Automatic breach detection during spine pedicle drilling based on vibroacoustic sensing

Massalimova, Aidana ; Timmermans, Maikel ; Cavalcanti, Nicola ; Suter, Daniel ; Seibold, Matthias ; Carrillo, Fabio ; Laux, Christoph J ; Sutter, Reto ; Farshad, Mazda ; Denis, Kathleen ; Fürnstahl, Philipp

Abstract: Pedicle drilling is a complex and critical spinal surgery task. Detecting breach or penetration of the surgical tool to the cortical wall during pilot-hole drilling is essential to avoid damage to vital anatomical structures adjacent to the pedicle, such as the spinal cord, blood vessels, and nerves. Currently, the guidance of pedicle drilling is done using image-guided methods that are radiation intensive and limited to the preoperative information. This work proposes a new radiation-free breach detection algorithm leveraging a non-visual sensor setup in combination with deep learning approach. Multiple vibroacoustic sensors, such as a contact microphone, a free-field microphone, a tri-axial accelerometer, a uni-axial accelerometer, and an optical tracking system were integrated into the setup. Data were collected on four cadaveric human spines, ranging from L5 to T10. An experienced spine surgeon drilled the pedicles relying on optical navigation. A new automatic labeling method based on the tracking data was introduced. Labeled data was subsequently fed to the network in mel-spectrograms, classifying the data into breach and non-breach. Different sensor types, sensor positioning, and their combinations were evaluated. The best results in breach recall for individual sensors could be achieved using contact microphones attached to the dorsal skin (85.8%) and uni-axial accelerometers clamped to the spinous process of the drilled vertebra (81.0%). The best-performing data fusion model combined the latter two sensors with a breach recall of 98%. The proposed method shows the great potential of non-visual sensor fusion for avoiding screw misplacement and accidental bone breaches during pedicle drilling and could be extended to further surgical applications.

Research paper

Automatic breach detection during spine pedicle drilling based on vibroacoustic sensing

Aidana Massalimova [a,1], Maikel Timmermans [b,*,1], Nicola Cavalcanti [a], Daniel Suter [a], Matthias Seibold [a], Fabio Carrillo [a], Christoph J. Laux [c], Reto Sutter [d], Mazda Farshad [c], Kathleen Denis [b], Philipp Fürnstahl [a]

[a] *Research in Orthopedic Computer Science (ROCS), University Hospital Balgrist, University of Zurich, Zurich, 8008, Switzerland*
[b] *KU Leuven, Department of Mechanical Engineering, BioMechanics (BMe), Smart Instrumentation Group, Leuven, 3001, Belgium*
[c] *Department of Orthopaedics, Balgrist University Hospital, University of Zurich, Zurich, 8008, Switzerland*
[d] *Department of Radiology, Balgrist University Hospital, Zurich, 8008, Switzerland*

ARTICLE INFO

ABSTRACT

Pedicle drilling is a complex and critical spinal surgery task. Detecting breach or penetration of the surgical tool to the cortical wall during pilot-hole drilling is essential to avoid damage to vital anatomical structures adjacent to the pedicle, such as the spinal cord, blood vessels, and nerves. Currently, the guidance of pedicle drilling is done using image-guided methods that are radiation intensive and limited to the preoperative information. This work proposes a new radiation-free breach detection algorithm leveraging a non-visual sensor setup in combination with deep learning approach. Multiple vibroacoustic sensors, such as a contact microphone, a free-field microphone, a tri-axial accelerometer, a uni-axial accelerometer, and an optical tracking system were integrated into the setup. Data were collected on four cadaveric human spines, ranging from L5 to T10. An experienced spine surgeon drilled the pedicles relying on optical navigation. A new automatic labeling method based on the tracking data was introduced. Labeled data was subsequently fed to the network in mel-spectrograms, classifying the data into breach and non-breach. Different sensor types, sensor positioning, and their combinations were evaluated. The best results in breach recall for individual sensors could be achieved using contact microphones attached to the dorsal skin (85.8%) and uni-axial accelerometers clamped to the spinous process of the drilled vertebra (81.0%). The best-performing data fusion model combined the latter two sensors with a breach recall of 98%. The proposed method shows the great potential of non-visual sensor fusion for avoiding screw misplacement and accidental bone breaches during pedicle drilling and could be extended to further surgical applications.

1. Introduction

Pedicle screw placement is a standard procedure in spinal surgeries performed in cases of scoliosis, trauma, spinal tumors, and degenerative spinal pathologies. However, due to adjacent vital anatomical structures, sub-optimal pedicle screw positioning can lead to neurological and vascular injuries. It was reported that the rate of pedicle screw malpositioning reaches up to 41% in the lumbar spine and 55% in the thoracic spine [1]. The optimal screw positioning is achieved when screws have maximum diameter and length within the pedicle, do not breach the pedicle's cortical layer or vertebral body, and follow a converging trajectory usually parallel to the upper endplate. In conventional freehand techniques, a profound understanding of the anatomical landmarks is essential to identify the entry point. After finding the entry point, a surgical drill or a probe is used to create the pilot hole inside the pedicle. This technique demands thorough anatomical knowledge and experience-based judgment from tactile feedback to avoid pedicle cortical wall perforation. The purpose of present study is to detect cortical wall perforation using a convolutional neural network (CNN) trained on vibroacoustic sensor data and to investigate whether data fusion from different sensors improves the breach detection rate during pedicle drilling.

With the advances in imaging and robotic technologies, imaging-based and robot-assisted techniques have been developed to navigate the surgeon during pedicle screw positioning and prevent pedicle

---

breaches. Several reviews show improved outcome for fluoroscopy-guided and optical marker-based navigation, both for robot-assisted and freehand operations [2,3]. However, the fluoroscopy-guided method is characterized by higher radiation exposure [1,4], and non-negligible patient movement during the operation introduces additional errors to optical marker-based navigation. These issues were tackled by a smart commercial surgical instrument called PediGuard Probe (SpineGuard, Vincennes, France). It enables detection of transition from one tissue to another by measuring the local electrical conductivity of the tissue in contact with the probe tip. When tested in combination with fluoroscopy by Chaput et al. and Bolger et al. the accuracy of the procedure remained 98% but the number of fluoroscopy shots reduced by 30% [5,6]. Guillen et al. [7] tested the PediGuard probe without any fluoroscopic guidance and achieved 90.06% accuracy. These results will be used as a target or benchmark for our proposed breach detection method.

Previously, we conducted a systematic study on the existing intra-operative tissue classification methods in orthopedic and neurological studies [8], where we reported the potential of sensor fusion and vibroacoustic sensing with transducers such as microphones and accelerometers in differentiating cortical and cancellous bone layers. Besides, the non-visual nature of the vibroacoustic sensing alleviates the limitation to the preoperative planning and exhibits less radiation exposure than the methods above. The use of microphones has been extensively investigated for this purpose [9–13].

