Informatics in Medicine Unlocked 25 (2021) 100669

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Automated classification of total knee replacement prosthesis on plain film radiograph using a deep convolutional neural network

Samuel C. Belete [a], Vineet Batta [b], Holger Kunz [a,*]

[a] Institute of Health Informatics, University College London, UK
[b] Luton & Dunstable University NHS Hospital Trust, UK

ARTICLE INFO

Keywords:
Automated detection
Convolutional neural network
Machine learning
Total knee replacement
X-ray
Plain film radiograph

ABSTRACT

The identification of the make and model of a total knee replacement (TKR) is a necessary step prior to revision surgery for periprosthetic fracture, loosening, wear or infection. Current methods may fail to correctly identify the implant up to 10% of the time. This study presents the training of a Convolutional Neural Network (CNN) to automatically identify the make and model of seven TKR implants or the absence of a TKR on plain-film radiographs. Our dataset consists of 588 anteroposterior (AP) X-rays of the knee. They were randomly divided into a train, validation and testing sets with a 50:25:25 split. A CNN based on the ResNet-18 architecture was trained with the best model selected using validation results. The final model was tested on the hold-out test dataset.

The trained network demonstrated perfect accuracy in classifying a hold-out test dataset of X-rays to one of the eight labelled classes. Saliency maps demonstrated the outlines of the implants are key to a given prediction.

Further research will benefit from larger datasets with more complete coverage of the possible implants. The ability to recognize that implants are outside the networks trained distribution is essential to such an algorithm operating safely in clinical practice. With these issues and limitations addressed there is potential that such an algorithm could save clinicians time and reduce instances where implants are not identified pre-operatively, simplifying re-operative cases and improving clinical outcomes.

1. Introduction

The identification of orthopaedic implants prior to revision surgery is key to allowing appropriate pre-operative planning and ordering of equipment [1,2]. Despite this being a longstanding problem, it can remain a challenge if the implant is not known to the clinical team and the primary operative records are not available. The automated detection of orthopaedic implants on plain film radiographs using machine learning and artificial intelligence algorithms could aid clinicians in real-time decision making. This study aims to develop a Convolutional Neural Network (CNN) that is able to identify the make and model of total knee replacements (TKR).

Osteoarthritis is a common condition that results in joint pain, deformity and reduced function in affected individuals [3,4]. The knee is often involved and osteoarthritis is responsible for around 98% of total and unilateral knee replacements in the UK [5]. There are over 90,000 of these completed each year, making it one of the most common elective surgeries in the UK [5]. It tends to be a hugely successful procedure that can alleviate pain and increase function, with implants that can survive for decades [6–8].

As with all operations, they are not without complications – some of which may necessitate further surgery (revision surgery). These include prosthetic joint infections (PJI), periprosthetic fractures and aseptic loosening [9]. Revision surgery is typically more complex and lengthier than the primary operation and as such tends to be associated with increased morbidity and mortality [10,11]. Pre-operative identification of the prosthesis is key to reduce complications and improve chances of successful surgery. It allows for appropriate operative planning and for the correct equipment to be present [1,2].

If the patient re-presents to their original hospital, or their operative note is easily accessible the make and model tends to be easy to identify. At other times the implant may be clearly recognisable to the team on X-ray, typically if it is a model used locally. However, this is often not the case. Complications may occur decades after the primary procedure with the patient moving about the country and computer/record systems changing. Patients who had their primary operation in another country represent an even greater challenge with both data protection laws and language barriers playing a factor. This can necessitate

[*] Corresponding author. University College London, Institute of Health Informatics, UK.
E-mail address: h.kunz@ucl.ac.uk (H. Kunz).

https://doi.org/10.1016/j.imu.2021.100669
Received 18 April 2021; Received in revised form 23 June 2021; Accepted 13 July 2021
Available online 21 July 2021

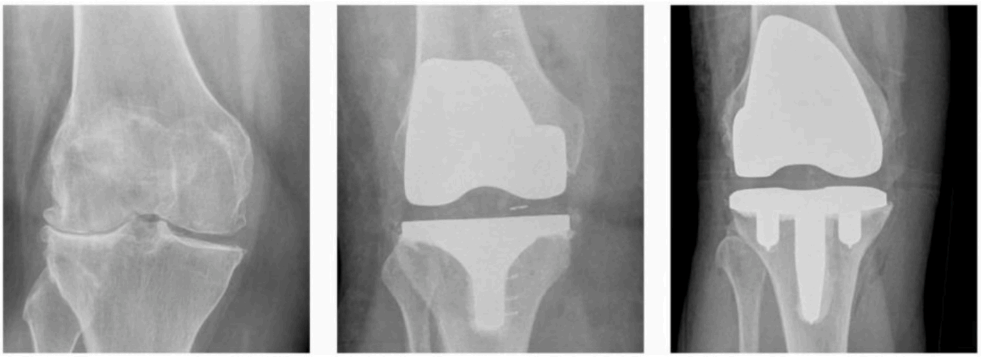


Fig. 1. Samples of primary dataset, anteroposterior (AP) film radiographs of the knee. A: No prosthesis, B; Columbus Knee System, C; Medial Rotation Knee.

prolonged searches for old records or to contact other hospitals. This is both a drain on clinicians already limited with time and can lead to delays in surgery [12]. This challenge is heightened by the vast number of possible implants used in both current and historical practice. Total knee replacements (TKR) have been performed since the 1960's and the majority of modern prosthesis will last longer than 20 years [13,14]. New implants are introduced at regular intervals – to improve patient's functional outcomes and the lifespan of the implant.

