JNS SPINE

CLINICAL ARTICLE

A computer vision approach to identifying the manufacturer of posterior thoracolumbar instrumentation systems

Adrish Anand, BA,[1] Alex R. Flores, MD,[1] Malcolm F. McDonald, BS,[1,2] Ron Gadot, BS,[1] David S. Xu, MD,[3] and Alexander E. Ropper, MD[1]

[1]Department of Neurosurgery, Baylor College of Medicine, Houston; [2]Medical Scientist Training Program, Baylor College of Medicine, Houston, Texas; and [3]Department of Neurosurgery, The Ohio State University School of Medicine, Columbus, Ohio

OBJECTIVE Knowledge of the manufacturer of the previously implanted pedicle screw systems prior to revision spinal surgery may facilitate faster and safer surgery. Often, this information is unavailable because patients are referred by other centers or because of missing information in the patients' records. Recently, machine learning and computer vision have gained wider use in clinical applications. The authors propose a computer vision approach to classify posterior thoracolumbar instrumentation systems.

METHODS Lateral and anteroposterior (AP) radiographs obtained in patients undergoing posterior thoracolumbar pedicle screw implantation for any indication at the authors' institution (2015–2021) were obtained. DICOM images were cropped to include both the pedicle screws and rods. Images were labeled with the manufacturer according to the operative record. Multiple feature detection methods were tested (SURF, MESR, and Minimum Eigenvalues); however, the bag-of-visual-words technique with KAZE feature detection was ultimately used to construct a computer vision support vector machine (SVM) classifier for lateral, AP, and fused lateral and AP images. Accuracy was tested using an 80%/20% training/testing pseudorandom split over 100 iterations. Using a reader study, the authors compared the model performance with the current practice of surgeons and manufacturer representatives identifying spinal hardware by visual inspection.

RESULTS Among the three image types, 355 lateral, 379 AP, and 338 fused radiographs were obtained. The five pedicle screw implants included in this study were the Globus Medical Creo, Medtronic Solera, NuVasive Reline, Stryker Xia, and DePuy Expedium. When the two most common manufacturers used at the authors' institution were binarily classified (Globus Medical and Medtronic), the accuracy rates for lateral, AP, and fused images were 93.15% ± 4.06%, 88.98% ± 4.08%, and 91.08% ± 5.30%, respectively. Classification accuracy decreased by approximately 10% with each additional manufacturer added. The multilevel five-way classification accuracy rates for lateral, AP, and fused images were 64.27% ± 5.13%, 60.95% ± 5.52%, and 65.90% ± 5.14%, respectively. In the reader study, the model performed five-way classification on 100 test images with 79% accuracy in 14 seconds, compared with an average of 44% accuracy in 20 minutes for two surgeons and three manufacturer representatives.

CONCLUSIONS The authors developed a KAZE feature detector with an SVM classifier that successfully identified posterior thoracolumbar hardware at five-level classification. The model performed more accurately and efficiently than the method currently used in clinical practice. The relative computational simplicity of this model, from input to output, may facilitate future prospective studies in the clinical setting.

https://thejns.org/doi/abs/10.3171/2022.11.SPINE221009

KEYWORDS machine learning; computer vision; spinal hardware; spinal revision surgery; lumbar; thoracic

Posterior thoracolumbar revision surgery is often assisted by preoperative identification of the implanted hardware, which allows selection of the correct tools for removal if indicated. Currently, the methodology for the identification of implanted hardware is rudimentary and inefficient. If the surgeon is unable to identify the hardware, they may use other expert opinions from colleagues or hardware representatives. If this does not yield a solution, they may then use crowdsourcing, for example, via private Facebook groups of surgeons and representatives.

ABBREVIATIONS AP = anteroposterior; SVM = support vector machine.

SUBMITTED September 6, 2022. **ACCEPTED** November 17, 2022.
INCLUDE WHEN CITING Published online December 27, 2022; DOI: 10.3171/2022.11.SPINE221009.

