



# A Machine Learning Framework for Intraoperative Segmentation and Quality Assessment of Pedicle Screw X-Rays

Hooman Esfandiari<sup>1,2</sup>, Carolyn Anglin<sup>3</sup>, John Street<sup>4</sup>, Pierre Guy<sup>2,4</sup>, Antony J Hodgson<sup>1,2</sup>.

<sup>1</sup>Department of Biomedical Engineering, University of British Columbia, Vancouver, Canada

<sup>2</sup>Centre for Hip Health and Mobility, University of British Columbia, Vancouver, Canada

<sup>3</sup>Department of Civil Engineering, University of Calgary, Calgary, Canada

<sup>4</sup>Department of Orthopaedics, University of British Columbia, Vancouver, Canada

[hooman.esfandiari@ubc.ca](mailto:hooman.esfandiari@ubc.ca) [Hooman.esfandiari@gmail.com](mailto:Hooman.esfandiari@gmail.com)

## Abstract

Pedicle screw fixation is a technically demanding procedure with potential difficulties and reoperation rates are currently on the order of 11%. The most common intraoperative practice for position assessment of pedicle screws is biplanar fluoroscopic imaging that is limited to two-dimensions and is associated to low accuracies. We have previously introduced a full-dimensional position assessment framework based on registering intraoperative X-rays to preoperative volumetric images with sufficient accuracies. However, the framework requires a semi-manual process of pedicle screw segmentation and the intraoperative X-rays have to be taken from defined positions in space in order to avoid pedicle screws' head occlusion. This motivated us to develop advancements to the system to achieve higher levels of automation in the hope of higher clinical feasibility.

In this study, we developed an automatic segmentation and X-ray adequacy assessment protocol. An artificial neural network was trained on a dataset that included a number of digitally reconstructed radiographs representing pedicle screw projections from different points of view. This model was able to segment the projection of any pedicle screw given an X-ray as its input with accuracy of 93% of the pixels. Once the pedicle screw was segmented, a number of descriptive geometric features were extracted from the isolated blob. These segmented images were manually labels as 'adequate' or 'not adequate' depending on the visibility of the screw axis. The extracted features along with their corresponding labels were used to train a decision tree model that could classify each X-ray based on its adequacy with accuracies on the order of 95%.

In conclusion, we presented here a robust, fast and automated pedicle screw segmentation process, combined with an accurate and automatic algorithm for classifying

views of pedicle screws as adequate or not. These tools represent a useful step towards full automation of our pedicle screw positioning assessment system.

## 1 Introduction

Pedicle screw fixation is a common practice in many surgical spinal procedures, particularly fusion and deformity correction. Metal screws are first placed into the vertebral body through the pedicle pathway and used as firm anchor points for rods connecting to adjacent spinal segments. Despite the obvious advancements that this method has brought to spinal surgeries, it is a technically demanding procedure with potential difficulties (Katonis 2003). Due to the proximity of the spinal cord, nerves and blood vessels, a misplaced pedicle screw can lead to serious clinical complications. Furthermore pedicle screw malplacement rates have been reported to be on the order of 20% while major violations can occur in roughly 4% of inserted screws (Nevzati 2014), resulting in reoperation rates of roughly 11% primarily attributable to screw malplacement (Christensen 2002).