A mel-spectrogram is a visual representation of the audio signal's amplitude over time across different frequency bins in the mel scale. The latter is a scale proportional to the (by humans) perceived difference, due to the cochlear anatomy, in magnitude of different frequencies that are factually at the same magnitude [14]. Several studies have shown how mel-spectrograms can be used for audio event detection and classification in a variety of applications [15–17]. Purwins et al. [18] reported in his review that the deep learning methods trained on mel-spectrograms have a comparable performance on a less amount of data as a method trained on raw audio data. Whereas, raw waveforms provide better exploitation of the deep learning methods without hand-crafted features, but with more data. Furthermore, Seibold et al. [19] showed that a convolutional neural network, using mel-spectrograms as input, can detect the breakthrough event during a femur drilling process with an accuracy of over 90% in a total execution time of less than 140 ms. Therefore, we propose a novel method of detecting pedicle breaches during the pedicle drilling step in spine surgery by leveraging the power of vibroacoustic sensing and CNNs. A breach event is considered as the penetration of the surgical instrument from the cancellous layer to the adjacent cortical of the vertebra. This study aims to analyze the performance of the CNN with different vibroacoustic sensors and to investigate whether and to which extent data fusion of those different sensors improves the breach detection rate during pedicle drilling. We hypothesize that the performance of the CNN can be boosted in the case of sensor fusion since they might provide complementary information. The following contributions can substantiate the novelty of this study:

1. Our work introduces an automatic labeling approach based on tracking data from the optical navigation system for non-visual signal data that greatly accelerates data collection for machine learning.
2. Multiple vibroacoustic sensors, including a contact microphone, a free-field microphone, a tri-axial accelerometer, and a uni-axial accelerometer, were integrated into the setup such that we could systematically investigate vibroacoustic sensing technologies both individually and fused.
3. A breach detection algorithm was developed by leveraging sensor fusion and a squeeze-and-excitation neural network.
4. The data collection was performed by conducting a series of ex-vivo cadaveric experiments to facilitate close to the actual clinical application scenario.

5. The concept of transfer learning in the mel-spectrogram domain was tested for the first time using pre-trained weights from a previous study and re-training the network on vibroacoustic sensor data.

2. Methods

The proposed method includes a definition of drilling trajectories, data collection, data processing, and deep learning, as shown in Fig. 1. Definition of drilling trajectories was performed first as described in Section 2.1 to simulate different breaches during pedicle drilling, which was later visualized in the surgical navigation setup (Section 2.2.2). Vibroacoustic data from contact microphones (Mic_1), a free-field microphone (Mic_2), uni-axial accelerometers (PCB_1 and PCB_2) and a tri-axial accelerometer (PCB_3) were collected during pedicle drilling along with surgical navigation data (Section 2.2). Collected data were processed (Section 2.3) and used to develop a CNN for breach detection in Section 2.4.

2.1. Definition of drilling trajectories

Four fresh-frozen human cadaver spines from T12 to coccyx were acquired from Science Care (Phoenix, USA) and used for data collection. Ethical approval was obtained from the ethical committee for conducting this cadaveric study (ID: 2021-01196). The spines were fixed onto a wooden board using surgical pins with a diameter of 3 mm to provide stability to the anatomy during experiments. The soft tissue on top of the sacrum was removed. A marker with four converging canals was designed, 3D-printed, and attached to the sacrum of each specimen to track relative motion between anatomy and marker. CT scans of each specimen with a marker and infrared reflecting spheres were taken with a slice thickness of 1 mm (SOMATOM Edge Plus, Siemens Healthcare, Erlangen, Germany). Afterward, all vertebrae, markers, and spheres were segmented and converted into 3D triangular surface models using the Materialise Mimics Innovation Suite software (version 19.0, Materialise NV, Leuven, Belgium). Drilling trajectories were planned on each spine, from T10 to L5 levels, and on both sides of each vertebra using the surgery planning tool of our institution. Trajectories were planned by resident surgeons (NC and DS) and confirmed by an experienced spine surgeon (CL). Along with optimal or breach-free trajectories, laterally- and medially-breaching trajectories were planned as shown in Fig. 2. In all cases, the anterior cortical wall of the vertebral body was deliberately breached.

Cadavers were thawed a day previous to the experiment, and the surgical approach was performed by a surgeon, providing midline access to the posterior elements of the thoracolumbar spine.

2.2. Data collection

2.2.1. Sensors

Fig. 3 shows all sensors in contact with the sample, the free-field microphone above the sample on the left, and the tri-axial accelerometer attached to the drill on the right. Three custom piezo-electric contact microphones (Mic_1) were attached to the skin directly above bony structures near the incision site using kinesiology tape and connected to an analog/digital converter (PreSonus Studio 68, PreSonus Electronics, Inc., Baton Rouge, LA, USA) through a 48 V phantom-powered impedance matching circuit,[2] to tackle the impedance mismatch issue (Fig. 3a). A PCB free-field condenser microphone (Mic_2) of model 378B02 (PCB Piezotronics, Depew, NY, USA) was placed at a distance of approximately ten to twenty centimeters from the drilling site using a standard tripod, as shown in Fig. 3c, to capture acoustic variations in
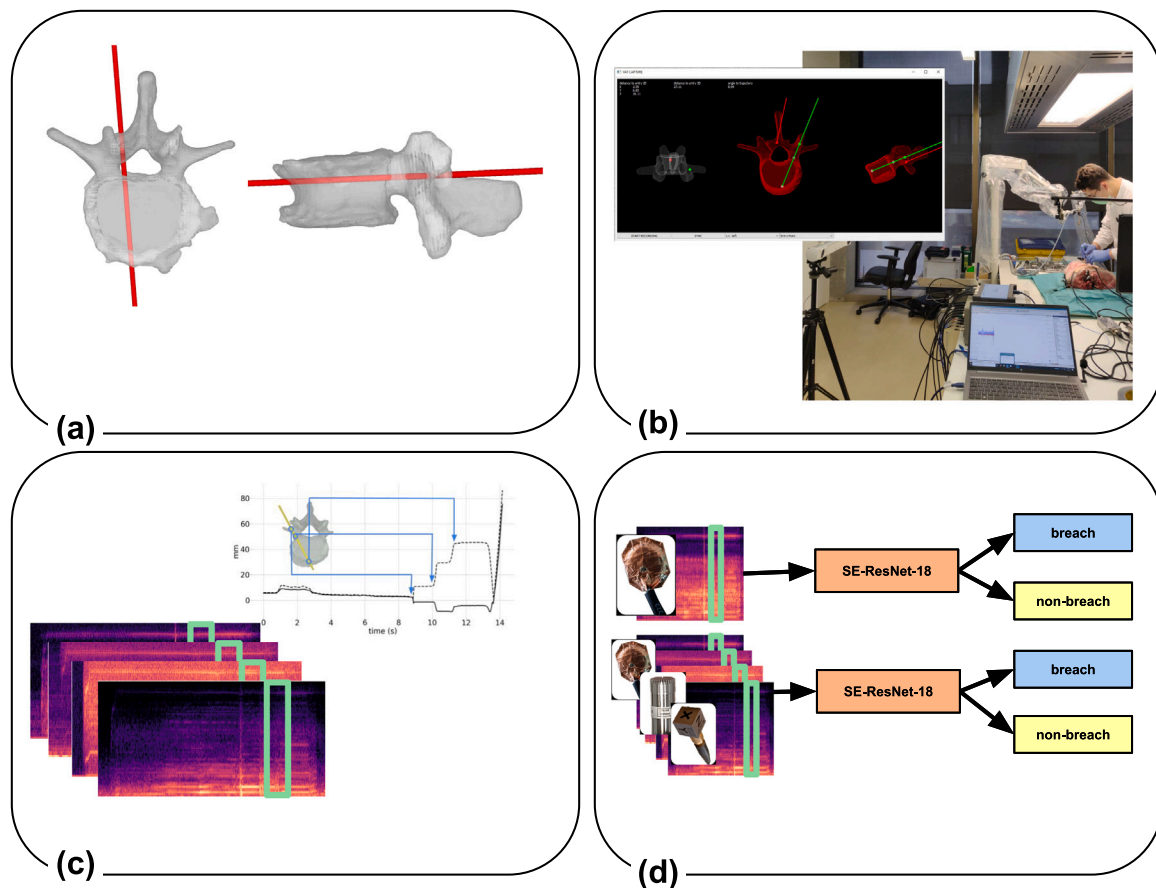

