Automated Identification of Make and Model of Total Wrist Replacement Implants using Deep Learning

Conference Paper · August 2023

DOI: 10.11159/mvml23.119

6 authors, including:

Saisha Shetty
University of California, Davis

Gayathri Mani
SRM Institute of Science and Technology

Malathy Chidambaranathan
SRM Institute of Science and Technology

Vineet Batta
Luton and Dunstable Hospital NHS Foundation Trust

Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23)
Brunel University, London, United Kingdom - August 03-05, 2023
Paper No. MVML 119
DOI: 10.11159/mvml23.119

Automated Identification of Make and Model of Total Wrist Replacement Implants using Deep Learning

Saisha Shetty[1], Naman Garg[1], Gayathri M[1], Malathy C[1], Vineet Batta[2*], A Ramanathan[1]
[1]School of Computing, College of Engineering and Technology, S.R.M Institute of Science & Technology
Kattankulathur, Chennai, India
ss5109@srmist.edu.in; ng7739@srmist.edu.in; gayathrm2@srmist.edu.in; malathyc@srmist.edu.in;
ramsaivigbala@yahoo.co.in
[2]Department of Orthopaedics and Trauma, Luton and Dunstable University Hospital NHS Trust,
Luton, United Kingdom
battavineet@doctors.org.uk

***Abstract* -** Accurately identifying orthopaedic implants is a crucial step in executing revision surgeries, as any misidentification can result in surgical delays and adverse outcomes. With the rising number of primary and revision surgeries for wrist replacement, there is a growing need for a reliable method to recognize the make and model of wrist implants depicted in X-ray images. This paper proposes an innovative approach that employs deep learning techniques to accurately identify wrist implants, potentially enhancing the precision and efficiency of revision surgeries. The study demonstrated that the utilisation of deep learning techniques was extremely effective in identifying the exact make and model of wrist implants from X-ray images, with a remarkable accuracy rate of 95.12% a superior Area Under Curve (AUC) of 0.9959 in identifying 3 models of total wrist prosthesis.

***Keywords*:** Wrist Implant, Classification, Deep learning, Machine learning, Computer vision, Identification

1. Introduction

Wrist replacement surgery, formally referred to as total wrist arthroplasty (TWA), is a surgical procedure that is performed to alleviate arthritis in the wrist joint.[1]. The TWA is more effective in preserving the motion of the wrist joint, helps in relieving pain [2], and provides strength and stability to perform day-to-day activities without much trouble [1].

However, replacement implants suffer from huge complications such as implant loosening, instability, periprosthetic fractures, and dislocations which demand the need for revision surgery. Sometimes the requirement for repeat revision arises because of carpal and radial component loosening and infection. On average, the revision rates of TWA vary between 0% to 60% depending on the above-mentioned factors [2]. Rates of failure of TWA depend on implant fixation methods and implant design. TWA tends to fail earlier than Knee and Hip arthroplasty [3]. According to the report of the Australian joint registry in 2022, there were 884 wrist replacements performed in Australia alone out of 116 revisions of wrist implants [4]. The Annual report of the Dutch Arthroplasty Register published in 2022 indicates that there were 576 recorded total wrist replacements from 2017 to 2021, out of which 145 were revisions of wrist implants [5]. Identification of the implant is the first and most important step when planning a revision surgery. Failure to identify an implant can cause higher blood loss, increased healthcare costs, and complex surgeries [6]. The paper proposes a novel and efficient framework that implements deep learning techniques to effectively analyse total wrist replacement implants, potentially leading to improved diagnostic and treatment strategies for patients. The work uses Data albumentations methods [7] and transfer learning-based pre-trained models such as VGG16 [8], Resnet50V2 [9], and MobilenetV2 [10] for the classification of implants

2. Literature Survey

Natalia et al investigated how much time surgeons spent identifying failed implants. According to a 2012 survey, 87% of surgeons used at least three methods to identify failed implants prior to surgery, with the median surgeon identification time being 20 minutes. The use of UDIs in TJA registries and Electronic Medical Records (EMR) has been identified as best practice to support implant identification and save time [11].

