Early Detection of Hip Periprosthetic Joint Infections Through CNN on Computed Tomography Images

Francesco Guarnera[1]([✉]) [iD], Alessia Rondinella[1] [iD], Oliver Giudice[1] [iD],
Alessandro Ortis[1] [iD], Sebastiano Battiato[1] [iD], Francesco Rundo[2] [iD],
Giorgio Fallica[3], Francesco Traina[4], and Sabrina Conoci[5] [iD]

[1] Department of Mathematics and Computer Science, University of Catania,
Catania, Italy
francesco.guarnera@unict.it, alessia.rondinella@phd.unict.it,
{giudice,ortis,battiato}@dmi.unict.it
[2] STMicroelectronics, ADG R&D Power and Discretes Division, Catania, Italy
francesco.rundo@st.com
[3] IBMTech s.r.l., via Napoli 116, Catania, Italy
giorgio.fallica@ibmtech.it
[4] University of Bologna, Rizzoli Orthopedic Institute of Bologna, Bologna, Italy
francesco.traina@ior.it
[5] Department ChBioFarAm, University of Messina, Messina, Italy
sabrina.conoci@unime.it

Abstract. Early detection of an infection prior to prosthesis removal (e.g., hips, knees or other areas) would provide significant benefits to patients. Currently, the detection task is carried out only retrospectively with a limited number of methods relying on biometric or other medical data. The automatic detection of a periprosthetic joint infection from tomography imaging is a task never addressed before. This study introduces a novel method for early detection of the hip prosthesis infections analyzing Computed Tomography images. The proposed solution is based on a novel ResNeSt Convolutional Neural Network architecture trained on samples from more than 100 patients. The solution showed exceptional performance in detecting infections with an experimental high level of accuracy and F-score.

Keywords: Periprosthetic Joint Infection detection · Hip Arthoplasthy · Medical Imaging · Artificial Intelligence

1   Introduction

The surgical replacement of human joints has become increasingly common in recent years as a treatment to pathologies like osteoarthritis or rheumatoid arthritis. However, the Periprosthetic Joint Infection (PJI) that unfortunately occurs around a joint implant, still represents a serious concern for patients

and physicians [1]. A PJI can lead to pain, joint dysfunction, and the need for revision surgery, which typically pose an increased potential for further complications for the patients, as well as additional costs [2]. To address this challenge, medical imaging plays a critical role in the early and accurate detection of PJI. Traditionally, PJI detection is carried out retrospectively by means of a combination of methods relying on biometric or other medical data [3] which could be time-consuming and limited in term of detection accuracy. To counter this, Machine Learning (ML) techniques could train on past data and be applied to PJI detection, with the aim of improving the accuracy of diagnosis [4,5]. Certainly, the use of Computed Tomography (CT) images provided several advantages in the diagnosis of infection [6], giving that the periosteal response should be a strong prior when the infection is present [7]. With this hypothesis, the information obtained from CT scans could be employed as input to supervised ML approaches to develop models able to accurately predict PJIs, reducing the risk of recurrent infections and improving patient outcomes. However, the PJI is particularly challenging to be diagnosed, even for physicians, but recent evolution in Convolutional Neural Network (CNN) architecture have proven great results in medical imaging [9] giving inspiration to this work. In this paper, a Convolutional Neural Network (CNN) solution for the classification of infected and aseptic patients with hip replacements by analyzing CT scans is presented [8].

The proposed solution exploits the attention mechanism of recent ResNeSt [10] architecture to proper extract features from CT scans and build a predictive model that can accurately differentiate between infected and non-infected (aseptic) patients. A private dataset acquired at Rizzoli Orthopedic Institute (IOR) was employed for training and final evaluation of the solution (approved by ethical committee). The dataset was manually labeled by expert radiologists. The obtained results showed the effectiveness of the proposed solution in the PJI detection task, demonstrating its potential to improve overall medical outcomes. This highlights the importance of medical imaging in the diagnosis of PJI and underscores the potential of machine learning techniques in this area.

The remainder of this paper is organized as follows: Sect. 2 briefly presents related works; Sect. 3 details the proposed solution and Sect. 4 shows the experimental results with discussion. Conclusions are summarized in Sect. 5.

