A CNN-Based Novel Approach for Classification of Sacral Hiatus with GAN-Powered Tabular Data Set

Ferhat Kilic[1], *, Murat Korkmaz[2], Orhan Er[3], Cemil Altin[4]
[1]Department of Mechatronics Engineering, Yozgat Bozok University,
66200 Yozgat, Turkey
[2]Department of Orthopedics and Traumatology, Yozgat Bozok University,
66200 Yozgat, Turkey
[3]Department of Computers Engineering, Izmir Bakircay University,
35665 Izmir, Turkey
[4]Department of Electrical and Electronics Engineering, Yozgat Bozok University,
66200 Yozgat, Turkey
*ferhatkilic@hitit.edu.tr, murat.korkmaz@yobu.edu.tr, orhan.er@bakircay.edu.tr,
cemil.altin@yobu.edu.tr

Abstract—**Caudal epidural anaesthesia is usually the most well-known technique in obstetrics to deal with chronic back pain. Due to variations in the shape and size of the sacral hiatus (SH), its classification is a crucial and challenging task. Clinically, it is required in trauma, where surgeons must make fast and correct selections. Past studies have focused on morphometric and statistical analysis to classify it. Therefore, it is vital to automatically and accurately classify SH types through deep learning methods. To this end, we proposed the Multi-Task Process (MTP), a novel classification approach to classify the SH MTP that initially uses a small medical tabular data set obtained by manual feature extraction on computed tomography scans of the sacrums. Second, it augments the data set synthetically through a Generative Adversarial Network (GAN). In addition, it adapts a two-dimensional (2D) embedding algorithm to convert tabular features into images. Finally, it feeds images into Convolutional Neural Networks (CNNs). The application of MTP to six CNN models achieved remarkable classification success rates of approximately 90 % to 93 %. The proposed MTP approach eliminates the small medical tabular data problem that results in bone classification on deep models.**

Index Terms—**Bone classification; Synthetic tabular data generation; Two-dimensional embedding; Transfer learning; Generative adversarial networks; Convolutional neural networks; Deep learning.**

I. INTRODUCTION

Caudal epidural block (CEB) is usually the most known technique in obstetrics, performed by inserting a needle through the inferior hiatus. This technique aims to gain entrance into the sacral epidural space to inject an anaesthesia solution to treat spinal deformities and manage chronic back pain [1]. The most facing difficulties during CEB are needle positioning, especially specifying the location of the apex, which is the most critical anatomical landmark of the sacral hiatus (SH) [2]. Because SH shape and size variations play an important role in the accuracy of needle position and feasibility, the success of the CEB technique decreases [3], [4]. Additionally, due to obesity, palpation may be complicated [5]. Therefore, classification of SH requires diagnosis, surgical planning, and postoperative assessment.

Ultrasonography or fluoroscopy can improve the position of the clinical needle [2]. Since minimally invasive surgeries are related to a lower risk of infection, less exsanguination, and a complete decrease in disorder, they are helpful for the patient. However, they increase the workload of a surgeon [6]. Also, they may be less preferred due to radiation exposure, time limitation, and lack of trained and informed personnel [7]–[12]. In addition, because computed tomography (CT) scans comprise many slices and each holds many data, they cannot be accurately analysed by investigation or conventional methods. Therefore, additional information about the anatomical structure of SH is highly beneficial to support surgeons during their intraoperative decision-making.

Many workers have performed morphometric studies to classify SH traditionally in many different geographical areas [13]–[17]. However, these studies have focused only on statistical data analysis and methods to increase the success of the CEB technique. Consequently, it is crucial to classify SH automatically and accurately via deep learning (DL) methods not studied so far. The need for robust data analysis has become more critical as the world becomes more data-driven for decision-making [18]. One of the most exciting developments in this field is the rise of DL to analyse large amounts of data.

Data come in two different forms: structured and unstructured data. Structured data are stored as tabular data, such as Excel sheets and databases, while unstructured data are stored in various forms, including text, email, images,

and videos. In structured data, each column represents a distinct feature and each row represents a different instance. Structured data are easy to analyse and process using traditional classifiers such as the Support Vector Machine (SVM) [19], Logistic Regression (LogReg) [20], and tree-based algorithms [21]. These models are designed to outperform DL models, mainly when working with small tabular data sets [22]. On the contrary, DL models are complex architectures with many hidden layers and nodes, allowing them to process and analyse large amounts of data accurately [23].

However, the Convolutional Neural Network (CNN) cannot be applied directly to tabular data. Therefore, several studies have recently introduced methods that make CNN viable for tabular data classification [24]–[28]. Converting tabular data sets to images makes it easier to analyse through CNNs. SuperTML [24] uses two algorithms: Equal Font-SuperTML (SuperTML_EF) and Variable Font-SuperTML (SuperTML_VF). In contrast, the SuperTML_EF algorithm provides equal space for each feature, and SuperTML_VF gives greater image space for more relevant features.

DeepInsight [25] uses t-distributed stochastic neighbour embedding (t-SNE) to project feature vectors onto a two-dimensional (2D) space while minimising the Kullback-Leibler divergence. The resulting 2D projection is compared with the original full-dimensional space, and a rectangle is identified that includes all projected feature points with a minimum area. This rectangle forms the image representation.

Representation of Features as Images with Neighborhood Dependencies (REFINED) [26] uses Bayesian multidimensional scaling. It can generate high-quality 2D images while preserving the original feature distribution. It is an effective means of minimising distortion and maintaining data integrity in a 2D space.

The Image Generator for Tabular Data (IGTD) [27] algorithm creates an image accurately, assigning every feature within the tabular data to a specific pixel, corresponding to its intensity reflected in the image.

The Dynamic Weighted Tabular Method (DWTM) [28] uses advanced statistical techniques such as Pearson's correlation and Chi-square to accurately compute the weights of each feature in tabular data. DWTM arranges them in descending order and assigns them space within the image canvas according to their significance.

In this study, we proposed the Multi-Task Process (MTP) classification approach, a crucial and challenging task and a novel way to classify SH using two-dimensional embedding with CNNs. As shown in Fig. 1, manual feature extraction is applied to two-dimensional (2D) sacrum images in CT in the first stage. The distance values are measured manually considering SH's shape and size between the priorly determined locations on sacrum bones. In the second stage, these distance values are obtained as tabular data. Due to the insufficient data set to train a data-hungry CNN, in the third stage of the MTP, the rows of the tabular data set are increased by a Generative Adversarial Network (GAN) synthetically. In the fourth stage, with an algorithm known as two-dimensional embedding, these values are projected

into images. In the last stage, after feeding these images into pre-trained CNN models, we achieved significant SH classification success rates of approximately 90 %–93 % after 30 epochs.
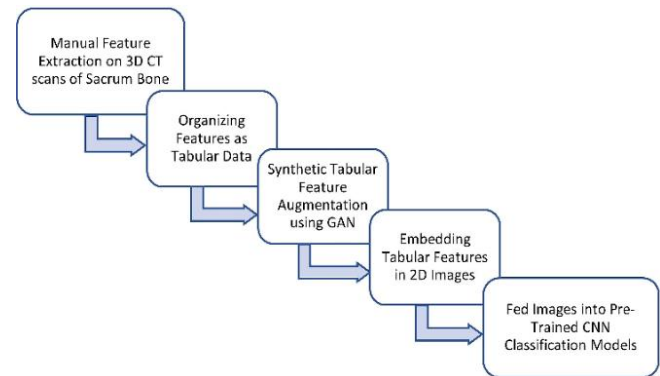

Fig. 1. MTP: Flow chart of the SH classification.