Sukrit et al used Deep Learning to accurately identify 6 total knee replacement implants with an Area under the Curve value of 0.985. The study employed both AP and LAT view images. DenseNet201 and VGG16 models were used in the study, with DenseNet201 outperforming the others [12].

Ravi et al proposed a convolutional neural network (CNN) that can recognize orthopaedic implant models from radiographs. 12 implant models from 650 patients were included in a dataset of 427 knee and 922 hip unilateral anteroposterior radiographs. On a balanced unseen dataset of 180 radiographs, the newly developed U-Net segmentation network was able to mask around the implants on radiographs, achieving a remarkable 98.9% accuracy and 100% top-three accuracy [13].

Zibo et al applied an AI-based method to identify femoral implant makers and models in total hip arthroplasty (THA) using anteroposterior plain radiographs. The study's deep learning system achieved accuracy rates of 97.9%, sensitivity rates of 88.6%, and specificity rates of 98.9% in the testing set of external images [14].

The automated identification of orthopaedic implants, with applications in knee and hip arthroplasty, has demonstrated great potential as a result of recent developments in deep learning and Artificial Intelligence (AI) techniques. These models have proven to be highly accurate and reliable, often even outperforming senior orthopaedic doctors. The proposed novel framework approach utilises deep learning's potential for identifying orthopaedic whole-wrist replacement implants.

3. Dataset Description

The Dataset gathered consists of three different wrist replacement models. Upon obtaining appropriate authorization, the acquired dataset encompassed X-ray images captured from two different angles, namely the front-to-back i.e. Anterior-Posterior (AP) and side i.e. Lateral (LAT) views, offering a complete and diverse set of data for the analysis. Due to the low count of LAT images, they were discarded and the study used only AP images. The table below contains the names of the make and model of the implants, and the number of available images for each view. The data has been completely anonymized, with all patient details removed to protect their privacy. The images were only labelled for the Make and Model of the respective implants. Table 1 below lists the implants used in the proposed work.

Table 1: Implants used in the proposed work.

MAKE	**MODEL**	**AP IMAGES**	**LAT IMAGES**
DEPUY	BIAX	22	9
INTEGRA	UNIVERSAL 2	49	4
ZIMMER BIOMET	MAESTRO	35	2

Due to the low count of LAT images, they were discarded and the study used only AP images.

A B C

Figure 1 (A-C): Raw images of Depuy Biax, Integra Universal 2, and Zimmer Maestro total wrist implants

Figure 1 shows the original images of Depuy Biax, Integra Universal 2, and Zimmer Maestro total wrist replacement implants

4. Methods and Methodology

4.1.1 Training Set - 70% of the total data for the purpose of training our models. Training of the models depends on the quality of the data that is used. To increase the capacity of the training and to make the model more robust to real-time external images, data augmentations are performed.

4.1.2 Validation Set / Internal Testing Set - The validation set comprises 30% of our unaugmented total data to assess the performance of our model after training. By setting aside this separate subset of data, one can evaluate the model's accuracy and identify any issues with overfitting or underfitting, ensuring that it can generalize effectively to new and unseen data.

4.1.3 Data Augmentation - Data augmentation is the technique of creating more images from existing ones by applying minor modifications to the original dataset. This method applies only to the training sample. To evaluate the effectiveness of our model, we maintain the testing data in its original form without making any alterations. For our method, we made use of the albumentations techniques [7] such as Vertical and Horizontal Flip, Random Rotation, and changing the sharpness and contrast of the images were the few techniques used in a random trial and error manner. The number of data samples increased from 106 to 1800 (each implant class consists of 600 images).