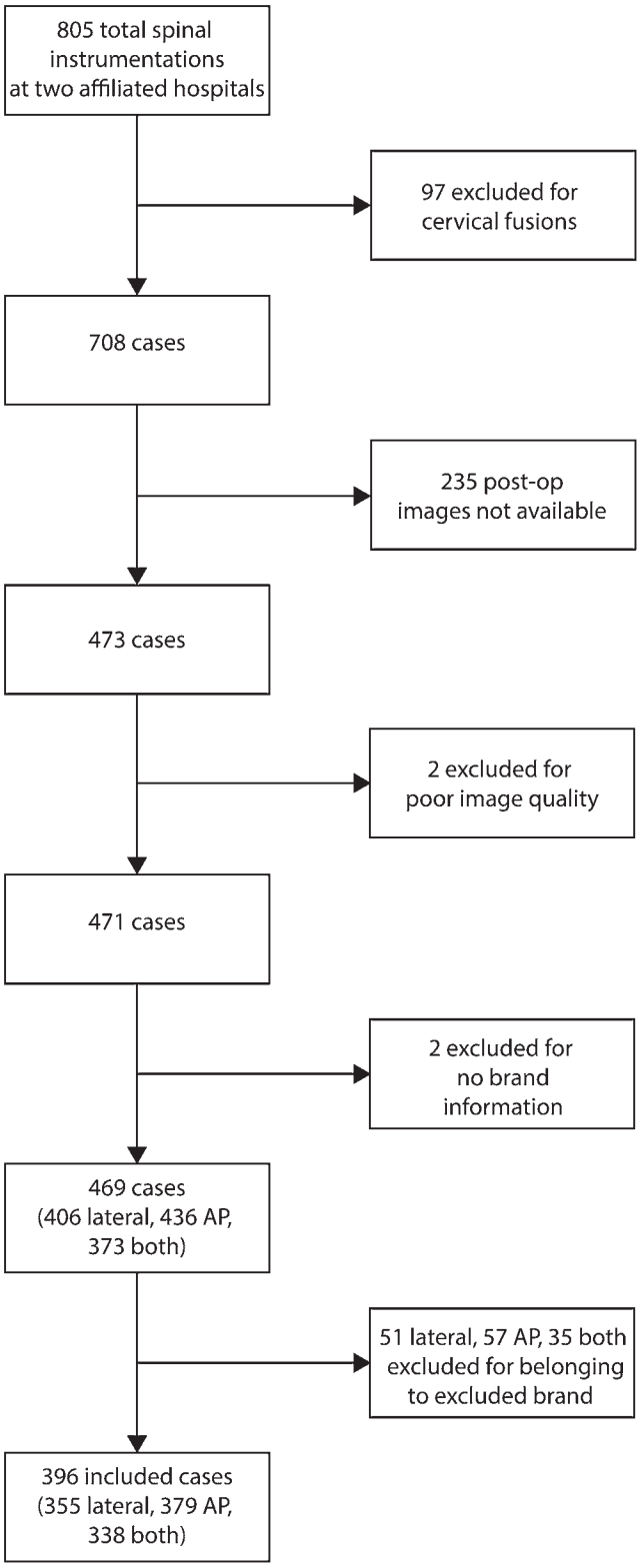

FIG. 1. Image and patient inclusion and exclusion flowchart.

In the current era of progressively advanced technology in medicine, the above methodology is due for improvement.[1,2] Hardware identification is based on neuroimaging, most commonly anteroposterior (AP) and lateral radiographs of the thoracolumbar spine. This problem naturally lends itself to a computer vision solution.[3,4]

Computer vision is a subset of machine learning that focuses on finding, characterizing, and learning data from digital images. Specifically, it can be harnessed for identification problems via medical imaging, including radiographs. Computer vision solutions can be achieved via a multitude of machine learning methods, including classic, well-studied, and well-understood algorithms such as support vector machines (SVMs), or newer but less-understood techniques, including neural networks and deep learning. The current state of applied machine learning in neurosurgery is mostly limited to retrospective proof-of-concept investigations.[5,6] Reviews of this literature consistently call for a move toward prospective use and validation of published classifier and prediction models and future investigations.[1,7,8] The consensus limitation of prospective use is the risk of the clinician-user not understanding the mechanics of the model and its output and therefore not being able to troubleshoot seemingly aberrant predictions and classifications.

To reduce the risk of poor user understanding and increase prospective use, computing power and simplicity must be balanced when choosing classifier and prediction models to solve clinical problems. A single prior publication exists discussing the use of machine learning in thoracolumbar spinal implant identification. The goal of that study was to demonstrate the clinical utility specifically of deep learning, with the example being thoracolumbar spinal implants.[9] The authors used AutoML via Google and Apple neural network platforms to achieve respectable results but at the price of "black box" mechanics, that is, poorly understood pathways from input to output.

The goal of the present study was to use computer vision techniques to achieve a powerful, relatively understandable, and clinically applicable classifier of posterior thoracolumbar spinal hardware, with the goal of future prospective use and further validation.

Methods

Patient Selection

We retrospectively analyzed all patients who underwent posterior thoracolumbar instrumentation for any indication at two institution-affiliated hospitals from 2015 to 2021. Institutional review board approval was obtained prior to the study. Patients with lateral or AP radiographs after implantation were identified. If a patient had multiple postoperative radiographs, only the first radiograph after implantation was included. The manufacturer of the implanted hardware was identified from the patients' operative records. Patients were excluded if there was no documentation of the manufacturer (n = 2) or if no radiographs were available (n = 235). Most of the cases with no available radiographs were the earliest cases in 2015. Figure 1 depicts a flowchart with image and patient exclusion criteria and numbers associated with each exclusion criterion.