2   Related Works

The state of the art (SOTA) methods for the detection of PJIs employ mainly statistical methods, such as regression and Fisher test. In [11] a risk prediction model is proposed for PJI detection within 90 days after surgery using Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis [12]. Other approaches propose the use of the Fisher's test, to detect the infection from CT images at the site of hip prosthesis before surgery [13,14]. While these methods demonstrated to be effective, they are limited in capturing complex relationships between imaging features thus better understanding the PJI status. For instance, many patients' problems, such as mechanical loosening, can

Coronal

Sagittal

Axial Bone-Tissue

Axial Soft-Tissue
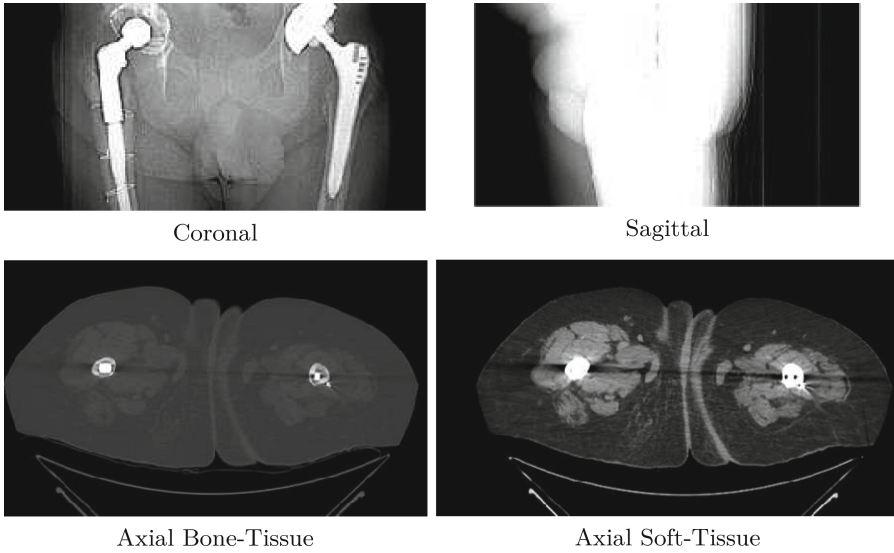
Fig. 1. Examples of acquisitions in CT.

be detected only retrospectively by studying the bone-prosthesis system and its evolution [15]. Only a limited number of studies exploited machine learning (ML) techniques. This is likely due to issue concerning datasets: availability of reliable labeling, variability in image quality and in patients samples, etc. However, ML approaches have the potential to significantly enhance the accuracy of PJI detection. Several studies have demonstrated the potential of deep learning approaches for PJI prediction, with promising results in terms of accuracy, sensitivity, and specificity. The authors of [16] proposed a ResNet architecture to classify pathological sections of patients with PJI achieving high accuracy. In [5], a ML approach was proposed to anticipate recurrent infections in patients who have undergone revision total knee arthroplasty due to a PJI. In some cases, the performance of these models has been shown to be superior to traditional diagnostic methods that rely solely on clinical inspection and laboratory tests. To the best of the authors' knowledge, as far as hip CT images are concerned, there are no SOTA methods able to early detect PJIs. This specific medical imaging problem is still in its early stages and there are several obstacles that need to be overcome: the most important one is the scarcity of large datasets. Given this, a specific dataset of hip CT scan images was collected and labeled by experts. Thus, a novel CNN solution was introduced in order to advance the state of the art in early PJI detection task, achieving the first automatic approach with promising high level of accuracy.

3   Methodology

The CT scans are a widely used medical imaging data providing detailed cross-sectional images of internal structures of a human body. In the context of hip replacements, they play an important role in detecting a PJI. The axial plane imaging mode has the ability to highlight both bone and soft tissue (see Fig. 1), thus facilitating the identification of a orthopedic joint and bone infections.

