Shoulder Implant Identification and Categorization with A Fully Implemented Convolution Neural Network

B. RAJENDRA PRASAD[1], PROF. M. JAMES STEPHEN[2], PROF. P.V.G.D. PRASAD REDDY[3]
[1] *Research Scholar, Andhra University, Visakhapatnam, Andhra Pradesh, India*
[2] *Chair Professor, Dr. B. R. Ambedkar Chair Cluster Andhra University, Visakhapatnam, Andhra Pradesh, India*
[3] *Professor, Andhra University, Visakhapatnam, Andhra Pradesh, India*

Abstract*— Shoulder implants usually need to be changed after a specific time has passed. However, establishing the implant maker or model during this transition may be a complex and error-prone procedure for medical practitioners. This research aims to determine which of four separate implant manufacturers produced each of the 597 X-ray photos of shoulder implants. In order to accomplish this goal, both pre-trained ESA architectures (DenseNet201, DenseNet169, InceptionV3, NasNetLarge, VGG16, VGG19, and Resnet50) and cascading models fed by the YOLOv3 detection algorithm were developed, and the classification performances of these models were compared with one another. The job of the YOLOv3 detection algorithm in the stepped models is to identify the head area of the shoulder implants and provide this region as an input to the ESA designs. This work is performed within the context of the stepped models. In addition, conventional machine learning techniques were integrated with the ensemble learning approach. This integration was analyzed using Fully Implemented Convolution Neural Network (FCNN model) to see how well they performed on the data set. With an accuracy of 84.76 percent, the stepped DenseNet201 model achieved the best classification performance. This rate is greater than the one found in another research that used a comparable data set. The categorization accuracy provided by ensemble models is noticeably lower than that provided by ESA models. Additionally, the classification accuracy achieved by YOLO-assisted cascade models is superior to that achieved by individual ESA models. That is to say, concentrating on the head area of the implant while utilizing the YOLOV3 detection algorithm helped boost the accuracy of the categorization. This methodology will motivate more research into this subject area.

Indexed Terms*— YOLO, Shoulder implant, object detection, deep learning, CNN

I. INTRODUCTION

In the Total Shoulder Arthroplasty, also known as TSA, the ball and socket joint of the shoulder is replaced with an implant as part of the therapy for injured shoulder joints [1-3]. This technique is performed routinely in the field of orthopedics. Shoulder injuries of a severe nature and significant joint inflammation are often the driving forces for patients seeking TSA surgery. The patient has minor discomfort due to the treatment, and the range of motion in their shoulder is improved. Several companies specialize in the production of implants [2].

Additionally, the implant may need to be updated a few months or even years after being implanted. In this particular scenario, it is essential to identify the manufacturer or model of the implant. Due to the lack of clarity in the patient's medical record, patients and their physicians could be in the dark. At this point, the work of identifying an implant maker or model in such circumstances depends on the meticulous analysis and visual comparison of X-ray pictures of the implant performed by medical specialists. Even though X-ray scans display the essential characteristics of the implants rather clearly, the identification process is notoriously time-consuming, prone to mistakes, and challenging for orthopedic surgeons and radiologists alike. It is a tedious process that must be completed carefully and thoroughly for each new patient [2, 4-6]. Suppose the implant design is not determined before

surgery. In that case, this may result in an increased operating time, more complex operations, more blood loss, increased bone loss, prolonged healing time, and an overall rise in the cost of healthcare [6]. In addition to this, it has been reported that ten percent of the implants cannot be recognised before the procedure, and that two percent of the implants cannot be found when they are being placed [7, 8]. X-ray pictures may be processed using Convolutional Neural Network (ESA) architectures that are trained with deep learning to handle this particular issue.