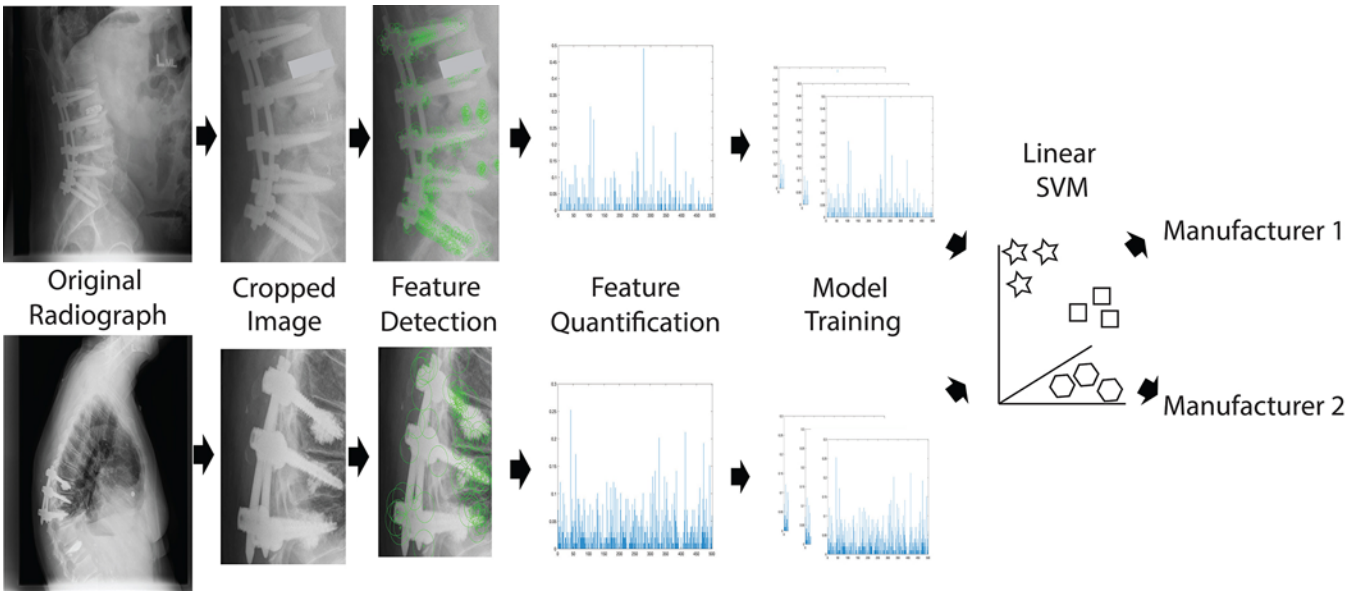

FIG. 2. Methodological workflow. From *left* to *right*, the process of image cropping, feature detection, feature quantification, model training, and classification is demonstrated. Interbody cages have been covered. Figure is available in color online only.

Image Processing

Initial DICOM images were extracted from the patients' records. Images were converted to 256-bit grayscale files and cropped using Adobe Photoshop. Cropped images included only the pedicle screws and rods. Images were not scaled, rotated, or reoriented in any way. In cases in which other hardware, such as interbody cages or other types of screws, was present but not overlapping with the desired hardware, the extraneous hardware was covered with a grayscale box. The rationale for this was that radiographs for one included manufacturer had a disproportionate number of interbody cages compared with other manufacturers, which could lead to bias in classification. In cases in which other hardware overlapped with the desired hardware, extraneous hardware was kept in the image.

For patients who had both lateral and AP images available, a third image category called "fused" was created. These images were created by initially resizing the AP image to the same dimensions as the lateral image and then scaling the contrast to be within the minimum and maximum values of the lateral image. The contrast of the AP image was scaled to the maximum and minimum values of the lateral image so that the classifier would not prefer one image over the other solely because one was brighter or darker. Finally, the two images were adjoined left to right. This image type was created because both lateral and AP images may provide important features for the classifier, and this format would allow the classifier to extract features from both images at the same time. Analysis was performed with and without this processing step.

Feature Extraction and Classifier Training

We applied a bag-of-visual-words feature extraction and SVM classification technique similar to the method of Huang et al.[10] These techniques are well established but are briefly described here. This approach was selected over other traditional computer vision techniques such as convoluted neural networks because the data sets were sparse.[5,9]

A KAZE feature extractor identified relevant radiographic features from training images. This algorithm was chosen because it is scale and rotation invariant and uses nonlinear scale-spaces to preserve important features while eliminating noise in the image. The KAZE algorithm operates independently of the angle, length, or number of screws in any given image.[11] Furthermore, other feature extractors (SURF, MSER, and Minimum Eigenvalue) were tested;[12] however, they had difficulty identifying relevant radiographic differences between screws and had lower classification accuracy. Once relevant features were extracted, they were clustered with similar features to form a "visual word." The feature space was capped at K = 500, and visual words were clustered using a K-means clustering algorithm. Once the feature space was defined from the training images, the strongest 80% of features were selected and used to train an SVM classifier. Subsequent test images were then described by the relevant features and classified by the SVM classifier. The full methodological workflow is shown in Fig. 2.

Classifier Performance

To test the accuracy and validity of our model, we used a bootstrapping method with an 80/20 training/testing split over 100 iterations. During each iteration, each image category was pseudorandomly partitioned, with 80% in the training set and 20% in the test set. Accuracy was calculated as the percentage of correct classifications in the test set. In this validation method, the model was tested on a set of images it was not trained with, so the validation is an accurate reflection of model performance. The means and standard deviations were recorded. We tested

TABLE 1. Distribution of manufacturers in the data set

	No. of Images (%)		
	Lateral	AP	Fused
Globus Medical Creo	95 (26.76)	118 (31.13)	88 (26.04)
Medtronic Solera	80 (22.54)	77 (20.32)	77 (22.78)
NuVasive Reline	65 (18.31)	66 (17.41)	65 (19.23)
Stryker Xia	65 (18.31)	63 (16.62)	59 (17.46)
DePuy Expedium	50 (14.08)	55 (14.51)	49 (14.50)
Total	355	379	338

the model accuracy in binary and multilevel classifications. This investigation was performed because we aimed to determine how classifier performance changed as we added manufacturers that were not as well represented in our data set. Additionally, we used a similar methodology to test how the classifier would perform if it were only trained on pedicle screws rather than the pedicle screw-and-rod system.