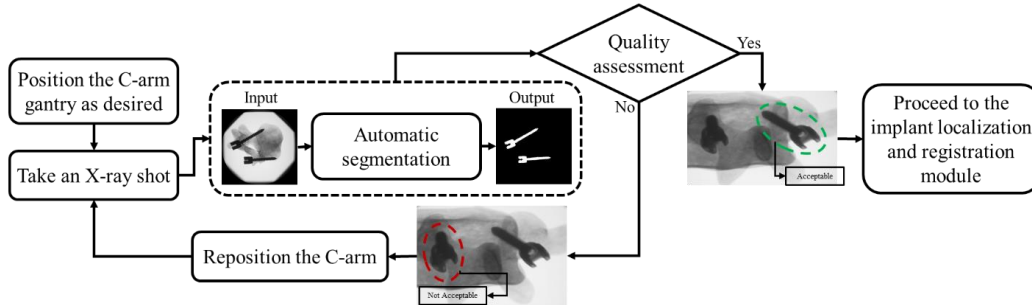
This experience highlights the importance of performing an intraoperative implant position assessment after screw insertion. In clinical practice, this is usually done using intraoperative biplanar X-rays, intraoperative cone-beam computed tomography (e.g. O-arm) or postoperative (postop) computed tomography (CT) modalities. While postoperative CT imaging is considered the gold standard in terms of identifying misplaced screws, it results in high levels of ionizing radiation, is prone to metal artifact issues and occurs too late to make changes during the surgery. Intraoperative X-ray assessment is the most common practice (due to its low cost, availability and ease of use), but, compared to the gold standard, it has limited sensitivity and specificity (70% and 83% respectively; Choma 2006). Intraoperative cone-beam CT can provide 3D spatial feedback on the screw's position, however this method attributed to the similar issues of postoperative CTs. Additionally cone-beam CT has been shown to have low rates of sensitivity (77% presumably due to metal artifact issues; Cordemans 2016) and is not universally available. In (Esfandiari 2016) we introduced an intraoperative (post-placement) pedicle screw position assessment system based on C-arm fluoroscopy. In this technique, we register the intraoperative fluoroscopic images (which generate relatively low radiation doses) to preoperatively-acquired CT volumes. Although we achieved acceptable levels of accuracy with our models (2mm and 1.3o on average), the need for manual intervention at several stages of the processing pipeline would constitute a significant barrier to clinical adoption. In this study, we therefore focus on automating several components of the system in the hope of achieving higher clinical acceptability.

Most of the pedicle screws currently used in clinical practice use a two-component design that consists of a shaft and an articulating (or fixed) head. The head component is relatively large and depending on the imaging configuration, can occlude a considerable portion of the screw's axis. Since this axis needs to be clearly visible for quantitative assessment, C-arm images should be acquired well away from the long axes of all screws being imaged. Once we have a pair of acceptable images, our current pipeline involves a semi-automated step in which the user identifies the screws on the acquired X-rays.

In this study, we aim to increase the level of automation in our processing pipeline by automatically identifying screws in the image and assessing the adequacy of the acquired intraoperative X-rays' view angle. We report on a first step towards this goal: using machine learning techniques to segment a single screw in a fluoroscopic image and to classify the perspective of the C-arm as adequate or not.

## 2 Materials and Methods

The proposed enhancement to our existing workflow has two key steps: implant segmentation and image quality assessment (Figure 1).



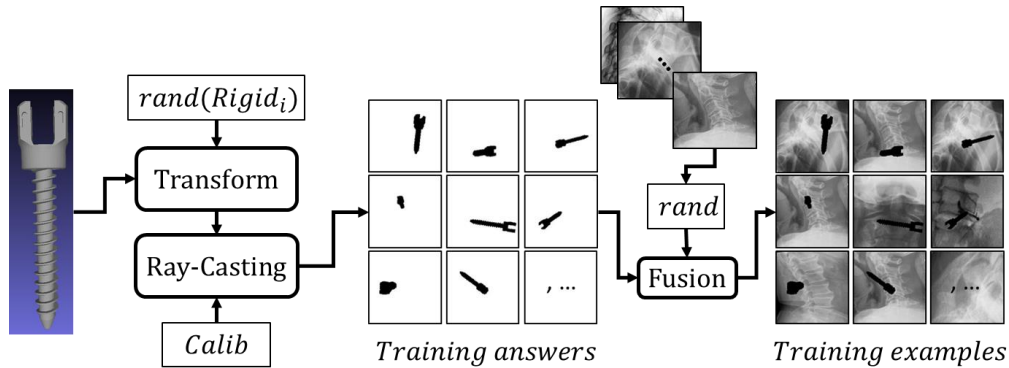
**Figure 1:** Proposed clinical workflow

Despite the relatively high contrast between the screws and the background in the fluoroscopic images, which facilitates segmentation, it can be surprisingly difficult to both identify all screws in the image and accurately segment them without requiring any manual input (e.g. specifying seed points for region growing methods or roughly tracing of the outlines for active contour-based methods). This motivated us to develop a machine learning model that can automatically identify and segment the projections of the pedicle screws, given an intraoperative X-ray as its input.

Since all machine learning algorithms require a large number of training examples, and since it is time-consuming to acquire a large number of real fluoroscopic images, we opted instead to create a synthetic training set by overlaying artificially-generated images of a pedicle screw from multiple perspectives with a dataset of real fluoroscopic images from clinical cases.