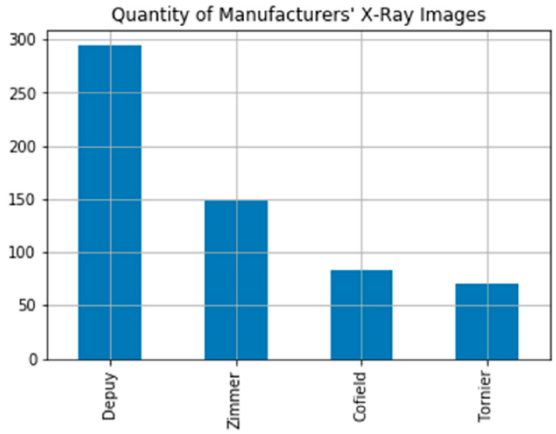

Figure 1: Typical counts of Shoulder implant manufacturer graph

Data is the starting point for anything in deep learning (DL). Biomedical data may be huge but also exceedingly diverse. The creation of algorithms that are able to handle several forms of data, in particular variable-dimensional structured data [9], is essential to the application of deep learning approaches to the analysis of data pertaining to biomedical research. Recent years have seen the effective use of DL models in a variety of domains [10–12], such as the categorization of medical images [13–15], and they have started to take the role of machine learning approaches [16]. However, determining the presence of shoulder implants from X-ray pictures is not a topic that has been thoroughly researched or addressed [2].

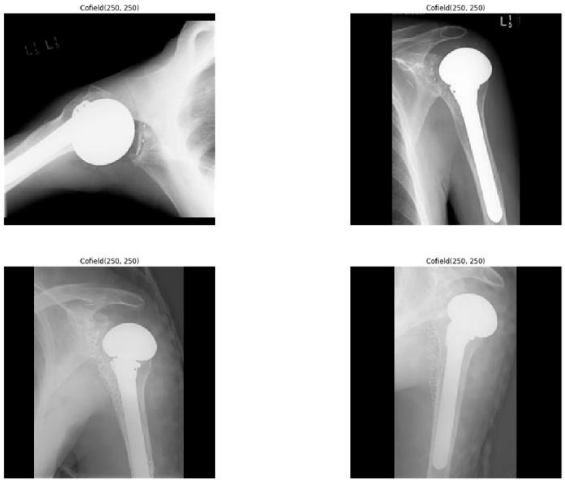

Figure 2: Sample X-Ray images of Shoulder Implant

There aren't many research that uses machine learning techniques to categorise the implant model or manufacturer. There are two studies in particular that categorise the shoulder implant. Researchers Urban et al. [2] identified four distinct implant manufacturers by analysing 597 X-ray pictures of shoulder implants. During the process of categorization, both conventional approaches to machine learning and a variety of various pre-trained ESA architectures were used, including VGG16, VGG19, ResNet-50, ResNet-152, NasNet, and DenseNet-201. Using a 10-fold cross-validation strategy, the NasNet ESA architecture was able to reach the greatest possible categorization accuracy. Yi et al. [4] analysed 482 X-ray pictures of the shoulder and categorised 5 distinct shoulder implant types. In order to complete the categorization procedure, they used the ResNet pre-trained ESA model. The most significant shortcoming of this research is that, rather than categorising all five distinct implant types using a single ESA model, the researchers developed a unique ESA model for each implant and then made dual classifications of the data. There is additional research that differentiates shoulder implants from hip and knee implants using deep learning. These studies focus on hip and knee implants. On a total of 274 X-ray pictures, Yi et al. [8] categorised two distinct knee implant types utilising the ResNet ESA framework. Ghose et al. [17] used 8 distinct pre-trained ESA topologies to classify 6 different orthopaedic knee implant types based on 878 X-ray pictures. Because to the MobileNetV2 design, they were able to achieve the best classification accuracy. Borja et al. [6] identified 3 distinct complete

hip replacement implants. On a total of 252 X-ray pictures of hip prostheses, they classified the data using a DenseNet-201 system that had been pre-trained.