The study contributions are as follows:
1. The novelty of this study is that the classification of SH using DL methods in CT has not previously been studied in the literature;
2. Deep models require many weights that need to be optimised due to their complexity during training to avoid overfitting. Unfortunately, the acquisition of many CT scans is challenging in the medical field. Also, training deep models with three-dimensional (3D) volumetric data are computationally intensive and exponentially increases the need for weights to be learnt. Additionally, annotated CT scans are limited because they are time consuming and varies from an expert's point of view. To do this, we manually transform the SH shape and size features on 2D sacrum CTs into tabular data and use a GAN to increase it synthetically;
3. We embed increased non-heterogeneous tabular data rows into images. Each image file with a 3D matrix of pixels represents an individual sample that a CNN can train. Thus, the computational cost is reduced since only 2D convolution is required;
4. We investigate the relative performances of the MTP on other pre-trained DL models. We demonstrate that MTP is a successful approach in terms of applicability with different DL models;
5. We investigate the relative results of the MTP classification approach and non-deep models, traditional machine learning (ML) classification algorithms. We do not observe a definitive superiority between the results, and they are close to each other;
6. We eliminate the difficulties that surgeons face in planning external pelvic injuries and enable them to do the planning in a much shorter time and with the correct diagnosis decisions;
7. Knowing the anatomical structure well of the SH helps surgeons prevent possible complications during operations to be performed in this area; therefore, MTP protects the neural structures and their surroundings in surgical interventions in the pelvic region;
8. MTP reduces the use of the tomography device, reducing the high radiation exposure rate of the operating team and the patients;

9. MTP contributes to the increase in the success rate of the operation and the patient's recovery in a shorter time, reducing the error rate during the pre-operation;

10. MTP allows the construction of multi-model pipelines.

II. MATERIALS AND METHODS

A. Data Acquisition

After the YOBU Clinical Research Ethics Committee approved this study, CT scans belonging to the pelvic region of 252 patients were retrospectively obtained between June 01, 2018 and October 15, 2020. Typically, participants are patients 20 years or older who come to the YOBU Research and Application Hospital for pelvic CT scans. Individuals' CT scans do not have pathological tumours or defects in the pelvic rings. We excluded CT scans of patients with too low image resolution from this study. Therefore, only 123 CT scans from patients for feature extraction remained of 252 CT scans.

B. Organising the SH Shapes

Figure 2 shows different SH shapes of sacrum bones [15]. Before GAN, we pre-processed SH shapes and reduced six types to three types to obtain more consistent results and ensure data integrity. SH shapes in Figs. 2(a), 2(d), and 2(f) are represented by one as type M because of their similarity, Fig. 2(c) - by two as type V, and in Fig. 2(b) - by three as type U. We ignored the SH shape in Fig. 2(e) because it is scarce in the population.
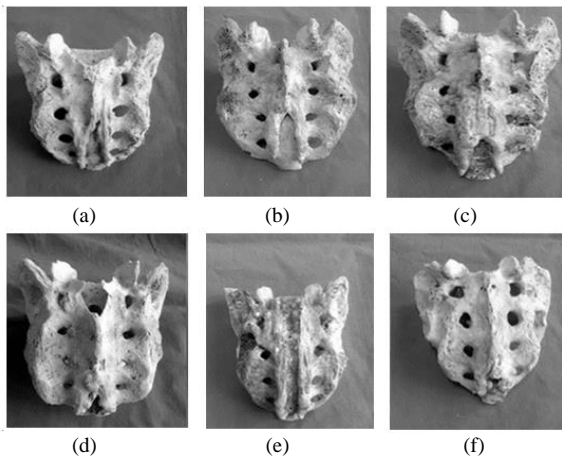
Fig. 2. Different hiatus shapes [15]: (a), (d) irregular hiatus shape, (b) inverted U shape, (c) inverted V shape, (e) incomplete formation of the posterior sacral canal wall, (f) dumple shape.

C. Feature Extraction

We use Radiant Dicom Viewer software to perform manual feature extractions in sacrum CT in Fig. 3. Nine distance values (features) are measured between the location of each patient's landmarks and are stored as comma-separated values (CSV) files.
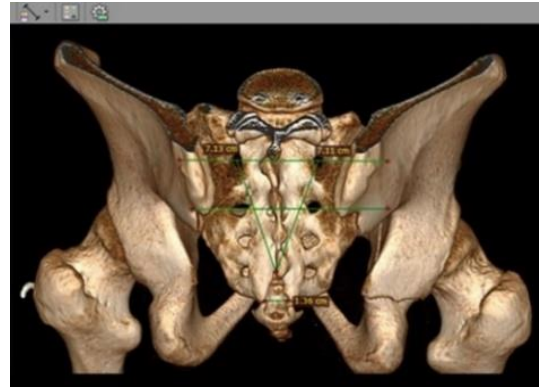
Fig. 3. Feature extractions in the Radiant Dicom Viewer.

Figure 4 shows the locations of the landmarks and their distances from each other, considering the apex of SH. Table I shows the original SH tabular data set (o-SHd) belonging to the first five patients.
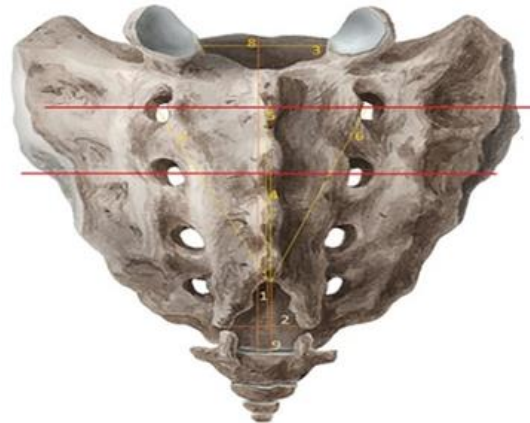
Fig. 4. Located landmarks and distance features on the sacrum.

A. Tabular Data Set Augmentation

In data science, data are categorised into structured and unstructured data. Structured data are also known as tabular data. Tabular data include a set of rows and columns consisting of features and are the most common data type in real-world applications. Many challenges arise when feeding them to deep neural networks, including lack of locality, missing values, mixed feature types, and unlike text or images, lack of prior knowledge of the structure of the data set [29].

The o-SHd used in the MTP was very small, with only 123 instances. When it splits into train, validation, and test sets with a 50:30:20 ratio, we only got 69 train sets, 29 validation sets, and 25 test sets.

TABLE I. o-SHd.