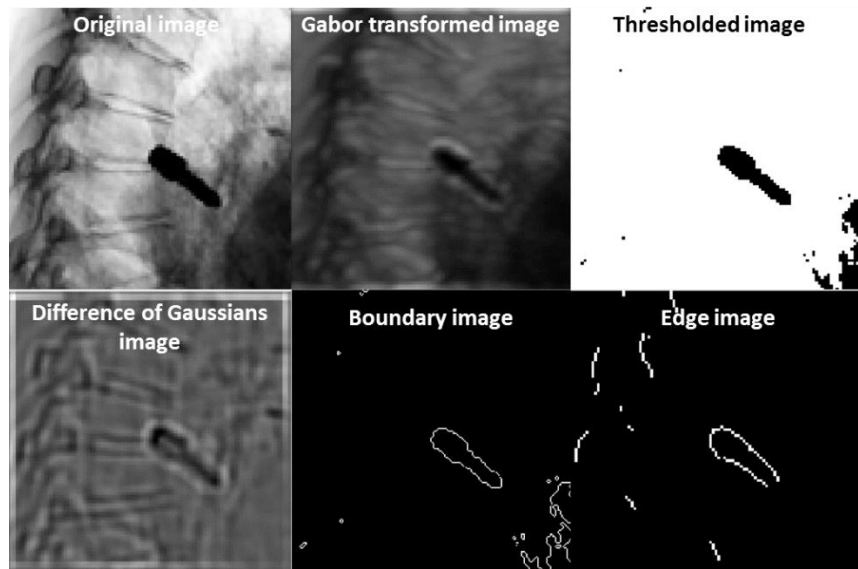
To generate the synthetic images, we used the process illustrated in Figure 2 – we generated a set of 2000 random positions for the pedicle screw within the imaging cone of a C-arm machine (using realistic intrinsic imaging parameters derived from a clinical C-arm (Arcadis Orbic, Siemens AG, Munich, Germany)) and applied a ray-casting algorithm to produce a digitally-reconstructed radiograph showing the pedicle screw on a blank background. We then overlaid each of these

screw images on one of ten clinical fluoroscopic images from our archive to produce a final training dataset of 2000 images.



**Figure 2:** Generation of the synthetic training dataset; Calib: intrinsic and extrinsic imaging parameters of the modeled C-arm;  $Rigid_i$ : the rigid transformation parameters assigned for each image

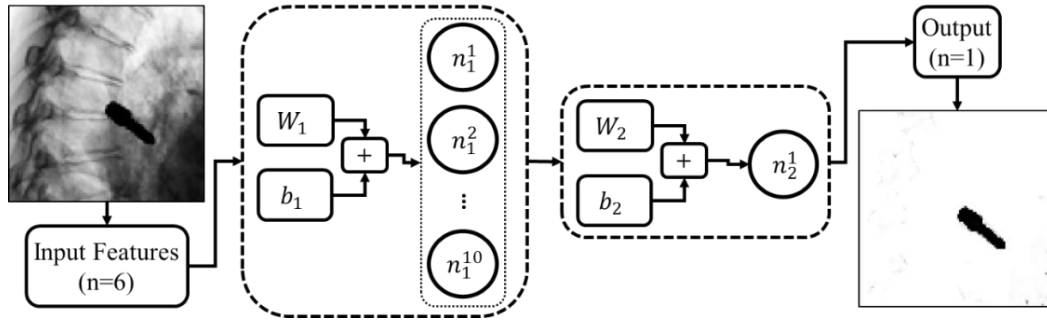
The first step in the machine learning pipeline was to recover the projection of the pedicle screw. To do this, we extracted a set of features from each training instance based on a series of processed images as shown in Figure 3: a Gabor Transformation (Soares 2006), difference of Gaussian kernels, global thresholding (threshold value interpreted from the ray-casted images), Canny edge detector (Canny 1986) and boundary estimation. The feature images were downsampled to 100 by 100 pixels to reduce the computational resources required.