Despite all efforts it may be impossible to identify the implant due to lost or inaccessible records in up to 10% of cases [2]. This prevents complete pre-operative planning and knowledge of what operating sets may be required for revision cases. This has been shown to result in longer operations and greater intra-operative blood loss, both factors that are associated with worse short and long-term outcomes [2,15]. An automated solution, taking advantage of advanced computer vision techniques may help to provide a fast and efficient method of identifying orthopaedic implants.

CNNs have become one of the premier modes of computer vision since AlexNet dominated the 2012 ImageNet Large Scale Visual Recognition Challenge [16]. They utilise convolutional and pooling layers to extract features and improve computational efficiency before the information is passed to a fully-connected layer prior to classification [17]. CNNs are now used in all aspects of computer vision, from simple image classification to the core-component of self-driving cars.

There are valid concerns about using complex mathematical models to assist clinical decisions or make diagnoses. The first is the 'black box' problem. Though the foundational concepts and mathematical principles on which CNNs are developed are relatively easy to understand, networks are increasingly complex, consisting of hundreds of layers and filters. As such they can seem like 'black-box' classifiers where the decision making cannot be fully understood by humans [18,19]. This criticism can be somewhat countered with feature visualisation. To help combat this concern we will utilise a feature visualisation technique called saliency maps to identify what areas of the image are most important for a given prediction [20].

Another area of concern is how CNNs handle uncertainty and data that is outside the distribution of which it was trained. CNNs are trained to classify, as such a network to identify orthopaedic implants on X-ray will still classify a picture of a different object to one of its trained classes [21,22]. More of concern in the medical field would the assessment of an implant it had not been trained on, yet still giving a prediction of one of the trained classes (i.e. false positive). To the CNN, all images outside its training distribution are unknown unknowns. This is a limitation that has not been addressed in previous studies looking at automated detection of orthopaedic implants. We will look to apply soft-max thresholding to the CNN output, to offer a measure of confidence in a given prediction and allow the algorithm to reject images that do not appear to fall within one of the trained classes.

In this study we will look to train a CNN to accurately differentiate between a number of orthopaedic implants on plain film radiography. As the model is developed, training and validation loss and accuracy will be monitored and plotted for evidence of overfitting. The final model will be tested on a hold-out test dataset. A confusion matrix will be created and the accuracy, F1-score, and ROC-AUC calculated. Features important to predictions will be visualised with saliency maps and we will attempt to identify and exclude images that are outside the training distribution.

2. Related works

There has been a great deal of research into the application of CNNs to healthcare challenges. The applications are diverse from automated recognition of skin lesions to pacemaker implants [23,24].

More recently a number of research groups have turned their attention to the problem described in this paper. To date there are three publications that utilise CNNs to identify orthopaedic prostheses. Borjali et al. [25] focused on differentiating three types of total hip replacements (THR) using 252 AP (anteroposterior) x-rays and the DenseNet architecture with 100% accuracy. Kang et al. [26] utilised a simple CNN to differentiate 29 different THRs from just 171 AP x-rays. Data augmentation was used to increase their dataset to 3606 and achieved and AUC 0.99. However there are questions about their methodology which did not utilise a validation set and indicates their test set consisted of augmented images as opposed to a 'hold-out' test set [26]. Yi et al. were the first to look at the categorisation of TKRs, though they were limited to just two models, they used ResNet-152 and achieved a perfect AUC [27].

3. Methods

3.1. Dataset

The primary dataset consists of 558 anterior-posterior (AP) plain film radiographs of the knee. 158 images were gathered from a single NHS orthopaedic centre. The remaining 430 were taken from a dataset gathered in India. Each radiograph was anonymised and exported to JPEG format (RGB colour) directly from PACS. The dataset comprised of eight groups; seven models of TKR and one set of radiographs without TKR (no prosthesis), a sample of the X-rays can be seen in Fig. 1. The models of TKR were:

1) *The Columbus Knee System* by Aesculap (Tuttlingen, Germany)
2) *Medial Rotation Knee (MRK)* by MatOrtho (Surrey, England)
3) *Optetrak Logic Primary System* by Exactech (Florida, United States)
4) *Legion Total Knee System* by Smith and Nephew (Watford, England)
5) *Scorpio NRG* by Stryker (Michigan, United States)
6) *Legacy Posterior Stabilised (LPS) Knee solution* by Zimmer Biomet (Indiana, United States)
7) *Persona* by Zimmer Biomet (Indiana, United States)

Table 1
Summary of the knee X-ray dataset.

Model	Manufacturer	Total number of radiographs
No prosthesis	n/a	42
Columbus Knee System	Aesculap	59
Medial Rotation Knee	MatOrtho	57
Optetrak Logic Primary System	Exactech	154
Legion Total Knee System	Smith and Nephew	73
Scorpio NRG	Stryker	76
LPS Knee solution	Zimmer Biomet	69
Persona	Zimmer Biomet	58