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7	Feature 8	Feature 9	Class
Patient 1	1,4	1,8	2,5	3,4	6,2	6,7	6,2	9,8	2,5	1
Patient 2	1,8	1,7	4,1	3,4	6,7	8,7	8,4	12,1	3,2	1
Patient 3	2,4	1,5	4,0	3,1	7,9	8,4	8,4	11,9	3,4	3
Patient 4	2,6	1,5	4,1	3,1	6,7	8,4	6,3	12,5	4,4	2
Patient 5	1,9	1,1	1,6	2,2	3,8	4,4	4,4	9,8	3,9	3

Due to the insufficient o-SHd for training a data-hungry CNN, we augmented it with respect to its feature distributions and their classes. We adopted the GAN for Tabular Data (TabGan) project [30] for this process. Therefore, we synthetically increased 123 in o-SHd to 874 instances of new data set named "GAN-powered SH tabular data set" (g-SHd). After splitting g-SHd with an 80:20 ratio using the Scikit-Learn library of Python programming language, we got a training set of 699 samples and a blind test set of 175.

Figure 5 shows the heat map of correlation of o-SHd and g-SHd on how strongly each feature relates to each other. White represents strong relations, whereas red represents weak relations. The correlation is apparent in both data sets, especially between the 6th-7th and 8th-9th features. Figure 6 shows the t-SNE visualisations of o-SHd and g-SHd.
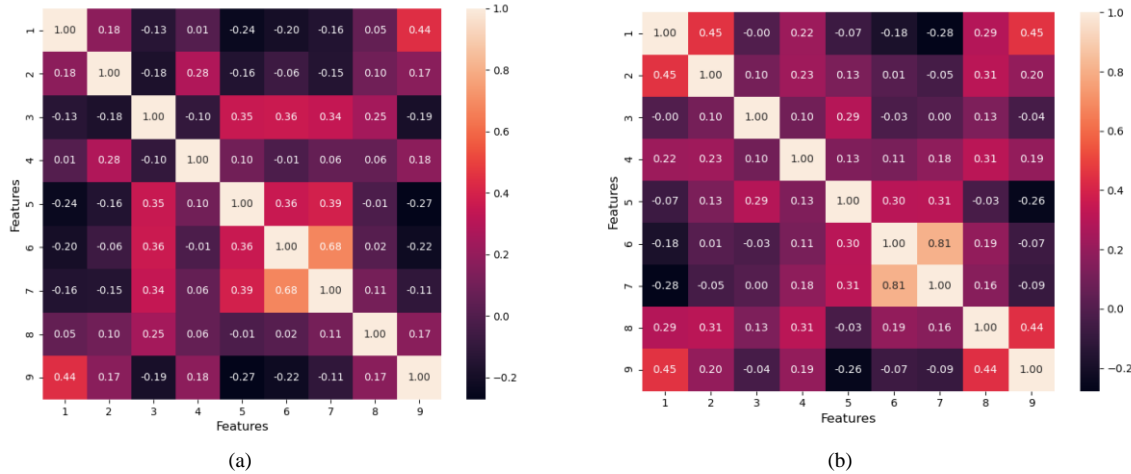

(a)

(b)

Fig. 5. Correlation heat maps of (a) o-SHd and (b) g-SHd.
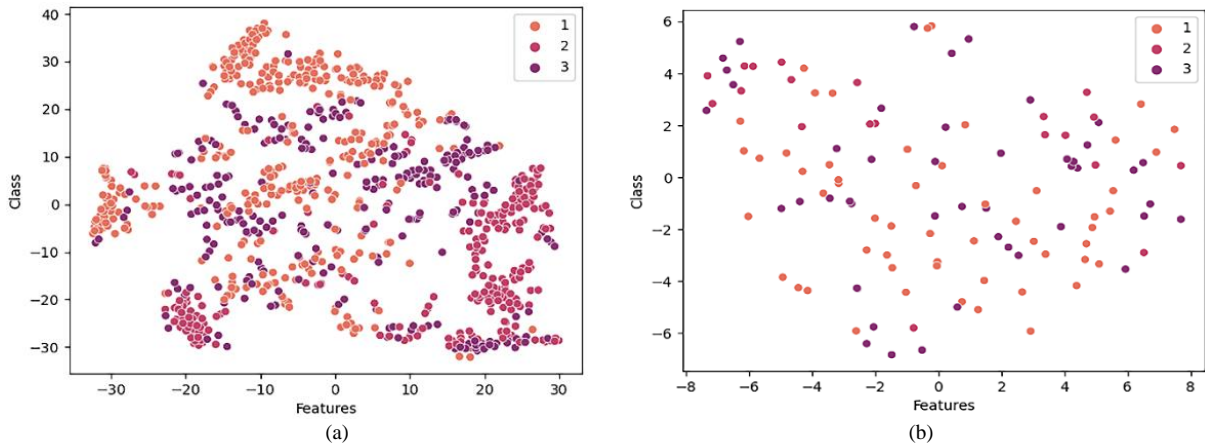

(a)

(b)

Fig. 6. t-SNE visualisations of (a) o-SHd and (b) g-SHd.

B. Embedding Features in the Images

We developed a new algorithm to generate an image data set and applied it to g-SHd using the SuperTML method [24]. Eight hundred seventy-four instances and their features were successfully embedded in the images using the OpenCV library. Due to the input size of the deep models, images were generated in 224×224 sizes. The SuperTML method is motivated by the analogy between Tabular data Machine Learning (TML) problems and text classification tasks. However, unlike the text classification problems studied in [31], the features of the tabular data are of separate dimensions. Therefore, generated images should have some space between their features in different sizes to ensure that they do not overlap in the generated images [24]. Because g-SHd has the same importance for features, the algorithm allocated the same size to every feature. So, g-SHd could be directly embedded into images without calculating its importance. As shown in Fig. 7, every instance, and their nine features in g-SHd train and test sets,

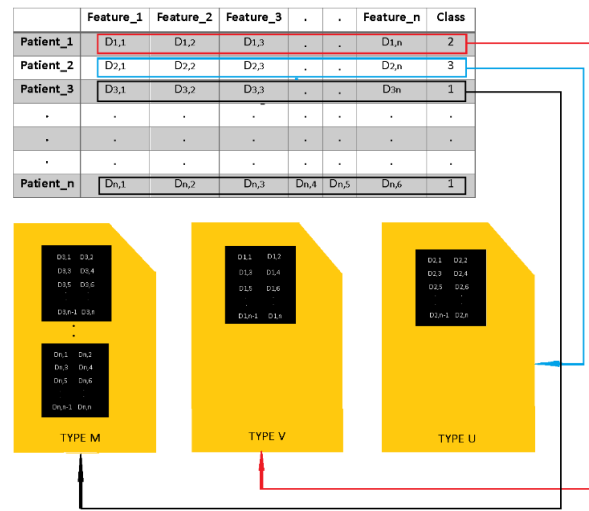was projected with an order to new images.

Fig. 7. Embedding the features into 224×224 images and distributing them into three folders.

By the algorithm saving images in class folders, we got g-SHd (eg-SHd) embedded. In Table II, the class distribution of the eg-SHd is shown. Figure 8 shows the flow chart of the proposed algorithm.