This study's objective is to categorise X-ray images of shoulder implants according to the manufacturer using a combination of the YoloV3 (You Only Look Once) object detection algorithm and the pre-trained state-of-the-art ESA architectures (DenseNet201, DenseNet169, InceptionV3, NasNetLarge, VGG16, VGG19, Resnet50). This will be accomplished by cascading models. to evaluate the effectiveness of various topologies in terms of categorization. The following is a list of the primary contributions that the study makes:

- It is suggested to categorise the shoulder implant using a novel stepwise deep learning model that was not used in any of the prior investigations.
- A greater classification accuracy was found in the research [2] that used a data set comparable to what was found in the literature.
- Using the YOLOV3 detection algorithm with a focus on the head area of the implant has been shown to boost classification accuracy. This was proved via a series of experiments.
- Seven distinct ESA models are used, and the output of the YoloV3 method as well as the raw X-ray pictures are used to analyse the data, after which the findings are analysed and compared. In addition, conventional techniques of machine learning are merged with the method of ensemble learning, and the combined approaches' respective outcomes are analysed.

II. PRELIMINARIES

The detection of the head region of the implants using YOLOv3, the creation of stepped models by combining the YOLOv3 model with pre-trained ESA models, the training of the models and their parameters, the shoulder implant X-ray image dataset that was used, and the performance metrics that were used in the evaluation of the models are all explained in this section.

The shoulder implant X-ray that was used by Urban et al. [2] was the dataset that was used for this investigation gained simply looking at the pictures. Some of the photos included in this dataset were collected from the University of Washington's shoulder website, while others came from individual surgeons and manufacturers. This dataset was compiled by Urban et al. [2]. The dataset includes 597 shoulder implant X-ray pictures from a variety of patients, each of which was taken using one of sixteen distinct models produced by four separate manufacturers. These photos include 83 implants manufactured by Cofield, 294 implants manufactured by Depuy, 71 implants manufactured by Tornier, and 149 implants manufactured by the firm Zimmer. Figure 1 displays a selection of representative X-ray pictures taken from the dataset. In addition, the dataset includes a few drawbacks, including the following: (i) The image resolution of X-rays is modest and might vary from picture to image. A good many of them have a resolution that is exactly 250 pixels on each side. (ii) Images may have a wide variety of aspect ratios. (iii) According to the manufacturer, the visual contrast is poor, and the classes are not spread equally. The performance of the models in terms of categorization is significantly impacted as a result of all of these drawbacks.

III. PROPOSED MODELS

There have been two alternative approaches suggested for classifying the manufacturer of the shoulder implant based on the X-ray pictures. Figure 2 illustrates these many approaches to the problem. In the first technique, pre-trained ESA models such as DenseNet201, DenseNet-169, Inception V3, NasnetLarge, VGG-16, VGG-19, and Resnet50 were cascaded using the YOLOv3 detection algorithm. These models were used to generate cascading models. First, the YOLOv3 recognition algorithm was trained to recognise the head area of the implants using pre-labeled X-ray pictures. This was the first phase in the process. The head area of the implants was located by feeding X-ray pictures into a trained version of the YOLO algorithm in the second phase of the process. In the third step, the identified head area was clipped and sent as an input to the pre-trained ESA models. This allowed for the processes of training and testing to take place. The pre-trained ESA architectures that were utilised in the first technique were trained and evaluated using the raw X-ray pictures that were

included in the dataset as part of the second procedure. In addition, the performance of conventional machine learning algorithms on the data set was evaluated, specifically with regard to classification accuracy. In order to accomplish this goal, the output data of the "avg pool" pooling layer of the DenseNet201 architecture (the layer that comes before the output layer) were retrieved. In order to carry out the classification procedure, the features that were obtained from this layer are provided as input to the ensemble classifier. This ensemble classifier is comprised of the Random Forest (RF), Extreme Gradient Boosting, K-Nearest Neighbor, and Multi-Layer Detector methods, and it uses the voting method to make its decisions. The performance of each approach in terms of categorization is compared head-to-head in the part devoted to the study's findings and debates. On the implant data set, the classification performances of several ESA models, cascade models, and standard machine learning models have thus been proven.