4.2. Deep Learning Methods

After using data augmentation methods, the next step would be to train the model using these augmented images. Convolutional neural networks (CNN) are particularly useful for this task because they can extract and learn important features from the images that are essential for accurate classification. CNNs are a highly effective deep learning technique for classifying wrist implants from their image. The fundamental structure of a CNN architecture comprises three primary categories of layers: convolutional, pooling, and fully connected. Convolutional layers are responsible for extracting significant features from the input image, whereas pooling layers assist in decreasing the dimensionality of the feature maps generated by the convolutional layers. Utilising the extracted features, the fully connected layer performs the classification of the wrist implant's class depicted in the image. To make a prediction, the fully connected layer produces class scores using the softmax activation function. The scores assigned to each class denote the probability of the image being classified under any of the possible categories. Subsequently, the class with the highest probability is designated as the inferred class of the image [15]. In deep learning, transfer learning has gained popularity as a technique that employs pre-trained models to initialise the training of new models, resulting in a reduction of the necessary training time required to achieve optimal performance.. In this paper, the authors focus on three models pre-trained on the ImageNet dataset - VGG16, ResNet50V2, and MobileNet. Transfer learning is a powerful tool that leverages the knowledge learned from large datasets to improve the performance of new models even when working with limited data [16].

4.3. Proposed Method:

The proposed model architecture uses ResNet50V2 [9] as the base model with additional fine-tuning techniques to improve accuracy.ResNet50V2 is a deep neural network architecture that includes residual connections, which help to address the issue of vanishing gradients when training very deep networks. Each ResNet50V2 block contains convolutional layers, batch normalisation, and activation layers, followed by an identity block with a skip connection. It also includes a max pooling layer with a 2x2 pool size and stride. The output of the convolutional layers is combined with the output of the identity block using skip connections. The resulting output is then passed through a nonlinear activation function, such as ReLU or sigmoid, to produce the block output. Additionally, the use of batch normalisation layers in ResNet50V2 helps to improve training by reducing internal covariate shifts. The 'softmax' activation function and 'Adagrad' optimizer were then used to optimise the model for the specific task at hand.

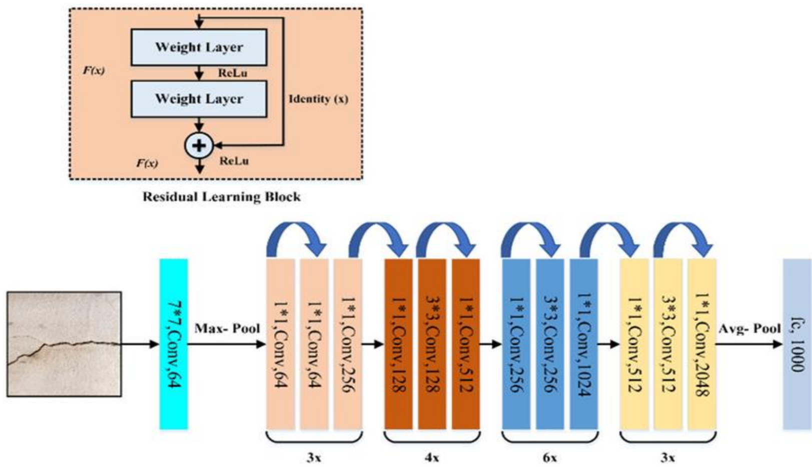

Figure 2: Schematic Diagram of ResNet50V2 Model [17]

Figure 2 shows the diagram of Resnet50V2 Model which had showed superiority in the proposed work

5. Results and Discussion

5.1. Data Augmentation Results:

Figure 3 shows the rotated Images of Depuy Biax, Integra Universal 2, and Zimmer Maestro total wrist implants in various angles thereby increasing the number of training images.

A **B** **C**

Figure 3 (A-C): Rotated Images of Depuy Biax, Integra Universal 2, and Zimmer Maestro total wrist implants

5.2 Deep Learning Results:

Table 2: Best obtained results in the classification of Implants across various Deep learning models

Model	Number of epochs	Training Accuracy (%)	Training Loss	Validation Accuracy (%)	Validation Loss
Resnet50V2	30	95.89	0.1944	95.12	0.2106
VGG16	30	99.81	0.0622	93.94	0.2577
MobileNet	30	99.10	0.1410	92.68	0.3831

Table 2 shows the best-obtained results in the classification of implants across 3 different deep learning models after rigorous fine tuning. Table 2 reveals that ResNet50V2 outperforms other models with a Validation Accuracy of 95.12%.