Finally, we compared our classifier with the current standard of practice using a reader study. In this study, we removed 100 images consisting of a nearly equivalent distribution of the five included manufacturers and devices in our data set. We trained a model on fused images with the same methodology previously described and then used the 100 images as the test set. We compared the model's performance with that of two surgeons and three manufacturer representatives on the same 100-image test set using a multiple-choice Google Forms quiz. The outcomes of interest were classification accuracy and quiz completion time. The study consisted of two surgeons and three manufacturer representatives. Both surgeons were regular users of the five included manufacturers. The three representatives were familiar with one or two of the included manufacturers. Included participants were not provided any specific teaching on how to identify unfamiliar hardware. The participants were presented an image and were instructed to select one of the five included manufacturers.

All analyses were performed using MATLAB (2020b, The MathWorks, Inc.) with the aid of Computer Vision, Image Processing, and Statistics and Machine Learning Toolboxes.

Results

In total, 406 lateral, 436 AP, and 373 fused images were identified for 9 different manufacturers. Among the lateral, AP, and fused images, 51, 57, and 35 images, respectively, were excluded from the analysis because they belonged to manufacturers that accounted for fewer than 50 images (Orthofix, K2M, Zimmer, and Corelink). The five manufacturers and devices included in this study were Globus Medical Creo (lateral: n = 95, 26.76%; AP: n = 118, 31.13%; and fused: n = 88, 26.04%), Medtronic Solera (lateral: n = 80, 22.54%; AP: n = 77, 20.32%; and fused: n = 77, 22.78%), NuVasive Reline (lateral: n = 65, 18.31%; AP: n = 66, 17.41%; and fused: n = 65, 19.23%), Stryker Xia (lateral: n = 65, 18.31%; AP: n = 63, 16.62%; and fused: n = 59, 17.46%), and DePuy Expedium (lateral: n = 50, 14.08%; AP: n = 55, 14.51%; and fused: n = 49, 14.50%) (Table 1). The included images contained hardware spanning 2–17 levels.

These five manufacturers were included in the classifier, which consisted of 355 lateral, 379 AP, and 338 fused images. Examples of included hardware can be seen in Fig. 3. Binary classification between the two most common manufacturers (Globus Medical and Medtronic) was performed with high accuracy among all three views (lateral: 93.15% ± 4.06%; AP: 88.98% ± 4.08%; and fused: 91.08% ± 5.30%). In the binary classification, lateral images showed better performance than AP images ($p < 0.001$) and fused images ($p = 0.04$). As more models were included, an approximately 10% decrease in classifier performance per added manufacturer was observed; however, all comparisons performed better than chance (Table 2). For three-, four-, and five-way classification, lateral and fused images performed significantly better than AP images. High accuracy was maintained among multilevel classification despite in-class variations in the number of screws, screw length, screw angle, and the presence of overlying metallic hardware or bony tissue (Fig. 4). Additionally, five-way classification showed good results despite low numbers relative to the other four models.

In five-way classification, across all image types, Globus Medical had the highest accuracy (lateral: 81%; AP: 72%; and fused: 77%). Medtronic (70%, 68%, and 66%, respectively) was followed by NuVasive (49%, 66%, and 58%, respectively), Stryker (64%, 38%, and 58%, respectively), and DePuy (44%, 48%, and 58%, respectively) in decreasing classification accuracy. Across all image types, non–Globus Medical instrumentation was misclassified as Globus Medical instrumentation at a higher rate than it was misclassified as other hardware (Fig. 4). In this model, all test sets had the same manufacturer distribution as the training set, which had the highest representation from Globus Medical (approximately 33%). Consequently, chance guessing would result in 33% accuracy despite including five manufacturers.

Reader Study

To assess the clinical utility of our model, we compared our results with the current practice of manufacturer identification. Our model performed with 79% accuracy on the 100-image test set of five different models and completed classification in 14 seconds. Two surgeons from our institution and three manufacturer representatives not affiliated with our institution completed the reader study. Their average accuracy was 44% (range 23%–66%), and the average time of quiz completion was 20 minutes (range 14–32 minutes). Random guessing by the representatives would have resulted in 20% accuracy, given that the test set included the five manufacturers in near-equal distribution. Next, we compared how individual manufacturer representatives performed on hardware they were familiar with compared with unknown hardware. Representative A (overall performance: 66%; completion time: 14 minutes) was familiar with Globus Medical and Medtronic hardware and performed better on those devices than on the three other devices (75% vs 60%). Representative B (overall performance: 23%; completion time: 17 minutes) was