3.1. PROPOSED FULLY CONVOLUTIONAL NEURAL NETWORK (FCNN MODEL)

The YOLO and ESA models were developed with the help of the Python programming language and the Tensorflow and Keras Library. The process of training ESA-based models without the use of a graphics processing unit (IOP) is one that is exceedingly challenging and time-consuming. Because of this, the training and testing of the models were executed on Google Colab, a product of Google Research, making use of T4 and P100 IOPs respectively. In the following paragraphs, a detailed presentation of the training of YOLO and ESA models as well as the model parameters employed in this process will be given.

Training of the YOLOv3 detection algorithm and automated trimming of the implant head region are included in section 2.3.1. (Training of YOLOv3 detection algorithm and automatic cropping of implant head area)YOLO is a method for detecting objects that is based on ESA. This algorithm reframes object detection as a single regression problem from image pixels to bounding box coordinates and class probabilities [23]. In this particular investigation, the YOLOv3 version was used to the task of identifying the implant's head location. YOLOv3 detects objects far more quickly than any other approach currently available. The DarkNet-53 ESA architecture is used in order to extract the features. DarkNet-53 is composed of 53 layers and employs convolutional layers with sizes of 3x3 and 1x1. [24]

Because normalisation was performed, the information about the height (H) and width (W) of the identified implant head area in the picture are equivalent. This is because the centre point coordinates (X, Y) were made equivalent. 1 and Eq. Obtainable in accordance with [25] 2.

$$F_{S_k} = \sum_{i=1,j=1}^{p} x_{i,j} * w_{i,j} \quad (1)$$

It is necessary to train the YOLOv3 model so that it can identify the implant head area. In order to properly complete this training technique, the head region on the X-ray pictures of the implant has to be tagged.

A tool for annotating was used. Using this programme, a text file was produced for each individual picture. This file contains the height (iY), width (iG), and position information of the implant head area (X0, Y0, X1, Y1) coordinates of the tagged X-ray pictures. The height information is measured in inches, while the width information is measured in pixels. Additionally, the developed dataset is compatible with the design of YOLOv3, which was another goal of this project.

Operations are carried out using bounding boxes that are greater than a predetermined IoU threshold while the mAP value is being calculated. Using 0.50 as the cutoff value, the mAP value for the test data was determined to be 99.99 percent, while the average IoU value was found to be 81.40 percent. In light of these performance figures, it is possible to assert that the head area of the implants is detected with a level of precision that is above average. As a result, the weights that are determined to be appropriate for these performance criteria are documented. After that, the trained YOLOv3 algorithm was provided with input from all of the X-ray pictures included in the dataset, and it was able to automatically locate the implant head area. OpenCV was used to do some trimming on the identified implant head area before it was presented as part of an introduction to ESA architectures and stepwise models were developed. The X-ray images shown in Figure 4 have had the

implanted head area recognised and have had the head clipped.

The initial step in the training of ESA models included scaling down X-ray pictures to a resolution of 224 by 224 by 3. Following that, each piece of input data was multiplied by 255 and then subjected to normalisation. In addition to that, the output class labels get one hot encoding. During the training as well as the testing phase, the 10-fold cross validation approach was used. Table 1 displays some of the training parameters utilised for cascade models as well as those used for individually employed pre-trained ESA models. This mAP is the outcome of values such as accuracy and sensitivity as well as the F1-score and the IoUoffers just one viewpoint or perspective. Spouse. The calculation looks as this: 4 [25].