5.3. Performance Metrics:

In order to assess the performance of the trained model, the authors employ a set of performance metrics. These metrics include precision, recall, F1 score, Accuracy, and Area Under Curve (AUC), Confusion Matrix [18] which were determined by the number of true positives, false positives, true negatives, and false negatives [19].

Table 3: Performance metrics for classification of Implants across various Deep learning models

Model	Number of epochs	AUC	Precision	Recall	F1 Score
Resnet50V2	30	0.9959	0.9508	0.9364	0.9427
VGG16	30	0.9921	0.9512	0.9285	0.9304
MobileNet	30	0.9886	0.9565	0.9061	0.9259

Table 3 shows the best obtained performance metrics in the classification of implants across 3 different deep learning models that are used in Table 2.

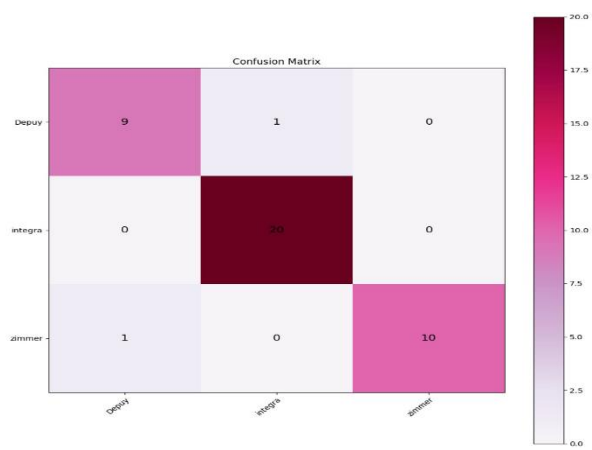

Figure 4 Confusion Matrix of ResNet50V2 Model

Figure 4 shows the plot of the confusion matrix of the Resnet50V2 deep learning model. The implants suffer very small misclassification within the 3 implant classes.

5.4. Training Loss & Accuracy plots:

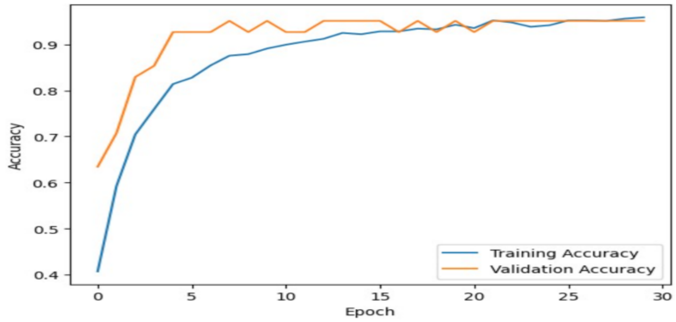

Figure 5: Plot of Accuracy during Training and Validation for ResNet50V2 Model

Figure 5 shows the plot of train and validation accuracy of the ResNet50V2 model. It is clearly seen that both train and validation accuracy increases when trained with 30 epochs.

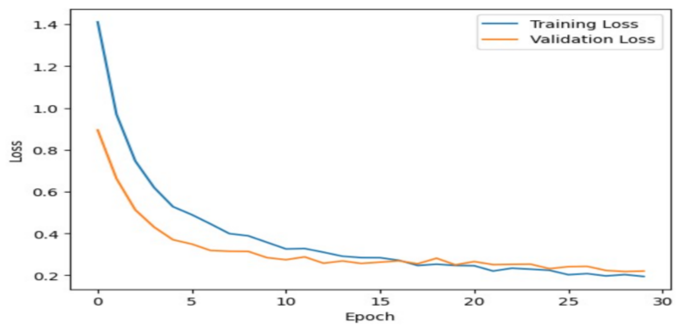

Figure 6: Plot of training and validation loss of ResNet50V2 Model

Figure 6 illustrates the train and validation loss of the ResNet50V2 model, demonstrating a decrease in both losses as the model was trained for 30 epochs.

It is evident from the tables and graphs that Resnet50V2 outperforms VGG16 and MobileNet. All three models were fine-tuned for multiple hyper parameters such as Batch Size, Regularizes, and Learning rates [20]**.** Optimizers such as Adam [21] and SGD [22] were used across the three deep-learning models. The best results were obtained with the Adam optimizer and with a learning rate of 0.0002.

