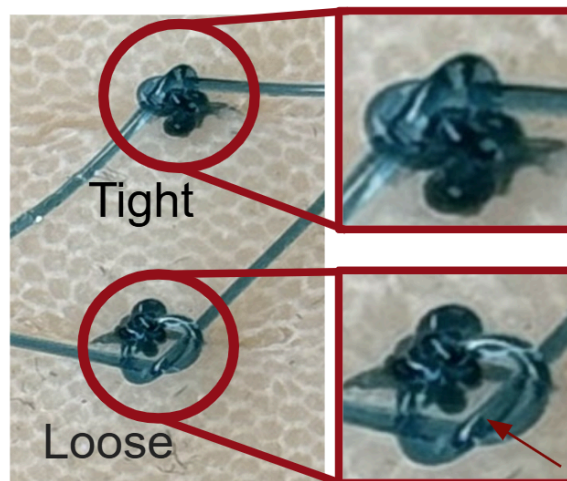
Figure 11. Comparison of square knots: the top knot demonstrates correction tension, while the bottom knot illustrates insufficient tightening.

Displacement as an Indicator of Plastic Deformation

Another method that can be used to determine when a suture knot is sufficiently tightened is to use displacement values as an indirect indicator of plastic deformation in the suture material. Displacement values are dependent on both the suture's material properties and diameter, and would first need to be characterized and validated through controlled mechanical testing. The proposed design concept, illustrated in Figure 12, incorporates a location sensor that tracks the position of the user's hands in a defined coordinate system before and after the knot of the final throw is pulled tight. This allows the system to calculate the total displacement and provide real-time feedback to the user with a visual indicator (e.g., a green light) when the target

displacement is reached. Compared to force-based measurement, displacement is a simpler and more accessible parameter to monitor; however, challenges such as grip slippage during knot tying may result in inaccurate readings.
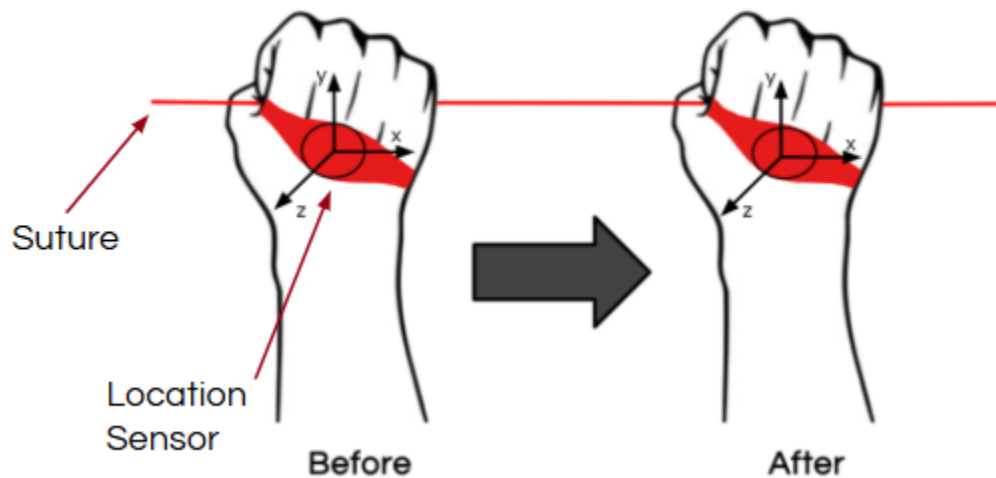
Figure 12. Location Sensor Design to Measure Displacement.

Tension or Force as an Indicator of Plastic Deformation

Force is the intensity of the push or pull of an object. Measuring the force of the trainee's pulling, pinching, or tightening motions are all methods of capturing and quantifying the suturing process and provides insight into their technique, precision, and control. Using force offers a means of assessing suturing proficiency in a measurable and repeatable way.

One simple way to determine the force of pinching or pushing on the suture is with the use of force sensor resistors (FSR). These sensors detect the force applied to them and respond by changing their resistance value. Some FSRs are commercially available with calibration charts that directly correlate specific resistance values with force values. Others, however, require manual calibration using standardized weights and a system to map resistance to force. This system would be incorporated into a wearable glove and wrist-mounted design that allows for minimal interference of the procedure, shown in Figure 13. In this concept, the force sensors would be sewn into the glove at critical contact points. Wires embedded in the glove would connect these sensors to a compact digital processing unit housed on the wrist. This unit would interpret sensor data in real time and provide immediate feedback to the user via visual indicators, such as LED lights. Testing and iteration would be necessary to determine the optimal placement of the FSRs on the glove, ensuring minimal interference with the trainee's hand movements and maximum accuracy in force capture.
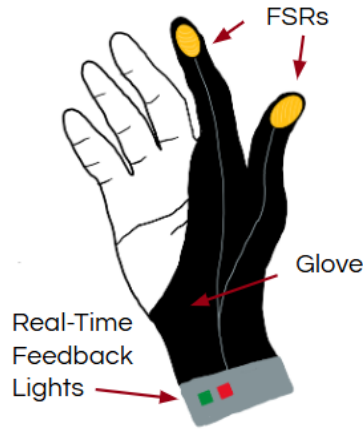
Figure 13. Wearable FSR Design.

A second critical aspect of evaluating suturing technique is the tension applied to the suture thread during the final knot-tightening phase. Excessive tension may damage the suture, while insufficient tension will cause the knot to slip. To measure this, a tensiometer can be incorporated into the training process as a temporary attachment to the suture.

One option is a three-pronged tensiometer, shown in Figure 14. This device can be clipped to the free end of the suture just before the knot is tightened. As the trainee pulls the thread, the downward movement of the bottom prong reflects the tension applied. An internal sensor measures the displacement and correlates it to a tension value.
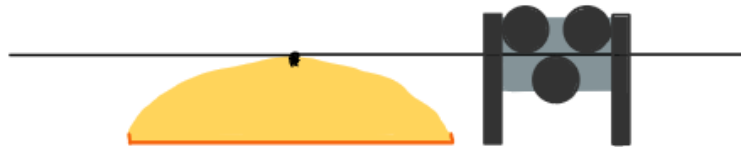
Figure 14. Three-Pronged Tensiometer Design.

Alternatively, a spring-loaded tensiometer offers a more dynamic approach. In this design, see Figure 15, the trainee loops the suture through the hook or pulley mechanism connected to a calibrated spring. As they pull the suture to tighten the knot, the spring stretches. The distance of spring displacement corresponds to the force exerted, which can then be converted to tension values using pre-established calibration data.
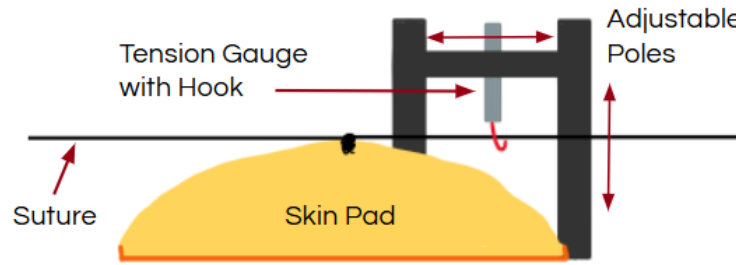
Figure 15. Spring-Loaded Tensiometer Design.

Both of the tensiometer designs require integration with a digital feedback system, ideally providing real-time feedback through visual cues, including LED lights or haptic feedback (like vibration). This would allow trainees to immediately understand if they are applying too much or too little tension during knot tying.

The integration of FSRs and tensiometers in surgical training tools offers a quantifiable data-driven approach to assessing and improving suturing skills. Future iterations could incorporate Bluetooth or Wifi modules for data logging and remote evaluation. By combining wearable sensor technology with intuitive feedback mechanisms, this system has the potential to significantly enhance training outcomes, promote muscle memory, and reduce variable performance across trainees.

Preliminary Design Evaluation

Table 1 presents the design matrix that was used to evaluate each of the design concepts against the criteria outlined in the product design specifications (PDS) in Appendix A. The highest-weighted criterion is functionality, defined as the device's ability to provide objective, real-time feedback on suture tension. The tool must provide feedback during the suturing process so that students can learn to recognize when proper tension has been achieved.

The second most important criterion is integration with the suturing workflow that students are currently taught. The device should simulate natural suture technique while providing feedback; if it interferes with existing methods or introduces non-standard movements movements, it risks reinforcing improper technique post-training

Durability is also a key consideration as the training tool will undergo repeated use by trainees. Additionally, because sutures vary in diameter and material, the design must be adaptable or able to be calibrated to accommodate these differences.

Finally, feasibility and affordability are included to ensure the selected design can be realistically developed within the team's available resources, time, and budget over the course of the school year.

	Visual Knot Characterization		Force Sensing Glove		Location Sensor Design	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Functionality (25)	3	15	5	25	3	15
Workflow Integration (20)	5	20	2	8	3	12
Durability (15)	4	12	1	3	4	12
Adaptability (15)	4	12	4	12	2	6
Feasibility (15)	3	9	5	15	5	15
Affordability (10)	2	4	4	8	4	8
Total (100)		72		71		68

Table 1. Design Matrix for evaluation of measurement methods.

Evaluation: Visual Knot Characterization

The visual method of knot characterization scored highly for workflow integration because it relies only on an overhead camera, introducing no physical obstacles and preserving natural suturing technique. Its durability is ranked highly, as the design is not subject to direct mechanical loads. The design's adaptability benefits from consistent visual characteristics of secure versus insecure knots across different suture types, unlike mechanically based designs that require calibration. However, this concept scored lower in functionality because providing real-time feedback is more complex than in other designs; the system must analyze multiple frames per second with a machine learning algorithm and translate the results into usable feedback. Feasibility and affordability were also ranked lower due to the need for an expensive, high-definition camera and extensive image training to support machine learning, an area the team currently has limited experience in.

Evaluation: Force Sensing Glove

The force sensing glove received varied scores across multiple categories, showing its strengths and weaknesses in comparison to the other designs. The glove scored highest in functionality and feasibility because it has strong practical utility in capturing force data during use and offers a technically achievable and implementable design. This design also scored low in workflow integration and durability, two categories that are codependent. With integration of the glove, this option causes the most disruptions to the suturing process. Consequently, because it is

constantly under a mechanical load, it might not withstand repeated or prolonged use without degradation and sensor impact. Lastly, the glove received a relatively high score in both adaptability and affordability. This indicates the design can be used across different suture sizes and materials. Additionally, the components to create the design are inexpensive and widely available.

Evaluation: Location Sensor Design

The location sensor design scored highly in the feasibility category due to its use of displacement as a measurement parameter, which is comparatively simple and accessible to implement. It also received strong scores in affordability and durability because the location sensor and visual feedback circuitry are generally low-cost and expected to remain reliable under normal operating conditions. However, functionality was rated lower because grip slippage during knot tying could lead to inaccurate displacement measurements. In terms of workflow integration, potential issues may arise from the sensor's placement on the user's hand, which could interfere with natural hand movements. Additionally, the adaptability of this design is more limited than other options, since displacement thresholds must be predetermined through mechanical testing for each specific suture material and size.

Final Proposed Design

The visual knot characteristics design was selected as the primary proposed solution. It was chosen due to its close alignment with current analysis techniques, minimal interference with suturing techniques, and advantageous adaptability and durability compared to other designs. Conceptually, it is the most promising option. However, the complexity of training a machine learning model and the significant coding required make it less feasible at this stage. To balance promise with practicality, the team decided to split efforts: half of the members began work on building a database for image training and beginning preliminary work on the knot characteristics design to ensure its viability before committing full resources, while the other half developed a prototype of the force-sensing glove. The glove design is more affordable, simpler to fabricate, and ranks highly in terms of functionality because of its ability to provide real-time feedback using a force-sensitive resistor.

Concept Exploration for Force Sensing Glove

Suture MTS Testing

To define the material properties of different suture sizes and brands, the team performed tensile strength testing using a MTS Criterion Model C42.104 and a TW elite data acquisition program. For each suture type, a knot-pull test was conducted to replicate the mechanical conditions experienced during the final throw in a square knot. The test set-up is represented clearly in Figure 18 below. For each test, 4 square knots were tied and positioned in the center of the load cells to ensure symmetrical loading [24].

The samples were then tested at a strain rate of 50 mm/min using a low capacity load cell (≤100 N) to ensure an accurate measurement at low tensions typical for suture materials. Three samples (Maxon 0, Wego 3-0, and Monocryl 0) were tested to gather mechanical data on a range of suture sizes, materials and manufacturers.
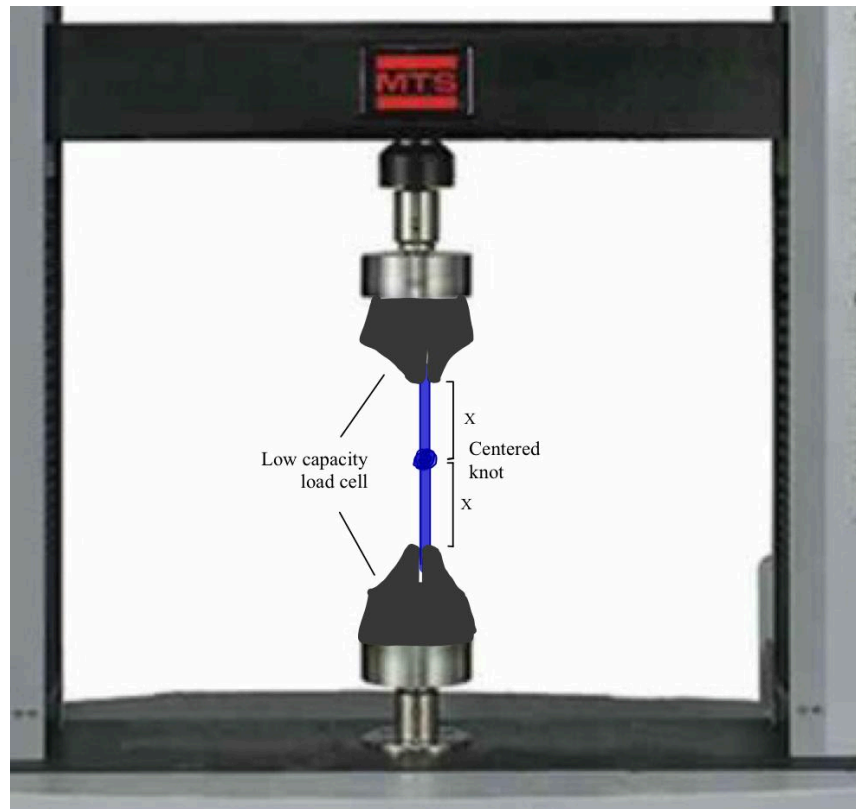
Figure 18. MTS Tensile Strength Test Set-Up [25].