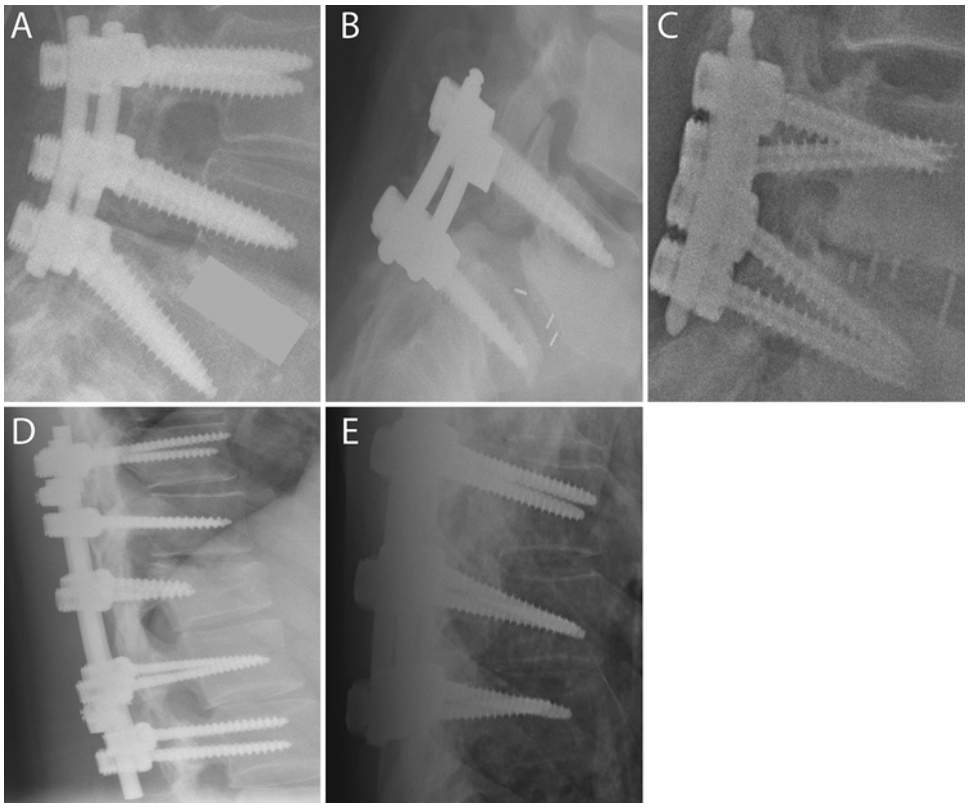

FIG. 3. Radiographs of the included hardware. **A:** Globus Medical Creo. **B:** Medtronic Solera. **C:** NuVasive Reline. **D:** Stryker Xia. **E:** DePuy Expedium. Interbody cages have been covered.

familiar with Globus Medical and performed better on Globus Medical hardware than on the other four hardware types (36% vs 17%). Finally, representative C (overall performance: 52%; completion time: 17 minutes) was familiar with Globus Medical and NuVasive but performed equally on familiar and unfamiliar hardware (53% vs 50%).

Discussion

We present a computer vision machine learning approach for identifying manufacturers of thoracolumbar screw-and-rod systems. This machine learning approach demonstrated good accuracy among five manufacturers in three radiographic views and showed superiority compared with expert human review in both accuracy and efficiency.

Our model performed well in five-level classification across all three image types. Notably, classification accuracy was significantly higher for lateral and fused images than for AP images. The higher accuracy for those two image types is likely because the feature detector is identifying features such as the screw diameter, frequency of screw threads (pitch), screw tapering at the end of the screw, and the screw-rod junction as salient (Fig. 5). On AP images, these features are not visible except for the screw-rod junction. Despite the lack of salient features, the accuracy for AP images was higher than chance. Additionally, there was a decrease in accuracy within specific manufacturer classification consistent with the number of images of that manufacturer included. For example, for lateral images, Globus Medical had the highest accuracy, followed by Medtronic, then NuVasive and Stryker, which were similar in accuracy, and finally DePuy with the lowest accuracy (Table 1 and Fig. 4). Furthermore, in cases

TABLE 2. Model performance for lateral, AP, and both images for two-, three-, four-, and five-way classification

	Lateral	AP	Fused
Globus Medical vs Medtronic	93.15 ± 4.06%	88.98 ± 4.08%**	91.08 ± 5.30%*
Globus Medical vs Medtronic vs NuVasive	82.35 ± 5.09%	71.79 ± 5.70%**	81.98 ± 4.80%
Globus Medical vs Medtronic vs NuVasive vs Stryker	71.51 ± 5.78%	66.65 ± 4.89%**	72.86 ± 5.51%
Globus Medical vs Medtronic vs NuVasive vs Stryker vs DePuy	64.27 ± 5.13%	60.95 ± 5.52%**	65.90 ± 5.14%

Values represent mean accuracy ± SD of the mean over 100 iterations.
* $p < 0.05$ compared with the lateral group. **$p < 0.001$ compared with the lateral group.