3.1   Hip CT Scans Image Dataset

At first, patients who underwent hip replacement surgery and subsequent CT scans for various reasons, including postoperative follow-up, suspected PJI, or routine monitoring were identified at Rizzoli Orthopedic Institute of Bologna. As infection is not easily detectable, and there is no standard protocol for medical experts to identify it. However, physicians are able to detect infection clues in proximity to the bone borders, with particular attention to the acetabular and femoral area. Finally, they are able to confirm the presence of infection by means of cultures: this is a retrospective method. In order to have a first early detection solution, a dataset consisting of 102 CT scans of patients with hip replacements was collected (52 samples of an infected patient and 50 aseptic ones). Given the before mentioned insight, only images regarding the axial bone tissue were employed. The labeling of this dataset posed a further challenge: the labels had to be assigned at patient level. In other words, each patient is labeled as either being infected or aseptic, rather than each individual CT scan image. Thus, all images from a specific patient will be labeled as infected or aseptic. Also each CT scan contains a different number of images per patient. These are big issues for ML solutions as it means that no explicit information is available about the presence or absence of a PJI on each individual images. All collected images are in Digital Imaging and Communication in Medicine (DICOM) format[1].
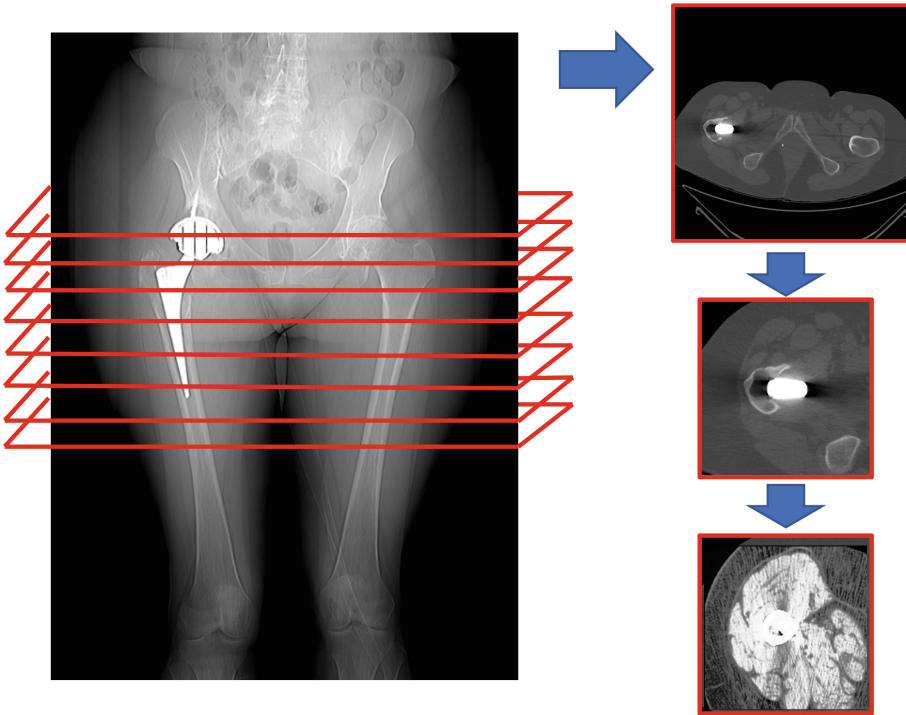
3.2   Pre-processing Pipeline

In order to identify what images contain the prosthesis, every pixel of the images have been converted through the Hounsfield scale, as follows:

$$h_p = p * s + i \tag{1}$$

where p is a pixel value, s and i are the slope and the intercept respectively contained in the DICOM metadata, and h_p is the value of the pixel in the Hounsfield scale. The Hounsfield scale is a quantitative measure for describing radio-density which is the property of relative transparency of a material to the passage of the X-ray portion of the electromagnetic spectrum. Images with $h_p > 3000$ were selected as being images presenting non-human material like a metallic prosthesis [17]. This allowed the selection of images coming from the upper part of the acetabulum to the lower part of the stem only.

---

[1] https://www.dicomstandard.org/.

Fig. 2. Pre-processing pipeline: Slice selection based on Hounsfield selection, contour-based ROI detection and final histogram equalization.

A further preprocessing operation was carried out extracting the contours of the bone relative to the prothesis [18] on images having $h_p > 1000$. Once the contours were extracted, the corresponding centroid was computed and a sub-image of 188×188 pixels was extracted. Finally, histogram equalization was carried out on these patches to adjust the contrast and to normalize different image histograms. The overall pre-processing operations are graphically summarized in Fig. 2.