After data collection, the raw force-displacement data was exported in text file format for post-processing in MATLAB. With this data, stress-strain and force-displacement graphs were generated to identify key mechanical properties, including Young's modulus, maximum strain, fracture stress, and yield stress. Protocol and collected results are documented in Appendix B. These suture measurements were used to better understand how suture size and material influence mechanical properties under tensile loading.

Fabrication

To fabricate the force-sensing glove concept established in the preliminary design phase, the team utilized a Force Sensing Resistor (FSR), an Arduino IDE microcontroller, and supporting electrical components. The FSR acquired for this prototyping did not include specific resistor value for integration into a voltage divider circuit. Therefore, testing was required to determine the optimal resistor pairing. A schematic of the voltage divider circuit is shown in

Figure 19. Several resistor values were tested, including 4.7 kOhm, 10 kOhm, and 47 kOhm, to assess sensor response under different pressure conditions. Lower resistor values selected to improve sensitivity at higher pressure ranges, which are anticipated when tying thicker suture materials. The 47 kOhm resistor was selected as the best resistor pair with the FSR.
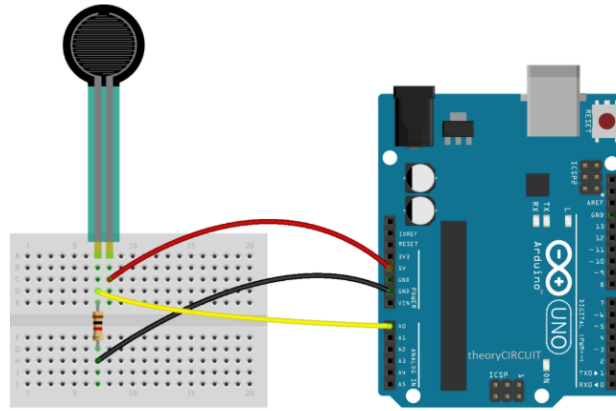
Figure 19. Voltage divider circuit schematic incorporating FSR used to determine correct resistor values. [26].

The circuit schematic shown in Figure 20 integrates an FSR, a fixed 47 kOhm resistor, an Arduino microcontrolled, and an LED to measure force and provide visualized feedback. The fixed 47 kOhm resistor and FSR are configured as a voltage divider, where the output voltage varies depending on the resistance of the FSR, which changes based on the applied pressure. When pressure is applied to the FSR, its resistance decreases. This change alters the voltage read at the Arduino's input analog pin A0. The Arduino converts the analog violated into a corresponding ADC value ranging from 0-1023, correlating applied pressure with a measurable voltage value. The LED is connected to digital pin 9 which provides real-time visual feedback based on the measurement force. Therefore, when more force is applied, the LED brightness increases. This circuit was not fabricated into a glove prototype as the team decided to move forward with the knot classification model prototype.
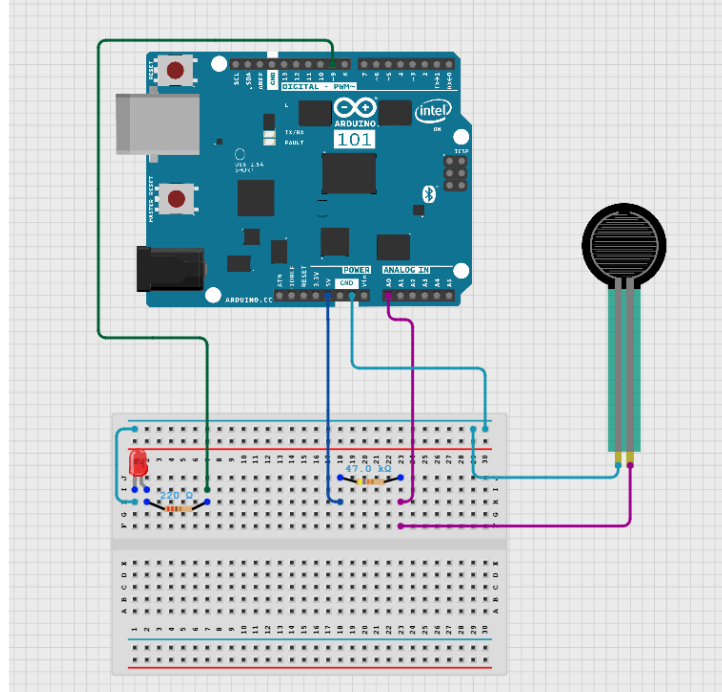
Figure 20. Final circuit schematic that incorporates FSR pressure sensing and LED real-time feedback.

Calibration

For the calibration process, the FSR component of the circuit was placed on a digital scale, and a series of downward forces ranging from 0 to 600 g, the upper limit of the scale, were applied. At each force level, the corresponding ADC, voltage, and FSR resistance values were recorded from the Arduino Serial Monitor. The collected data was then plotted, and exponential curve fits were generated for the relationships between Force (g) and Resistance, as well as Force (g) and Voltage (V), as shown in Figures 21 and 22 below.
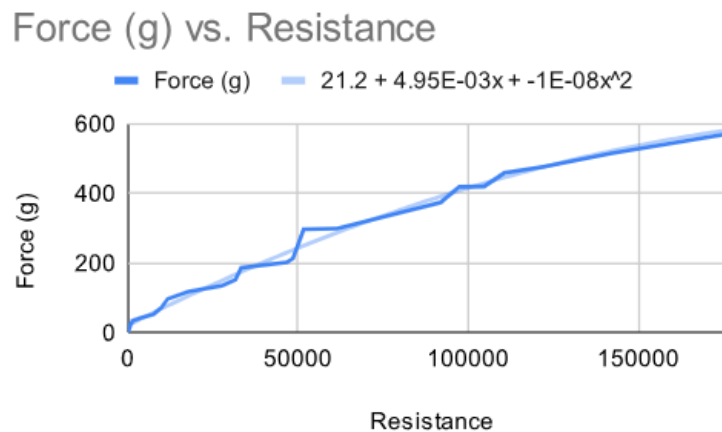
Figure 21. FSR calibration curve showing the relationship between force (g) and resistance using an exponential curve fit.
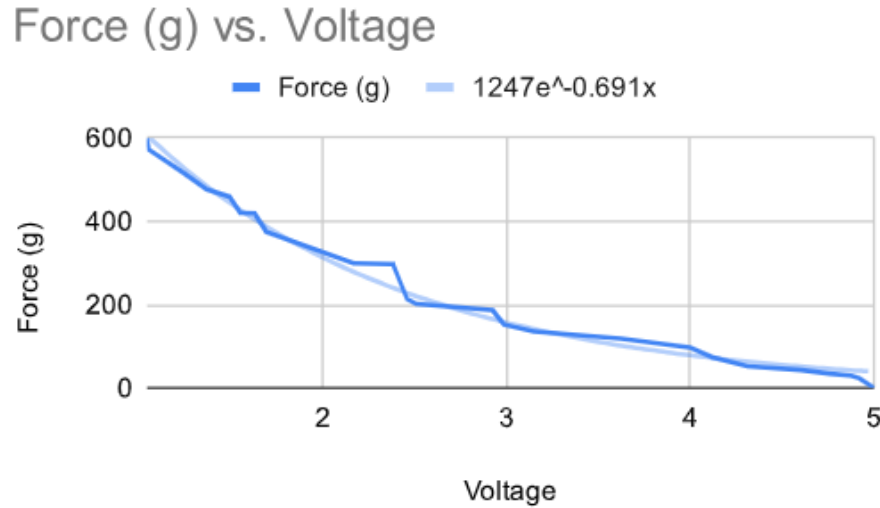
Figure 22. FSR calibration curve showing the relationship between force (g) and voltage using an exponential curve fit.

Ergonomics

The device chosen should accommodate for bias and inclusivity within its design. Namely, trainees' hand sizes, grip strength, and dexterity may affect how users interact with the device. The knot characteristics machine learning model will accommodate this due to limited human interaction with the device. The force sensor glove would need to be one size fits all or be commercially available in different sizes. Additionally, the sensor placements would need to vary depending on the size of the glove to address variabilities in the suture procedure due to hand size and dexterity. Another key consideration of the design would be novice versus advanced trainees. The designs need to be intuitive and not reliant upon expected force or speed values that are expressed by experienced surgeons. Next, the design must accommodate a multitude of training contexts. While the School of Veterinary Medicine at the University of Wisconsin - Madison hosts training in a clean, large, and resourceful room, to address widespread suture applications, the design should be adaptable to varying training environments. This involves utilizing low-cost materials, having offline functionality, and modular components for use in developing areas. Lastly, the interface should accompany a culturally neutral system that provides consistent feedback. It needs to avoid idiomatic language, confusing graphics, and assumptions about prior knowledge.

Rationale for Exclusion

A key limitation in the calibration process was the restricted force range imposed by the scale's 600 g maximum capacity. FSRs exhibit a nonlinear response, particularly at higher loads, so the calibration curve generated from this limited range will not accurately represent the sensor's behavior during actual use. As a result, the fitted exponential relationship becomes less reliable at higher force levels, which is important for ensuring that the suture knot is tightened

sufficiently. This limitation is especially significant for larger suture sizes, which require greater force to secure properly.

In the FSR glove, most of the sources of error involve the sensor itself and the integration into the suturing task.The FSR itself requires precise calibration, and without an accurate calibration curve, force readings become unreliable. Additionally, sensor placement will greatly vary the force output results. Incorrect sensor placement can fail to capture key pinch or pull forces or falsely record pressure from non-relevant movements. Optimal sensor placement may vary between users, hand size, and technique. Lastly, movement artifacts from unrelated hand movements can introduce noise or unintended pressure in dynamic suturing tasks.

Beyond these circuitry-related issues, measured force from the FSR is not directly proportional to actual suture tension due to the direction of the applied force differing from the direction of suture tension [27]. Combined with the high variability in suturing technique, these limitations prevent reliable assessment of knot tightness from FSR data alone. As a result, efforts were shifted from a force-based method to the machine learning model.

Model Fabrication

Materials

Machine learning models heavily reflect the data that they are trained on. In order to calibrate the image classification model thoroughly, a large portion of the team's effort went into knot tying and image dataset compilation. The veterinary school provided the group with suture varying from size 1 (largest) to size 3-0 (smallest). The clients also provided four skin pads and three sets of suturing materials, including needle drivers, forceps, and surgical scissors. Due to a lack of consistent suture sizes provided, the team utilized all sizes and colors to train the model.

The team evaluated camera options to ensure sufficient resolution for capturing clear images of knots for training. A digital camera was tested for preliminary data collection, but offered no advantages to an IPhone camera and could not focus on singular knots. Therefore, an IPhone camera was used to capture each of the knot images, top and side. Implementation of a real camera is anticipated and outlined in the future work section of this report.

Some image classification frameworks are provided within Python deep-learning libraries such as PyTorch and TensorFlow/Keras. Two well-known convolutional neural network (CNN) architectures included in these frameworks are VGG (Visual Geometry Group) and ResNet (Residual Network). Both VGG and ResNet are CNNs, meaning that they are designed to process data in grid-like structures (such as images) using convolutional layers that extract hierarchical features. As an image passes through successive convolutional layers, its spatial dimensions typically decrease while the number of feature channels increases when pooling convolution layers are applied. The VGG (e.g. VGG16 and VGG19) rely on stacking multiple 3x3 convolutional layers in sequence, usually followed by max-pooling layers. This design is conceptually simple, but deep. Advantages of VGG include its simplicity, uniform layer design,

and strong performance on classification tasks. This is outweighed, however, by having a large number of parameters that result in high storage and computational expense. ResNet introduces residual (skip) connections that allow the network to learn identity mappings and enable the training of extremely deep architectures (e.g. ResNet50, ResNet101, ResNet152) without vanishing gradients. ResNet also replaces some of the fully connected layers at the end of the network with a global average pooling layer to reduce parameters and mitigate overfitting. One downside to ResNet applications is the hypersensitivity of hyperparameters like learning rates and batch normalization, so the model requires fine tuning for optimal perform[28] [28].

Lastly, the RoboFlow platform was used to prototype a model that is more user-friendly and intuitive to new programmers. The RoboFlow system uses the same ResNet50 backbone that the Python models, but limits the user's ability to fine tune the models.

Methods

The initial phase of model prototyping began with building a dataset of square knot suture images. After tying the knots, the team met with the client to classify them as "Tight" or "Loose" and organized them into training folders accordingly. Initially, the images were taken from the top view of the knots, depicted in Figure 23. The client expressed her preference of capturing knots from the side for better identification capabilities, so the team added side images into the batch as well.

Figure 23. Two loose knots depicted from the top view (left) and side view (right). A mixture of these two datasets was used to train and differentiate the five models presented below.

In order to allow for thorough evaluation of model efficacy, the team split into groups and trained five different models. These models are differentiated by their training data and coding backbones. Maddie trained two models utilizing Python's ResNet and VGG capabilities trained on side and top images combined. Presley trained two models using side images only. Lastly, Sadie trained one model on RoboFlow using both side and top images. Both side and top images were used for the fifth model because the group hypothesized that this would be the most accurate training data to classify knots because it accounts for the most variation in knot tying angles. These distinctions are outlined below in Table 2.