TABLE II. THE CLASS DISTRIBUTION OF eg-SHd.

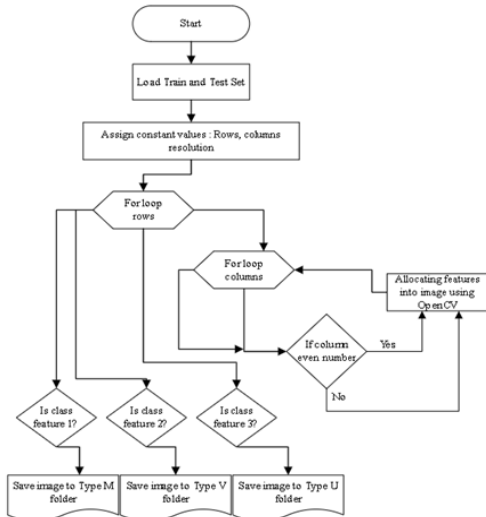| Type M	Type V	Type U	Train data set
291	212	196	699
Type M	**Type V**	**Type U**	**Test data set**
77	47	51	175

Fig. 8. The flow chart of the proposed algorithm.

C. Deep Convolutional Neural Networks

CNNs have achieved tremendous image classification and recognition performance. A CNN recognises an object in detail by looking for low-level features, such as edges, lines, and curves, and then builds up more features with an overall outlook [32].

DL performance depends on the amount and quality of the data gathered in a learning model for the target application [33]. Accuracy scores are higher when more sample images are used for training on the success of a CNN [34]. When the training data set is enlarged, the diversity of the data increases, and a better generalisation is achieved. However, the necessary amount of data to build deep models are hard to find, especially in the medical field. The transfer learning technique deals with similar problems [32]. The image data augmentation technique is another procedure for creating various transformed images such as rotation, flips, zooms, etc. This technique expands the training data sets to improve the CNNs' ability to generalise their learnings to new images to reduce overfitting.

D. Experimental Environment

This study uses a machine with Intel(R) Core (TM) i7-8750H CPU @2.20 GHz processor, 32 GB SDRAM, and 8 GB of NVIDIA Geforce RTX 2070 graphical processing unit with CUDA 11.5 CuDNN 8.1 for the experimental environment.

Because they provide a huge set of ML libraries and are very convenient for DL, all training steps were conducted in Python (Python 3.8.13). Keras 2.10 [35] with TensorFlow 2.10 [36] backend was used for training CNN models. Figure 9 shows the general architecture of the MTP approach.
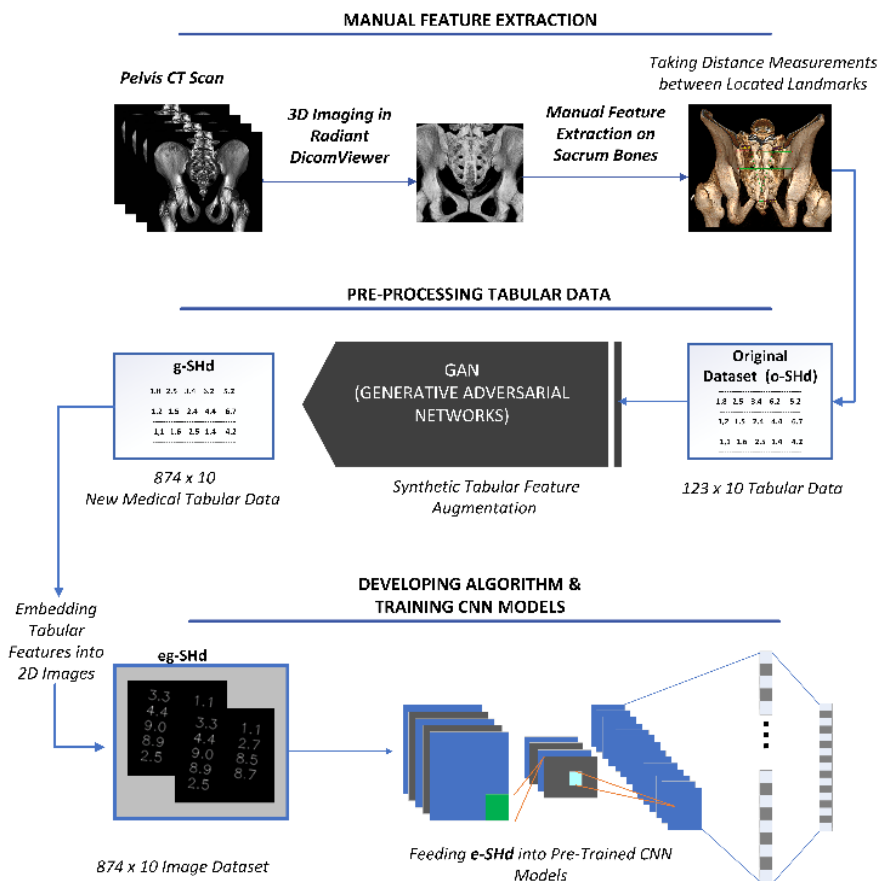

Fig. 9. Overall architecture of the MTP approach.

E. Pre-Processing of eg-SHd with Keras

The eg-SHd, including 699 images, was split into the validation data set using the rates of 25 % with the Keras image data generator pipeline. After the process, 525 train sets and 174 validation sets were obtained, almost the same numbers as 175 in the blind test set. Table III shows the distribution of the eg-SHd.

Data augmentation techniques are also applied to increase the generalisability of the models and prevent them from overfitting. Thus, deep models constantly saw slightly modified versions of new input data and were able to learn more robust features. In the training set, the rotation range of 30 degrees was performed. Figure 10 shows modified versions of the eg-SHd after applying Keras's augmentation techniques.

TABLE III. DISTRIBUTION OF eg-SHd.

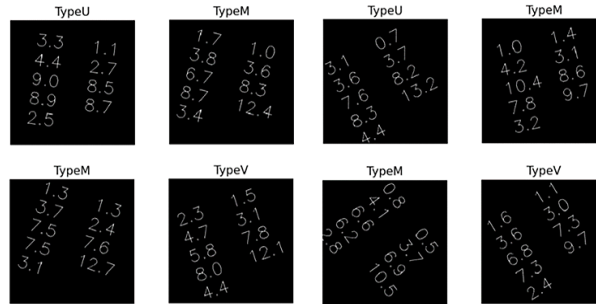| eg-SHd	Total Images
Train	525
Validation	174
Test	175
Total	874

Fig. 10. Transformed eg-SHd.

F. Training CNN Models

As a technique of ML, during training, transferring the level of learning acquired from one task to another related task is known as transfer learning [37]. In this study, we loaded the pre-trained CNN models with ImageNet weights to transfer learning and set the top property of the model to false. We added Gaussian noise with 0.8 and a dropout rate of 0.2 to prevent the model from overfitting before the fully connected layers with 4096 neurons.

CNN models (Inception-v3, Inception ResNet-v2, ResNet50-v2, ResNet152-v2, DenseNet121, and DenseNet169) employed in this study, were trained with Keras in 30 epochs. Unfroze all convolutional layers, implement time-based decay, and set the initial learning rate to 0.004. The decay rate is proportional to the learning rate, the epoch value and the momentum. Momentum applied 0.3 within the Stochastic Gradient Descent (SGD) optimiser, which tends to converge to better results.