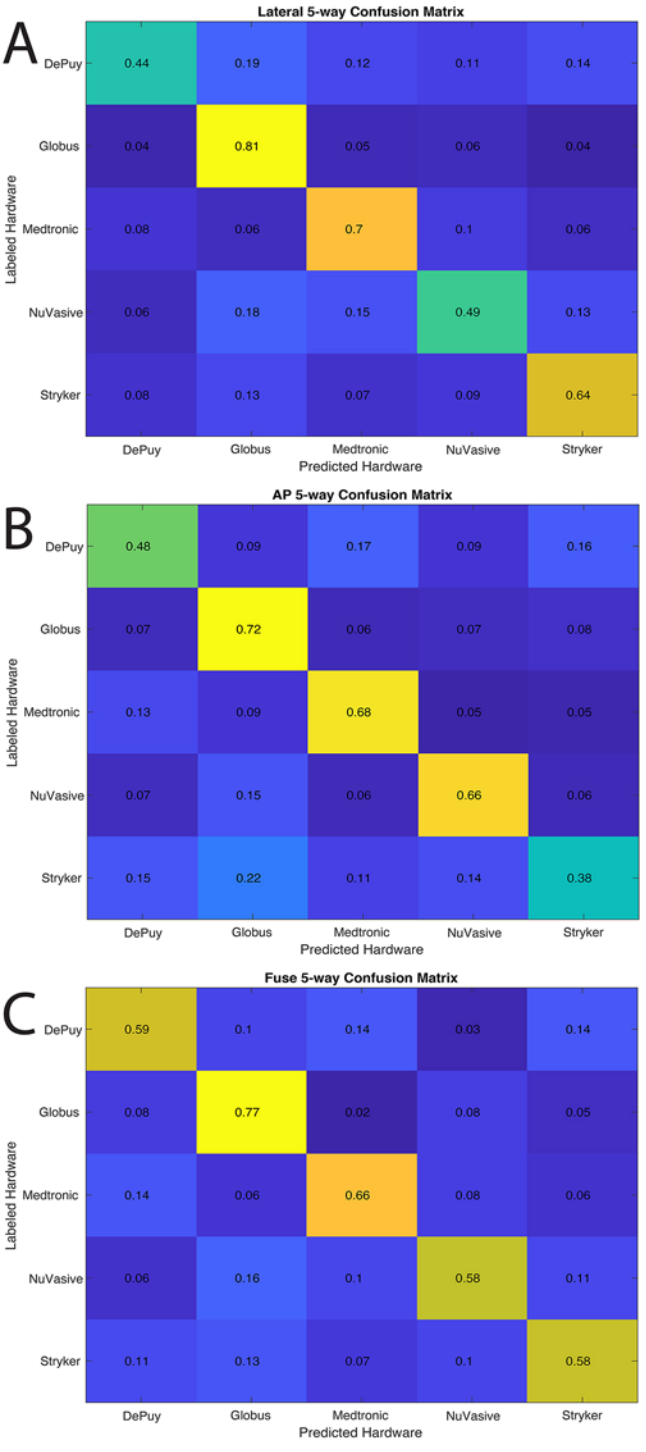

FIG. 4. Confusion matrices for five-level classification. **A:** Lateral images. **B:** AP images. **C:** Fused images. The y-axis indicates the image labels, and the x-axis indicates the predicted hardware. Values range from 0 to 1 and represent the fraction assigned to that category. The values are means over 100 iterations. The values in each row should sum to approximately 1. Figure is available in color online only.

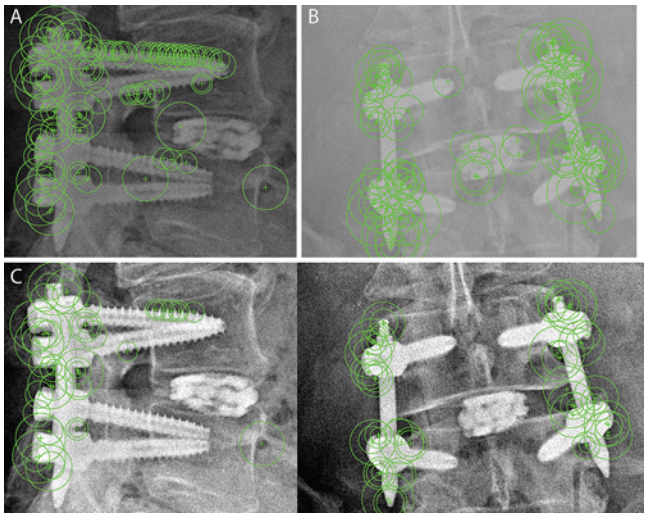

FIG. 5. Examples of KAZE feature detection on different image types. **A:** Lateral image. **B:** AP image. **C:** Fused image. *Green crosses* represent salient features, and *green circles* represent the strength of the feature. Smaller circles represent stronger features, while larger circles indicate weaker features. The top 100 features are plotted on the image. Figure is available in color online only.

of misclassification, the model was biased and misclassified images as Globus Medical or Medtronic more so than the three other manufacturers. The drop in accuracy with fewer images in the model and the bias in misclassification indicate that the model is biased by how many images of each model were included in the data set. If the model had to choose between Globus Medical and DePuy and was unsure, it erred on choosing Globus Medical, solely because it had a higher likelihood of being Globus Medical since there were more images in the model. At worst, the model would perform at an accuracy that is equal to the percentage of the most represented manufacturer in the test set. In Fig. 4 and Table 2, the test set had the same distribution as the training set (approximately 33% Globus Medical), so chance guessing would result in 33% accuracy; however, in the reader study, the test set consisted of close to even distribution of all five manufacturers, so guessing would result in approximately 20% accuracy. Equalizing the number of included images across manufacturers will likely improve the overall accuracy for less-represented manufacturers and reduce the misclassification bias.