The optimization method is considered to be the most critical parameter. This method provides a description of the learning process as well as an explanation of how millions or even billions of parameters may be updated [27]. When training models, it might be challenging to decide which optimization methods to use. As a result of this, the Adam, Adadelta, Sgd, Rmsprop, Adamax, and Nadam optimization algorithms were tested, and the table below presents the results of those tests, along with the algorithms that provided the best classification performance according to each ESA model. The learning rate is yet another significant characteristic that affects the overall performance of the learning process. In the layers that are not the layer, the activation function is represented by a rectified linear unit, abbreviated as RELU. The activation functions of Relu and Softmax in respective order. It is defined in the same manner as in 5 and 6. Spouse. xj in 5; refers to the jthoutput in the most recent layer, (the pace of learning) If the learning rate parameter is given big values, there is a possibility that the estimator would fluctuate, which is another way of saying that it will wander between feasible local solutions. When tiny values are picked, the amount of time needed to learn rises. Different learning rate values were experimented with when ESA models were being trained in order to get the highest possible classification performance. The Core Components The convolution and pooling layers are connected to the trainable parameter in some way. If this option is set to "False," the layers in question will be preserved in their current state, training will not be carried out, and the "imagenet" weights will be applied without modification. This parameter value is set to "True" in the DenseNet201 model, "True" in the DenseNet169 model, and "False" in the Inception V-3 ESA model; however, in other models, it is assigned to "False." In addition, the categorical cross entropy was used as a loss function in each and every model. Additionally, the training procedure was finished at 70 epochs on the DenseNet-169 and Inception V-3 ESA models, but it was finished at 50 epochs on the other models.

The activation function is another key parameter that is considered during the training of models. The output layer is the very last layer in ESA models, and the activation function used there is called Softmax. This layer has a fixed number of four neurons in its composition. Figure 3 depicts the training loss and accuracy curves for the cat-2 step of the stepped DenseNet201 model, which had the greatest classification performance. Calculations were also made for curves that belonged to other models; however, these calculations were not included in the research in order to make the study easier to read. As can be seen in Figure 4, the training process shown by the model was both rapid and consistent. The subduction value was very near to zero at a level that was close to 10 epochs, then it peaked a bit, and then it continued to be relatively close to zero at a level that was close to 20 epochs. The accuracy curve displays a scenario quite similar to this one as well.

3.1 METRICS FOR MEASURING PERFORMANCE

Metrics like precision (K), sensitivity (D), accuracy (F1), and F1-score (F1) were used in the assessment of the models. Following are the mathematical formulae that correspond to each of these measures. In these equations, "DP" stands for "True Positive," "DN" stands for "True Negative," "FP" stands for "False Positive," and "YN" stands for "False Negative."

- Precision: This parameter determines the ratio of actual positive results to the total number of anticipated positive results. As a result, it is contingent upon the values of DP and FX. Spouse. The calculation looks as this:

  $$K = DP / (DP + FP)$$

- Sensitivity: The sensitivity of a model may be defined as the fraction of true positives that are

properly categorised by it. Sensitivity Co. for all samples that tested positive. The calculation for it is as in 8.

D = DP⁄(DP + YN)

- F1-Score: The F1 score is a measurement that combines the model's precision and sensitivity characteristics in order to determine the overall accuracy value of the whole model. Comparable to the harmonic mean of the sensitivity values and the sensitivity values. The calculation looks as this: 9 [25].

$$Avg = \frac{\sum_{k=1}^{N} F_{S_k}}{N} \quad (2)$$

- Accuracy: Equi. The calculation looks as this:

$$Std = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(F_{S_k} - Avg\right)^2}, S_{F_{S_k}} = \frac{F_{S_k} - Avg}{Std} \quad (3)$$

Accuracy may be calculated as (DP+DN)/(DP+DN+YP+YN) (10) Densely Connected Neural Networks may be found in section 2.4. (Densely Connected Neural Network). Because the highest classification performance is achieved in the DenseNet201 model, which is a densely connected neural network, it is crucial to discuss the architecture of this model within the context of the research because it is one of the models being investigated. Huang et al. have built neural networks that feature dense connections to one another. It was produced by member number 18. Figure 3 provides an overview of the network's general design. To guarantee that the greatest amount of information is sent through from one layer to the next in the network, all of the layers have identical feature map sizes and are directly linked to one another in a feed-forward manner. In this manner, each layer was given extra inputs from all of the layers that came before it, and it also transmitted its own feature maps to all of the layers that came after it. That would be Eq. It is stated in the 11th paragraph. In this case, [X0, X1,..., Xl–1]; 0,... refers to the merging of feature maps that were formed in l-1 different layers.