Model Number	1	2	3	4	5
Training Data	Side and Top	Side and Top	Side Only	Side Only	Side and Top
Model Type	ResNet	VGG	ResNet	VGG	ResNet
Layers	50	16	50	16	50
Dataset Size	424	424	309	309	424
Tight : Loose	1.24	1.24	1.22	1.22	1.24

Table 2. Models 1, 2, 3, 4, and 5 specifications. Model 1 is a ResNet model trained on 424 top and side knot images. Model 2 is a VGG model trained on 424 top and side knot images. Model 3 is a ResNet model trained on 309 side knot images. Model 4 is a VGG model trained on 309 side knot images. Lastly, Model 5 is a ResNet model trained on 424 side and top images.

The ratios in Table 2 highlight the number of tight images to loose images. Ideally, this ratio would be a 1, representing that an equal number of tight and loose images were provided for the model to train. An equal proportion of tight and loose images ensures that minimal bias is introduced to the model.

Image classification model training followed the diagram shown in Figure 24. First, the full, pre-classified training dataset was parsed through by image cleaning code. This code was responsible for checking the file types of the input data. If any images were in an incompatible format, the code would output the number and names of the images being rejected. Image cleaning also included duplicate removal via mean squared error comparison and dataset size comparisons. Next, image augmentation transpired. In this stage, the images were augmented via random rotations, grey scaling, cropping, blurring, image saturation and brightness adjustments, and image flipping. In the case of the first four models, the classification with the least amount of images (loose) was augmented until it matched the dataset size of the tight dataset. Image augmentation was followed by training, validation, and testing splitting. For these models, the data was split based on ratios, so 80% of the images went to training the model, and 10% went to validation and testing each. These values were chosen because it is critical to ensure that the training set is diverse and representative of the data to prevent overfitting. All three sets of data were provided to the validation and training code that fit the ResNet and VGG models to the data input. Using the model's internal testing, the team acquired a classification report that outlined the required metrics to do external testing on all five of our models.
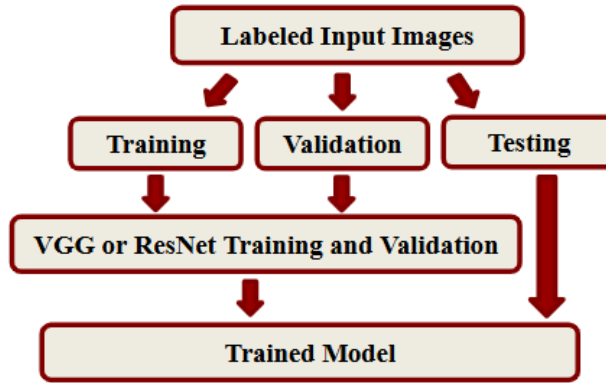
Figure 24. Diagram of the image classification model training process. The process begins with the input of pre-classified training data, which is then split into three sets, training, validation, and testing. Training and validation data is passed through a ResNet or VGG model framework and tested internally with the unused set of testing data.

Lastly, the RoboFlow model, Model 5, utilized top and side training data. This is an open source image classification site that allows for cropping and classification of images. After initial fabrication, the cropping mechanism of RoboFlow was not properly identifying the top knot. For this model, the images were input onto the software and augmented from 424 images to 730 images. The augmentations included horizontal and vertical flipping, rotation, shearing, grey scale, hue changes, and saturation. Next, the images are split in the same fashion as the other four models and internal testing of the model took place.

Model Testing and Results

Testing Plan

To ensure fair evaluation of all five models, a new test dataset of 80 knot images, 40 tight and 40 loose, was created for consistent analysis. This number was chosen due to the availability of suture materials. All images were pre-processed using the same procedure as the original dataset. Each image was cropped to a uniform size, and it was verified that each image contained only one knot. The dataset was then labeled by sorting the images into "tight" and "loose" folders.

Because the models were developed across different platforms, separate scripts were written for each model type to load the same image set and compute identical performance metrics including accuracy, precision, recall, and F1 score. All testing code and detailed procedures are documented in Appendix C. Each trained model was then used to make predictions on the 80 images by looping through the dataset and recording the true label and predicted label for every image.

Using scikit-learn from the Python library, two confusion matrices were generated for each model, one treating tight knots as the positive class and one treating loose knots as the positive class. A confusion matrix displays the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), as shown in Figure 25. Evaluating both labels as a positive condition provides a more complete view of model performance. This approach reveals whether a model is particularly strong at identifying tight knots but weaker at identifying loose knots, or the opposite. Ideally, the model should perform consistently well with both classes.

	Actually Positive	Actually Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Figure 25. Confusion matrix example. True positives and true negatives represent correct classifications, while false positives and false negatives represent incorrect classifications.

Following creation of a confusion matrix for each positive case, additional performance metrics were calculated to draw more meaningful conclusions from the data. The equations in Figure 26 show how each metric was calculated.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Figure 26. Equations used to calculate performance metrics, including accuracy, precision, recall, and F1-Score. Variables include true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) [29].

Accuracy is the proportion of all predictions that the model classifies correctly and is consistent regardless of which class is labeled as positive, since it is calculated by summing all true classifications and dividing by the total number of inputs. Precision, recall, and F1 score, on the other hand, depend on the choice of positive class. Precision, or positive prediction value, measures the proportion of predicted positives that are actually correct, which provides insight into the quality of positive predictions. High precision means when the model predicts "positive," it is usually correct. Recall, or sensitivity, measures the proportion of actual positives that the model correctly identifies. This provides insight into the coverage of the actual positives. High recall means that the model catches most of the true positives. The F1 score combines precision and recall into a single metric that balances both, achieving a high value only when both precision and recall are high.

Testing Results

Figure 27 below displays the confusion matrix for each of the five models tested, with a tight knot designated as the positive condition.

	Actually Tight	Actually Loose	Model 1: TensorFlow ResNET (Side & Top)		Model 2: TensorFlow VGG (Side & Top)		Model 3: TensorFlow ResNET (Side Only)		Model 4: TensorFlow VGG (Side Only)		Model 5: Roboflow ResNET (Side & Top)	
Predicted Tight	True Positive (TP)	False Positive (FP)	34	4	36	11	36	15	40	20	30	4
Predicted Loose	False Negative (FN)	True Negative (TN)	6	36	4	29	4	25	0	20	10	36

Figure 27. Confusion matrix for models 1–5 *(left to right)* with tight knots as the positive condition. A false positive is highlighted in red because it represents the worst-case scenario. This occurs when a loose knot is predicted as tight. In this situation, students would believe the knot is correctly tensioned when it is not, which leads to poor training outcomes. False negatives are highlighted in yellow and represent tight knots predicted as loose. Although not ideal, this is less concerning because a knot predicted as loose encourages students to continue tightening it.

Similarly, Figure 28 below shows the confusion matrix for each of the five models, this time with a loose knot designated as the positive condition. The values remain the same, but the classifications of true positives, false positives, false negatives, and true negatives change according to the positive class. These differences account for the variations in performance metrics (accuracy, recall, and F1) between the two positive cases.

	Actually Loose	Actually Tight	Model 1: TensorFlow ResNET (Side & Top)		Model 2: TensorFlow VGG (Side & Top)		Model 3: TensorFlow ResNET (Side Only)		Model 4: TensorFlow VGG (Side Only)		Model 5: Roboflow ResNET (Side & Top)	
Predicted Loose	True Positive (TP)	False Positive (FP)	36	6	29	4	25	4	20	0	36	10
Predicted Tight	False Negative (FN)	True Negative (TN)	4	34	11	36	15	36	20	40	4	30

Figure 28. Confusion matrix for models 1–5 *(left to right)* with <u>loose knots as the positive condition</u>. A false negative is highlighted in red because it represents the worst-case scenario: a loose knot is predicted as tight. False negatives are highlighted in yellow and represent tight knots predicted as loose.

As described in the testing plan above, performance metrics for each model were calculated using the values from their respective confusion matrices. The raw data used in these calculations is provided in Appendix D, where the highest score for each metric across all five models is highlighted in green. Graphical summaries of these metrics are presented in Figures 29-32 below.

Model 1 is a TensorFlow-based ResNet model trained on images taken from both side and top views of the knots. Model 2 is a TensorFlow-based VGG model trained on the same combination of side and top views. Model 3 is a TensorFlow-based ResNet model trained solely on side-view images. Model 4 is a TensorFlow-based VGG model trained solely on side-view images. Finally, Model 5 is a ResNet model created in RoboFlow and trained on side- and top-view images.
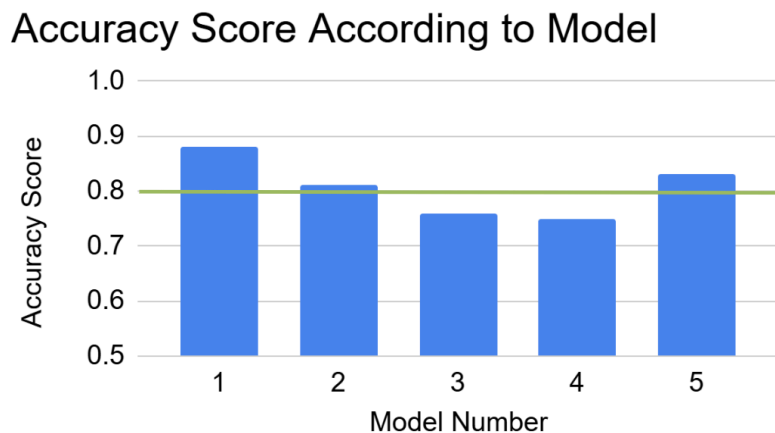

Figure 29. Accuracy scores for each model. Accuracy represents the proportion of all predictions that were correct and is the same regardless of which class is labeled as positive.

The green line in Figure 29 represents the desired 80% accuracy threshold defined in the PDS, which is a reasonable initial benchmark for the first pass of a machine learning model. The graph shows that Models 1, 2, and 5 met this requirement, with Model 1 achieving the highest overall accuracy at 88%. Models 3 and 4, on the other hand, fell below the threshold and will not be considered for further refinement.
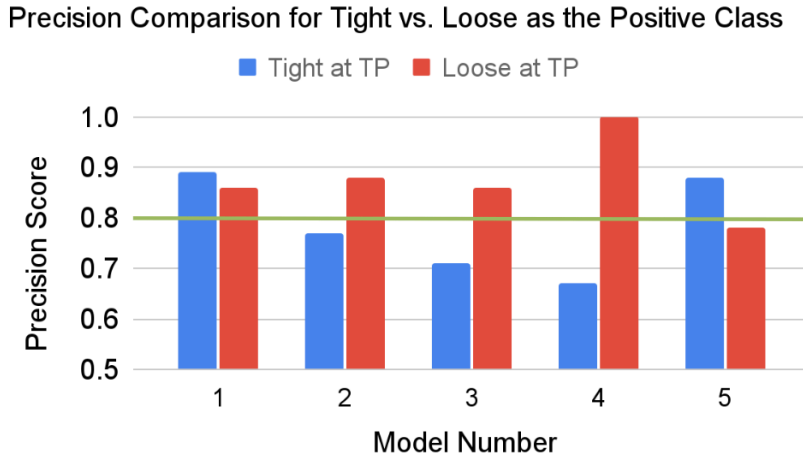
Figure 30. Precision scores for each model. Precision is defined as the proportion of predicted positive cases that were actually positive. *As indicated by the legend, the blue bars represent the precision when tight knots are designated as the positive class, and the red bars represent the precision when loose knots are designated as the positive class.*

Precision offers a more detailed view because it differentiates performance depending on whether tight or loose knots are defined as the positive condition. As outlined in the PDS, a minimum precision score of 0.8 was required, which is represented by the green line in Figure 30. When tight is the positive condition (shown in blue), precision represents the proportion of predicted tight knots that were actually tight compared to those that were loose but predicted as tight. Misclassifying a loose knot as tight is the project's worst case scenario, which means that high precision for the blue bars is especially important. Models 1 and 5 both exceed the 0.8 precision threshold for tight knots; however, only Model 1 achieves this level of precision for both tight and loose knots. This indicates that Model 1 is generally well balanced in its ability to predict both categories with similar precision. Models 2, 3, and 4 are noticeably stronger at predicting loose knots than tight knots while Model 5 shows the opposite pattern.
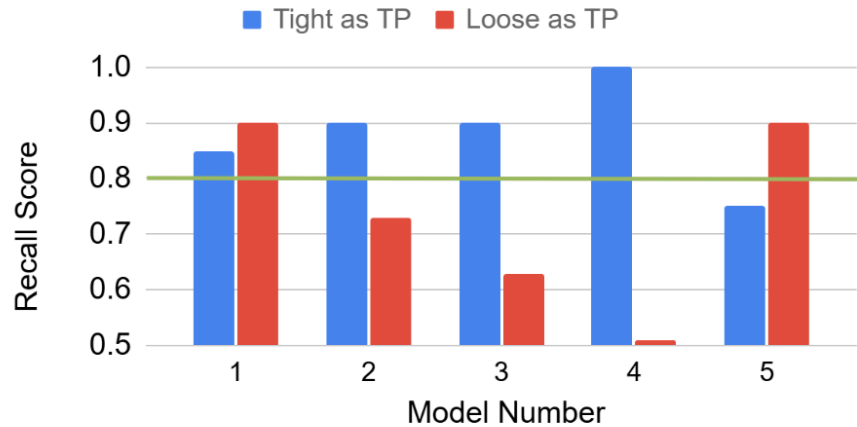
Figure 31. Recall scores for each model. Recall is defined as the proportion of all actual positive cases that were correctly identified. *As indicated by the legend, the blue bars represent the recall when tight knots are designated as the positive class, and the red bars represent the recall when loose knots are designated as the positive class.*

Recall also differentiates performance depending on which class is defined as positive. As outlined in the PDS, a minimum recall score of 0.8 was required, represented by the green line in Figure 31. When loose is the positive condition (shown in red), recall represents the proportion of actual loose knots that were correctly identified compared to those that were loose but predicted as tight. Because this encompasses the project's worst case scenario, high recall for the red bars is especially important in this graph. Again, Models 1 and 5 both exceed the 0.8 recall threshold for loose knots; however, only Model 1 achieves this level of recall for both positive conditions. This indicates that Model 1 is well balanced in its ability to correctly identify both classes of knots. Models 2, 3, and 4 tend to identify tight knots more reliably than loose knots, leading to frequent misclassification of loose knots as tight. Model 5 shows the opposite pattern.