3.3   The Proposed Approach for PJI Detection

In this section the proposed CNN-based solution for the early detection of hip PJI detection is described. A ResNeSt architecture [10] was employed to classify each data sample (CT scan images from a patient with hip replacements) between infected or aseptic classes. The ResNeSt architecture is an evolution of the ResNet neural network [19], incorporating split-attention blocks instead of residual blocks, thus emphasizing the correlation of informative feature maps. ResNeSt was selected for being the best on the specific task, after testing other similar architectures. More specifically, performances obtained by the ResNet and the ResNeXt architectures [20] were compared, However, the ResNeSt

network architecture, incorporating split-attention mechanisms, demonstrated superior results in both the validation set and the test set. The other evaluated architectures, although achieving similar levels of accuracy, demonstrated poor generalization performances, as indicated by their "randomic behaviour" on heatmap analysis. The proposed approach takes input images of size 188×188 that have undergone pre-processing, as described in Sect. 3.2. The images are then processed within the architecture and the loss function is computed based on the comparison between the output predictive values and truth labels of the data. The loss function employed was the cross entropy and moreover, in order to improve the stability of the model, the Jacobian regularization was applied [21]. Jacobian regularization permitted to reduce the impact of input perturbations caused by metallic artifacts in the prostheses. The use of ResNeSt with the addition of regularization techniques allowed to exploit the strengths of this state-of-the-art network architecture to extract features from CT scans and accurately classify infected and aseptic patients with hip replacements (with explainable non-randomic heatmaps).

4   Experiments and Results

As outlined in Sect. 3.1, the dataset collected and employed for experiments is composed of 50 aseptic and 52 infected patients. To evaluate the generalizing properties of the proposed solution, a balanced test set (D_x) of 12 patients (6 aseptic and 6 infected) was extracted from the full dataset. Additionally, to effectively control overfitting during the training process, we employed four distinct balanced validation sets $(D_{v1}, D_{v2}, D_{v3}, \text{and } D_{v4})$, each consisting of 12 different patients. The remaining 42 patients were considered as D_t. The following configurations were considered for the training phase:

– C_1: train $\{D_t \cup D_{v2} \cup D_{v3} \cup D_{v4}\}$ and valid D_{v1};
– C_2: train $\{D_t \cup D_{v1} \cup D_{v3} \cup D_{v4}\}$ and valid D_{v2};
– C_3: train $\{D_t \cup D_{v1} \cup D_{v2} \cup D_{v4}\}$ and valid D_{v3};
– C_4: train $\{D_t \cup D_{v1} \cup D_{v2} \cup D_{v3}\}$ and valid D_{v4}.

Training was performed for 100 epochs, employing ADAM as optimizer with a starting learning rate of $1e{-}4$, a weight decay equal to $1e{-}4$ and a batch size fixed at 4. The proposed approach was implemented in Python language (version 3.9.7) using the Pytorch package. All experiments were done on a NVIDIA Quadro RTX 6000 GPU. The overall training procedure took 50 h.

Tables 1 and 2 shows the results obtained applying the best model obtained for each configuration and tested on D_x. As explained in Sect. 3.3 the proposed technique makes a classification (positive or infected vs. negative or aseptic) for each image of a patient; this implies that the model is independent to the number of images per patient, as each patient has a different number of images based on their length of the prosthesis (third column of Tables 1) and 2. It has to be noted that, in order to accurately classify a patient as infected or aseptic through the images predicted by the model, a threshold strategy must be employed. The

**Table 1.** Accuracy and F-score obtained on C_1,C_2 configurations employing the same test set D_x.

Patient number	Patient type	Number images	Accuracy / F-score configuration C_1	Accuracy / F-score configuration C_2
1	aseptic	67	0.88/0.94	0.88/0.93
2	aseptic	39	0.66/0.8	0.41/0.58
3	aseptic	69	0.85/0.85	0.76/0.87
4	aseptic	68	0.52/0.69	0.85/0.92
5	aseptic	77	0.79/0.88	0.79/0.88
6	aseptic	53	0.96/0.98	1/1
7	infected	80	0.46/0.63	0.51/0.68
8	infected	115	1/1	1/1
9	infected	191	0.98/0.99	0.91/0.96
10	infected	71	0.18/0.31	0.13/0.20
11	infected	77	0.57/0.72	0.89/0.94
12	infected	117	0.95/0.98	0.96/0.98

Table 2. Accuracy and F-score obtained on C_3,C_4 configurations employing the same test set D_x.