Fig. 1. Overall pipeline of the proposed method for breach detection during pedicle drilling: (a) definition of drilling trajectories in dedicated software, (b) drilling by experienced surgeon and data collection, (c) processing of the vibroacoustic data and the optical tracking data, (d) training the deep learning model on individual sensor data and fused sensor combination data.

bone properties similar to the work of Goossens et al. [20]. One PCB uni-axial lightweight accelerometer (PCB_1) of model 352A24 (PCB Piezotronics, Depew, NY, USA) was attached to the spinous process of the vertebra being drilled, utilizing an in-house manufactured 3D-printed clamp (Fig. 3b). Another uni-axial accelerometer (PCB_2) of the same model was attached to a surgical pin (Fig. 3b), and drilled through the sacral bone to fix the spine to the wooden board. On the surgical power drill, a tri-axial accelerometer (PCB_3) of model J356A45 (PCB Piezotronics, Depew, NY, USA) was glued with beeswax and taped over to prevent accidental detachment (Fig. 3d). All PCB sensors were connected to a custom signal conditioner via BNC cables, and the signal was put through to a PicoScope 4824A oscilloscope (Pico Technology, Cambridgeshire, UK) for data recording.

The recording of all sensors was started, and the surgeon performed the entire drilling procedure for the selected pedicle at the maximum drilling speed of 3500 rpm. The surgeon started the drilling process guided by the surgical navigation system, as described in the next section. Post-operative CT scans were analogous to the preoperative CT scans after each experiment to compare the drilled trajectories marked with k-wires with the planned trajectories.

2.2.2. Navigation setup

An optical tracking system (FusionTrack 500, Atracsys LLC, Puidoux, Switzerland) was used to precisely navigate the planned screw trajectories. Similar to the markers designed for the anatomy, markers for the surgical drill sleeve and for a power drill (Colibri, Depuy Synthes, Oberdorf, Switzerland) were modeled. The diameter of the drill bit was 3 mm. The drill sleeve was used to find the entry point and direction of the planned trajectory because it provides easier targeting

and stable drilling. At the same time, the drill marker allowed for tracking the progression of the drill in depth. Fig. 4 depicts the designed markers for the drill, drill sleeve, sacrum, and their frames (F_d, F_{ds} and F_{ct}), respectively. Markers were CT scanned with the infrared reflecting spheres attached. Each tool, sphere, and marker were segmented from preoperative CT scans similarly to the anatomy segmentation, and a 3D mesh per each object was generated. The segmented tools with spheres were used to extract the transformation matrices (T_d^{dt}, T_{ds}^{dst}) shown in Fig. 4 from the marker space to the corresponding tooltip space. Each marker was re-calibrated with FusionTrack 500 software to generate the new geometry file for the tracking. Geometry files contain information on the spatial position of each fiducial in the camera coordinate space, which was used to register 3D meshes of the anatomy or tool to the camera coordinate space using the iterative closest point (ICP) registration method for surgical navigation in the custom-developed software. The software is python-based and allows visualization of the tool position and orientation in real-time with planned entry points and trajectories on 3D meshes of the vertebra. A GUI has been implemented to show the three anatomical views (coronal, axial, and sagittal) of the drilled vertebra (Fig. 1b). In this manner, the surgeon could meticulously control the drilling procedure such that the drill is positioned in the correct location (i.e., the entry point at the pedicle) and orientation that would result in the planned drilling path. Moreover, the software can record tracking and vibroacoustic data in a synchronized fashion.

2.3. Data processing

Since all sensor modalities are synchronized, data labeling was performed based on the timestamps of the events retrieved from the
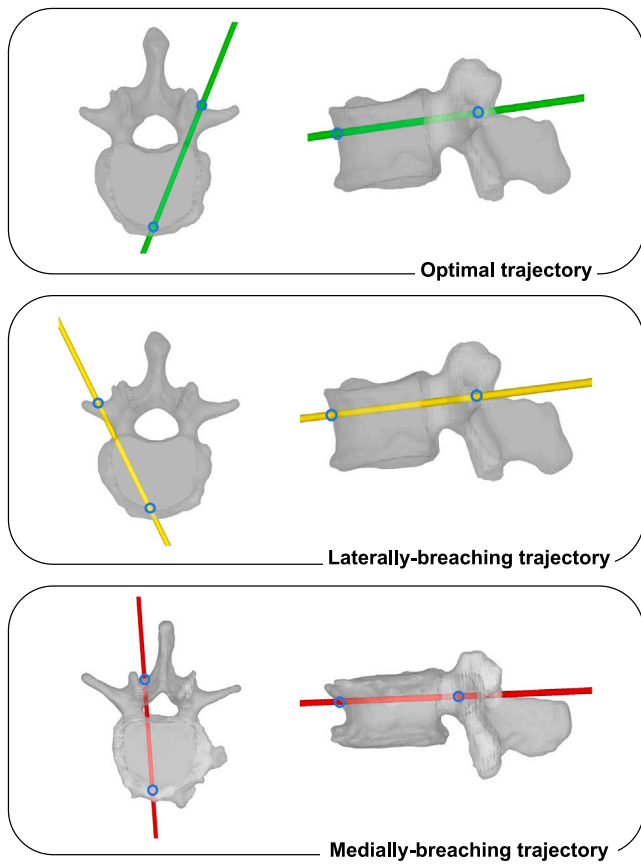
Fig. 2. Optimal, laterally-breaching, and medially-breaching trajectories overlaid on 3D meshes of the vertebra in axial and sagittal views.

Table 1
Data split per subject and per class before augmentation in each sensor dataset.

Subject	Train		val		Test	
	Breach	Non-breach	Breach	Non-breach	Breach	Non-breach
S1	16	314	5	78	3	45
S2	54	469	14	117	9	66
S3	32	392	8	98	5	55
S4	20	234	5	58	4	34

transforming it to the camera space using T_{CT}^{cam} as in Eq. (4).

$$P_{breach} = \arg\min_{P_{dt}}(\|P_{dt} - Mesh_v T_{CT}^{cam}\|) \tag{4}$$

The entry, exit, and breaching points are visually illustrated in Fig. 5 along with their corresponding points on the plots of distance from the drill tip position to the entry point and to mesh vertices in camera space. This visualization of the tracking data was also used to fine-tune timestamp labeling. On top of that, recorded tracking data was replayed, and timestamps of the breaching events were visually inspected thoroughly. For example, in Fig. 6, labels from Fig. 5 were overlaid in the same time points on mel-spectrograms from each vibroacoustic sensor using blue arrows.