Figure 32. F1 scores for each model. The F1 score combines precision and recall into a single balanced metric. *As indicated by the legend, the blue bars represent the F1 score when tight knots are designated as the positive class, and the red bars represent the F1 score when loose knots are designated as the positive class.*

Finally, the F1 score provides a metric that balances both precision and recall, offering a more well-rounded view of the model's performance. From the graph above, in Figure 32, both Models 1 and 5 achieve an F1 score greater than 0.8 for both posit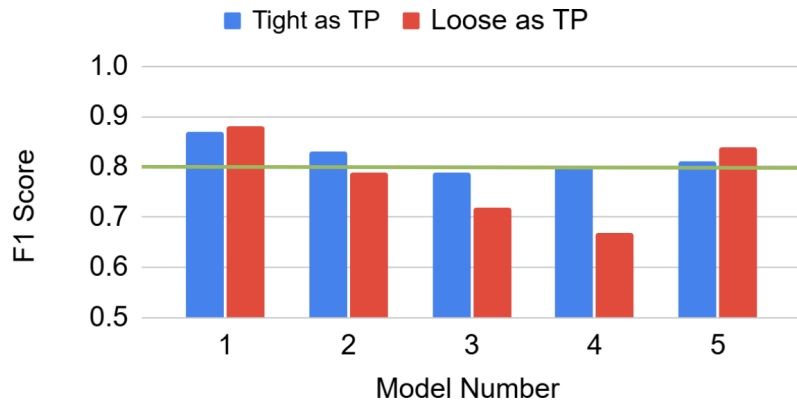ive conditions, indicating that they meet the product design specifications outlined at the beginning of the semester. However, Model 1 achieved higher F1 scores overall for both positive cases.

Given that Models 1 and 5 were the only models to meet PDS requirements, a McNemar's test was used to statistically compare their performance and determine whether the observed differences were significant. McNemar's test is used in situations when two classifiers are evaluated on the same set of instances. In this case, both models were tested on the same set of images, with each prediction classified as correct or incorrect relative to the ground truth labels [30]. Unlike other tests that require repeated sampling, McNemar's test focuses solely on the instances where the two models disagree, making it suitable for this shared testing set. Using the prediction correctness from Models 1 and 5, a 2x2 contingency table was constructed. This table is shown below in Table 3.

	Model 5 Correct (0)	Model 5 Incorrect (1)
Model 1 Correct (0)	66	4
Model 1 Incorrect (1)	0	10

Table 3. McNemar 2x2 contingency table. The cell *00* represents the number of images that both Model 1 and Model 5 classified correctly, while cell *11* represents the number of images that both models classified incorrectly. Cell *01* counts the images where Model 1 was correct but Model 5 was incorrect, and cell *10* counts the images where Model 1 was incorrect but Model 5 was correct.

The test was implemented in Python using the 'statsmodels' library, as outlined in the testing procedure in Appendix E. The null hypothesis states that Models 1 and 5 have equal accuracy on the testing images. This resulting p-value was 0.125, which reflects the probability of observing the measured disagreement if both models truly have equal accuracy. Since the p-value is greater than 0.05, the null hypothesis cannot be rejected. Although Model 1 produced slightly more correct predictions than Model 5, this difference is not statistically significant and may be due to random chance in this sample of images. Because of this, additional testing with a larger dataset is recommended to more reliably assess differences in performance.

Discussion

The results of the model evaluation provided insight into the relative strengths and limitations of the five models in distinguishing between tight and loose knots. Across all metrics, Model 1 and Model 5 demonstrated the most balanced and reliable behavior. Based on this analysis, the team decided to continue to refine the TensorFlow ResNet 50 model trained on top and side images (Model 1) as well as the RoboFlow model (Model 5) because they both demonstrated strong results. Both models achieved the original design specifications of an accuracy and precision above 80% and an F1-score greater than 0.8 when the tight classification was the true positive case.

An important aspect was the interpretation of the confusion matrices under two different positive class definitions. When tight knots were treated as the true positive case, false positives represented the worst-case scenario because they identified loose knots incorrectly predicted as tight. This misclassification is particularly harmful for training applications because it might lead students to believe a poorly tied knot is acceptable. When the labeling convention was reversed, a false negative was then the worse-case scenario and identified loose knots as tight. While the assignment of the true positive case did not change numerical values in the confusion matrices, it helped show the asymmetries in model behavior observed in precision, recall and F1 scores. Based on precision asymmetry, both models 1 and 5 demonstrated discrepancies in their ability to identify tight vs loose knots. While both models achieved higher precision on tight knots, their performance decreased when predicting loose knots.

While model 1 produced slightly more correct knot predictions, according to McNemar's test, differences between model 1 and 5 are not statistically significant. This highlights a need to further develop both models on a larger image set before determining the most robust model.

Sources of Error

While the knot identification model underwent training that met the team's initial design requirements, there are potential sources of error that could impact its testing reliability and real-world performance. In the image set for training purposes, both variability in suture type and photo quality can lead to a misclassification of secure and loose knots. Suture type variability

includes differences in material, diameter, texture and color, all of which can alter the visual characteristics the model relies on for classification. Photo variability refers to background lightning, shadows, color, glare, and pixel count. When these elements are inconsistent across images, the model may inadvertently focus on irrelevant visual cues rather than the structural characteristics of the top suture knot, increasing the risk of false positives or false negatives.

Although image consistency is necessary to reduce misclassifications of knots, the team also has to be mindful of avoiding overfitting the model. Overfitting occurs when the model becomes too tailored to the training dataset, causing it to perform poorly when exposed to new or varied images. The team was conscious of finding a balance between image set variability and uniformity to create a robust and generalizable model. However, due to limited suture supplies, the team was unable to produce a fully uniform dataset with equal representation of each suture material across all knot classifications.

Additionally, as the team explored different knot views to input into the model, we did not have access to a camera stand to capture perfectly consistent camera angles. For both RoboFlow and TensorFlow models to successfully classify the knot, it must focus on the top knot of a suture. If the model were to identify other locations on the skin pad or down the suture line, misclassification becomes more likely.

Finally, in a future user-friendly system, latency in providing feedback could confuse trainees and provide incorrect associations between their actions and feedback received. Ensuring responsive feedback will therefore be an important consideration as the team begins developing the system.

Limitations

As the team begins to further improve and refine the model, the team might run into some software and hardware limitations. Roboflow is a user-friendly platform that simplifies model development, but it also restricts the number of parameters the user can adjust and edit. As a result, the roboflow model might be more difficult to fine-tune beyond its default settings. Additionally, the team currently relies on personal devices with limited computing power. Training models on a larger dataset may exceed the processing capacity of our personal computers, requiring access to more powerful hardware.

Ethical Considerations

Although the device will not be incorporated into patient healthcare settings, ethical concerns can arise in how the device is marketed, adopted, and relied upon in training environments. Firstly, if the device gives inaccurate feedback, it can reinforce poor technique. The training tool must undergo rigorous validation and calibration. To increase device accessibility and equity, the system should maintain a low-cost structure, low maintenance, open-source software, and proper licensing. To reduce the probability of trainees' dependence

upon the device for feedback, the solution should build judgment and encourage reflective practices, not just mechanical repetition.

Future Work

       The team plans to further improve and fine-tune both the TensorFlow ResNet50 Model trained on top and side images as well as the RoboFlow Model. The team hopes to elevate the model from good to excellent by improving its accuracy and precision to above 90%. This would involve meeting with a machine learning expert to assist the team in editing the parameters within the model such as adjusting the amount of layers and hidden layers. The team will then retrain the fine-tuned model on a larger and more consistent dataset. This database will include images captured using controlled camera angles, a digital camera, and a standardized photo environment to ensure uniform images. Additionally, the team will photograph equal numbers of tight and loose knots across all different types of suture materials. These improvements will hopefully reduce discrepancies between tight and loose identifications and improve the overall robustness of the models.

       Once additional testing is conducted on the improved models using a larger testing set, the model that demonstrates the highest precision, accuracy and F1-score with the fewest discrepancies in successful identifications between tight and loose knots will be integrated into a user-friendly system to provide real-time feedback. This system will incorporate a camera stand designed to hold a digital camera and a stationary skin pad, minimizing variability in angle, lighting and background. A proposed camera stand design is shown in Figure 33.
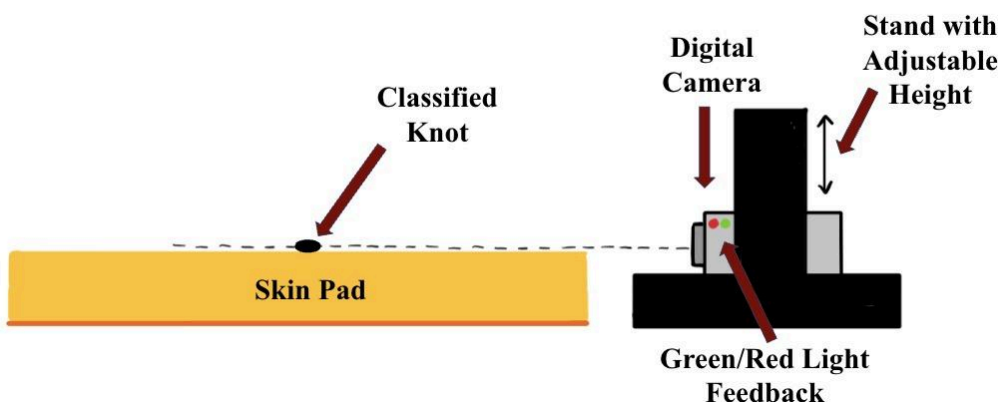
Figure 33. Potential camera stand design incorporating a digital camera aligned with the side view of knots on a skin pad. The design includes a green and red light feedback system and an option to adjust camera height.

The team is currently considering multiple options for incorporating a real-time feedback system that provides the user with immediate identification of tight or loose knots. Potential platforms including a mobile application, website, and microcontroller-based systems. The final choice will be based on feasibility, usability, and overall effectiveness. Depending on the selected platform, the knot feedback could involve visual cues (such as lights or word prompts), auditory signals, or verbal cues.

Conclusion

In conclusion, the proposed suturing training device addresses a critical gap in veterinary education by providing real-time, objective feedback on suture tension. This skill is essential for surgical success but difficult for novices to master. By utilizing optical analysis through a machine learning model, the proposed designs offer versatile approaches to skill development that mirror instructor assessments while reducing subjectivity and feedback delays. These innovations have the potential to improve learning efficiency, reduce material waste, and promote animal welfare by ensuring students develop proper technique before operating on live patients. While technical and logistical challenges remain, continued development and validation of these initiatives could significantly enhance the quality and consistency of surgical training in veterinary programs

References

[1] "What Is Wound Dehiscence?," Cleveland Clinic. Accessed: Oct. 07, 2025. [Online]. Available: https://my.clevelandclinic.org/health/diseases/wound-dehiscence

[2] "The ideal surgical knot (according to science)." Accessed: Oct. 07, 2025. [Online]. Available: https://healthcare-in-europe.com/en/news/ideal-surgical-knot.html

[3] W. Y. Rung, "Biomechanical Investigation of Different Tying Forces on Knot Security," *Biomed. J. Sci. Tech. Res.*, vol. 17, no. 2, Apr. 2019, doi: 10.26717/BJSTR.2019.17.002985.

[4] S. Neuhofer, K. Wieser, G. Lajtai, D. Müller, C. Gerber, and D. C. Meyer, "Surgical knot tightening: how much pull is necessary?," *Knee Surg. Sports Traumatol. Arthrosc. Off. J. ESSKA*, vol. 22, no. 11, pp. 2849–2855, Nov. 2014, doi: 10.1007/s00167-013-2452-9.

[5] M. C. Jordan, S. Boelch, H. Jansen, R. H. Meffert, and S. Hoelscher-Doht, "Does plastic suture deformation induce gapping after tendon repair? A biomechanical comparison of different suture materials," *J. Biomech.*, vol. 49, no. 13, pp. 2607–2612, Sept. 2016, doi: 10.1016/j.jbiomech.2016.05.023.

[6] R. K. Elmallah *et al.*, "Economic evaluation of different suture closure methods: barbed versus traditional interrupted sutures," *Ann. Transl. Med.*, vol. 5, no. Suppl 3, pp. S26–S26, Dec. 2017, doi: 10.21037/atm.2017.08.21.

[7] "Visual Force Feedback Improves Knot-Tying Security - ClinicalKey." Accessed: Oct. 07, 2025. [Online]. Available: https://www-clinicalkey-com.ezproxy.library.wisc.edu/#!/content/playContent/1-s2.0-S1931720413001852?returnurl=null&referrer=null

[8] MediShield BV, *ForceTRAP training system for open suture tasks*, (Oct. 16, 2014). Accessed: Oct. 07, 2025. [Online Video]. Available: https://www.youtube.com/watch?v=w3Bi8t8SqYg

[9] T. Horeman, E. Meijer, J. J. Harlaar, J. F. Lange, J. J. van den Dobbelsteen, and J. Dankelman, "Force Sensing in Surgical Sutures," *PLOS ONE*, vol. 8, no. 12, p. e84466, Dec. 2013, doi: 10.1371/journal.pone.0084466.

[10] T. Horeman, E. Meijer, J. J. Harlaar, J. F. Lange, J. J. van den Dobbelsteen, and J. Dankelman, "Force Sensing in Surgical Sutures," *PLOS ONE*, vol. 8, no. 12, p. e84466, Dec. 2013, doi: 10.1371/journal.pone.0084466.