Independent of misclassification bias, errors in classification may be due to variability in individual images. For example, there was no standardization in image acquisition parameters, so images containing the same hardware may have different contrast levels, even for AP and lateral images of the same exact hardware. This variability may make it difficult to clearly see the screw pitch consistently or distinguish the tapering of the screw from surrounding bone. Additionally, the angle of the screws in the bone is variable within the same type of hardware. This introduces bias because in some images in which the screws are parallel, they may appear thicker on a lateral image, since we cannot tell which screw is right or left and consequently the screws may appear on top of each other. In

cases in which the right and left screws are not parallel, the image has more salient data points to extract features (Fig. 3). Despite these sources of variability, the KAZE feature detector is good at ignoring size and rotational differences for feature extraction.

Recently, machine learning has gained wider use in various healthcare fields, especially neurosurgery.[7,8,13,14] Primarily, these advanced statistical techniques are used to examine large sets of clinical data and predict outcomes after surgery or for diagnosis in neuroimaging.[3,15] There has been limited application of computer vision in surgical planning.[16] Recently, three studies have applied computer vision techniques for hardware identification for the planning of revision surgery.[9,10,17] Two of the studies used computer vision SVMs and deep learning to develop classifiers to identify anterior cervical discectomy and fusion hardware, showing good results.[10,17] The third study, similar to our study, aimed to identify thoracolumbar screw-and-rod systems; however, they utilized deep learning, demonstrating classifier accuracy ranging from 73% to 98% for five-level classification.[9]

Although Yang et al. built an automated process to identify posterior thoracolumbar screws and rods, further exploration of the topic is still necessary.[9] The prior study only included images with one spinal level instrumented, which may limit the applicability of predictive models for larger instrumentation surgeries. Additionally, they excluded images that contained other types of hardware, which may serve as distracting information. Training a classifier that is impervious to "visual noise" is a useful feature. In our study, we included all examples of thoracolumbar screws and rods, regardless of the levels instrumented or the presence of overlapping hardware. Notably, in some cases, we occluded extraneous hardware that was not overlapping with the desired hardware.

The study by Yang et al. used deep learning, while we used SVMs for classification. Recently, neural nets and deep learning have become more common in machine learning applications as data sets become larger and more complicated.[18] One major disadvantage of these methods is that they are often seen as "black boxes" in which an image is input and an output class is obtained, without really understanding the complex computations used in classification. These classifiers are often widely generalizable but come at the cost of explainability, which is critical for clinical applications.[13] Our algorithm uses a KAZE feature detector, which allows detection of salient features on radiographs prior to model training and classification. With these two distinct steps, we can develop distinct radiographic profiles for each type of screw and rod and understand how the classifier is representing each image. For example, in Fig. 5, the feature detector determines the frequency of the screw threads (or pitch), tapering at the end of the screw, and the junction between the screw and the rod as salient features on lateral radiographs. Additionally, Yang et al. utilized a data set of nearly 1500 (300/class) images with even distributions among all classes for both lateral and AP radiographs, while our data set consisted of approximately 400 (25–100/class) unevenly distributed images for each view (Table 1). Differences in data set size, image distribution, and machine learning algorithms all contribute to the difference in performance observed between the current study and that by Yang et al.[9]

Despite a lower accuracy than in prior studies, we believe that the current study still has clinical utility. Our reader study demonstrated that expert human review by surgeons and manufacturer representatives resulted in an accuracy rate of 44% on a 100-image test set, while our model performed at 79% accuracy on the same test set. This surgeon-representative blinded assessment most accurately captures the real-world workflow for surgeons preparing for revision of pedicle screw constructs. Additionally, our model performed classification in less than 1 minute, while an average expert required 20 minutes for 100 images. Furthermore, our model outperformed manufacturer representatives on their own company's hardware. The greater accuracy and efficiency of our model indicate that our methodology can be used in conjunction with human review to expedite the surgical planning process. The current workflow of revision surgery planning consists of surgeons communicating with multiple manufacturer representatives to eventually identify the implanted screw or using some type of crowdsourced method. This practice is slow and based on guesswork (unless operative reports can be located). Our method takes a data-driven approach to reduce the time and uncertainty of revision surgery planning.

Limitations and Future Work

We demonstrated promising results and a robust study design; however, many improvements must be made prior to clinical use. Primarily, improvement in accuracy is imperative. Since this study was limited to a single institution, we had access to a limited set of manufacturers and a low number of images. Expanding our image database to other institutions will allow for more images and manufacturers, which will provide more cases for the algorithm to use for learning. We observed a close to 10% decrease in accuracy with the addition of each manufacturer. This was likely due to an uneven distribution of images for each included manufacturer as well as overall low numbers of included images. The most represented manufacturer had close to 100 images, while the least represented had just over 50 images. Typical computer vision algorithms require several hundred images per each included class. In clinical practice, there are dozens of possible vendors, so a clinically useful algorithm must have strong results despite adding many more manufacturers. If we were to add more manufacturers to our model without balancing the distribution of images for each manufacturer, then the algorithm would likely classify at a lower accuracy for each added manufacturer. If the distribution is even and many more images were added for each manufacturer, then the algorithm would be expected to perform at a high accuracy. Often surgeons have difficulty identifying instrumentation that is not common or is obsolete, which, currently, our algorithm would not help. By expanding this study to other institutions, we can improve our model and broaden its use for identifying less common or more obsolete models. In our study, we excluded rarely used instrumentation, which limits the model's generalizability. Additionally, developing an application in which the user can upload the

raw image and get an immediate result would allow for ease of use in a clinical setting. Once the two prior steps are completed, we can effectively assess the prospective use of computer vision technology in our clinical practice.