Patient number	Patient type	Number images	Accuracy / F-score configuration C_3	Accuracy / F-score configuration C_4
1	aseptic	67	0.91/0.95	0.89/0.94
2	aseptic	39	0.92/0.96	0.56/0.72
3	aseptic	69	0.76/0.87	0.67/0.8
4	aseptic	68	0.70/0.83	0.57/0.73
5	aseptic	77	0.96/0.98	0.82/0.9
6	aseptic	53	1/1	0.92/0.96
7	infected	80	0.39/0.55	0.65/0.79
8	infected	115	0.98/0.99	1/1
9	infected	191	0.82/0.90	0.92/0.96
10	infected	71	0.03/0.05	0.13/0.22
11	infected	77	0.52/0.68	0.91/0.95
12	infected	117	0.99/0.99	0.91/0.95

choice of the threshold could be done considering the right trade-off between false positive and false negative.

Accuracy (Eq. 2) and F-score (Eq. 3) were employed as performance metrics as shown in Tables 1 and 2.
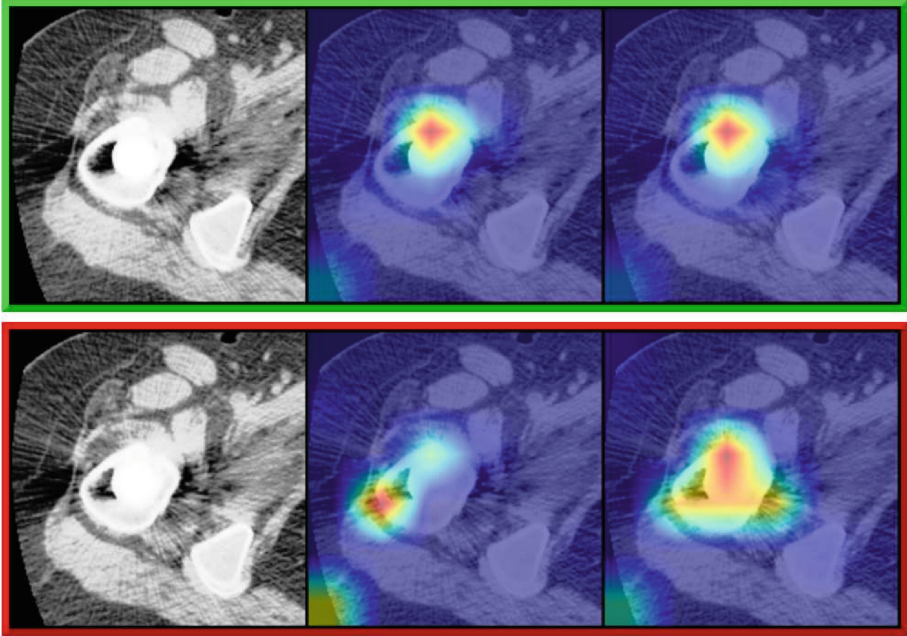
Fig. 3. Input images and corresponding heatmaps of trained model attention produced by means of GradCam and GradCam++ libraries. Green border indicates a correct prediction while the red border shows a wrong one. (Color figure online)

Accuracy is defined as follows:

$$ACC_y = Y_{rp}/PNI \tag{2}$$

where y is the class, Y_{rp} represents the right prediction for that class and PNI is the number of images per patient.

F-score is defined as follows:

$$F = 2 * (P * R)/(P + R) \tag{3}$$

where P and R are precision and recall respectively. Results shown in Tables 1 and 2 show how the model generalizes with respect to the training set: accuracy and F-score are similar for each one of the cross-validation configurations.

In the task of PJI detection, experts in the field emphasize the significance of focusing on regions in images adjacent to the bones, as these areas often hold discriminatory information. Building upon this expert knowledge, a Gradient-weighted Class Activation Mapping (Grad-Cam) analysis [22] was conducted on each test image using the trained model, aiming to enhance the interpretability of the model's decisions. Through this analysis, deeper insights into the interpretability and clinical relevance of the heatmaps generated were gained. The examination of Grad-Cam heatmaps enabled to identify discriminative features

near the bones, which exhibited a strong correlation with accurate predictions. This correlation is exemplified by the green-bordered example shown in Fig. 3. Conversely, when the model placed emphasis on multiple features dispersed throughout the image, the prediction was incorrect.

5   Conclusion and Future Works

In this paper an early detection technique for hip periprosthetic joint infections on computed tomography images was presented. This was a first in the state of the art to the best of author's knowledge. To this aim, hip CT images were collected and labeled by experts. A dedicated pre-processing pipeline was developed and a ResNeSt CNN solution was trained. The proposed pipeline demonstrated to achieve strong performances (in terms of accuracy and F-score) in detecting infected and aseptic images of a specific patient.