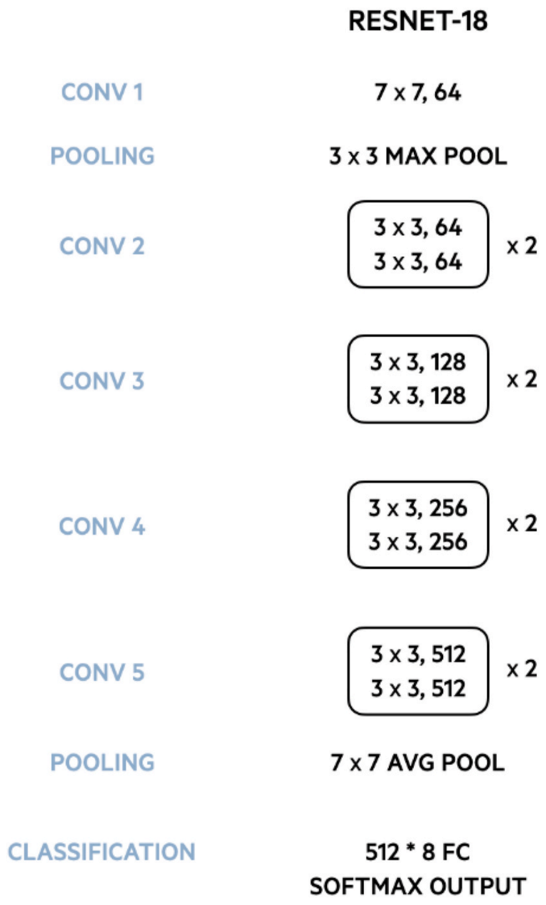


Fig. 2. Schematic diagram of ResNet-18.

A ground-truth of operative notes was used. Quality control was conducted by two clinicians: a senior orthopaedic surgeon experienced with TKR and the implants used and an additional clinician. A number of re-operative cases were found with revision stems or repaired periprosthetic fractures, these were removed from the dataset. Finally, on occasion 'sizing balls' for calibration were present, if they were within the image crop they were removed. A summary of the dataset is shown in Table 1.

3.2. *Technical specifications*

The machine learning library used was Pytorch, based upon the Torch library. Pytorch is a high-level python library developed by Facebook's AI Research lab. Released in 2016, it is open source and flexible, allowing CUDA-enabled GPU acceleration. This model was developed and trained using Pytorch version 1.5.1.

The model was trained on a MacBook Pro 2017 (Apple Inc. Cupertino, USA), 3.1 GHz Intel Dual Core i5, Intel Iris Plus Graphics 650 1.5 GB (Intel Corporation, Santa Clara, USA) and Google Colab (Google, California, USA) with NVIDIA Tesla K80 (NVIDIA Corporation, Santa Clara, USA) accessed via the described MacBook Pro.

3.3. *Convolutional neural network*

As CNNs became more complex with increasing depth, vanishing gradients become a significant issue [28]. Vanishing gradients refer to the loss of gradients during backpropagation through many layers. Developed in 2015, residual networks (ResNet) are specialised CNNs that utilise identity shortcut connections. These connections skip one or more layers helping to combat the vanishing gradients problem. This enables to develop increasingly deep networks that avoided earlier problems with sudden degradation of accuracy [29]. Identity short-cut connections add neither extra parameters nor computational complexity [29].

ResNet-18 is an 18-layer network that accepts 224 × 224 × 3 (i.e. 224 pixels by 224 pixels by 3 channels) input images. It begins with a simple convolution layer (7 × 7) followed by a max pooling layer. After this it has four convolutional blocks each composed of four 3 × 3 filters, with a residual connection between each set of two filters. It is then flattened into a fully connected layer with a soft-max activation function and 1000 output features.

For the purposes of this research, we update the output layer from 1000 features to the desired number of outcomes (eight). The decision was made to stick with a 224 × 224-pixel input layer – this offers a good balance between detail and computational efficiency. A schematic diagram of the ResNet-18 architecture is shown in Fig. 2.

Transfer learning is a powerful and increasingly used technique in machine learning and CNNs. Rather than initialising a CNN with random weights in each filter, we use the weights from the pre-trained network. This may reduce training time as certain common filter types (such as edge detection) may help in a myriad of tasks [30]. All layers of the network will be updated with respect to the loss function and gradient descent.

3.4. *Hyperparameters*

ResNet-18 has had a number of its hyperparameters optimised using the ImageNet dataset. The optimisers stochastic gradient descent (SGD)

Table 2
Summary of the training, validation and test datasets.

Model	Class	Total number of radiographs	Train (original + augmented)	Validation (all original)	Test (all original)
No prosthesis	0	42	21 + 21	10	11
Columbus Knee System	1	59	28 + 28	14	16
Medial Rotation Knee	2	57	28 + 28	14	15
Optetrak Logic Primary System	3	154	77 + 77	38	39
Legion Total Knee System	4	73	36 + 36	18	19
Scorpio NRG	5	76	38 + 38	19	19
LPS Knee solution	6	69	34 + 34	17	18
Persona	7	58	29 + 29	14	15