Data from all sensors were processed similarly to keep the data format consistent across vibroacoustic sensors. Each recording per each sensor was broken down into smaller windows using the sliding window. The window length of 100 ms and window step size of 25 ms, leading to overlapping windows with an overlap of 75%, was chosen to facilitate lower latency. The number of mels, hop size, and maximum frequency included in the spectrogram were set to 128, 32 samples, and 2 kHz, respectively, to generate mel-spectrograms. Table 1 shows the number of window samples per class. However, we could note the variability in the number of breach windows of 100 ms per subject. Breach duration is influenced by factors such as breach type, operator technique, drilling force, and bone quality. For example, drilling perpendicular to cortical bone results in immediate penetration, while drilling at an angle leads to gradual penetration until full breach. Additionally, because a complete breach of one bone layer can consist of small subsequent breaching events, the drilling force and bone quality affect breach duration as they can speed up or slow down such breaching sequences. Thus, these factors directly affected breach counts per subject.

However, the number of breach samples was considerably smaller, necessitating the application of data augmentation strategies. Pitch shifting and loudness changing are the standard audio signal transformations used for audio augmentation purposes [19,21]. Therefore, the training set of the breach class was augmented by varying the gain (−5 dB, −3 dB, 3 dB, 5 dB) and pitch shifting (−2, −1, 1, 2 semitones). All spectrograms were generated using *librosa 0.9.2* and have a size of $128 \times 126 \times 1$. They were further used as input to the deep learning network.

From visual inspection of CT scans, we noticed differences in bone mineral density (BMD). Therefore, we estimated BMD from the preoperative CT scans to investigate the correlation between BMD and the performance of the proposed method using the Materialise Mimics software. No density calibration was done for the CT images, but all preoperative scans were taken consecutively to minimize deviations. A constant elliptical area of 450 mm^2 was selected, from which the average Hounsfield unit (HU) was calculated. For each drilled vertebra, the area was swept through all axial CT slices of the vertebral body. All calculated HU values were averaged over the number of axial slices within the vertebral body to estimate the average BMD of the specified vertebra. In a study of Schreiber et al. [22], HU values calculated from CT scans with the Mimics software were correlated to T-scores generated from DEXA scans. Based on the results of that study, the HU values obtained in the present study were related to a value of the BMD,

tracking data. Timestamps were determined for each event to create two data subsets for each sensor: **breach** and **non-breach**. A breach event has a duration within a range of 100–300 ms. Following this concept, the closest drill tip positions to the vertebra mesh vertices were used to identify breach timestamps from the tracking data. First, the tracking data was mapped as follows to compensate for non-negligible errors resulting from the navigation setup. A postoperative CT was registered to the preoperative CT scan using ICP registration. Recorded tracking data from the anatomy marker was used to transform 3D meshes from postoperative CT and estimate the distances between the drill tip position and 3D mesh of the vertebra. Entry (EP_{CT}) and exit (SP_{CT}) points in the CT space were extracted by finding the intersection points between the vertebra and the surgical pin. Afterward, the entry point (EP) and exit point (SP) in the tracking data of the drill tip were estimated by following Eqs. (1)–(2), where P_{dt} is a set of drill tip position in camera space, T_{CT}^{cam} is a transformation matrix from CT space in $F_C T$ to camera space.

$$EP = \arg\min_{P_{dt}}(\|P_{dt} - EP_{CT} T_{CT}^{cam}\|) \tag{1}$$

$$SP = \arg\min_{P_{dt}}(\|P_{dt} - SP_{CT} T_{CT}^{cam}\|) \tag{2}$$

These two points were further used to map the tracked drill tip positions between entering and exiting timestamps to the straight line connecting the entry and exit points (Eq. (3)).

$$P_{dt} \mapsto \langle P_{dt}[t_{EP} : t_{SP}], EP - SP \rangle \tag{3}$$

To find the breaching point P_{breach}, the mapped data was used to find the closest points to the mesh vertices of the vertebra ($Mesh_v$) by

Fig. 3. Sensors positioning. Left: (a) contact microphones, (b) uni-axial accelerometers, and (c) free-field microphones. Right: (d) tri-axial accelerometer.
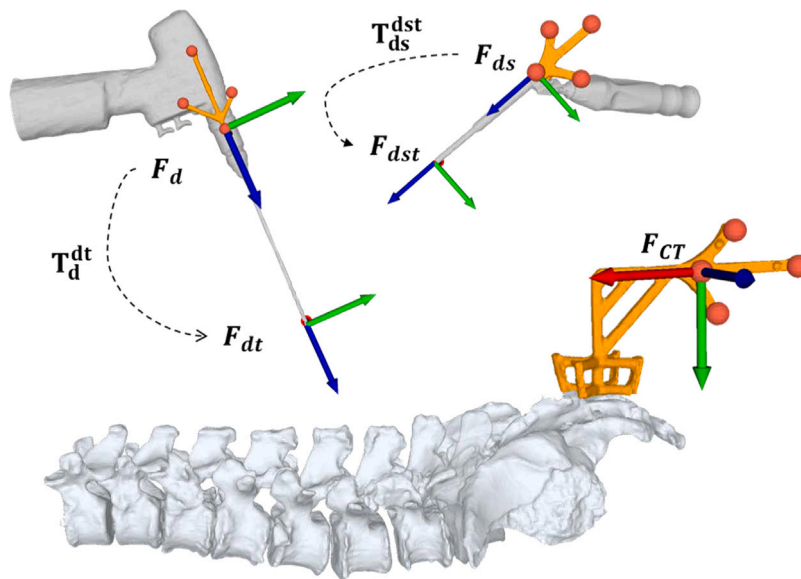

Fig. 4. Navigation setup overview with frames (F) and transformations (T). F_d and F_{ds} are the frames of markers attached to the drill and drill sleeve. F_{ct} is a marker frame attached to the anatomy representing CT space. F_{dt} and F_{dst} are tip frames of each tool, which are defined by applying respective transformation matrices (T_d^{dt}, T_{ds}^{dst}).
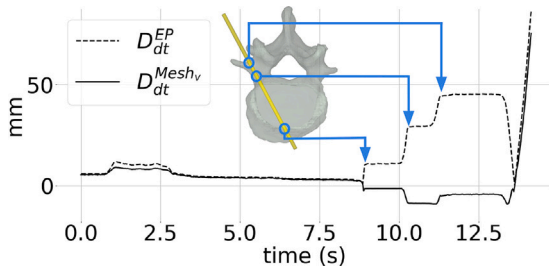

Fig. 5. The arrows indicate the correspondence of the start, breach, and stop events between a drill position in a segmented vertebra from postoperative CT and temporal tracking data information. D_{dt}^{EP} is the distance between the drill tip position and the entry point, whereas $D_{dt}^{Mesh_v}$ is the distance between the drill tip position and the closest vertex of the vertebra mesh.

Table 2
Test split describing the number of the normal and abnormal samples.