[11] "Basic Knot | Surgery." Accessed: Oct. 08, 2025. [Online]. Available: https://www.bumc.bu.edu/surgery/training/technical-training/basic-knot/

[12] "FIGURE 1. Examples of a square knot (A) and a slip knot (B) constructed...," ResearchGate. Accessed: Oct. 08, 2025. [Online]. Available: https://www.researchgate.net/figure/Examples-of-a-square-knot-A-and-a-slip-knot-B-constructed-with-Prolene-2-0-suture-as_fig1_234823099

[13] "Simple continuous suture: Exploring variations and their clinical benefits." Accessed: Oct. 08, 2025. [Online]. Available: https://medicosutures.com/simple-continuous-suture/

[14] "Maxon - an overview | ScienceDirect Topics." Accessed: Oct. 08, 2025. [Online]. Available: https://www.sciencedirect.com/topics/medicine-and-dentistry/maxon

[15] "Absorbable Suture with Needle Biosyn™ Polyester - McKesson." Accessed: Oct. 08, 2025. [Online]. Available: https://mms.mckesson.com/product/533503/Covidien-SM-693

[16] "DSI PDS Polydioxanone Sutures | DSI Dental Solutions | 2023." Accessed: Oct. 08, 2025. [Online]. Available: https://www.dsisrael.com/surgical/pds

[17]   S. Thongchamrat, "World Small Animal Veterinary Association World Congress Proceedings, 2015," *VIN.com*, Mar. 2016, [Online]. Available: https://www.vin.com/doc/?id=7259440

[18]   O. M. Education, "Suture sizes and suggested indications for their use," Oxford Medical Education. Accessed: Oct. 08, 2025. [Online]. Available: https://oxfordmedicaleducation.com/surgery/suture-sizes-and-suggested-indications-for-their-use/

[19]   "Nonabsorbable Sutures Nylon Blue Monofilament With Needle Medical Disposable Pekan Nanas, Johor, Malaysia Supplier, Supplies, Supply," Zipper Resources Sdn. Bhd. Accessed: Oct. 08, 2025. [Online]. Available: https://www.zipperresources.com/showproducts/productid/5036537/cid/524640/nonabsorbable-sutures-nylon-blue-monofilament-with-needle/

[20]   "F1 Score in Machine Learning Explained | Encord." Accessed: Dec. 09, 2025. [Online]. Available: https://encord.com/blog/f1-score-in-machine-learning/

[21]   N. Marsidi, S. A. M. Vermeulen, T. Horeman, and R. E. Genders, "Measuring Forces in Suture Techniques for Wound Closure," *J. Surg. Res.*, vol. 255, pp. 135–143, Nov. 2020, doi: 10.1016/j.jss.2020.05.033.

[22]   "IEC 62368-1: Ask the Engineers, Question-and-Answer Page," UL Solutions. Accessed: Sept. 18, 2025. [Online]. Available: https://www.ul.com/resources/iec-62368-1-ask-engineers-question-and-answer-page

[23]   "IEC 61010-1:2010." Accessed: Sept. 16, 2025. [Online]. Available: https://webstore.iec.ch/en/publication/4279

[24]   D. Yalcin, "Suture Tensile Strength Testing," ADMET. Accessed: Oct. 08, 2025. [Online]. Available: https://www.admet.com/blog/suture-tensile-strength-testing/

[25]   "MTS Criterion® Electromechanical Test Systems." Accessed: Oct. 08, 2025. [Online]. Available: https://www.mts.com/en/products/materials/static-materials-test-systems/criterion-electromechanical

[26]   T. Digital, "Interface Force Sensing Resistor (FSR) with Arduino," theoryCIRCUIT - The Online Community for Electronics and Circuit Design. Accessed: Dec. 09, 2025. [Online]. Available: https://theorycircuit.com/arduino-projects/interface-force-sensing-resistor-fsr-with-arduino/

[27]   K.-T. von Trotha *et al.*, "Surgical sutures: coincidence or experience?," *Hernia*, vol. 21, no. 4, pp. 505–508, Aug. 2017, doi: 10.1007/s10029-017-1597-8.

[28]   "(PDF) Comparative analysis of VGG, ResNet, and GoogLeNet architectures evaluating performance, computational efficiency, and convergence rates," *ResearchGate*, July 2025, Accessed: Dec. 01, 2025. [Online]. Available: https://www.researchgate.net/publication/378739119_Comparative_analysis_of_VGG_ResNet_and_GoogLeNet_architectures_evaluating_performance_computational_efficiency_and_convergence_rates

[29]   B. A. Newsletter, "F1-score, AUC-ROC, Precision, Recall," Business Analytics Review. Accessed: Oct. 08, 2025. [Online]. Available: https://businessanalytics.substack.com/p/f1-score-auc-roc-precision-recall

[30]   "9.6: McNemar's test," Statistics LibreTexts. Accessed: Dec. 07, 2025. [Online]. Available: https://stats.libretexts.org/Bookshelves/Applied_Statistics/Mikes_Biostatistics_Book_(Dohm

)/09%3A_Categorical_Data/9.6%3A_McNemar s_test cNemar's test," Statistics LibreTexts. Accessed: Dec. 07, 2025. [Online]. Available: https://stats.libretexts.org/Bookshelves/Applied_Statistics/Mikes_Biostatistics_Book_(Dohm)/09%3A_Categorical_Data/9.6%3A_McNemar s_test call

Appendices

Appendix A - Product Design Specifications
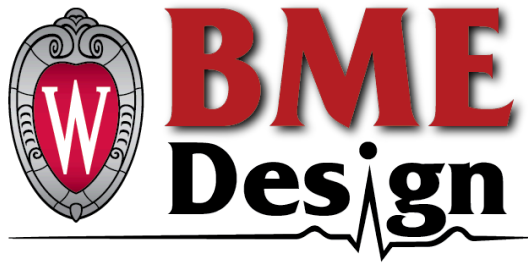
Knot too Tight - Not too Loose

PRODUCT DESIGN SPECIFICATIONS (PDS)

BME 400

Team Name:

The Knotorius Five

Team Members:

Madison Michels (Team Leader)

Lucy Hockerman (Communicator)

Kate Hiller (BSAC)

Sadie Rowe (BPAG)

Presley Hansen (BWIG)

Client:

Dr. Margene Anderson

Advisor:

Professor Walter Block; University of Wisconsin - Biomedical Engineering Department

September 18th, 2025

Function/Problem Statement:

In veterinary training, mastering the skill of applying appropriate suture tension is essential for successful wound closure and patient recovery, However, novice practitioners often struggle to judge the correct amount of force needed, leading to either insufficient tension or excessive tension, which can cause failure of the suture material or tissue damage. Currently, the evaluation of suture technique relies heavily upon subjective instructor feedback, lacking objective, real-time metrics to guide learners. This gap hinders consistent skill development and increases the risk of procedural errors. There is a critical need for a real-time suture tension measurement and feedback system to help students learn to apply optimal tension, prevent material or tissue compromise, and improve surgical outcomes through data-driven training.

Client requirements:

- Provide real-time feedback.
- Give veterinary students an indication when suture is pulled with correct tension and speed.
- Prefers the device to be minimally disruptive to the suturing process.
- Measure tension during the final throws of the square knots.

Force Sensor Design Requirements:

1. Physical and Operational Characteristics
   a. Performance Requirements
      i. The sensor must correctly measure the speed and tension of a suture.
      ii. The device must contain a display that identifies a range of force and speed required to plastically deform a suture in order to secure a square knot.
      iii. Sutures vary in both material and width to accommodate the different force required to close wounds on animals of various sizes, therefore, each suture type demands a unique force and speed to secure a square knot [1]. The device should allow for calibration to different sutures, allowing it to be used across multiple suture types.
   b. Safety

i. The device shall not interfere with the ability of the user to perform a suture technique. Interference can pose a safety risk as the user is handling many surgical tools, including sharp objects such as the suture needle.

ii. The device shall be made with a material that is smooth, disinfectant-resistant, and easy to clean. Ensure compliance with ISO10993 if in contact with live tissue.

iii. There shall be electrical safety in the device:

1. Voltage and current values shall be well below shock hazard thresholds per IEC 62368 [2] [3].

2. All exposed wire (if any) shall be insulated.

3. There shall be a current limiter to prevent device overheating.

4. Compliance to IEC 61010-1. This international standard outlines safety requirements for electrical equipment used in measurement, control, and laboratory applications [4].

5. If the device requires a charging component, the charger shall be compliant to IEC 62368-1 international safety standard for ICT and AV equipment [2].

iv. There shall be mechanical safety in the device:

1. There shall be a maximum force limit, the load applied shall not exceed safe values.

2. The device shall be stable to prevent tipping or moving unexpectedly during use. The electrical components shall be secured in or on the device.

3. The moving parts of the device shall be enclosed or guarded to ensure user safety.

c. Accuracy and Reliability

i. The tensioning device must withstand forces up to 30 N, with a working range of 0–20 N [5].

ii. The force sensor must have an accuracy of ±0.5 N to ensure measurements are close to the actual force within the working range [5], [6].

iii. The force sensor must have a sensitivity of 0.1 N to reliably detect small changes in tension, giving users realistic feedback for training purposes [6].

d. Shelf Life

i. The device will be used each semester in a training environment. During storage over break periods, it must maintain full functionality when returned to normal operating temperature and humidity conditions.

ii. Shelf life is not a priority constraint for this device.

e. Life in Service

i. The sensing element should withstand repeated use during training sessions. Additional discussions with the client will be conducted to determine the required number of tensioning cycles the device must endure without performance degradation, where each cycle consists of one application and release of tension in the suture.

ii. The device must withstand cleaning after each training session involving biological material, including a thorough wipe-down with a disinfectant solution, without loss of functionality.

f. Operating Environment

i. Non-electronic components must tolerate disinfectant wipes and incidental contact with biological material without structural or functional degradation.

ii. The device must maintain functionality under normal operating temperatures and humidity.

g. Ergonomics:

i. Users should be able to tie knots with their hands close to the practice pad, mimicking real surgical technique. The device must not cause awkward reaches or unnatural hand angles.

ii. The device should not noticeably alter the resistance or feel of knot tying.

iii. Real-time feedback should be intuitive and immediate, delivered through vibration, LED, or audio cues.

iv. The device should integrate smoothly with existing training methods so users do not have to adjust their typical working posture.

h. Size:

i. The sensing element should remain compact to ensure users can practice in a realistic, unobstructed space.

ii.　If used in future surgical use, the device must maintain a low-profile form that does not obstruct access to the wound site. The device should not be larger than a table-top device to maintain these requirements.

i.　Weight

   i.　The device should be lightweight to avoid fatigue or unreasonable movement during suturing practice.

   ii.　The device should feel unobtrusive in use, so users focus on knot tension rather than compensating for added weight.

j.　Materials

   i.　The device should be made of lightweight, rigid materials to maintain low weight and structural integrity.

   ii.　Internal electronics should be housed in a compact, lightweight, and protective enclosure to prevent damage.

   iii.　If used for future clinical applications, materials should be biocompatible and sterilizable.

k.　Aesthetics, Appearance, and Finish

   i.　The device should have a clean and professional appearance appropriate for training and potential clinical environments.

   ii.　No specific color, style, or finish is required, but finishes should support durability and user comfort.

   iii.　All surfaces should be smooth and free of sharp edges to ensure safe handling during practice or surgical use.

   iv.　Visual indicators, such as LEDs or labels, should be clearly visible without distracting from the suturing task.

2. Production Characteristics

   a.　Quantity

   i.　According to the client, one device is required for teaching purposes. After fabrication, testing, and approval from the client more devices could be created to upscale the number of students that can learn at a time during a lab section.

Additionally, this product could be applied to industries aside from veterinary applications, so more could be replicated for teaching in human applications.

b. Target Product Cost

    i. The client has provided $250 for the project budget with room for negotiation and increases as necessary. This cost, however, will include testing materials, sutures, and the design itself.