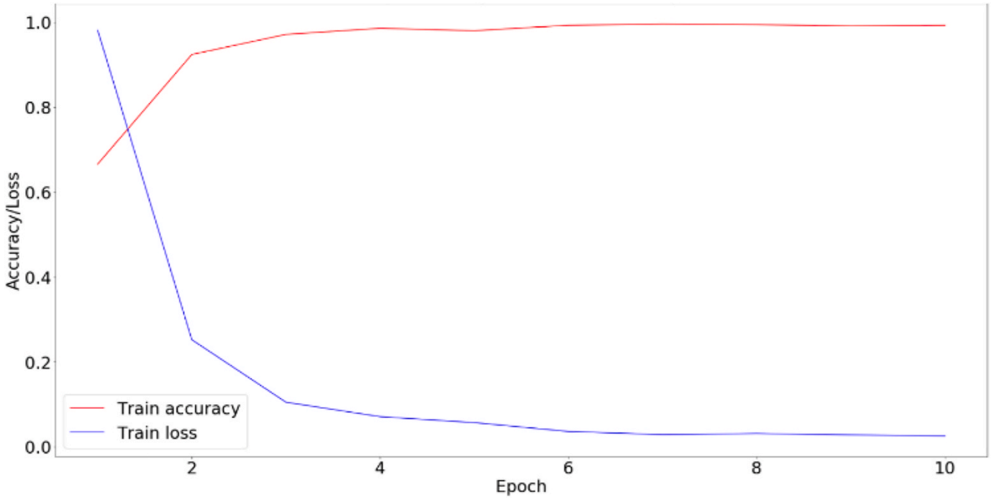


Fig. 3. Training accuracy and loss across epochs.

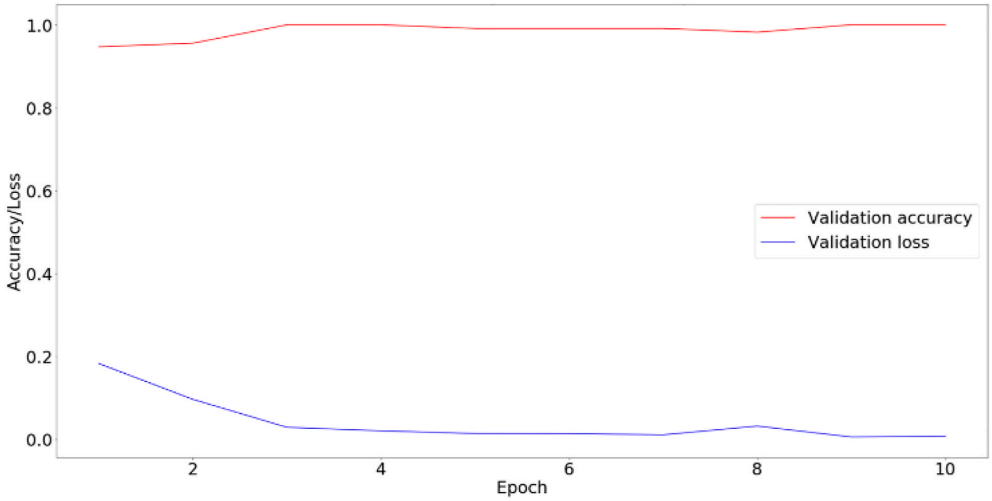


Fig. 4. Validation accuracy and loss across epochs.

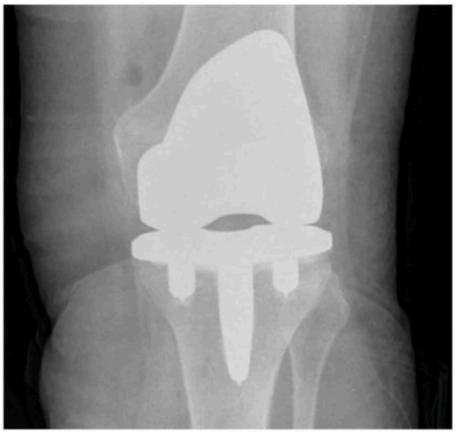


Fig. 5. An example test image with the algorithm prediction and softmax output.

and Adam will be compared. Two forms of SGD will be trialled, one with a fixed learning rate and one with step-wise learning rate decay. The stepwise function decays the learning rate from a starting point 0.1 by factor 'gamma' (=0.1) between each epoch.

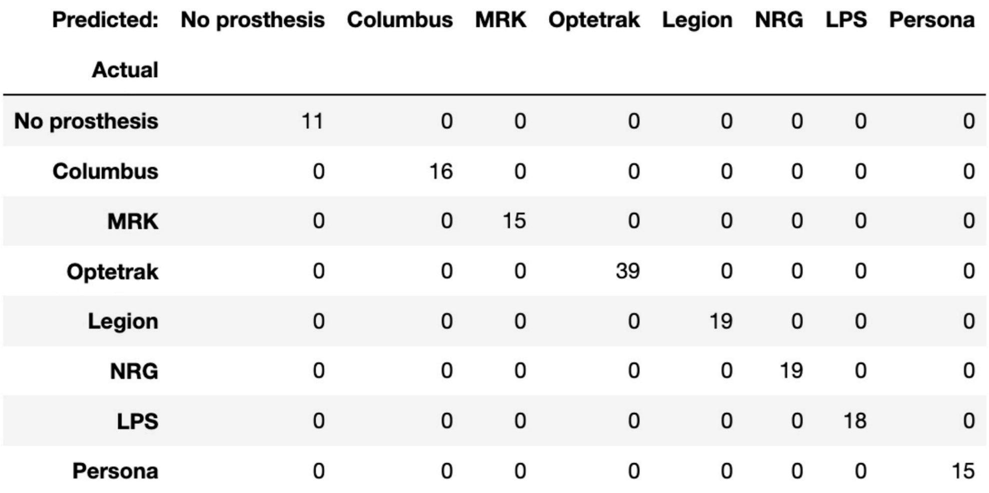
By contrast the Adam optimiser allows for efficient stochastic optimisation. Learning rates are adapted for individual parameters using estimates of first and second moments of gradients. It has been shown to perform favourably when compared to other optimisers [31].