Subject	Normal		Abnormal	
	Breach	Non-breach	Breach	Non-breach
S1	2	7	1	38
S2	5	18	4	48
S3	0	5	5	50
S4	1	9	3	25

abnormal in the case of osteopenia or osteoporosis, and the data was classified as such (see Table 2).

2.4. Deep learning

This work used a residual network integrated with squeeze and excitation (SE) blocks, particularly SE-ResNet-18 as shown in Fig. 7. SE blocks enhance the performance of the ResNet networks widely used for classification tasks owing to their ability to model the interdependencies between the channels of convolutional features. The SE blocks can be integrated into standard architectures such as ResNet.
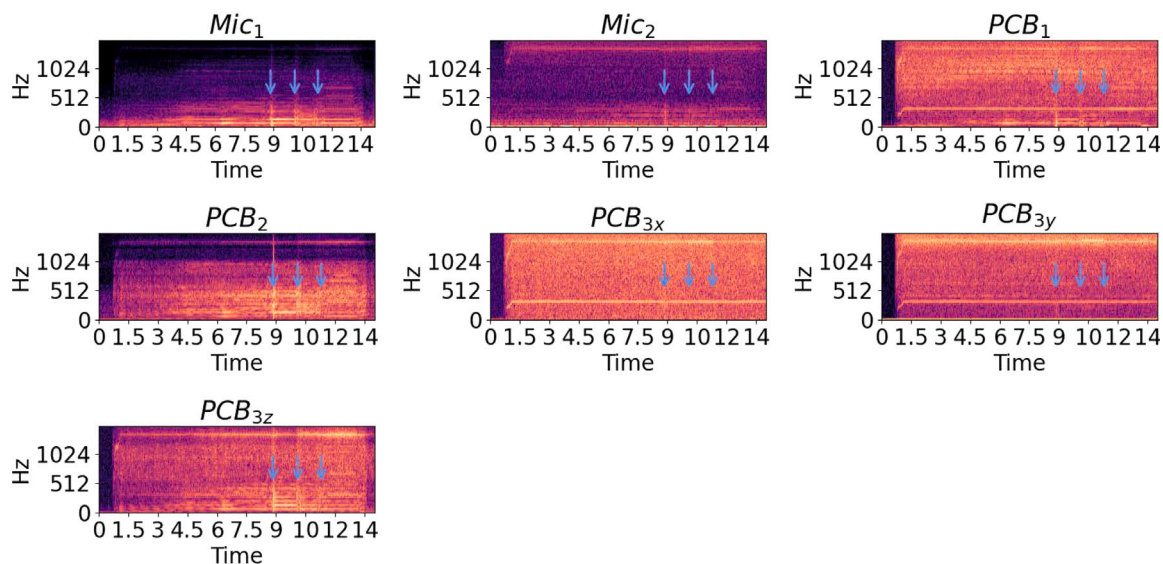
a range of T-scores, and a qualitative description by interpolation on the trend lines. Afterward, test sets were split into either normal or

Fig. 6. Mel-spectograms for the whole drilling session from each sensor overlaid with labels from tracking data (Subject S2, vertebra L4, left pedicle). PCB_{3x}, PCB_{3y}, and PCB_{3z} are separated x (axial), y (radial), and z-directional (tangential) values from the tri-axial accelerometer on the drill (PCB_3).
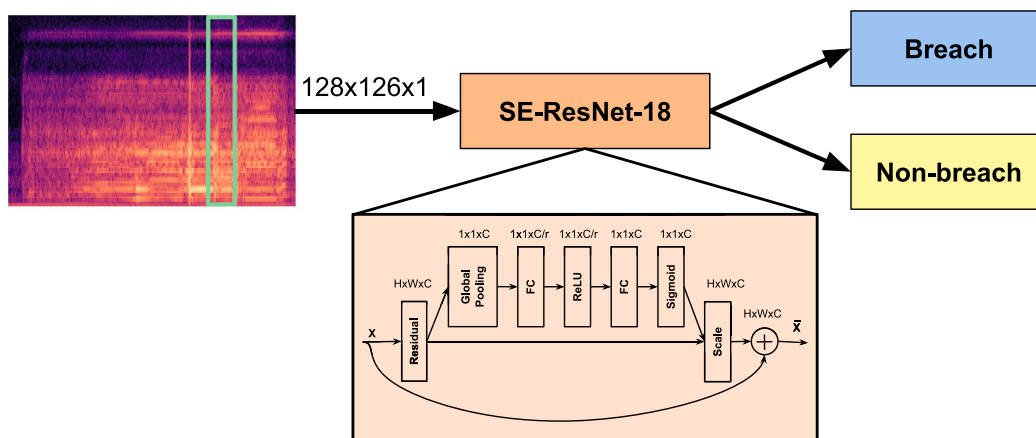

Fig. 7. Mel-spectrogram with a window size (indicated by box) of 100 ms is an input to the SE-ResNet-18, which outputs breach/non-breach. The topology of the squeeze and excitation block is shown in the bottom center box.

Hu et al. [23] reported that ResNet-50 with SE blocks (SE-ResNet-50) outperformed ResNet-50 in different classification tasks, e.g., by 0.86% single-crop top-5 error with ImageNet 2012 dataset. Since we aim to develop a robust and real-time system, the most compact model of ResNet with 18 layers was used to achieve the fastest inference time. Because of the dataset imbalance, we used the Focal Loss function as a loss function [19]. Further, the Adam optimizer was used with a learning rate of 1e−6. The training and testing were performed on an NVIDIA GeForce RTX 6000 GPU machine using *Tensorflow 2.4.1*. In this work, we used a nested cross-validation method to overcome the problem of overfitting the training dataset. We kept 10% of the dataset for testing, and the remaining 90% was divided into 5-folds. As mentioned in the previous section, each experiment was run five times per cross-validation. Thus, results are reported in terms of mean, standard deviation, and 95% confidence interval (CI) of breach recall (sensitivity).

We ran three different experiments: training the model from scratch based on individual sensor data (Experiment I), training the model using pre-trained weights (Experiment II), and training the model using fused data from different sensor combinations (Experiment III). The classification results are computed for each individual window independently. This means that the network output is obtained for each window separately, without averaging over a specific number of windows. By

employing this approach, we can capture temporal dynamics and variations within shorter time intervals. It allows us to detect and classify events accurately, enabling real-time monitoring and prompt response. In Experiment I, we additionally tested with the data from Table 4 to investigate whether the performance of each network is correlated with BMD. The flow diagram of the model is shown in Fig. 7. Experiment II was based on the pre-trained weights using the same model generated based on the data from Seibold et al. [19], and targeted to test the concept of transfer learning from the same domain. Experiments I and II were ablation studies that served as a foundation for finding the optimal sensor combinations in Experiment III, where the spectrograms from different varieties of sensors were concatenated or fused and used as an input to the network (Fig. 1d). The sensor combinations were formed by combining the two best-performing sensor modalities with the rest sensor modalities. The free-field microphones (Mic_2) were not included in this evaluation, as they had performed too poorly overall.