3. Miscellaneous

    a. Standards and Specifications

        i. This device shall comply with FDA standards as it is, by definition, a medical device as it is "intended to affect the structure or any function of the body or other animals…" [7] The device is utilized in conjunction with the user to complete suture knot tying, which affects the structure of the body.

        ii. ISO 14971:2019 Risk Management [6]

            1. Risk analysis through Failure Modes and Effects Analysis (FMEA) should be completed to identify potential risks for the patient, operator, and property. This includes gathering data and reviewing literature about the risks of similar medical devices. This standard states that the concept of risk involves the probability of the occurrence of harm and the severity of its consequences.

        iii. Code of Federal Regulations, Title 21, Chapter 1, Part 803 [9]

            1. Manufacturers and facilities that use the device must report deaths and serious injuries that the device has caused or contributed to through a Medical Device Report (MDR).

        iv. IEC 61010-1:2010 Safety Requirements for Electrical Equipment for Measurement, Control, and Laboratory Use [10]

            1. Surface temperature limits for protection against burns

            2. Protection against electrical shock

            3. Resistance to mechanical stress

            4. Protection against the spread of fire

            5. Protection against hazards from fluids

v. ISO 12100:2021 Safety of machinery - General principles for design - Risk assessment and risk reduction [11]

1. This standard establishes a strategy for risk assessment, identifying hazards, and risk reduction in machinery design. If the device incorporates any mechanical or machinery components, compliance with the standard is required.

vi. ASTM F2458-05:2024 Standard Test Method for Wound Closure Strength of Tissue Adhesives and Sealants [12]

1. This test method shall be used to evaluate the force values of a suture, allowing for the assessment of material properties and failure of the suture material. This test method is for comparing bonding processes for susceptibility to fatigue, mode of failure, and environmental changes.

vii. ISO 10993-1:2018 Biological Evaluation of Medical Devices [13]

1. The device shall be biocompatible as the device is in contact with live tissue (user and potentially patient).

2. Further biocompatibility testing shall be completed if the device is to be used during surgical procedures to ensure hemocompatibility.

viii. IEC 60601:2015 General requirements for medical electrical equipment [4]

1. This standard outlines the basic safety and essential performance of medical electrical equipment.

2. This standard includes identifying and testing the operating temperature of the device.

b. Customer

i. The initial device is intended for students at the University of Wisconsin-Madison School of Veterinary Medicine. The device will primarily be used in the curriculum of the first and second year students, as it is intended for beginners in suture techniques. However, the device has the potential to benefit students across broader areas of medicine, including medical schools, dental schools, and other healthcare training institutions.

c. Patient-Related Concerns

    i. If the device were to display inappropriate feedback for the correct tension or speed while securing a knot, students would learn incorrect suturing techniques that could translate to the quality of suture on live animals later in the student's career. This puts animals at risk for wound dehiscence after surgery if the square knots are not secured correctly.

    ii. Surgical wound dehiscence can lead to infection, excessive scarring, and necrosis on the wound site [14], all of which compromise the welfare of animals.

    iii. Additionally, poorly thrown sutures on cadavers limit their future use for training and increase the demand of additional animal cadavers in academic settings.

d. Economic Impact

    i. The cost of the suture varies depending on suture material, but the average cost of an individual monofilament absorbable suture is $1.75 - $1.83 per stitch [15]. Real-time feedback on tension and speed can decrease learning time and ultimately lower the amount of suture material needed for practice.

    ii. If improper tension is applied to sutures, stitches have a higher chance of breaking or unraveling. Depending on the size of the open wound, additional procedures might be necessary. Suture procedures cost can range from $50 - $1,000 to cover expenses for personnel, medicine, and surgical instruments during an operation [16].

e. Competition

    The use of real-time force analysis in surgical settings is scarce, but drawing recent attention due to its effectiveness in reducing heal time and improving aesthetics after incision. One notable research study conducted by T. Horseman et. al. developed three force sensor recording techniques that highlight unique approaches to installation, force detection, and functionality [6].

    i. Hook in Force (HIF) Sensor

        1. Shown in Figure 1, this sensor is a U-shaped device composed of two spring blades (feature C) and four plastic discs (feature A) that guide the suture material through the machine. Two discs are lined with silicone

(feature B) to prevent damage of the thread material. A magnet (feature E) and sensor (feature D) work in conjunction to measure the displacement of the device once the string is fed into the system. The design allows for a maximum detectable displacement of 3 millimeters (mm) and a minimum detectable displacement of 1 mm.
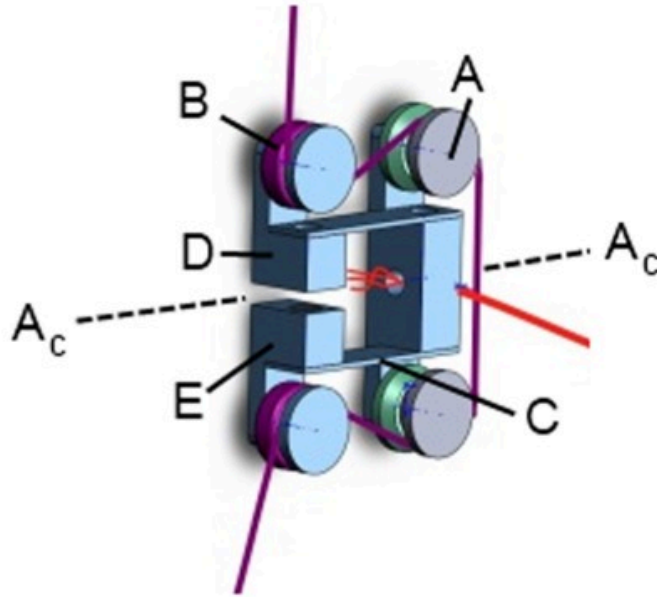
Figure 1. HIF sensor labeled diagram.

The free body diagram shown in Figure 2 displays the pulling nature of the system once a suture is threaded through. The spring blade counteracts the movement of the suture being pulled, leaving a displacement for the sensor and magnetic to measure.
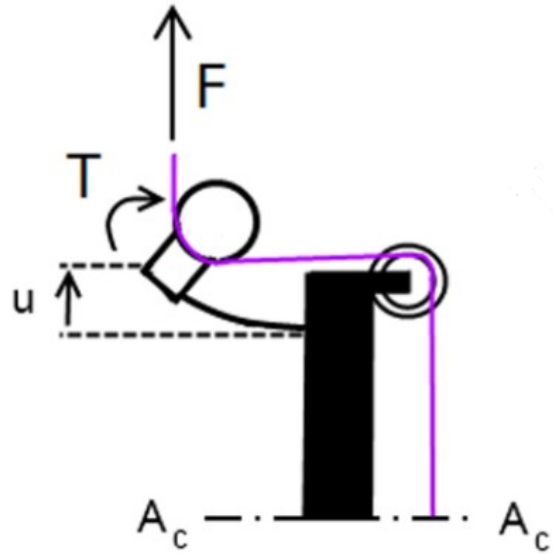
Figure 2. HIF Sensor Free Body Diagram.

ii. Stitch Force (SF) Sensor

    1. The SF sensor measures the forces required to close the tissue around a wound. It is in continuous contact with the skin and requires 2.5 mm of the incision to be positioned. Shown in Figure 3, the SF sensor includes a housing (feature A), spring blades oriented in a circle (feature B), a fissure tip (feature C), and a hall sensor and magnet (features E, G, and D). Once inserted, the thread creates a torque in the tip and relies upon the hall sensor and magnet to read the displacement of the spring blades. The resulting force is output in volts (V) [6].
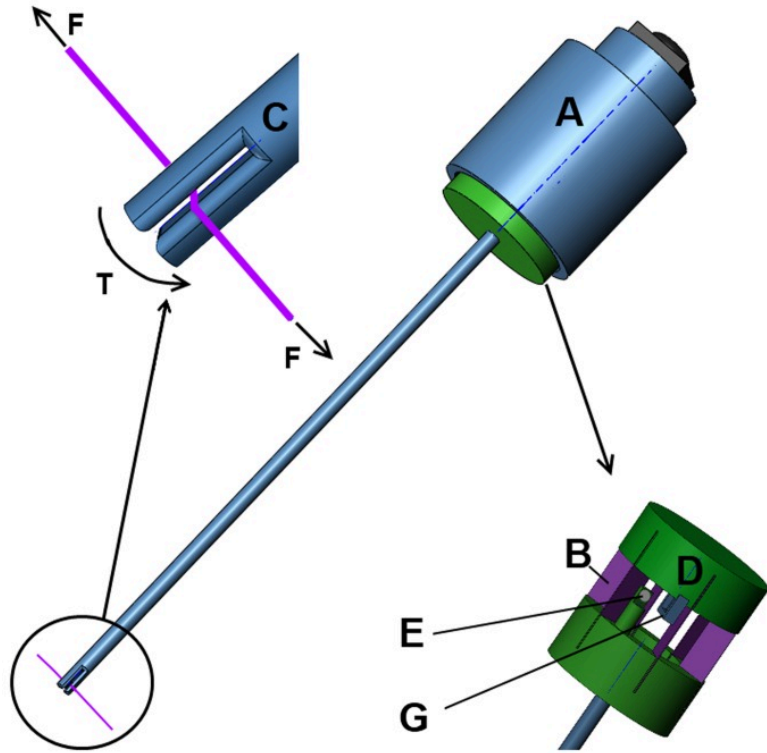
Figure 3. SF sensor labeled diagram and free body diagram.

iii.   Wheel Sensor

1. The wheel sensor was designed to address training needs for students and practitioners. The system is designed to be simple, small, lightweight, and cheap. The wheel was laser cut with medical grade plastic, and three metal pins were placed in through holes along the edge. The wheel supports itself between two tensioned threads. The pulling force can be related to the output signal of the hall sensor after calibration. An ATtiny85 micro controller controls the system and operates at 100 hz. The entire design weighs 11.3 grams (g), but can be reduced to 8 g if a custom circuit board is used. Two LED lights guide the accuracy of the trainee, green symbolizing a safe working range for the pulling force and red depicting that the surgeon has exceeded a predefined threshold [6].
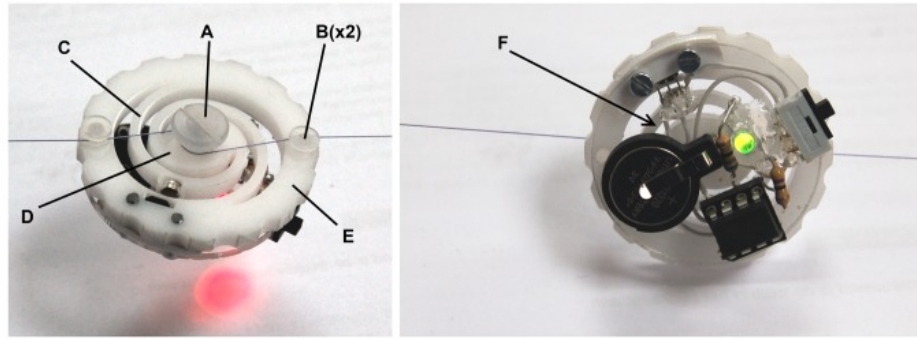
Figure 4. Wheel Sensor Diagram. A - inner pin, B - external pins, C - spiral-shaped bar, D - inner ring, E - external ring, F - embedded electronics for force feedback [6].

Machine Learning Model Design Requirements:

1. Physical and Operational Characteristics

    l.   Performance Requirements

        i.   The model must accurately predict a knot as "tight" or "loose."

        ii.   The device should allow for calibration or application to different sutures, allowing it to be used across multiple suture types.

    m. Safety

        i.   The device shall not interfere with the ability of the user to perform a suture technique. Interference can pose a safety risk as the user is handling many surgical tools, including sharp objects such as the suture needle.

        ii.   The device shall be made with a material that is smooth, disinfectant-resistant, and easy to clean. Ensure compliance with ISO10993 if in contact with live tissue.

        iii.   There shall be electrical safety in the device:

               1.   Voltage and current values shall be well below shock hazard thresholds per IEC 62368 [2] [3].

               2.   All exposed wire (if any) shall be insulated.

               3.   There shall be a current limiter to prevent device overheating.

               4.   Compliance to IEC 61010-1. This international standard outlines safety requirements for electrical equipment used in measurement, control, and laboratory applications [4].

5. If the device requires a charging component, the charger shall be compliant to IEC 62368-1 international safety standard for ICT and AV equipment [2].

    iv. There shall be mechanical safety in the device:

1. The device shall be stable to prevent tipping or moving unexpectedly during use. The electrical components shall be secured in or on the device.

2. The moving parts of the device shall be enclosed or guarded to ensure user safety.

n. Accuracy and Reliability

    i. The machine learning model must have an accuracy of 80% or greater [17].

    ii. The model must have a precision value of 80% or greater for a "tight" knot as a true positive to minimize false positives [17].

    iii. The model must achieve a recall value of 80% or greater for a "loose" knot as the true positive value to minimize false negatives [17].

    iv. The F1 score of the model must be greater than or equal to 0.8 [17].

o. Shelf Life

    i. The device will be used each semester in a training environment. During storage over break periods, it must maintain full functionality when returned to normal operating temperature and humidity conditions.

    ii. Shelf life is not a priority constraint for this device.

p. Life in Service

    i. The sensing element should withstand repeated use during training sessions.

    ii. The device must withstand cleaning after each training session involving biological material, including a thorough wipe-down with a disinfectant solution, without loss of functionality.

q. Operating Environment

    i. Non-electronic components must tolerate disinfectant wipes and incidental contact with biological material without structural or functional degradation.

    ii. The device must maintain functionality under normal operating temperatures and humidity.

r. Ergonomics:

    i.   Real-time feedback should be intuitive and immediate, delivered through vibration, LED, or audio cues.

    ii.   The device should integrate smoothly with existing training methods so users do not have to adjust their typical working posture.

s. Size:

    i.   If used in future surgical use, the device must maintain a low-profile form that does not obstruct access to the wound site. The device should not be larger than a table-top device to maintain these requirements.

t. Weight

    i.   The device should be able to rest on a tabletop without weight concerns.

    ii.   Weight is not a priority constraint for this design.

u. Materials

    i.   The device should be made of lightweight, rigid materials to maintain low weight and structural integrity.

    ii.   Internal electronics should be housed in a compact, lightweight, and protective enclosure to prevent damage.

    iii.   If used for future clinical applications, materials should be biocompatible and sterilizable.

v. Aesthetics, Appearance, and Finish

    i.   The device should have a clean and professional appearance appropriate for training and potential clinical environments.

    ii.   No specific color, style, or finish is required, but finishes should support durability and user comfort.

    iii.   All surfaces should be smooth and free of sharp edges to ensure safe handling during practice or surgical use.

    iv.   Visual indicators, such as LEDs or labels, should be clearly visible without distracting from the suturing task.

2. Production Characteristics

    c. Quantity

        i. According to the client, one device is required for teaching purposes. After fabrication, testing, and approval from the client more devices could be created to upscale the number of students that can learn at a time during a lab section. Additionally, this product could be applied to industries aside from veterinary applications, so more could be replicated for teaching in human applications.

    d. Target Product Cost

        i. The client has provided $250 for the project budget with room for negotiation and increases as necessary. This cost, however, will include testing materials, sutures, and the design itself.