6. Conclusion

The study proposes a novel framework to recognize the make and model of total wrist replacement implants from plain X-ray images. Three popular deep learning models were used in the study and the best results were obtained with the Resnet50V2 model with an accuracy rate of 95.12% and with an F1 Score of 0.9427 respectively. The work can be further extended by adding more implant classes and by exploring higher-end deep learning models.

References

[1] Cooney, W., Manuel, J., Froelich, J. and Rizzo, M., 2012. Total wrist replacement: a retrospective comparative study. *Journal of wrist surgery*, *1*(02), pp.165-172.

[2] Wagner, E.R., Srnec, J.J., Fort, M.W., Barras, L.A. and Rizzo, M., 2021. Outcomes of Revision Total Wrist Arthroplasty. JAAOS Global Research & Reviews, 5(3).

[3] Nair, R., 2014. Total wrist arthroplasty. Journal of Orthopaedic Surgery, 22(3), pp.399-405.

[4] Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR). Hip, Knee & Shoulder Arthroplasty: 2022 Annual Report, Adelaide; AOA, 2022: 1-487. [Accessed from: https://aoanjrr.sahmri.com/annual-reports-2022

[5] LROI, 2022. Annual report 2022. Retrieved from https://www.lroi-report.nl/

[6] Borjali, A., Chen, A.F., Muratoglu, O.K., Morid, M.A. and Varadarajan, K.M., 2020. Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural networks. Journal of Orthopaedic Research®, 38(7), pp.1465-1471.

[7] Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M. and Kalinin, A.A., 2020. Albumentations: fast and flexible image augmentations. Information, 11(2), p.125.

[8] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[9] He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J. and Li, M., 2019. Bag of tricks for image classification with convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 558-567).

[10] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[11] Wilson, N.A., Jehn, M., York, S. and Davis III, C.M., 2014. Revision total hip and knee arthroplasty implant identification: Implications for use of unique device identification 2012 AAHKS member survey results. The Journal of Arthroplasty, 29(2), pp.251-255.

[12] Sharma, S., Batta, V., Chidambaranathan, M., Mathialagan, P., Mani, G., Kiruthika, M., Datta, B., Kamineni, S., Reddy, G., Masilamani, S. and Vijayan, S., 2021. Knee Implant Identification by Fine-Tuning Deep Learning Models. Indian Journal of Orthopaedics, 55, pp.1295-1305.

[13] Patel, R., Thong, E.H., Batta, V., Bharath, A.A., Francis, D. and Howard, J., 2021. Automated identification of orthopedic implants on radiographs using deep learning. Radiology: Artificial Intelligence, 3(4), p.e200183.

[14] Gong, Z., Fu, Y., He, M. and Fu, X., 2022. Automated identification of hip arthroplasty implants using artificial intelligence. Scientific Reports, 12(1), pp.1-8.

[15] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), pp.84-90.

[16] Kornblith, S., Shlens, J. and Le, Q.V., 2019. Do better imagenet models transfer better?. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2661-2671).

[17] Ali, L., Alnajjar, F., Jassmi, H.A., Gocho, M., Khan, W. and Serhani, M.A., 2021. Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. Sensors, 21(5), p.1688.

[18] Vakili, M., Ghamsari, M. and Rezaei, M., 2020. Performance analysis and comparison of machine and deep learning algorithms for IoT data classification. arXiv preprint arXiv:2001.09636.

[19] Ramanathan, A. and Christy Bobby, T., 2020. Classification of Corpus Callosum Layer in Mid-saggital MRI Images Using Machine Learning Techniques for Autism Disorder. In Modeling, Machine Learning and Astronomy: First International Conference, MMLA 2019, Bangalore, India, November 22–23, 2019, Revised Selected Papers 1 (pp. 78-91). Springer Singapore

[20] Yu, T. and Zhu, H., 2020. Hyper-parameter optimization: A review of algorithms and applications. arXiv preprint arXiv:2003.05689.

[21] Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[22] Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.