**Figure 3:** Extracted features from each training example

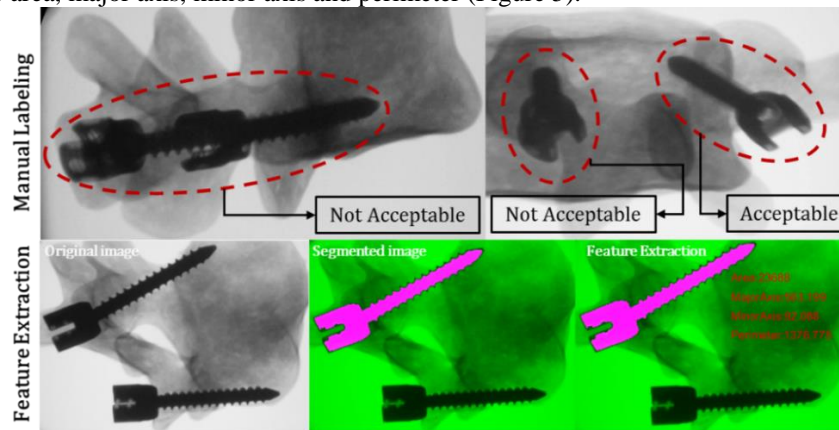
An artificial neural network (ANN) with 2 hidden layers and 10 + 1 neurons was trained on the generated learning dataset (Figure 4). Having a learning dataset of size 2000 images (100 by 100

pixels) and randomly dividing it into training (70%), validation (15%) and testing datasets (15%) resulted in 14 million training examples, 3 million validation examples and 3 million testing examples (each pixel is considered as an individual learning example with 6 features and a binary label).



**Figure 4:** A schematic overview of the implemented neural network;  $w_i$  and  $b_i$  are the weights and biases of each hidden layer;  $n_i^j$  is the  $j^{\text{th}}$  neuron in the  $i^{\text{th}}$  hidden layer

The perspective assessment module was based on a decision tree developed using manually-assigned labels specified by the first author on a set of fifty fluoroscopic images of two artificial vertebral models (Ammolite BioModels, Calgary, Canada; these models have realistic X-ray properties) with stainless steel or titanium pedicle screws inserted (Medtronic, Dublin, Ireland). The X-rays images were acquired using a mobile C-arm (Arcadis Orbic, Siemens AG, Munich, Germany) from different locations in space representing realistic clinical scenarios. All images were processed with the aforementioned ANN segmentation algorithm to isolate the projection of the inserted implants. The instances for which the segmentation algorithm did not produce sufficient results (i.e. high rates of misclassification) were excluded from the rest of the process. Then, the isolated screws were labeled as “Acceptable” or “Not acceptable” based on parameters like screw shaft length and occlusion by the screw head (Figure 5). We then extracted a set of descriptive features from the isolated blobs in the segmented images to characterise the geometric properties of the projected screws. The employed features were area, major axis, minor axis and perimeter (Figure 5).



**Figure 5:** Manual labeling and feature extraction for X-ray adequacy assessment

We trained a decision tree using the extracted descriptive features and the manually identified labels and applied the trained model to the 50-image dataset described earlier. By comparing model’s output

to the manually labeled adequacy measures we could calculate the accuracy of the quality assessment module. For this purpose a 5-folds cross validation routine was incorporated. This classifier will be able to generate a quality measure (acceptable or not acceptable) for every intraoperative X-ray that is ran through the segmentation process beforehand.

Custom codes were developed using Matlab R2016b (Mathworks, Natick, MA) for ray-casting and image processing. The neural network design and analysis was performed using the Neural Net toolbox in Matlab R2016b. The decision tree classifier was trained using the Classification Learner toolbox in Matlab R2016b. Training and cross validation of the machine learners was performed on a computer with a 2.60 GHz Core i7 processor running a 64-bit operating system (windows 10) with 64 GB of RAM.

### 3 Results

The implemented neural networks achieved an encouraging cross-entropy value of 0.12 on the testing dataset and only 0.53% of the test data was misclassified (i.e. only 0.53% of the total pixels were misclassified). Sensitivity and specificity values of 99.7% and 93.3% were achieved. This means that 93.3% of all the pixels in the test dataset (3 million) that belonged to the projected screw were correctly identified (the screws were projected in black as the foreground). The entire training process took 30 minutes of computation time and classifying an unseen image using the trained model was almost instantaneous. As the result of 5-folds cross validation, the adequacy assessment module was shown to achieve accuracies on the order of 95%.

### 4 Discussion

This study presented a novel framework for intraoperative segmentation and adequacy assessment of pedicle screw X-rays. We demonstrated promising ability to distinguish between pixels belonging to pedicle screws and to anatomy, and showed that a set of geometric features of the resulting blobs could accurately classify a view of a pedicle screw as adequate or not.