3. Miscellaneous

    f. Standards and Specifications

        i. This device shall comply with FDA standards as it is, by definition, a medical device as it is "intended to affect the structure or any function of the body or other animals…" [7] The device is utilized in conjunction with the user to complete suture knot tying, which affects the structure of the body.

        ii. ISO 14971:2019 Risk Management [6]

            1. Risk analysis through Failure Modes and Effects Analysis (FMEA) should be completed to identify potential risks for the patient, operator, and property. This includes gathering data and reviewing literature about the risks of similar medical devices. This standard states that the concept of risk involves the probability of the occurrence of harm and the severity of its consequences.

        iii. Code of Federal Regulations, Title 21, Chapter 1, Part 803 [9]

            1. Manufacturers and facilities that use the device must report deaths and serious injuries that the device has caused or contributed to through a Medical Device Report (MDR).

        iv. IEC 61010-1:2010 Safety Requirements for Electrical Equipment for Measurement, Control, and Laboratory Use [10]

            1. Surface temperature limits for protection against burns

2. Protection against electrical shock

3. Resistance to mechanical stress

4. Protection against the spread of fire

5. Protection against hazards from fluids

v. ISO 12100:2021 Safety of machinery - General principles for design - Risk assessment and risk reduction [11]

1. This standard establishes a strategy for risk assessment, identifying hazards, and risk reduction in machinery design. If the device incorporates any mechanical or machinery components, compliance with the standard is required.

vi. ASTM F2458-05:2024 Standard Test Method for Wound Closure Strength of Tissue Adhesives and Sealants [12]

1. This test method shall be used to evaluate the force values of a suture, allowing for the assessment of material properties and failure of the suture material. This test method is for comparing bonding processes for susceptibility to fatigue, mode of failure, and environmental changes.

vii. ISO 10993-1:2018 Biological Evaluation of Medical Devices [13]

1. The device shall be biocompatible as the device is in contact with live tissue (user and potentially patient).

2. Further biocompatibility testing shall be completed if the device is to be used during surgical procedures to ensure hemocompatibility.

viii. IEC 60601:2015 General requirements for medical electrical equipment [4]

1. This standard outlines the basic safety and essential performance of medical electrical equipment.

2. This standard includes identifying and testing the operating temperature of the device.

g. Customer

i. The initial device is intended for students at the University of Wisconsin-Madison School of Veterinary Medicine. The device will primarily be used in the curriculum of the first and second year students, as it is intended for beginners in

suture techniques. However, the device has the potential to benefit students across broader areas of medicine, including medical schools, dental schools, and other healthcare training institutions.

h. Patient-Related Concerns
   i. If the device were to display inappropriate feedback, students would learn incorrect suturing techniques that could translate to the quality of suture on live animals later in the student's career. This puts animals at risk for wound dehiscence after surgery if the square knots are not secured correctly.
   ii. Surgical wound dehiscence can lead to infection, excessive scarring, and necrosis on the wound site [14], all of which compromise the welfare of animals.
   iii. Additionally, poorly thrown sutures on cadavers limit their future use for training and increase the demand of additional animal cadavers in academic settings.

i. Economic Impact
   i. The cost of the suture varies depending on suture material, but the average cost of an individual monofilament absorbable suture is $1.75 - $1.83 per stitch [15]. Real-time feedback on tension and speed can decrease learning time and ultimately lower the amount of suture material needed for practice.
   ii. If improper tension is applied to sutures, stitches have a higher chance of breaking or unraveling. Depending on the size of the open wound, additional procedures might be necessary. Suture procedures cost can range from $50 - $1,000 to cover expenses for personnel, medicine, and surgical instruments during an operation [16].

j. Competition
   There are no competing designs that utilize machine learning to evaluate the top knot's tightness.

Resources

[1] H. Reuss-Lamky, LVT, VTS, E. FFCP, and CFVP, "Oh, Sew Easy," AAHA. Accessed: Sept. 17, 2025. [Online]. Available: https://www.aaha.org/trends-magazine/october-2022/gs-sutures/

[2] "Product Safety & Use," Oura Help. Accessed: Sept. 16, 2025. [Online]. Available: https://support.ouraring.com/hc/en-us/articles/360025428394-Product-Safety-Use

[3] "IEC 62368-1: Ask the Engineers, Question-and-Answer Page," UL Solutions. Accessed: Sept. 18, 2025. [Online]. Available: https://www.ul.com/resources/iec-62368-1-ask-engineers-question-and-answer-page

[4] "IEC 60601-1-11:2015," ISO. Accessed: Sept. 17, 2025. [Online]. Available: https://www.iso.org/standard/65529.html

[5] N. Marsidi, S. A. M. Vermeulen, T. Horeman, and R. E. Genders, "Measuring Forces in Suture Techniques for Wound Closure," *J. Surg. Res.*, vol. 255, pp. 135–143, Nov. 2020, doi: 10.1016/j.jss.2020.05.033.

[6] T. Horeman, E. Meijer, J. J. Harlaar, J. F. Lange, J. J. van den Dobbelsteen, and J. Dankelman, "Force Sensing in Surgical Sutures," *PLoS ONE*, vol. 8, no. 12, p. e84466, Dec. 2013, doi: 10.1371/journal.pone.0084466.

[7] C. for D. and R. Health, "How to Determine if Your Product is a Medical Device," *FDA*, Aug. 2023, Accessed: Sept. 18, 2025. [Online]. Available: https://www.fda.gov/medical-devices/classify-your-medical-device/how-determine-if-your-product-medical-device

[8] "ISO 14971:2019," ISO. Accessed: Sept. 17, 2025. [Online]. Available: https://www.iso.org/standard/72704.html

[9] "21 CFR Part 803 -- Medical Device Reporting." Accessed: Sept. 16, 2025. [Online]. Available: https://www.ecfr.gov/current/title-21/part-803

[10] "IEC 61010-1:2010." Accessed: Sept. 16, 2025. [Online]. Available: https://webstore.iec.ch/en/publication/4279

[11] *Safety of machinery: general principles of design : risk assessment and risk reduction (ISO 12100:2010)*, English version. London, UK, Brussels: BSI, CEN Management Centre, 2011.

[12] "Standard Test Method for Wound Closure Strength of Tissue Adhesives and Sealants." Accessed: Sept. 16, 2025. [Online]. Available: https://store.astm.org/f2458-05r24.html

[13] "ISO 10993-1:2018," ISO. Accessed: Sept. 17, 2025. [Online]. Available: https://www.iso.org/standard/68936.html

[14] "The Veterinary Nurse - An overview of postoperative wound care: surgical wound dehiscence." Accessed: Sept. 17, 2025. [Online]. Available: https://www.theveterinarynurse.com/content/clinical/an-overview-of-postoperative-wound-care-surgical-wound-dehiscence/

[15] R. K. Elmallah *et al.*, "Economic evaluation of different suture closure methods: barbed versus traditional interrupted sutures," *Ann. Transl. Med.*, vol. 5, no. Suppl 3, p. S26, Dec. 2017, doi: 10.21037/atm.2017.08.21.

[16] Medico, "Veterinary Suture by MedicoGrp: Dog Stitches Cost Guide," Medico. Accessed: Sept. 17, 2025. [Online]. Available: https://medicogrp.com/dog-stitches-cost/

[17] "F1 Score in Machine Learning," GeeksforGeeks. Accessed: Dec. 03, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/f1-score-in-machine-learning/

Appendix B - Suture MTS Results

Protocol
Testing objective: Understand how varying suture size and material effects the mechanical properties under tensile loading

1.) Obtain the following suture material:
- Maxon 3.5 Metric Monofilament absorbable suture
- Monocryl 3.5 metric Monofilament Synthetic Absorbable suture
- Wego Nylon Monofilament Nonabsorbable surgical suture
2.) Cut each suture material into 4 inch strips, tying 4 knots directly over top each other in the center of the suture line (2 inches on either side of the knot)
3.) Attach to MTS machine with a 100 N load cell grip
4.) Set system to the correct parameters:
- Test Rate: 50mm/min
- Data Acquisition Rate: 100.0Hz
5.) Run the machine until failure
6.) Save out results for later MATLAB analysis
* NOTE: width and thickness listed as 0.5 mm for both sutures because of machine limitations

Sample specific testing set up:

Sample	Size	Diameter	Grip Separation
Maxon 3.5 Metric Monofilament absorbable suture EXP June 2015	0	0.43 mm **Area: 0.145 mm^2**	57.73 mm
Monocryl 3.5 metric Monofilament Synthetic Absorbable suture EXP July 2019	0	0.43 mm **Area: 0.145 mm^2**	62.12 mm
Wego Nylon Monofilament Nonabsorbable surgical suture EXP November 2026	3-0	0.22 mm **Area: 0.038 mm^2**	68.22 mm

Results:
Force vs Displacement Graph

Force vs. Displacement (50 mm/min)

Stress vs Strain Graph
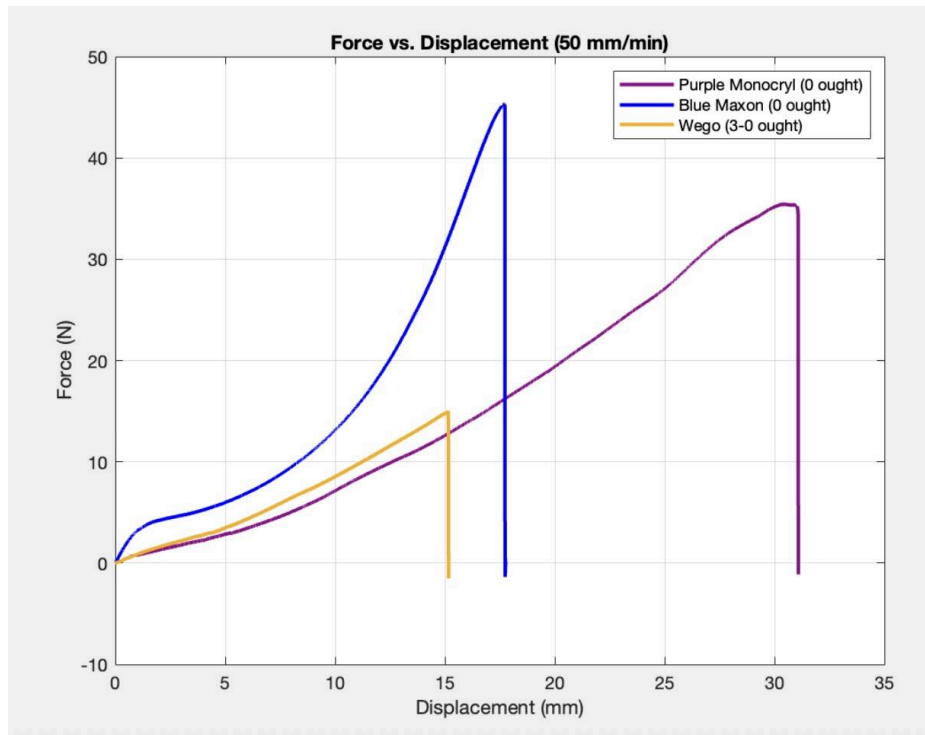
Stress vs. Strain (50 mm/min)
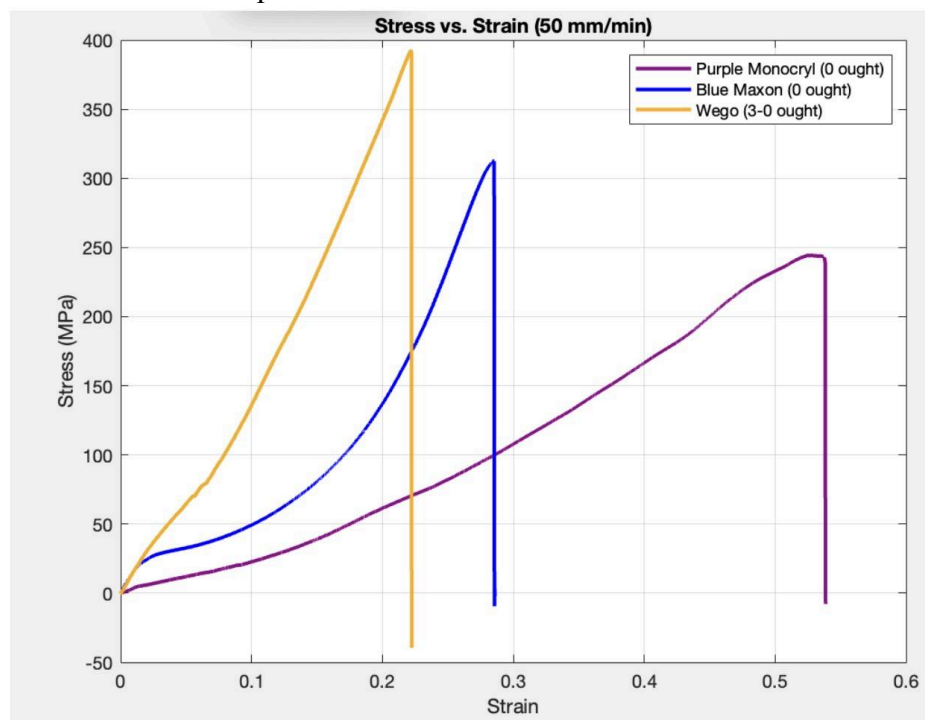
Suture Material Properties

Suture Material	Youngs Modulus (Mpa)	Max Strain	Fracture Stress (Mpa)	Yield Stress (Mpa)
Purple Monocryl (0 ought)	568.04	0.54	244.21	189.89
Blue Maxon (0 ought)	2184.17	0.29	310.07	N/A*
Wego (3-0 ought)	2019.32	0.22	390.41	N/A*

* No visible plastic deformation on Maxon and Wego suture material

Code Used for Data Analysis

```
% Close figures and clear out other variables that have been assigned
close all;
clear all;
%% Data Set 1
[file, path] = uigetfile('*','/Users/lucyhockerman/Documents/MATLAB/MATLAB2');
% prompts user to select the data file they wish to load
data1 = importdata([path,filesep,file]);
% Extract the columns of interest from your data
disp1=data1.data(:,1);
force1=data1.data(:,2);
time1=data1.data(:,3);
% Calculate stress and strain
stress1 = force1 ./ (0.145);
strain1 = disp1 ./ (57.73);
%% Data Set 2
[file, path] = uigetfile('*','/Users/lucyhockerman/Documents/MATLAB/MATLAB2');
% prompts user to select the data file they wish to load
data2 = importdata([path,filesep,file]);
% Extract the columns of interest from your data
disp2=data2.data(:,1);
force2=data2.data(:,2);
time2=data2.data(:,3);
% Calculate stress and strain
stress2 = force2 ./ (0.145);
strain2 = disp2 ./ (62.12);
%% Data Set 3
[file, path] = uigetfile('*','/Users/lucyhockerman/Documents/MATLAB/MATLAB2');
% prompts user to select the data file they wish to load
data3 = importdata([path,filesep,file]);
% Extract the columns of interest from your data
disp3=data3.data(:,1);
force3=data3.data(:,2);
time3=data3.data(:,3);
% Calculate stress and strain
stress3 = force3 ./ (0.038);
strain3 = disp3 ./ (68.22);
```

```
%% Plotting
% Stress vs Strain
figure(1)
plot(strain1, stress1, 'color', [0.5 0 0.5], 'LineWidth', 2);
hold on
plot(strain2, stress2, 'color', 'b', 'LineWidth', 2);
plot(strain3, stress3, 'LineWidth', 2);
title("Stress vs. Strain (50 mm/min)");
xlabel("Strain");
ylabel("Stress (MPa)");
legend('Purple Monocryl (0 ought)', 'Blue Maxon (0 ought)', 'Wego (3-0
ought)');
grid on
% Force vs Displacement
figure(2)
plot(disp1, force1,'color', [0.5 0 0.5], 'LineWidth', 2);
hold on
plot(disp2, force2, 'color', 'b', 'LineWidth', 2);
plot(disp3, force3, 'LineWidth', 2);
title("Force vs. Displacement (50 mm/min)");
xlabel("Displacement (mm)");
ylabel("Force (N)");
legend('Purple Monocryl (0 ought)', 'Blue Maxon (0 ought)', 'Wego (3-0
ought)');
grid on;
```

Appendix C - Model Comparison Testing Protocol

Testing Objective: Compare multiple model performance on a consistent dataset

Procedure:
1. Create a Balanced Test Dataset
    ○ Collect 40 tight and 40 loose knot images (total 80 images) that the models have not seen before. This testing dataset must be a pure test set.
    ○ Ensure diversity in lighting, angle, background, and material to avoid bias.