Apart from comparing experimental results in terms of mean, standard deviations, and 95% CI, we conducted statistical analyses. The performances of individual sensors-and sensor fusion-based models were first evaluated using the one-way analysis of variance (ANOVA) to test the hypothesis that there are statistically significant differences between multiple groups. Afterward, the best-performing model was compared to other models using a pair-wise t-test to verify if

Table 3
Performances of the individual sensors- and sensors fusion-based models. P values are from the pair-wise t-test results with respect to Mic_1 & PCB_1.

Sensor	Breach Re. (95% CI) [%]	Non-breach Re. [%]	P Value
Mic_1	**85.8 ± 3.19 (83.0–88.6)**	91.8 ± 2.60	<.001
Mic_2	63.8 ± 5.36 (59.1–68.5)	91.0 ± 3.39	<.001
PCB_1	81.0 ± 3.54 (77.9–84.1)	87.6 ± 7.02	<.001
PCB_2	79.8 ± 8.65 (72.2–87.4)	92.0 ± 0.71	.002
PCB_{3x}	75.2 ± 6.18 (69.8–80.6)	91.0 ± 1.22	<.001
PCB_{3y}	73.4 ± 7.96 (66.4–80.4)	87.6 ± 2.30	<.001
PCB_{3z}	77.2 ± 7.05 (71.0–83.3)	93.0 ± 1.58	<.001
Sensor combination			
Mic_1 & PCB_1	**98.0 ± 2.74 (95.6–100)**	96.2 ± 1.30	
Mic_1 & PCB_1 & PCB_{3z}	97.0 ± 4.47 (93.1–100)	96.0 ± 0.79	.681
Mic_1 & PCB_1 & PCB_2	94.0 ± 4.18 (90.3–97.7)	96.4 ± 0.55	.111
PCB_{3x} & PCB_{3y} & PCB_{3z}	84.8 ± 3.83 (81.4 −88.2)	84.6 ± 0.55	<.001
Mic_1 & PCB_1 & PCB_2 & PCB_{3x}	95.2 ± 5.72 (90.2–100)	97.8 ± 0.84	.352
Mic_1 & PCB_1 & PCB_2 & PCB_{3y}	96.4 ± 0.55 (97.0–100)	96.4 ± 0.55	.545
Mic_1 & PCB_1 & PCB_2 & PCB_{3z}	96.0 ± 2.24 (94.0–98.0)	96.9 ± 0.89	.242
Mic_1 & PCB_1 & PCB_2 & PCB_{3x} & PCB_{3y} & PCB_{3z}	92.0 ± 2.74 (89.6–94.4)	97.4 ± 0.55	.009

Table 4
Performances of individual sensors-based models with respect to the bone density. P values are from the pair-wise t-test between the results obtained from testing on normal and abnormal sets.

Sensor	Normal - Breach Re. (95% CI) [%]	Abnormal - Breach Re. (95% CI) [%]	P value
Mic_1	86.2 ± 6.22 (80.7–91.6)	85.2 ± 10.52 (76.0–94.4)	.859
Mic_2	58.6 ± 8.8 (50.9–66.3)	72.6 ± 18.5 (56.4–88.8)	.165
PCB_1	84.8 ± 5.31 (80.1–89.5)	75 ± 0.00 (75.0–75.0)	.003
PCB_2	72.6 ± 11.7 (62.4–82.8)	**92.8 ± 6.57 (87.0–98.6)**	.009
PCB_{3x}	**87.8 ± 3.83 (84.4–91.2)**	55.4 ± 11.24 (45.5–65.3)	<.001
PCB_{3y}	73.8 ± 11.5 (63.7–83.9)	72.8 ± 10.4 (63.7–81.9)	.889
PCB_{3z}	77.0 ± 8.0 (70.0–84.0)	77.8 ± 16.1 (63.6–92.0)	.923

the best-performing model was significantly better than the counterpart. Similarly, a pair-wise t-test was further used to compare each transfer learning-based model with its corresponding individual sensor-based model and the performances of individual sensors-based models with respect to BMD. P values <.05 are considered to be statistically significant.

3. Results

Experiment I showed that detection using contact microphones (Mic_1) showed the highest mean breach recall (85.8%) compared to other sensors (Table 3). In each table, the highest-performing sensor is indicated in bold. It was followed by another contact sensor (PCB_1) connected to the vertebra (81% average breach recall, 77.9–84.1% 95% CI). In contrast, a free-field microphone performed the worst with 63.8% average breach recall (59.1–68.5% 95% CI). Of the tri-axial accelerometer on the drill (PCB_3), the direction along the drilling axis (Z-axis) showed the highest performance (77.2% average breach recall, 71.0–83.3% 95% CI).

The inference of each model was tested with respect to BMD. According to Table 4, the results show significant differences in the models trained with PCB_1, PCB_2, and PCB_{3x}. PCB_1 and PCB_{3x} based models performed significantly better on the normal test set, whereas PCB_2 performed better on abnormal data (92.8% average breach recall, 87.0–98.6% 95% CI).

The follow-up experiment II results, investigating the power of transfer learning in the acoustic spectrogram domain, are presented in Table 5. PCB_1 showed significantly improved performance with 93.0% mean breach recall compared to the models trained from scratch. In contrast, using pre-trained weights on PCB_{3y} deteriorated its average breach recall from 73.4% to 62.8%. Meanwhile, there were no significant differences in the case of other sensors, e.g., contact and free-field microphones.

The main experimental results on sensor fusion were integrated into Table 3. Eight different combinations were tested in total. We first combined the best-performing sensors (Mic_1 and PCB_1). We achieved

Table 5
Performances of transfer learning-based models. P values are from the pair-wise t-test with respect to its corresponding model results trained from scratch.

Sensor	Breach Re. (95% CI) [%]	Non-breach Re. [%]	P value
Mic_1	85.6 ± 4.51 (81.7–89.6)	89.6 ± 1.66	.937
Mic_2	60.0 ± 5.70 (55.0–65.0)	88.2 ± 1.08	.309
PCB_1	**93.0 ± 2.74 (90.6–95.4)**	87.6 ± 1.32	<.001
PCB_2	79.0 ± 2.74 (76.6–81.4)	92.6 ± 0.8	.849
PCB_{3x}	74.0 ± 2.74 (71.6–76.4)	91.2 ± 0.75	.702
PCB_{3y}	62.8 ± 5.07 (58.4–67.2)	87.0 ± 1.05	.036
PCB_{3z}	83.8 ± 2.33 (81.6–86.4)	94.3 ± 1.47	.079

the highest performance, particularly 98.0% and 95.6%–100% 95% CI. Then, we added consecutive sensor types in the next experiments to test the combination of three sensor modalities. As a result, adding more sensor data (e.g., PCB_2) deteriorated the sensitivity (94.0% average breach recall, 90.3–97.7% 95% CI). We also combined X-, Y-, and Z-directional data from PCB_3 and increased average breach recall up to 84.8%. Concatenating 4 sensor data improved the breach recall by 2% in the case of Mic_1 & PCB_1 & PCB_2 & PCB_{3z} compared to Mic_1 & PCB_1 & PCB_2. However, combining all sensor data dropped the average breach recall by 4% compared to Mic_1 & PCB_1 & PCB_2 & PCB_{3z}. Nevertheless, compared to individual sensor-based models, the performance of data fusion is higher: e.g., Mic_1, PCB_1, and PCB_2 show 85.8%, 81.0%, and 79.8% average breach recall, while their fusion resulted in 94.0%. Furthermore, we conducted a one-way ANOVA on all individual sensors and their combinations, which showed significant differences between the groups (Table 3). Afterward, pairwise t-tests between the best combination (Mic_1 and PCB_1) with all individual sensors and their combinations were performed. As a result, the fusion of Mic_1 and PCB_1 was significantly better than all individual sensors-based models, and the combination of PCB_{3x} & PCB_{3y} & PCB_{3z} and the combination of all sensors according to Table 3.