In order to strengthen the robustness and generalizability of our proposed model, future research will focus on expanding the dataset through collaboration with other institutions to obtain a more diverse and representative dataset, enabling our model to perform effectively across various clinical settings.

Furthermore, our research will explore alternative forms of input, such as compressed images, and the implementation of additional automated techniques to further enhance the prediction of prosthesis-related joint infections. Additionally, we will investigate the outliers in the collected dataset to better understand their impact on the overall results.

References

1. Workgroup Convened by the Musculoskeletal Infection Society and others: New definition for periprosthetic joint infection. J. Arthroplast. **26**(8), 1136–1138 (2011)
2. Sculco, T.P.: The economic impact of infected total joint arthroplasty. Instr. Course Lect. **42**, 349–351 (1993)
3. Ting, N.T., Della Valle, C.J.: Diagnosis of periprosthetic joint infection-an algorithm-based approach. J. Arthroplast. **32**(7), 2047–2050 (2017)
4. Kuo, F.-C., Hu, W.-H., Hu, Y.-J.: Periprosthetic joint infection prediction via machine learning: comprehensible personalized decision support for diagnosis. J. Arthroplast. **37**(1), 132–141 (2022)
5. Klemt, C., et al.: Machine learning models accurately predict recurrent infection following revision total knee arthroplasty for periprosthetic joint infection. Knee Surgery, Sports Traumatology, Arthroscopy. 1–9 (2021)
6. Cyteval, C., Hamm, V., Sarrabère, M.P., Lopez, F.M., Maury, P., Taourel, P.: Painful infection at the site of hip prosthesis: CT imaging. Radiology **224**(2), 477–483 (2002)

7. Kapadia, B.H., Berg, R.A., Daley, J.A., Fritz, J., Bhave, A., Mont, M.A.: Periprosthetic joint infection. The Lancet **387**(10016), 386–394 (2016)
8. Conoci, S., Traina, F.: Image classification method, in particular medical images, for example radiographic images. I.T. Patent 102020000031289, June 2022
9. Rondinella, A., et al.: Boosting multiple sclerosis lesion segmentation through attention mechanism. Comput. Biol. Med. **161**, 107021 (2023)
10. Zhang, H., et al.: ResNeSt: split-attention networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 2736–2746, June 2022
11. Bülow, E., Hahn, U., Andersen, I.T., Rolfson, O., Pedersen, A.B., Hailer, N.P.: Prediction of early periprosthetic joint infection after total hip arthroplasty. Clin. Epidemiol. 239–253 (2022)
12. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. Roy. Stat. Soc. Ser. B (Methodol.) **58**(1), 267–288 (1996)
13. Galley, J., Sutter, R., Stern, C., Filli, L., Rahm, S., Pfirrmann, C.W.A.: Diagnosis of periprosthetic hip joint infection using MRI with metal artifact reduction at 15 T. Radiology **296**(1), 98–108 (2020)
14. Isern-Kebschull, J., et al.: Value of multidetector computed tomography for the differentiation of delayed aseptic and septic complications after total hip arthroplasty. Skeletal Radiol. **49**, 893–902 (2020)
15. Andrä, H., et al.: Structural simulation of a bone-prosthesis system of the knee joint. Sensors **8**(9), 5897–5926 (2008)
16. Tao, Y., Hu, H., Li, J., Li, M., Zheng, Q., Zhang, G., Ni, M.: A preliminary study on the application of deep learning methods based on convolutional network to the pathological diagnosis of PJI. Arthroplasty **4**(1), 49 (2022)
17. Morar, L., et al.: Analysis of CBCT bone density using the Hounsfield scale. Prosthesis **4**(3), 414–423 (2022)
18. Suzuki, S., et al.: Topological structural analysis of digitized binary images by border following. Comput. Vision Graph. Image Process. **33**(1), 32–46 (1985)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
20. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500 (2017)
21. Hoffman, J., Roberts, D.A., Yaida, S.: Robust learning with Jacobian regularization. arXiv preprint arXiv:1908.02729 (2019)
22. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Gradcam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the Ieee International Conference on Computer Vision, pp. 618–626 (2017)