Cross-entropy loss will be used by the CNN in training to measure the difference between the predicted outcomes and the ground truth labels. During training the CNN will aim to minimise the cross-entropy loss through backpropagation and updating of weights. Training accuracy will also be calculated to give a more interpretable idea of change in performance during training.

3.5. Image pre-processing

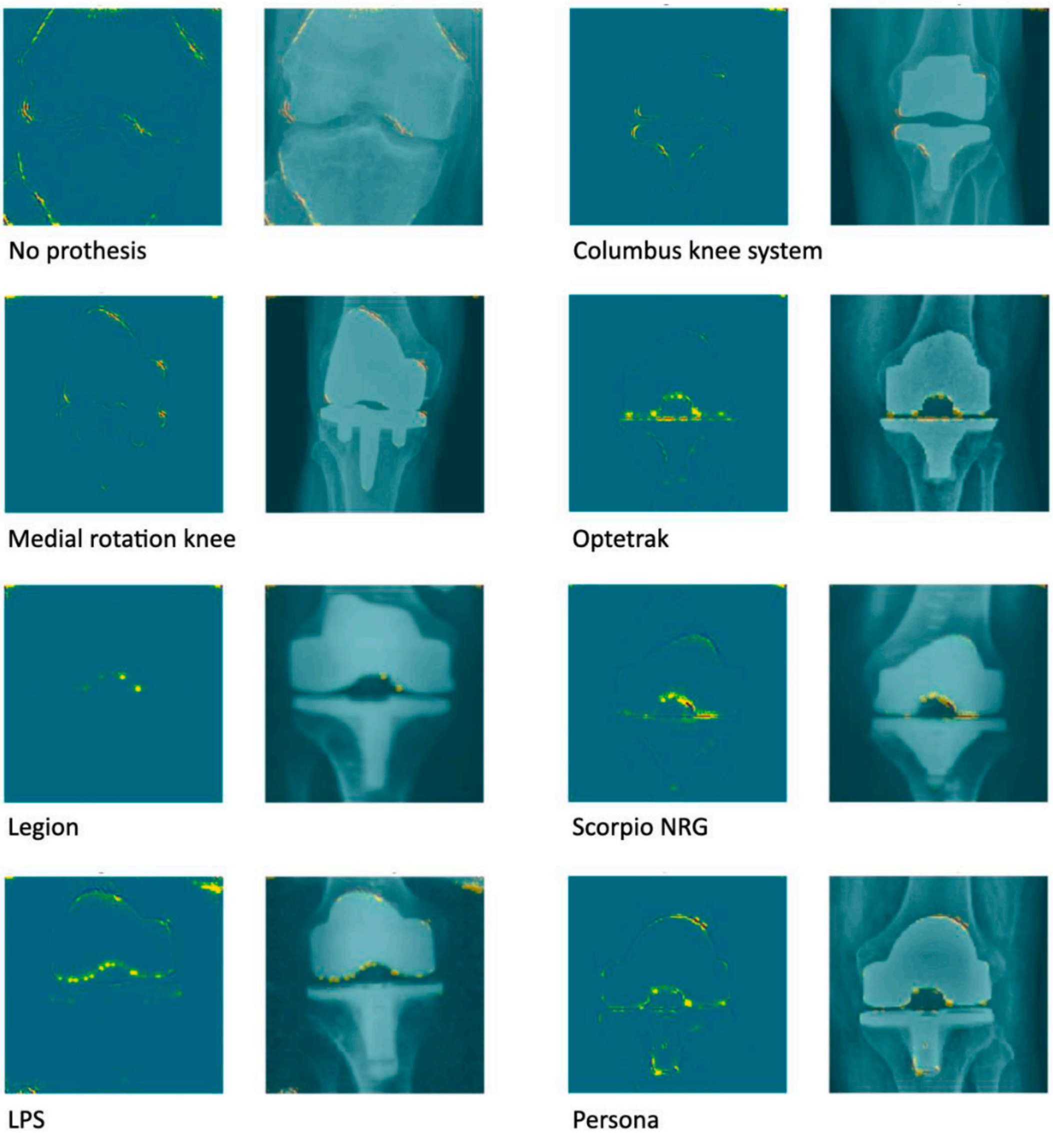
Manual segmentation was completed through cropping each image around the knee joint by a clinician, capturing the entire implant and joint. This removes parts of the image not relevant for implant classification reducing the computational load. This was relevant to achieve a dataset of a high quality for the training of the algorithms. The data cleansing of the dataset was informed by the expert opinions of two clinicians, where one clinician is a senior orthopaedic surgeon.

Each image was resized to 224 by 224 pixels. Each image was converted to PNG with an RGB colour profile. Images (that were not



Predicted: Actual	No prosthesis	Columbus	MRK	Optetrak	Legion	NRG	LPS	Persona
No prosthesis	11	0	0	0	0	0	0	0
Columbus	0	16	0	0	0	0	0	0
MRK	0	0	15	0	0	0	0	0
Optetrak	0	0	0	39	0	0	0	0
Legion	0	0	0	0	19	0	0	0
NRG	0	0	0	0	0	19	0	0
LPS	0	0	0	0	0	0	18	0
Persona	0	0	0	0	0	0	0	15

Fig. 6. Confusion matrix for test dataset results.



No prothesis

Columbus knee system

Medial rotation knee

Optetrak

Legion

Scorpio NRG

LPS

Persona

Fig. 7. Sample saliency maps for each class. Left shows maximum gradient, right shows maximum gradient overlay to the original image.

Table 3
Comparison of soft-max prediction output for different datasets.