Conclusions

We developed a computer vision model that successfully classified posterior thoracolumbar hardware at five-level classification for the following manufacturers: Globus Medical Creo, Medtronic Solera, NuVasive Reline, Stryker Xia, and DePuy Expedium. Our model performed more accurately and efficiently than the method currently used in clinical practice. We utilized a KAZE feature detector and an SVM learning algorithm, which provided a simple methodological workflow from input to output. The relative computational simplicity of our model may help facilitate future studies by prospectively analyzing the efficacy of machine learning in clinical settings.

References

1. Celtikci E. A systematic review on machine learning in neurosurgery: the future of decision-making in patient care. *Turk Neurosurg*. 2018;28(2):167-173.
2. Khan O, Badhiwala JH, Grasso G, Fehlings MG. Use of machine learning and artificial intelligence to drive personalized medicine approaches for spine care. *World Neurosurg*. 2020;140:512-518.
3. Zaharchuk G, Gong E, Wintermark M, Rubin D, Langlotz CP. Deep learning in neuroradiology. *AJNR Am J Neuroradiol*. 2018;39(10):1776-1784.
4. Li Z, Zhang X, Müller H, Zhang S. Large-scale retrieval for medical image analytics: a comprehensive review. *Med Image Anal*. 2018;43:66-84.
5. Arora A, Lin JJ, Gasperian A, et al. Comparison of logistic regression, support vector machines, and deep learning classifiers for predicting memory encoding success using human intracranial EEG recordings. *J Neural Eng*. 2018;15(6):066028.
6. Awais M, Chiari L, Ihlen EAF, Helbostad JL, Palmerini L. Classical machine learning versus deep learning for the older adults free-living activity classification. *Sensors (Basel)*. 2021;21(14):4669.
7. Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev*. 2020;43(5):1235-1253.
8. Senders JT, Staples PC, Karhade AV, et al. Machine learning and neurosurgical outcome prediction: a systematic review. *World Neurosurg*. 2018;109:476-486.e1.
9. Yang HS, Kim KR, Kim S, Park JY. Deep learning application in spinal implant identification. *Spine (Phila Pa 1976)*. 2021;46(5):E318-E324.
10. Huang KT, Silva MA, See AP, et al. A computer vision approach to identifying the manufacturer and model of anterior cervical spinal hardware. *J Neurosurg Spine*. 2019;31(6):844-850.
11. Okawa M. Synergy of foreground–background images for feature extraction: offline signature verification using Fisher vector with fused KAZE features. *Pattern Recognit*. 2018;79:480-489.
12. Zhong Z, Zhu J, Hoi SCH. Fast object retrieval using direct spatial matching. *IEEE Trans Multimed*. 2015;17(8):1391-1397.
13. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24-29.
14. Staartjes VE, Stumpo V, Kernbach JM, et al. Machine learning in neurosurgery: a global survey. *Acta Neurochir (Wien)*. 2020;162(12):3081-3091.
15. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*. 2018;392(10162):2388-2396.
16. Mandal S, Greenblatt AB, An J. Imaging intelligence: AI is transforming medical imaging across the imaging spectrum. *IEEE Pulse*. 2018;9(5):16-24.
17. Schwartz JT, Valliani AA, Arvind V, et al. Identification of anterior cervical spinal instrumentation using a smartphone application powered by machine learning. *Spine (Phila Pa 1976)*. 2022;47(9):E407-E414.
18. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88.

Disclosures

Mr. Gadot reported a patent for 63/374,877 pending. Dr. Xu reported personal fees from NuVasive and Providence Medical Technologies, and grants from Medtronic outside the submitted work. Dr. Ropper reported personal fees from Globus Medical and Stryker outside the submitted work; in addition, Dr. Ropper had a patent for computer vision systems and methods for classifying implanted thoracolumbar pedicle screws pending.

Author Contributions

Conception and design: Ropper, Anand, Flores, Gadot, Xu. Acquisition of data: all authors. Analysis and interpretation of data: Anand, Flores, McDonald, Gadot, Xu. Drafting the article: Ropper, Anand, Flores, Xu. Critically revising the article: Ropper, Anand, Flores, McDonald, Gadot. Reviewed submitted version of manuscript: Ropper, Anand, Flores, McDonald. Approved the final version of the manuscript on behalf of all authors: Ropper. Statistical analysis: Anand. Administrative/technical/material support: Ropper, Flores. Study supervision: Ropper, Flores.

Supplemental Information

Previous Presentations

This work was presented as a poster at the 90th AANS Annual Scientific Meeting, Philadelphia, Pennsylvania, May 1, 2022.

Correspondence

Alexander E. Ropper: Baylor College of Medicine, Houston, TX. alexander.ropper@bcm.edu.