There appears to be relatively little literature to date addressing this specific issue. Popescu 2012 evaluated a machine learning approach to segment pedicle screws from axial fluoroscopy images acquired from an artificial vertebral model, but only validated their approach on models with no soft tissue component and from only one viewing angle. As another limitation associated to that study, only 34 instances were used for training the segmentation model, hence the generalization ability of their model is questionable.

Although our results were encouraging, several aspects of the proposed framework need to be further improved before this technique can be used in practice. Most significantly, our current synthetic training set may not be sufficiently realistic – it does not include multiple screws, nor additional hardware commonly encountered in clinical images (eg, rods or tools). In addition, the screws used all had fixed heads, but screws with articulating heads are common in practice and should be accounted for. It is possible that we will need to adapt our segmentation step to incorporate additional local context in order to be more robust to these situations. In addition, we currently need to manually verify the segmented blobs prior to adequacy classification; this step will have to be automated in future.

In conclusion, we presented here a robust, fast and automated pedicle screw segmentation process, combined with an accurate and automatic algorithm for classifying views of pedicle screws as adequate or not. These tools represent a useful step towards full automation of our pedicle screw positioning assessment system (Esfandiari 2016).

## Disclosure

No relevant disclosures.

## Acknowledgements

This work has been supported by the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Canadian Institute of Health Research (CIHR). We thank the Centre for Hip Health and Mobility for providing the lab facilities used in this study and the Institute for Computing, Information and Cognitive Systems for program support. We also thank Canadian Institutes of Health Research for their support.

## References

- Canny J (1986): A Computational Approach to Edge Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, pp: 679–698.
- Choma TJ, Denis F, Lonstein JE, Perra JH, Schwender JD, Garvey TA, Mullin WJ. (2006): Stepwise Methodology for Plain Radiographic Assessment of Pedicle Screw Placement: A Comparison with Computed Tomography, *Journal of Spinal Disorders & Techniques*, 19 (8), pp: 547–553.
- Christensen FB, Hansen ES, Laursen M, Thomsen K, Bünger CE (2002): Long-Term Functional Outcome of Pedicle Screw Instrumentation as a Support for Posterolateral Spinal Fusion: Randomized Clinical Study with a 5-Year Follow-Up, *Spine*, 27 (12), pp: 1269–77.
- Cordemans V, Kaminski L, Banse X, Irda N, Francq B, Cartiaux O (2016): Accuracy of Pedicle Screw Insertion Using a New Intraoperative Cone-Beam Ct Imaging Technique: Retrospective Analysis of 586 Screws, In 16th Annual meeting of the International Society for Computer Assisted Orthopaedic Surgery, Osaka, Japan.
- Esfandiari H, Newell R, Bernard R, Anglin C, Street J, Swamy G, Hodgson AJ (2016). An Intraoperative (Post-Placement) Pedicle Screw Position Assessment System, In 16th Annual meeting of the International Society for Computer Assisted Orthopaedic Surgery, Osaka, Japan.
- Katonis P, Christoforakis J, Aligizakis AC, Papadopoulos C, Sapkas G, Hadjipavlou A (2003): Complications and Problems Related to Pedicle Screw Fixation of the Spine, *Clinical Orthopaedics and Related Research*, 411, pp: 86–94.
- Nevzati E, Marbacher S, Soleman J, Perrig WN, Diepers M, Khamis A, Fandino J (2014): Accuracy of Pedicle Screw Placement in the Thoracic and Lumbosacral Spine Using a Conventional Intraoperative Fluoroscopy-Guided Technique: A National Neurosurgical Education and Training Center Analysis of 1236 Consecutive Screws, *World Neurosurgery* 82 (5), pp: 866–871.
- Popescu D, Amza CG, Lăptoiu D, Amza G (2012): Competitive Hopfield Neural Network Model for Evaluating Pedicle Screw Placement Accuracy, *Strojniški Vestnik – Journal of Mechanical Engineering*, 58 (9), pp 509–516.
- Soares JVB, Leandro JJG, Cesar RM, Jelinek HF, Cree MJ (2006): Retinal Vessel Segmentation Using the 2-D Gabor Wavelet and Supervised Classification, *IEEE Transactions on Medical Imaging*, 25 (9), pp: 1214–1222.