Dataset	Range	Mean	Standard deviation
Test dataset	0.70–1.00	0.98	0.04
Out of distribution dataset	0.24–0.94	0.52	0.16
Maxx freedom dataset	0.21–0.82	0.48	0.17

Table 4
The effects of different soft-max prediction cut-offs on the rejection of images both in and out of the trained distribution.

Soft-max cut-off	Percentage of images in test dataset correctly classified	Percentage of test dataset rejected (not classified)	Percentage of out of distribution dataset incorrectly classified	Percentage of Maxx Freedom dataset incorrectly classified
0.5	100	0.00	48.53	33.33
0.75	98.41	1.69	0.05	13.33
0.90	94.44	5.56	0.01	0.00
0.95	88.89	11.11	0.00	0.00

already) were converted to grayscale retaining three channels for functionality with the pre-built ResNet architecture.

Each image was normalised prior to being provided as an input. The standard normalisation for a network built on ImageNet was updated to reflect the images used in this dataset. Mean and standard deviation for a representative subset of the dataset was calculated giving a mean of 0.191 and standard deviation of 0.253.

3.6. Data augmentation

Data augmentation was used to increase the dataset and to improve the generalisability of a model reducing the risk of overfitting to the training dataset. The total number of x-rays for each joint was relatively low, around 40–80 per class with the exception of Optetrak with 154.

First the dataset was randomly split into train, validation and test in a ratio of 50:25:25. Offline data augmentation was then applied only to the training dataset with one augmented image produced for each original image (Table 2). The following techniques of augmentation were applied:

I. **Random horizontal flipping:** Images flipped along horizontal axis, applied to 50% of augmented images.
II. **Gaussian blur:** Images blurred with gaussian kernels. The standard deviation of the gaussian kernel was randomly assigned between 0.0 (no blur) and 0.5 (mild blur).
III. **Gaussian noise:** Noise added from normal distribution to N(0, s) where s is sampled for each individual image and is between 0.0 and 0.2 × 255.

Online data augmentation was also utilised with random rotation between -5 and 5° applied to the training data. Online augmentation allows the images to appear differently between each batch further guarding against overfitting. No data augmentation was completed on the validation or test dataset.

3.7. Performance metrics

The evaluation of model performance is an essential step in understanding and developing a machine learning algorithm.

Accuracy is the most commonly used metric, giving the ratio of correct to incorrect predictions.

$$Accuracy = \frac{True\ positives + True\ negatives}{Positives + Negatives} \tag{1}$$

F1-score gives the harmonic mean of recall and precision (PPV: Positive predictive value, TPR: True positive rate).

$$F1\ Score = 2 \times \frac{PPV \times TPR}{PPV + TPR} \tag{2}$$

Receiver Operating Characteristic (ROC) curves will be plotted for dichotomisations of the outcome with Area Under the Curve (AUC) calculated.

3.8. Saliency mapping

One of the major criticisms of CNNs and neural networks in general is that they represent a ‘black box’ of decision making. It can be difficult or impossible to appreciate exactly how a complex, deep network of thousands of weights comes to classification decisions.

Saliency maps identify areas on an input image that the CNN is using to make its decisions, with the most important areas highlighted. It can offer a visual summary of the systems underlying logic – and importantly is instantly understandable to the human eye.

3.9. Out of distribution detection

A soft-max activation function was run on the output of the convolutional neural network. The soft-max function rescales the outputs by calculating the exponential of the inputs to that neuron, and dividing by the total sum of the inputs to all the neurons, so that the activations sum to 1 and all lie between 0 and 1. The activation function can be written as [32]:

$$y_K = g(h_K) = \frac{\exp(h_K)}{\sum_{k=1}^{N} \exp(h_K)} \tag{3}$$

The soft-max output is taken as a proxy for the probability or confidence of the prediction. The soft-max output of the test dataset was compared to two datasets composed of images that were outside of the training distribution. The first (named ‘out of distribution dataset’) is composed of 171 X-rays of other parts of the body (e.g. chest, ankle, pelvis) whilst the second (named ‘Maxx Freedom dataset’) is composed of 15 AP x-rays of knees with Maxx Freedom implants. The aim was to develop a soft-max cut-off that allows the rejection of images outside of the training distribution.

Table 5
Summary of papers that utilise CNNs for the automated detection of orthopaedic implants.

Paper	Arthroplasty	implants	number of radiographs	CNN	Results
Kang et al, 2020	Hip	29	170 (data augmentation used to increase to 3606)	Simple CNN (2 conv layers, 1 max pool, 2 FC layers)	AUC: 0.99
Yi et al, 2019	Knee	2	374 (data augmentation used to increase to 3080)	ResNet-152	AUC: 1.0
Borjali et al, 2020	Hip	3	252	DenseNet	Accuracy: 100%
This paper	Knee	7 (8 classes including ‘no prosthesis’)	588 (data augmentation used to increase to 936)	ResNet-18	Accuracy: 100% F1 score: 1.0 AUC: 1.0

4. Results

Validation results were used to help select the optimiser. Though Adam and stepwise SGD performed reasonably, simple SGD with a fixed learning rate of 0.001 and momentum of 0.9 offered the most stable performance with rapidly falling loss and improved accuracy across epochs. As such for the network SGD with a learning rate 0.001 and momentum 0.9 with no decay was used.