2. Standardize Image Preprocessing
    ○ Crop all images to a consistent size.
    ○ Ensure each image contains only one knot and similar framing.
    ○ Apply the same preprocessing steps used during model training (e.g., normalization, resizing).

3. Label the Dataset
    ○ Assign a class label to each image: "tight" or "loose".
    ○ Double-check labels to prevent label noise.

4. Load the Model in Python: Load trained model weights.
    ○ Code included below

5. Use the trained model to make predictions on the test images to see how well it performs
   ○ Loop through the dataset.
   ○ For each image, record:
      ■ Model prediction (tight/loose)
      ■ True label

6. Generate the Confusion Matrix: Use Python libraries (scikit-learn) to compute the confusion matrix
   ○ Verify interpretation with a tight knot as the positive condition
      ■ True Positive (TP): tight predicted as tight
      ■ True Negative (TN): loose predicted as loose
      ■ False Positive (FP): loose predicted as tight
      ■ False Negative (FN): tight predicted as loose
   ○ Verify interpretation with a loose knot as the positive condition
      ■ True Positive (TP): loose predicted as loose
      ■ True Negative (TN): tight predicted as tight
      ■ False Positive (FP): tight predicted as loose
      ■ False Negative (FN): loose predicted as tight

7. Compute Performance Metrics: Calculate the following metrics with both tight and loose as the positive conditions. The accuracy values will remain constant but precision, recall, and F1 score will depend on which condition is chosen as positive. Record values by explicitly state whether *tight* or *loose* is treated as positive
      ■ Accuracy
      ■ Precision
      ■ Recall
      ■ F1 Score

11. Document and Save Results: Save confusion matrix, metrics, and example misclassifications

Code:

Tensorflow Model
```
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score,
recall_score, f1_score
true_labels = []
preds_resnet = []
preds_vgg = []
label_mapping = {'Loose': 0, 'Tight': 1}
reverse_mapping = {0: "Loose", 1: "Tight"}
def quick_predict(image_path, model, class_mapping, device="cpu"):
    image_pil = Image.open(image_path).convert("RGB")
    img = load_image(image_pil).to(device)
    model.eval()
    with torch.no_grad():
        out = model(img)
        probs = F.softmax(out, dim=1)
```

```
            pred_idx = torch.argmax(probs, dim=1).item()
    return pred_idx
for folder, images in img_data.items():
    true_class = label_mapping[folder]   # Ground truth label
    for img_path in images:
        p1 = quick_predict(img_path, resnetmodel, class_mapping1)
        preds_resnet.append(p1)
        p2 = quick_predict(img_path, vggmodel, class_mapping2)
        preds_vgg.append(p2)
        true_labels.append(true_class)

print("\n====== RESNET METRICS ======")
cm_resnet = confusion_matrix(true_labels, preds_resnet)
print("Confusion Matrix:\n", cm_resnet)

print("Accuracy:", accuracy_score(true_labels, preds_resnet))
print("Precision:", precision_score(true_labels, preds_resnet,
average='binary'))
print("Recall:", recall_score(true_labels, preds_resnet, average='binary'))
print("F1 Score:", f1_score(true_labels, preds_resnet, average='binary'))

print("\n====== VGG METRICS ======")
cm_vgg = confusion_matrix(true_labels, preds_vgg)
print("Confusion Matrix:\n", cm_vgg)

print("Accuracy:", accuracy_score(true_labels, preds_vgg))
print("Precision:", precision_score(true_labels, preds_vgg, average='binary'))
print("Recall:", recall_score(true_labels, preds_vgg, average='binary'))
print("F1 Score:", f1_score(true_labels, preds_vgg, average='binary'))
```

Roboflow Model

```
import os
from inference_sdk import InferenceHTTPClient
from sklearn.metrics import confusion_matrix, classification_report
# ---- Initialize Roboflow Client ----
CLIENT = InferenceHTTPClient(
    api_url="https://serverless.roboflow.com",
    api_key="******************"
)
# ---- Paths ----
BASE_DIR = r"" ## update path to folder of testing images
CLASSES = ["tight", "loose"]  # ensure lowercase matches Roboflow output
y_true = []
y_pred = []
# ---- Loop through each class folder ----
for label in CLASSES:
    folder_path = os.path.join(BASE_DIR, label)
    for filename in os.listdir(folder_path):
        if not filename.lower().endswith((".png", ".jpg", ".jpeg")):
            continue
        img_path = os.path.join(folder_path, filename)
        # Run inference (classification)
        result = CLIENT.infer(img_path, model_id="tightvsloose-88zqq/2")
        # Parse classification result
```

```
        if len(result["predictions"]) == 0:
            pred_class = "no_prediction"
        else:
            pred_class = result["predictions"][0]["class"].lower()
        # Store values
        y_true.append(label)
        y_pred.append(pred_class)
# ---- RESULTS ----
print("\n=== Classification Report ===")
print(classification_report(y_true, y_pred, labels=CLASSES +
["no_prediction"]))
print("\n=== Confusion Matrix ===")
print(confusion_matrix(y_true, y_pred, labels=CLASSES + ["no_prediction"]))
```

Appendix D - Raw Performance Metric Data

Positive Condition: Tight Knot

Model Type	Accuracy $= \dfrac{TP + TN}{TP + TN + FP + FN}$	Precision $= \dfrac{TP}{TP + FP}$	Recall $= \dfrac{TP}{TP + FN}$	F1 Score $= 2 \times \dfrac{Precision \times Recall}{Precision + Recall}$
TensorFlow: Side and Top Data				
ResNET	0.88	0.89	0.85	0.87
VGG	0.81	0.77	0.90	0.83
TensorFlow: Side Only				
ResNET	0.76	0.71	0.90	0.79
VGG	0.75	0.67	1.00	0.80
Roboflow: Side and Top Data				
ResNET	0.83	0.88	0.75	0.81

Positive Condition: Loose Knot

Model Type	Accuracy $= \dfrac{TP + TN}{TP + TN + FP + FN}$	Precision $= \dfrac{TP}{TP + FP}$	Recall $= \dfrac{TP}{TP + FN}$	F1 Score $= 2 \times \dfrac{Precision \times Recall}{Precision + Recall}$
TensorFlow: Side and Top Data				
ResNET	0.88	0.86	0.90	0.88
VGG	0.81	0.88	0.73	0.79
TensorFlow: Side Only				
ResNET	0.76	0.86	0.63	0.72
VGG	0.75	1.00	0.5	0.67
Roboflow: Side and Top Data				
ResNET	0.83	0.78	0.90	0.84

Appendix E - McNemar Statistical Test

Testing Objective: Determine whether the difference in performance between Models 1 and 5 is statistically significant, based on predictions on the same set of test images

Procedure:
1. Ensure both models were evaluated on the exact same set of test images in the same order

2. Store columns in the same length and order with
   a. true: ground truth labels (one per image)
   b. pred1: Model 1 predictions (one per image)
   c. pred5: Model 5 predictions (one per image)

3. Construct a 2x2 McNemar contingency table.
   a. For each sample, i, compute:
      i. $c_1[i]$ = ($pred_1[i]$ == $true[i]$): boolean: Model 1 correct
      ii. $c_5[i]$ = ($pred_5[i]$ == $true[i]$): boolean: Model 5 correct

   b. Count:
      i. a = # samples where c_1 == True and c_5 == True
      ii. b = # samples where c_1 == True and c_5 == False
      iii. c = # samples where c_1 == False and c_5 == True
      iv. d = # samples where c_1 == False and c_5 == False

   c. Place these numbers into the contingency table as follows:

	Model 5 Correct	Model 5 Incorrect
Model 1 Correct	a	b
Model 1 Incorrect	c	d

4. Since b+c is small (<25 in this study), use the exact McNemar test rather than a chi-square approximation

5. Perform the test in Python with the following code:
```
import numpy as np
from statsmodels.stats.contingency_tables import mcnemar
# McNemar table (Model1 vs Model5)
# rows: Model 1 Correct/Wrong
# cols: Model 5 Correct/Wrong
table = np.array([[66, 4],
                  [0, 10]])
# run exact McNemar test (recommended when n01+n10 < 25)
result = mcnemar(table, exact=True)
print("McNemar p-value:", result.pvalue)
```

6. Interpret test results
   a. Null hypothesis: models 1 and 5 have equal accuracy
   b. Alternative hypothesis: models 1 and 5 have different accuracy

c. p ≤ 0.05, reject null hypothesis
d. p > 0.05, fail to reject null hypothesis

Appendix F - Code For Resistor and FSR Pairing Determination

```
const int force_pin = A0;
const float VREF = 5.0;
const float R_FIXED = 47000.0; // this value is varied based on restor tested
float readVoltageAverage(int pin, int samples=8) {
long sum = 0;
for (int i=0; i<samples; ++i) {
sum += analogRead(pin);
delay(2);  }
float avgADC = sum / (float)samples;  return avgADC * (VREF / 1023.0);}
void setup() {
        Serial.begin(9600);}
void loop() {
        float voltage = readVoltageAverage(force_pin, 16); // smoother
        int adc = round(voltage * 1023.0 / VREF);
        Serial.print("ADC: ");
        Serial.print(adc);
        Serial.print("\tV: ");
        Serial.print(voltage, 3);
         if (voltage < 0.02) { // threshold — treat as open / very large
                Serial.print("\tR_FSR: >");
                Serial.print(R_FIXED * 1000.0);
                Serial.println(" ohm (open)");
        }
        else {
                float r_fsr = R_FIXED * voltage / (VREF - voltage);
                Serial.print("\tR_FSR: "); Serial.print(r_fsr, 1);
                Serial.println(" ohm");  }
  delay(500);
}
```

Appendix G - FSR Sensor Code With LED

```
const int force_pin = A0;
const int led_pin = 9;        // LED connected to digital pin 9 (PWM)
const float VREF = 5.0;
const float R_FIXED = 47000.0;
float readVoltageAverage(int pin, int samples = 8) {
long sum = 0;
for (int i = 0; i < samples; ++i) {
        sum += analogRead(pin);
        delay(2);    }
float avgADC = sum / (float)samples;
return avgADC * (VREF / 1023.0);}
void setup() {
        Serial.begin(9600);
        pinMode(led_pin, OUTPUT);}


void loop() {
        float voltage = readVoltageAverage(force_pin, 16); // smoother
        int adc = round(voltage * 1023.0 / VREF);
        Serial.print("ADC: ");
        Serial.print(adc);
        Serial.print("\tV: ");
        Serial.print(voltage, 3);
        float r_fsr;
        if (voltage < 0.02) { // threshold — treat as open / very large
                Serial.print("\tR_FSR: >");
                Serial.print(R_FIXED * 1000.0);
                Serial.println(" ohm (open)");
                r_fsr = R_FIXED * 1000.0; // approximate very large value  }   else {
                r_fsr = R_FIXED * voltage / (VREF - voltage);     Serial.print("\tR_FSR: ");
                Serial.print(r_fsr, 1);
                Serial.println(" ohm");   }
 // --- LED FEEDBACK ---  // Map voltage (or resistance) to LED brightness
        int brightness = map(adc, 0, 1023, 0, 255);
        analogWrite(led_pin, brightness);
 delay(100);
}
```

Appendix G - FSR Sensor Calibration Code

```
const int force_pin = A0;  // FSR connected to A0
const float R_FIXED = 47000.0;  // 47kΩ resistor
const float VCC = 5.0;
void setup() {
        Serial.begin(9600);
        Serial.println("ADC_Value\tVoltage(V)\tResistance(Ohms)");}
void loop() {
        int force_Value = analogRead(force_pin);
        float voltage = force_Value * (VCC / 1023.0);
        float force_Resistance = (VCC - voltage) * R_FIXED / voltage;
        Serial.print(force_Value);
```

```
        Serial.print("\t");
        Serial.print(voltage, 3);
        Serial.print("\t");
        Serial.println(force_Resistance, 1);
  delay(200); // adjust for smoother plots
}
```