After employing eg-SHd with the MTP, accuracy and the loss performances of all models are shown after 30 epochs in Figs. 11–16, respectively. The confusion matrix was computed to understand which classes were correctly labelled (true-positives) diagonally and mislabelled (false-negatives) off-diagonally as boxes in the matrix. As can be seen in Fig. 11, 161 of 175 test data were correctly predicted, and 14 were incorrectly predicted for Inception-v3.

(a)   (b)   (c)

Fig. 11. Relatively evaluated performance of MTP (Inception-v3): (a) Accuracy, (b) Loss, (c) Confusion Matrix.

(a)   (b)   (c)

Fig. 12. Relatively evaluated performance of MTP (Inception ResNet-v2): (a) Accuracy, (b) Loss, (c) Confusion Matrix.

Fig. 13. Relatively evaluated performance of MTP (ResNet50-v2): (a) Accuracy, (b) Loss, (c) Confusion Matrix.
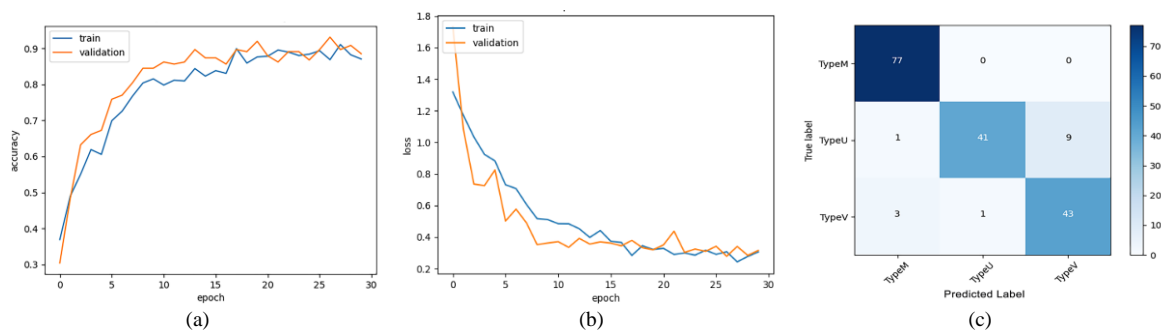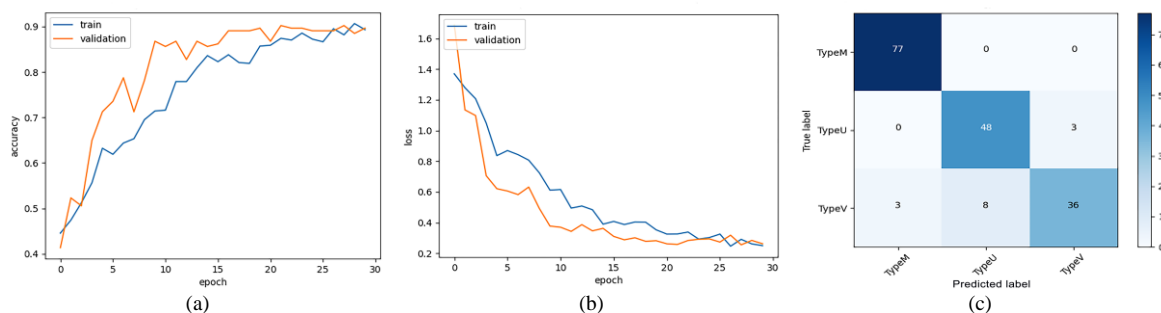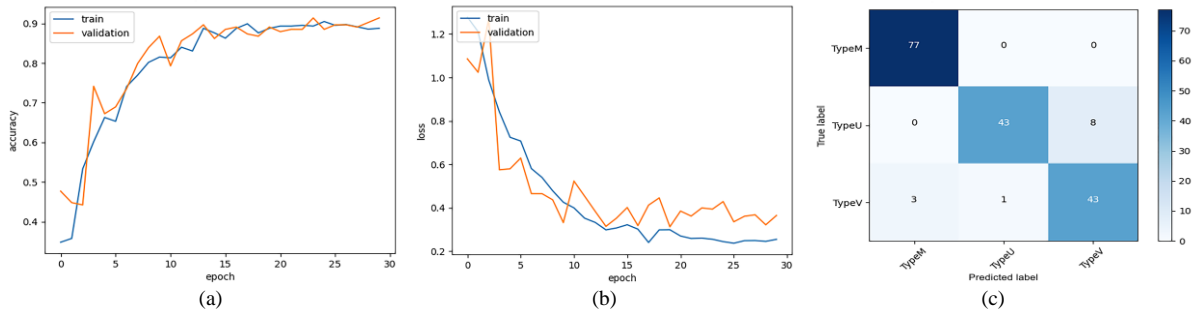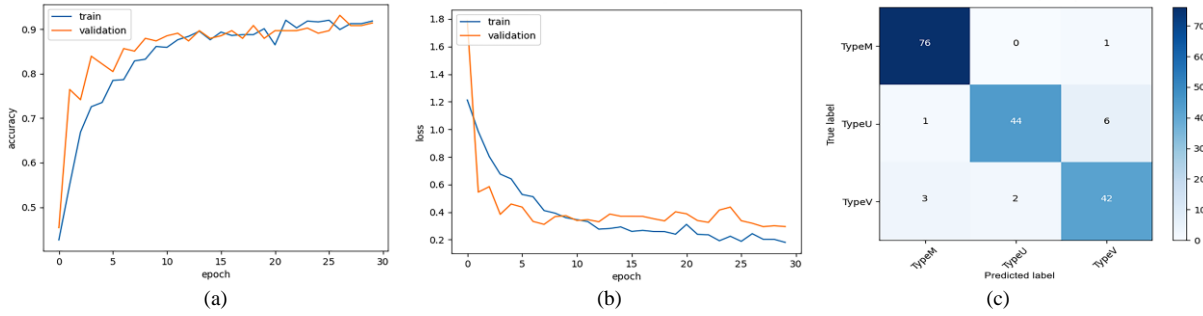

Fig. 14. Relatively evaluated performance of MTP (ResNet152-v2): (a) Accuracy, (b) Loss, (c) Confusion Matrix.
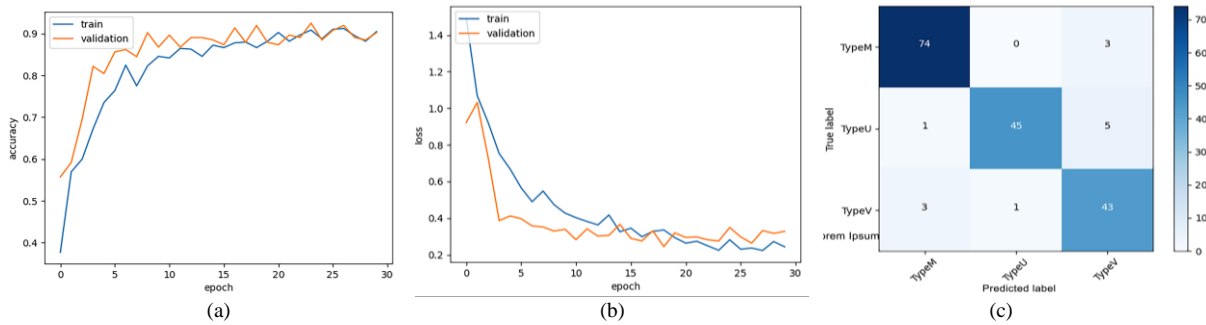