4. Discussion

This study investigated the potential of vibroacoustic sensors in detecting pedicle breaches. The contact between the drill and bone produces vibrations that vibroacoustic sensors can capture, and it has been proved that they can discern between cortical and cancellous layers of the bone [9–13]. In our work, we physically generated breaches controlled and accurately by integrating a surgical navigation system to translate the preoperative planning to the anatomy. Subsequently, we developed an automatic data labeling pipeline based on the tracking data obtained from the navigation setup. We explored the prospect of a contact microphone, a free-field microphone, a uni-axial accelerometer, and a tri-axial accelerometer enabled by the squeeze-and-excitation network (SE-ResNet-18) in detecting breaches.

Among all vibroacoustic sensing methods, contact microphones showed the highest sensitivity of $85.8 \pm 3.19\%$, which was followed by the uni-axial accelerometer connected to the bone ($81.0 \pm 3.54\%$) (Table 3). This tendency can be understood from Fig. 6. One can notice that spectrograms from contact microphones and uni-axial accelerometers attached to the bone and surgical pin provide more distinctive features for significant events. These sensors are the closest to the actual location of possible breach events. They are less susceptible to environmental noise than the free-field microphone, which performed poorly (63.8% mean breach recall). The sound waves generated in the bone also had to travel ten to twenty centimeters through the air before reaching the free-field microphone. Regarding the axes of the tri-axial accelerometer, the signal from the Z-direction (along the direction of drilling) is the one performing strongest. This fact can be justified by its being a direction of signal propagation, thus providing stronger signals.

The evaluation revealed the differences in bone quality. Determining bone quality based on vibroacoustic signals is a promising topic for future research. In our opinion, state-of-the-art lacks methods estimating bone density. Therefore, we tested the networks with respect to BMD, characterized as normal and abnormal. Overall, no statistical significance could be shown except for uni-axial accelerometers and the tri-axial accelerometer in the X-direction. Regarding contact microphones, the average sensitivity was similar for both types, around 85%. It shows that contact microphones are indifferent to bone quality for breach detection. Sensors showing significant differences between breach recall for normal and abnormal bone, such as the Z-axis of the tri-axial accelerometer (PCB_{3x}), could be due to the prevalence of a lower signal-to-noise ratio in the abnormal test set. However, further investigations are necessary to define the cause and why some sensor configurations are more susceptible to the BMD.

Furthermore, we implemented the concept of transfer learning by using the weights trained on the breakthrough detection data in the same mel-spectrogram by Seibold et al. [19]. It was used as an initialization for our models. We expected to see an added performance on the contact microphones since pre-trained weights are also trained on the data from contact microphones. However, statistical testing with respective models trained from scratch (Table 5) did not reveal significant general differences except for the uni-axial accelerometer on the bone and the Y-axis of the tri-axial accelerometer. However, the trend for these two sensors is controversial.

Apart from inputting individual sensor data to the network, their different combinations were examined to prove the hypothesis that sensor fusion boosts performance and robustness. Table 3 shows how much the fusion of several sensor data improves the accuracy of the network. For most sensor combinations, the accuracy is pushed well above 90% for both breach and non-breach classification. The high variability for some sensors in Table 3 is likely because the signal of breach events largely varies depending on the breach location. Due to differences in geometry, bone thickness, drill angle, and other parameters, a medial breach can cause a different response compared to a lateral breach. Thus, these variations perplex the network to generalize from the limited data. Owing to the direct contact of sensors with the

tissue, the best-performing combination is contact microphones with the uni-axial accelerometer on the bone (98% breach recall), which is the same as the performance of the PediGuard probe with fluoroscopic guidance [5]. It shows that using vibroacoustic sensors with a deep learning algorithm is a reliable alternative to fluoroscopy-supported interventions to minimize radiation exposure.

We believe the proposed method is a promising approach to navigate either hand-held or robotic surgical drilling to facilitate safe pedicle screw placement. However, this study still has several limitations before translation to a real clinical scenario. First of all, in this study, the network is trained for the detection of breach events with the purpose of stopping the drill before any vital structures are harmed. According to Li et al. a medial cortical breach is considered acceptable when the penetration depth is smaller than 2 mm [24]. As stressed by [19], automatic and fast breach detection approaches lead to a significantly shorter response time compared to a human operator. Integration of breach detection in a robotic system can warrant this fast response time to avoid damage to vital structures. In a clinical setting, it is most beneficial to predict an impending breach event to stop the drilling action before perforation of the cortical wall. Qualitative investigation showed promising features in the spectral content indicating intrusion of the drill into the cortical wall before breach but appeared to be dependent on the bone quality. This will be further investigated in future research where the next step involves training a model to predict an imminent breach. Further, the results of this study are based on ex-vivo experiments. Before discerning the clinical feasibility, the developed technology should be tested on in-vivo animal experiments for a close-to-real setup. The number of samples was limited due to the time frame, efforts, and costs of performing cadaver experiments. It is important to note that the collection of "breach samples" is hardly possible and unethical, as breaching incidents occur rarely in real patients and should be avoided at all costs. Thus, ex-vivo data collection, such as utilizing cadavers, becomes essential to create a relevant number of breach samples. Previous studies were mainly done using porcine specimens [25,26] or on an even smaller number of cadaveric specimens, namely 2 [7]. The higher costs associated with data collection and ethical considerations are the major challenge of such studies. Therefore, the research necessitates a step-wise evaluation. The performance could be further improved by having more samples and investigating synthetic data generation techniques for CNN training. In terms of the direct translation of the sensor technology into the operating room, the positioning of accelerometers and contact microphones are reproducible. Contact microphones can be easily placed on the skin of the patient. A uni-axial accelerometer can be attached to the spinous process of the vertebrae in open spine surgery, whereas a tri-axial accelerometer can be attached to the surgical drill. However, sterilizability of these sensors or the use of sterile drapes over the sensors has to be investigated and validated. In contrast, a free-field microphone should be positioned further away from the operation site, which may cause lower sensitivity to event detection. It is essential to emphasize that further investigation is required to determine the optimal positioning of individual sensors. We hypothesize that the detection accuracy will deteriorate if the position or direction of the sensors will vary from the current setup as the signal quality might change. Future work will focus on systematically exploring the effects of different sensor placements on signal quality and detection accuracy. Finally, we only analyzed the spectrograms over a frequency range from 0 to 2 kHz. However, some investigated sensors might give better results in higher frequency ranges. According to Dai et al. [25], high-frequency components in tactile vibration feedback contain the most critical information about transient contact events. This study represents the first effort to translate non-visual sensing technology to the operating field as a local sensing method for hand-held and robotic orthopedic surgery applications.