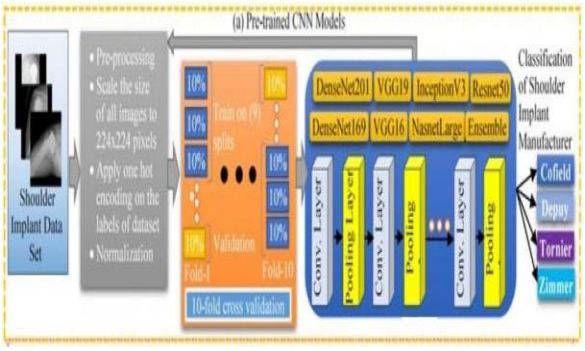

Figure 3: Proposed model architecture

The resulting function that is obtained by merging these feature maps is denoted by the letter Hl and is defined as the union of three operations in sequence (YN-Relu-Conv.). The stack normalisation (YN), the RELU [6], and the 3 x 3 convolution are the sequential processes that are being discussed here.

IV. RESULTS AND DISCUSSIONS

For the purpose of shoulder implant manufacturer categorization, the performances of the DenseNet201, DenseNet-169, InceptionV3, NasnetLarge, VGG-16, VGG-19, Resnet50, and Community architectures in individual and cascade models are discussed in this section in further detail. An overlapping complexity matrix (CM) was produced for each of the models, and performance metrics (D, F, F1-score, and accuracy) that are representative of the entire model were computed using this matrix. This was done so that a general assessment of the models could be made. When comparing the performance of different models, it is essential to make use of measures that are derived from overlapping KM in order to ensure that the overall performance is properly shown. The overlapping KM is created by combining together the individual KMs that are collected from each level [34].

Table 1: No data augmentation comparison of applied deep learning models.

Criteria	Accuracy	Sensitivity	Specificity	Precision	RMSE	F_1
AlexNet	94.41	0.959	0.041	0.972	2.44	0.965

Dark-Net-53	93.85	0.965	0.034	0.958	3.09	0.962
Insepti on-ResNet	95.53	0.965	0.034	0.979	1.93	0.972
ResNet-50	92.73	0.95	0.05	0.956	3.93	0.953
VGG-19	91.06	0.95	0.048	0.937	4.87	0.944
Proposed FCNN model	97.2	0.98	0.019	0.986	1.32	0.983

Table 2 provides a summary of the models' overall performance in terms of categorization. With an average classification accuracy of 84.76 percent, the stepped DenseNet201 model has the greatest classification performance. This model ranks the Cofield manufacturer with a score of 77.10 percent (D=0.7710), the Depuy manufacturer with a score of 93.20 percent (D=0.9320), the Tornier manufacturer with a score of 73.24 percent (D=0.7324), and the Zimmer manufacturer with a score of 77.85 percent (D=0.7785) for correctly classifying its manufacturer. The Depuy manufacturer has the greatest correct categorization rate of all of the manufacturers. This is due to the fact that the quantity of implant X-ray pictures provided by the producer of Depuy products is about three times greater. In accordance with the growing amount of X-ray pictures of implants, classification performance is also improving across various manufacturers. Figure 7 illustrates the overlapping KM that results from using the stepped DenseNet201 model. According to the model, Cofield has a score of 64 true and 19 false, whereas Depuy has a score of 274 true and 20 false, Tornier has a score of 52 true and 19 false, and Zimmer has a score of 116 true and 33 false.