Learning curves, plotting loss and accuracy for both the training (Fig. 3) and validation dataset (Fig. 4) across epochs were plotted to aid in the assessment of model training and performance. Across 10 epochs, loss is seen to fall rapidly in both the training and validation dataset. The curves demonstrated no evidence of overfitting during the epochs the model was trained. For both the training and validation set accuracy rapidly approached 100% and remained steady at this level. The final model selected was that with the best accuracy on the validation set.

4.1. Test results

The selected final trained ResNet-18 model was evaluated on the hold-out test dataset. This dataset was not seen by the model at any time during training. A total of 126 images composed the test dataset from eight classes. An example of a correctly classified test image and the algorithm output is shown in Fig. 5.

The model showed 100% accuracy in classifying all models of prosthesis and no prosthesis as demonstrated in the confusion matrix (Fig. 6). The model achieved an F1-score of 1.0. To calculate the AUC-score the classification problem was dichotomised using a one vs. all method. An AUC score of 1.0 was achieved for each dichotomisation.

4.2. Saliency maps

Saliency maps were obtained to identify which areas of the image were important to its classification (Fig. 7). They demonstrate primarily that the outline of the implants was important to classification, in most cases the maximal weights show an outline of the implant. For images with no prosthesis it is the joint line and cortical margins of the bone that contribute to classification.

4.3. Out of object detection

There is no official framework to define a representative dataset of images that are out of the distribution for which the CNN was trained. By definition it includes any and all images that would not be correctly classified as an AP X-ray for one of the trained classes. To provide the most rigorous challenge, images that were similar to the training dataset were selected. Two datasets were created to test out of distribution detection, one consisted of 171 X-rays of other parts of the body (chest X-rays, pelvic X-rays, hip X-rays and ankle X-rays) and lateral views of the knee that were declared the distribution dataset. The second dataset consists of 15 AP X-rays of the Maxx Freedom knee implant dataset; an implant not included in training of the model.

The CNN output was converted from its output to a soft-max ‘probability’. The class with the highest probability is given by the CNN as its prediction. We recorded the soft-max output for each image in the test dataset as well as the out of distribution dataset and Maxx Freedom dataset. The range, mean and standard deviation for each is shown in Table 3.

Images in the test dataset were correctly classified. By definition all results for the other datasets are false positives. It can clearly be seen that images within the trained distribution (i.e. the test dataset) have a significantly higher mean softmax output. However, on the lower end of the range are values that cross-over with the ranges of the other datasets.

Table 4 shows the effect of different cut-offs on the accuracy of the test dataset and the incorrect classification of images not within the trained distribution of the CNN. A cut-off of 0.95 correctly rejects all images that were not one of the classes the CNN was trained to identify. A recent review of a number of out-of-distribution detection techniques found setting a soft-max threshold one of the more effective techniques [33] (see Table 5).

It is of the utmost importance to avoid falsely classifying an implant that it has not been trained to identify. As such a relatively strict cut-off must be chosen. The testing found a cut-off of 0.95 produced reasonable results with the rejection of a range of images outside the training distributions from other types of X-rays (e.g. chest X-ray) to AP X-rays of the knee for an arthroplasty the CNN was not trained on (Maxx Freedom.

This study shows that the chosen CNN-architecture achieved a superior performance for the selected orthopaedic implants. One reason could be that the geometrical characteristics of the implants has a typical geometrical shape and can easily be identified by a CNN. Future work could improve and build on this study in a number of ways. Manual segmentation could be replaced by automated segmentation techniques like YOLOv3 [26]. The dataset could be improved in terms of size and the inclusion of all implants used across the NHS and possibly from other national healthcare systems. This study proves the feasibility and the possibility of the extension in terms of scale and scope.

5. Conclusion

This study has presented the training and testing of a CNN based on the ResNet-18 architecture that can accurately and automatically differentiate between seven classes of implants for total knee replacements and no prosthesis. Two clinicians including one senior orthopaedic consultant guided the data cleansing and data pre-processing. This was a safeguard so ensure that a clinically trustworthy dataset was used for the training of the algorithm. The next step is to develop and deploy an algorithm, based on similar principles but extended in scale and scope. The deployment of an extended system could save clinicians time and reduces occasions where lack of implant identification leads to operative delays, pro-longed revision surgery and increased blood loss.

Ethical statement

Ethical approval of the UK-data for this study has been approved by the NHS (see IRAS project ID: 264110/1400, identification of make and model of orthopaedic implant using artificial intelligence, sponsor: Luton & Dunstable University Hospital NHS Foundation). Ethical approval for the Indian dataset has been approved by the Institute Ethics Committee All India, Institute of Medical Sciences (IEC-496/August 02, 2019, RP-32/2019).

Declaration of competing interest

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

Acknowledgment

The authors want to thank the Institute of Health Informatics at University College London and Luton & Dunstable University Hospital NHS Foundation.

References

[1] Baré J, MacDonald SJ, Bourne RB. Preoperative evaluations in revision total knee arthroplasty,. Clin Orthop Relat Res 2006;446:40–4. https://doi.org/10.1097/01.blo.0000218727.14097.d5.

[2] Wilson NA, Jehn M, York S, Davis CM. Revision total hip and knee arthroplasty implant identification: implications for use of unique device identification 2012 AAHKS member survey results. J Arthroplasty 2014;29(2):251–5. https://doi.org/10.1016/j.arth.2013.06.027.

[3] Glyn-Jones S, et al. Osteoarthritis. Lancet 2015;386(9991):376–87. https://doi.org/10.1016/S0140-6736(14)60802-3.