5. Conclusion

In this paper, we conducted the first ex-vivo study on a deep learning-driven non-vibroacoustic sensing approach for breach detection during pedicle screw placement in spine surgery. The non-visual nature of vibroacoustic sensing eliminates errors caused by non-negligible patient movements during operation and radiation exposure to the patient and medical staff. The results are similar to those of the PediGuard probe, the only non-visual sensing device on the market. The proposed method has great potential to increase the autonomy of robotic surgery by introducing non-visual physical intelligence. We also proposed a new method of automatic data labeling based on surgical navigation, which can be beneficial in labeling a large amount of data in the future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] Perna F, Borghi R, Pilla F, Stefanini N, Mazzotti A, Chehrassan M. Pedicle screw insertion techniques: an update and review of the literature. Musculoskelet Surg 2016;100(3):165–9.

[2] Gelalis ID, Paschos NK, Pakos EE, Politis AN, Arnaoutoglou CM, Karageorgos AC, Ploumis A, Xenakis TA. Accuracy of pedicle screw placement: a systematic review of prospective in vivo studies comparing free hand, fluoroscopy guidance and navigation techniques. Eur Spine J 2012;21(2):247–55.

[3] Peng Y-N, Tsai L-C, Hsu H-C, Kao C-H. Accuracy of robot-assisted versus conventional freehand pedicle screw placement in spine surgery: a systematic review and meta-analysis of randomized controlled trials. Ann Trans Med 2020;8(13).

[4] Dusad T, Kundnani V, Dutta S, Patel A, Mehta G, Singh M. Comparative prospective study reporting intraoperative parameters, pedicle screw perforation, and radiation exposure in navigation-guided versus non-navigated fluoroscopy-assisted minimal invasive transforaminal lumbar interbody fusion. Asian Spine J 2018;12(2):309.

[5] Chaput CD, George K, Samdani AF, Williams JI, Gaughan J, Betz RR. Reduction in radiation (fluoroscopy) while maintaining safe placement of pedicle screws during lumbar spine fusion. Spine (Phila Pa 1976) 2012;37(21):E1305–9.

[6] Bolger C, Kelleher M, Mcevoy L, Brayda-Bruno M, Kaelin A, Lazennec jy, Huec J, Logroscino C, Mata P, Moreta P, Saillant G, Zeller R. Electrical conductivity measurement: A new technique to detect iatrogenic initial pedicle perforation. Eur Spine J : Official Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 2007;16:1919–24. http://dx.doi.org/10.1007/s00586-007-0409-8.

[7] Guillen PT, Knopper RG, Kroger J, Wycliffe ND, Danisa OA, Cheng WK. Independent assessment of a new pedicle probe and its ability to detect pedicle breach: a cadaveric study. J Neurosurg: Spine 2014;21(5):821–5.

[8] Massalimova A, Timmermans M, Esfandiari H, Carrillo F, Laux CJ, Farshad M, Denis K, Fürnstahl P. Intraoperative tissue classification methods in orthopedic and neurological surgeries: A systematic review. Front Surg 2022;9.

[9] Zakeri V, Hodgson AJ. Classifying hard and soft bone tissues using drilling sounds. In: 2017 39th annual international conference of the ieee engineering in medicine and biology society. EMBC, IEEE; 2017, p. 2855–8.

[10] Guan F, Sun Y, Qi X, Hu Y, Yu G, Zhang J. State recognition of bone drilling based on acoustic emission in pedicle screw operation. Sensors 2018;18(5). http://dx.doi.org/10.3390/s18051484, URL https://www.mdpi.com/1424-8220/18/5/1484.

[11] Sun Y, Jin H, Hu Y, Zhang P, Zhang J. State recognition of bone drilling with audio signal in robotic orthopedics surgery system. In: 2014 IEEE/RSJ international conference on intelligent robots and systems. 2014, p. 3503–8. http://dx.doi.org/10.1109/IROS.2014.6943051.

[12] Torun Y, Öztürk A, Hatipo˜ glu N, Öztemür Z. Breakthrough detection for orthopedic bone drilling via power spectral density estimation of acoustic emission. In: 2018 electric electronics, computer science, biomedical engineerings' meeting. EBBT, 2018, p. 1–5. http://dx.doi.org/10.1109/EBBT.2018.8391464.

[13] Torun Y, Pazarci Ö. Parametric power spectral density estimation-based breakthrough detection for orthopedic bone drilling with acoustic emission signal analysis. Acoust Aust 2020;48(2):221–31.

[14] Stevens SS, Volkmann J, Newman EB. A scale for the measurement of the psychological magnitude pitch. J Acoust Soc Am 1937;8(3):185–90.

[15] Poblete V, Espejo D, Vargas V, Otondo F, Huijse P. Characterization of sonic events present in natural-urban hybrid habitats using UMAP and sednet: The case of the urban wetlands. Appl Sci 2021;11(17):8175.

[16] Ostler D, Seibold M, Fuchtmann J, Samm N, Feussner H, Wilhelm D, Navab N. Acoustic signal analysis of instrument–tissue interaction for minimally invasive interventions. Int J Comput Assist Radiol Surg 2020;15:771–9.

[17] Papadimitriou I, Vafeiadis A, Lalas A, Votis K, Tzovaras D. Audio-based event detection at different SNR settings using two-dimensional spectrogram magnitude representations. Electronics 2020;9(10):1593.

[18] Purwins H, Li B, Virtanen T, Schlüter J, Chang S-Y, Sainath T. Deep learning for audio signal processing. IEEE J Sel Top Sign Proces 2019;13(2):206–19.

[19] Seibold M, Maurer S, Hoch A, Zingg P, Farshad M, Navab N, Fürnstahl P. Real-time acoustic sensing and artificial intelligence for error prevention in orthopedic surgery. Sci Rep 2021;11(1):1–11.

[20] Goossens Q, Pastrav L, Roosen J, Mulier M, Desmet W, Vander Sloten J, Denis K. Acoustic analysis to monitor implant seating and early detect fractures in cementless THA: An in vivo study. J Orthop Res 2021;39(6):1164–73.

[21] Schlüter J, Grill T. Exploring data augmentation for improved singing voice detection with neural networks.. In: ISMIR. 2015, p. 121–6.

[22] Schreiber JJ, Anderson PA, Rosas HG, Buchholz AL, Au AG. Hounsfield units for assessing bone mineral density and strength: a tool for osteoporosis management. JBJS 2011;93(11):1057–63.

[23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 7132–41.

[24] Li G, Lv G, Passias P, Kozanek M, Metkar US, Liu Z, Wood KB, Rehak L, Deng Y. Complications associated with thoracic pedicle screws in spinal deformity. Eur Spine J 2010;19:1576–84.

[25] Dai Y, Xue Y, Zhang J. Human-inspired haptic perception and control in robot-assisted milling surgery. IEEE Trans Haptics 2021;14(2):359–70. http://dx.doi.org/10.1109/TOH.2020.3029043.

[26] Dai Y, Xue Y, Zhang J. Noncontact vibration measurement based thoracic spine condition monitoring during pedicle drilling. IEEE/ASME Trans Mechatronics 2014;19(5):1532–40. http://dx.doi.org/10.1109/TMECH.2013.2287305.