Fig. 15. Relatively evaluated performance of MTP (DenseNet121): (a) Accuracy, (b) Loss, (c) Confusion Matrix.
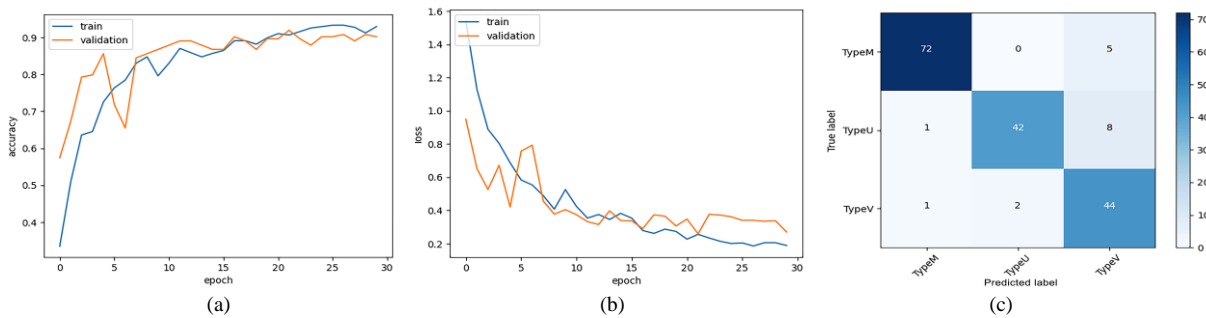

Fig. 16. Relatively evaluated performance of MTP (DenseNet169): (a) Accuracy, (b) Loss, (c) Confusion Matrix.

III. RESULTS

Table IV shows the performance metrics evaluated for the MTP approach on CNN models. Accuracy is the most intuitive performance measure and is simply a ratio of correctly predicted observations to the total observations in an experiment. The Inception-v3 model reached a 0.92 accuracy score with the MTP approach, which means that the model is approximately 92 % accurate.

High precision relates to the low false-positive rate, which also indicates whether the predicted positive observations are truly predicted as correct. Performance metrics obtained a 0.95 precision score for type M, 0.98 for type U, and 0.83 for type V class.

The arithmetic mean of all precision scores belonging to these classes is 0.92, which is known as the macro-average precision score and is used for needs to be treated equally to evaluate the overall performance of the model related to its class labels. The score for each class by the number of actual instances when calculating the average weighting score is 0.92. The weighted macro-averaging score is used in the case of the model imbalance classes. The ratio of correctly predicted positive observations to all observations in the actual class: 1.00 for type M, 0.80 for type U, and 0.91 for type V class:

$$Precision = \frac{TP}{TP + FP}, \qquad (1)$$

$$Recall = \frac{TP}{TP+FN}, \qquad (2)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}. \qquad (3)$$

TP, *TN*, *FP*, and *FN* denote the true positive, true negative, false positive, and false negative cases, respectively. The *F1-score* is the harmonic mean of precision and recall

$$F1-score = \frac{Precision Recall}{Precision+Recall}. \qquad (4)$$

The *F1-score* is the weighted average of *Precision* and *Recall* and is more beneficial than accuracy, which works best if false positives and false negatives have similar costs. In our case, the *F1-score i*s the same with a precision and recall value of 0.92 in the case of the weighted average, 0.97 for the type M class, 0.88 for the type U class, and 0.87 for the type V class.

A. Evaluation of the MTP Performances with eg-SHd and o-SHd Data Sets

The same tuning hyperparameters are used to evaluate MTP's performances on the most accurate deep models to make a consistent and accurate comparison. During training CNNs, only changes are applied to the batch size and the learning rate depending on the model's architecture.

Figures 11-16 shows the relatively evaluated performance of the MTP on all models using the eg-SHd data set. The models have all performed robustly with the MTP approach. We can understand from Table IV that the MTP has almost similar classification performance success rates of approximately 90 %–93 % in all models.

In addition, the results were compared by employing the o-SHd data set, both the most accurate ML classification

algorithms and the MTP. Table V shows the performance metrics relatively evaluated between the MTP and traditional ML classification algorithms. As can be seen from the table, a very significant performance improvement has been observed with the MTP approach employed with the eg-SHd data set.

B. Evaluation of the MTP's Performances Using Standard Data Sets

We used three standard data sets to test the MTP approach and compared its results with those obtained with the ML classifiers. Details of the selected data sets and comparison of the MTP results with the ML classifiers are given in Table VI. The objective is to evaluate performance of the MTP approach and test its robustness on small standard tabular data sets, not just for o-SHd.

The first standard data set is the Iris data set [38], which is a public data set containing 150 samples, four features and three classes. It is the most popular data set for pattern recognition.

The second is the Wine data set [39]. It contains 178 samples and 13 dimensions and is the UCI repository's third most popular data set.

Both data sets have three classes. The last data set is a medical data set, Breast Cancer Wisconsin [40] (BCW), which contains data from 699 patients. It has ten features and two classes. These three tabular data sets are preferred to test the classification performances of the MTP approach. All tasks of the MTP are considered for these classifications, as mentioned in Fig. 1.

The experiments show that the CNN-based MTP approach provides better results than the traditional ML classifiers on the medical o-SHd data set. The proposed approach also proves robust and successful in disease diagnosis in the BWC data set. The results on both the o-SHd and BWC data sets prove that MTP is a viable option for performing state-of-the-art results on medical data sets.

TABLE IV. PERFORMANCE METRICS OF MTP ON CNNS.

MTP (Inception-v3)				MTP (Inception ResNet-v2)			
	Precision	Recall	F1-score		Precision	Recall	F1-score
Type M	0.95	1.00	0.97	Type M	0.96	1.00	0.98
Type U	0.98	0.80	0.88	Type U	0.86	0.94	0.90
Type V	0.83	0.91	0.87	Type V	0.92	0.77	0.84
Accuracy			0.92	Accuracy			0.92
Macro Avg	0.92	0.91	0.91	Macro Avg	0.91	0.90	0.91
Weighted Avg	0.92	0.92	0.92	Weighted Avg	0.92	0.92	0.92
MTP (Resnet 50-v2)				MTP (Resnet 152-v2)			
	Precision	Recall	F1-score		Precision	Recall	F1-score
Type M	0.96	1.00	0.98	Type M	0.95	0.99	0.97
Type U	0.98	0.84	0.91	Type U	0.96	0.86	0.91
Type V	0.84	0.91	0.88	Type V	0.86	0.89	0.88
Accuracy			0.93	Accuracy			0.93
Macro Avg	0.93	0.92	0.92	Macro Avg	0.92	0.91	0.92
Weighted Avg	0.93	0.93	0.93	Weighted Avg	0.93	0.93	0.93
MTP (DenseNet 121)				MTP (DenseNet 169)			
	Precision	Recall	F1-score		Precision	Recall	F1-score
Type M	0.95	0.96	0.95	Type M	0.97	0.94	0.95
Type U	0.98	0.88	0.93	Type U	0.95	0.82	0.88
Type V	0.84	0.91	0.88	Type V	0.77	0.94	0.85
Accuracy			0.93	Accuracy			0.90
Macro Avg	0.92	0.92	0.92	Macro Avg	0.90	0.90	0.89
Weighted Avg	0.93	0.93	0.93	Weighted Avg	0.91	0.90	0.90

TABLE V. RELATIVELY EVALUATED PERFORMANCE METRICS OF MTP AND TRADITIONAL ML CLASSIFICATION ALGORITHMS.