[4] Vina ER, Kwoh CK. Epidemiology of osteoarthritis: literature update. Curr Opin Rheumatol 2018;30(2):160–7. https://doi.org/10.1097/BOR.0000000000000479.

[5] NJR. "17th annual report 2020: national joint registry for England, Wales, Northern Ireland and the Isle of Man," *NJR 17th.* Annu Rep 2019;2019.

[6] Price AJ, et al. Knee replacement. Lancet 2018;392(10158):1672–82. https://doi.org/10.1016/S0140-6736(18)32344-4.

[7] Ferket BS, Feldman Z, Zhou J, Oei EH, Bierma-Zeinstra SMA, Mazumdar M. Impact of total knee replacement practice: cost effectiveness analysis of data from the Osteoarthritis Initiative. BMJ 2017;356. https://doi.org/10.1136/bmj.j1131.

[8] Kane RL, Saleh KJ, Wilt TJ, Bershadsky B. The functional outcomes of total knee arthroplasty. J. Bone Jt. Surg. - Ser. A 2005;87(8):1719–24. https://doi.org/10.2106/JBJS.D.02714.

[9] Cheung A, Goh SK, Tang A, Keng TB. Complications of total knee arthroplasty. Curr Orthop 2008. https://doi.org/10.1016/j.cuor.2008.07.003.

[10] Stirling P, Middleton S, Brenkal I, Walmsley P. Revision total knee arthroplasty versus primary total knee arthroplasty. Bone Jt. Open 2020;1(3):29–34.

[11] Meneghini RM, Vince KG, Waddell BS, Westrich G. Revision total knee arthroplasty. in Orthopaedic Knowledge Update: Hip and Knee Reconstruction 2018;5.

[12] Wilson N, Broatch J, Jehn M, Davis C. National projections of time, cost and failure in implantable device identification: consideration of unique device identification use. Healthcare 2015;3(4):196–201. https://doi.org/10.1016/j.hjdsi.2015.04.003.

[13] Evans JT, Walker RW, Evans JP, Blom AW, Sayers A, Whitehouse MR. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. Lancet 2019; 393(10172):655–63. https://doi.org/10.1016/s0140-6736(18)32531-5.

[14] Yong TM, Young EC, Molloy IB, Fisher BM, Keeney BJ, Moschetti WE. Long-term implant survivorship and modes of failure in simultaneous concurrent Bilateral total knee arthroplasty. J Arthroplasty 2020;35(1):139–44. https://doi.org/10.1016/j.arth.2019.08.011.

[15] Ross D, Erkocak O, Rasouli MR, Parvizi J. Operative time directly correlates with blood loss and need for blood transfusion in total joint arthroplasty. Arch. Bone Jt. Surg. 2019;7(3):229–34. https://doi.org/10.22038/abjs.2019.28534.1736.

[16] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90. https://doi.org/10.1145/3065386.

[17] Sakib S, Ahmed, Jawad A, Kabir J, Ahmed H. An overview of convolutional neural network: its architecture and applications. *ResearchGate*; 2018 [Online]. Available, https://www.researchgate.net/publication/329220700.

[18] Nicholson Price W. Big data and black-box medical algorithms. Sci Transl Med 2018;10:471. https://doi.org/10.1126/scitranslmed.aao5333.

[19] Rai A. Explainable AI: from black box to glass box. J Acad Market Sci 2020;48(1): 137–41. https://doi.org/10.1007/s11747-019-00710-5.

[20] Zintgraf LM, Cohen TS, Adel T, Welling M. Visualizing deep neural network decisions: prediction difference analysis. 2017.

[21] Hendrycks D, Gimpel K. A Baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv 2016 [Online]. Available, http://arxiv.org/abs/1610.02136.

[22] Karimi D, Gholipour A. Improving calibration and out-of-distribution detection in medical image segmentation with convolutional neural networks. arXiv, [Online]. Available, http://arxiv.org/abs/2004.06569; 2020.

[23] Esteva A, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115–8. https://doi.org/10.1038/nature21056.

[24] Howard JP, et al. Cardiac rhythm device identification using neural networks. JACC Clin. Electrophysiol. 2019;5(5):576–86. https://doi.org/10.1016/j.jacep.2019.02.003.

[25] Borjali A, Chen AF, Muratoglu OK, Morid MA, Varadarajan KM. Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. J Orthop Res 2020;38(7):1465–71. https://doi.org/10.1002/jor.24617.

[26] Kang YJ, Il Yoo J, Cha YH, Park CH, Kim JT. "Machine learning–based identification of hip arthroplasty designs. J. Orthop. Transl. 2020;21:13–7. https://doi.org/10.1016/j.jot.2019.11.004.

[27] Yi PH, et al. Automated detection {\&} classification of knee arthroplasty using deep learning. Knee 2020;27(2):535–42. https://doi.org/10.1016/j.knee.2019.11.020.

[28] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. 2013.

[29] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016. https://doi.org/10.1109/CVPR.2016.90.

[30] Li X, Pang T, Xiong B, Liu W, Liang P, Wang T. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. 2018. https://doi.org/10.1109/CISP-BMEI.2017.8301998.

[31] Kingma Diederik, Ba Jimmy Lei. A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. ICLR; 2015. https://arxiv.org/pdf/1412.6980.pdf.

[32] Marsland Stephen. Machine Learning An Algorithmic Perspective. CRC Press; 2015. p. 81.

[33] Shafaei A, Schmidt M, Little JJ. A less biased evaluation of out-of-distribution sample detectors. 2020.