Models	Data set	#Samples	#Features	Precision	Recall	F1-score	Accuracy
Multi-Task Process - Resnet 50-v2 (MTP)	eg-SHd	874	9	0.93	0.92	0.92	0.93
Extreme Gradient Boosting (XGBoost)	o-SHd	123	9	0.67	0.60	0.60	0.60
Random-Forests (RF)	o-SHd	123	9	0.67	0.60	0.57	0.60
Linear Discriminant Analysis (LDA)	o-SHd	123	9				0.57
(k-Nearest Neighbors Algorithm) k-NN	o-SHd	123	9	0.71	0.48	0.52	0.48
Naive Bayes Classifier	o-SHd	123	9	0.61	0.52	0.52	0.52
Support Vector Machines (SVM)	o-SHd	123	9	0.46	0.44	0.41	0.44

TABLE VI. COMPARISON OF MTP AND OTHER ML CLASSIFIERS PERFORMANCES ACROSS STANDARD DATA SETS.

Data sets	Data representation	Data augmentation	#Samples	#Features	Models	Precision	Recall	F1-score	Accuracy
eg-Iris	SuperTML_EF	GAN Keras	899	4	MTP (CNN)	0.92	0.91	0.91	0.91
eg-Wine			1666	13		0.91	0.91	0.91	0.91
eg-BCW			1664	10		0.96	0.96	0.96	0.96
Iris	Tabular Data	-	150	4	XGBoost	0.90	0.90	0.90	0.90
					RF	0.97	0.97	0.97	0.97
					SVM	0.97	0.97	0.97	0.97
Wine			178	13	XGBoost	0.95	0.94	0.94	0.94
					RF	0.97	0.97	0.97	0.97
					SVM	1.00	1.00	1.00	1.00
BCW			699	10	XGBoost	0.96	0.96	0.96	0.96
					RF	0.94	0.94	0.94	0.94
					SVM	0.98	0.98	0.98	0.98

IV. DISCUSSION

In tabular data analysis, traditional classifiers typically outperformed CNN models. However, the results of the experiments in this study demonstrated that CNN models are more effective when handling tabular data. This study shows their effectiveness in medical data sets, showing that CNN models are consistently better than traditional classifiers. Additionally, this study shows that the MTP can perform well with small and multiclass data sets.

Although previous studies of SuperTML [24], DeepInsight [25], IGTD [27], and DWTM [28] focus on the idea of transforming tabular data into images, none of them used GANs for tabular data. MTP is a novel approach that uses its medical tabular data set, creating it with manual feature extraction and synthetically increasing it. MTP performs remarkably well on small data sets compared to the IGTD method and the DeepInsight technique. In the future, combining different methods and techniques and using them with the MTP approach can produce the best method for all kinds of tabular data classification tasks.

V. CONCLUSIONS

In this study, we have presented a novel approach to SH classification, MTP, which considers landmarks located on sacrum bones and their distances from each other to classify SH; in this way, it represents medical images with numbers in tabular data.

This attitude allows us to use the GAN, which makes the tabular data set synthetically generable. In this way, it eliminates significant problems in the medical field in case of insufficient or limited data sets. Also, transforming the tabular data set into new image forms makes training processes with CNNs possible using transfer learning techniques even in the face of the limited amounts, poor, and varying qualities of medical data.

The experiments demonstrate that the proposed approach achieved significant performance and usually outperformed results of traditional ML classifiers. In future work, we plan to improve the MTP classification approach by automating feature extraction and extending it to other medical applications.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] F. A. R. Stammers, "PERIPHERAL CIRCULATION IN MAN. A CIBA FOUNDATION SYMPOSIUM. Editors, for the Ciba Foundation, G. E. W. Wolstenholme, O. B. E., M. A., M. B., B. Ch.; and Jessie S. Freeman, M. B.; D. P. H.; assisted by Joan Etherington. 8½ x 6 in. Pp. xi+2l9, with 72 figures. Index. 1954. London: J. & A. Churchill Ltd. Price 25s", *J. Bone Jt. Surg. Br.*, vol. 36-B, no. 4, pp. 705–706, 1954. DOI: 10.1302/0301-620X.36B4.705-a.

[2] A. A. Shinde, P. R. Manvikar, and V. K. Bharambe, "Morphometric study of sacral hiatus and its significance in caudal epidural anesthesia", *Sahel Med. J.*, vol. 18, no. 3, pp. 134–138, 2015. DOI: 10.4103/1118-8561.169281.

[3] V. S. Lanier, H. E. McKnight, and M. Trotter, "Caudal analgesia: An experimental and anatomical study", *Am. J. Obstet. Gynecol.*, vol. 47, pp. 633–641, 1944. DOI: 10.1016/S0002-9378(16)40318-2.

[4] S. J. Zito, "Adult caudal anesthesia: A reexamination of the technique", *AANA J.*, vol. 61, no. 2, pp. 153–157, 1993.

[5] A. Singh, R. Gupta, and A. Singh, "Morphological and morphometrical study of sacral hiatus of human sacrum", *Natl. J. Integr. Res. Med.*, vol. 9, no. 4, pp. 65–73, 2018. [Online]. Available: http://nicpd.ac.in/ojs-/index.php/njirm/article/view/2385

[6] B. Bier *et al.*, "Learning to detect anatomical landmarks of the pelvis in X-rays from arbitrary views", *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 9, pp. 1463–1473, 2019. DOI: 10.1007/S11548-019-01975-5.

[7] Y. Abukawa *et al.*, "Ultrasound versus anatomical landmarks for caudal epidural anesthesia in pediatric patients", *BMC Anesthesiol.*, vol. 15, pp. 1–4, 2015. DOI: 10.1186/S12871-015-0082-0.

[8] C. P. C. Chen *et al.*, "Ultrasound guidance in caudal epidural needle placement", *Anesthesiology*, vol. 101, no. 1, pp. 181–184, 2004. DOI: 10.1097/00000542-200407000-00028.

[9] S. Mourgela, S. Anagnostopoulou, A. Sakellaropoulos, A. Koulousakis, and J. P. Warnke, "Sectioning of filum terminale externum using a rigid endoscope through the sacral hiatus. Cadaver study", *J. Neurosurg. Sci.*, vol. 52, no. 3, pp. 71–74, 2008.

[10] F. Yalcinkaya and A. Erbas, "Convolutional Neural Network and Fuzzy Logic-based Hybrid Melanoma Diagnosis System", *Elektron Elektrotech*, vol. 27, no. 2, pp. 55–63, 2021. DOI: 10.1016/S0034-7094(11)70011-3. DOI: 10.5755/j02.eie.28843.

[11] M. Nikooseresht, M. Hashemi, S. Amir Mohajerani, F. Shahandeh, and M. Agah, "Ultrasound as a screening tool for performing caudal epidural injections", *Iran. J. Radiol.*, vol. 11, no. 2, p. e13262, 2014. DOI: 10.5812/iranjradiol.13262.

[12] L.-Z. Wang, X.-X. Hu, Y.-F. Zhang, and X.-Y. Chang, "A randomized comparison of caudal block by sacral hiatus injection under ultrasound guidance with traditional sacral canal injection in children", *Paediatr. Anaesth.*, vol. 23, no. 5, pp. 395–400, 2013. DOI: 10.1111/pan.12104.

[13] F. Bonnet, "Principles & practice of regional anaesthesia, 3rd Edn. J. A. W. Wildsmith, E. N. Armitage and J. H. McClure (editors). Published by Churchill Livingstone, Edinburgh. Pp. 346; ISBN 0-443-06226-9", *Br. J. Anaesth.*, vol. 91, no. 3, pp. 453–454, 2003. DOI: 10.1093/bja/aeg608.

[14] S. K. Nagar, "A study of sacral hiatus in dry human sacra", *J. Anat. Soc. India*, vol. 53, no. 2, pp. 18–21, 2004.

[15] S. Yilmaz *et al.*, "Sakrum kemiğinin morfometrik değerlendirilmesi", *Bozok Tıp Derg.*, vol. 8, no. 4, pp. 13–17, 2018. DOI: 10.16919/bozoktip.397912.

[16] D. Sarika H. and B. Pritha S., "Morphometric study of first sacral vertebra in dry human sacra", *Indian J. Anat.*, vol. 7, no. 2, pp. 183–187, 2018. DOI: 10.21088/ija.2320.0022.7218.13.

[17] B. Chaudhary, A. Asghar, S. Naaz, and A. Satyam, "Estimation of the relationship between the sacral hiatus and other dorsal sacral parameters using principle component analysis", *Surg. Radiol. Anat.*, vol. 43, no. 9, pp. 1545–1554, 2021. DOI: 10.1007/s00276-021-02794-7.

[18] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making", *Big Data*, vol. 1, no. 1, pp. 51–59, 2013. DOI: 10.1089/big.2013.1508.

[19] W. S. Noble, "What is a support vector machine?", *Nat. Biotechnol.*, vol. 24, pp. 1565–1567, 2006. DOI: 10.1038/nbt1206-1565.

[20] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression: Third Edition*. Wiley, 2013. DOI: 10.1002/9781118548387.

[21] L. Breiman, "Random forests", *Mach. Learn.*, vol. 45, pp. 5–32, 2001. DOI: 10.1023/A:1010933404324.

[22] M. Iman, A. Giuntini, H. R. Arabnia, and K. Rasheed, "A comparative study of machine learning models for tabular data through challenge of monitoring Parkinson's disease progression using voice recordings", in *Advances in Computer Vision and Computational Biology. Transactions on Computational Science and Computational Intelligence*. Springer, Cham, 2021, pp. 485–496. DOI: 10.1007/978-3-030-71051-4_38.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nat.*, vol. 521, pp. 436–444, 2015. DOI: 10.1038/nature14539.

[24] B. Sun *et al.*, "SuperTML: Two-dimensional word embedding for the precognition on structured tabular data", in *Proc. of IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, 2019, pp. 2973–2981, 2019. DOI: 10.1109/CVPRW.2019.00360.

[25] A. Sharma, E. Vans, D. Shigemizu, K. A. Boroevich, and T. Tsunoda, "DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture", *Sci. Reports*, vol. 9, art. no. 11399, 2019. DOI: 10.1038/s41598-019-47765-6.

[26] O. Bazgir, R. Zhang, S. R. Dhruba, R. Rahman, S. Ghosh, and R. Pal, "Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks", *Nat. Commun.*, vol. 11, art. no. 4391, 2020. DOI: 10.1038/s41467-020-18197-y.

[27] Y. Zhu *et al.*, "Converting tabular data into images for deep learning with convolutional neural networks", *Sci. Reports*, vol. 11, art. no. 11325, 2021. DOI: 10.1038/s41598-021-90923-y.

[28] M. I. Iqbal, M. S. H. Mukta, A. R. Hasan, and S. Islam, "A dynamic weighted tabular method for convolutional neural networks", *IEEE Access*, vol. 10, pp. 134183–134198, 2022. DOI: 10.1109/ACCESS.2022.3231102.

[29] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need", *Inf. Fusion*, vol. 81, pp. 84–90, 2022. DOI: 10.1016/j.inffus.2021.11.011.

[30] I. Ashrapov, "Tabular GANs for uneven distribution", 2020. DOI: 10.48550/arXiv.2010.00638.

[31] E. Bataa and J. Wu, "An investigation of transfer learning-based sentiment analysis in Japanese", 2019. DOI: 10.18653/v1/P19-1458.

[32] C. A. Ferreira *et al.*, "Classification of breast cancer histology images through transfer learning using a pre-trained Inception Resnet V2", *Image Anal. Recognition. ICIAR 2018. Lect. Notes Comput. Sci.*, vol 10882. Springer, Cham, 2018, pp. 763–770. DOI: 10.1007/978-3-319-93000-8_86.

[33] J. Chang, J. Yu, T. Han, H.-j. Chang, and E. Park, "A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer", in *Proc. of 2017 IEEE 19th Int. Conf. e-Health Networking, Appl. Serv.*, 2017, pp. 1–4. DOI: 10.1109/HealthCom.2017.8210843.

[34] M. Hussain, J. J. Bird, and D. R. Faria, "A study on CNN transfer learning for image classification", in *Adv. Comput. Intell. Syst. UKCI 2018. Adv. Intell. Syst. Comput.*, vol. 840. Springer, Cham, 2019, pp. 191–202. DOI: 10.1007/978-3-319-97982-3_16.

[35] Keras: (About 2023). [Online]. Available: https://keras.io/about

[36] TensorFlow: (About 2023). [Online]. Available: https://www.tensorflow.org/about

[37] Y. Gao and K. M. Mosalam, "Deep transfer learning for image-based structural damage recognition", *Comput. Civ. Infrastruct. Eng.*, vol. 33, no. 9, pp. 748–768, 2018. DOI: 10.1111/mice.12363.

[38] UCI Machine Learning Repository: Iris Data Set. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/iris

[39] UCI Machine Learning Repository: Wine Data Set. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/wine

[40] UCI Machine Learning Repository: Breast Cancer Wisconsin (Original) Data Set. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+%28 original%29