Table 2: With data augmentation comparison of applied deep learningmodels.

Criteria	Accuracy	Sensitivity	Specificity	Precision	RMSE	F_1
AlexNet	93.88	0.96	0.036	0.955	3.02	0.959
Dark-Net-53	90.95	0.95	0.047	0.927	5.07	0.939
Inseptio n-ResNet	93.13	0.95	0.044	0.957	3.46	0.956
ResNet-50	88.69	0.92	0.07	0.923	6.12	0.926
VGG-19	86.51	0.93	0.066	0.891	8.47	0.91
Propose d FCNN model	96.31	0.97	0.021	0.974	1.46	0.976

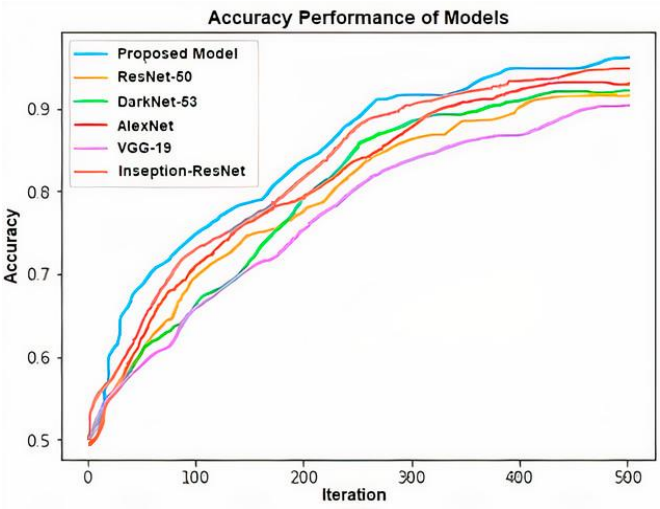

Figure 4: Accuracy performance of applied models' graph

The DenseNet201 model has the greatest classification performance, with an accuracy value of 76.73 percent, when the individual model performances that are not fed with YOLO are assessed. On the other hand, the accuracy of this model's classification is noticeably worse compared to that of the cascading DenseNet201 model that was fed with YOLO. The same may be said about the many other models. To put it another way, the classification performance has been significantly improved thanks to the development of stepwise models, which were accomplished by providing

YOLO to the ESA models. DenseNet169 models have the ability to classify data with the second-best performance, according to both the cascade and individual models.The accuracy of these models in terms of categorization was 81.74 percent, and it is now at 72.29 percent. On this particular data set, DenseNet models demonstrated superior classification performance compared to that of other models.Figure 8 presents a comparative analysis of the classification accuracy values obtained from the stepwise and individual modelling approaches. Classification performances of the models, ordered from largest to smallest, for cascade models DenseNet201 (accuracy = 0.8476), DenseNet169 (accuracy = 0.8174), InceptionV3 (accuracy = 0.8042), Ensemble (accuracy = 0.7688), NasnetLarge (accuracy = 0.7421), VGG16 (accuracy = 0.6600), VGG19 (accuracy = 0.6082) and Resnet50 (acc DenseNet201 (accuracy = 0.7673), DenseNet169 (accuracy = 0.7523), InceptionV3 (accuracy = 0.7254), NasnetLarge (accuracy = 0.6414), VGG19 (accuracy = 0.5125), Community (accuracy = 0.4927), Resnet50 (accuracy = 0.4302), and VGG19 (accuracy = 0.4099) are the individual models with the highest accuracy. When compared to both cascading and individual models, the Resnet50 and VGG19 models had the weakest classification performance. In addition, the Community model has a classification performance that is only modest when applied to the cascade structure, but it shows a classification performance that is much worse when applied to the individual structure. The classification performance of the ensemble model is greatly improved when fed with YOLO in comparison to the performance of other models.

As was said previously, there are relatively few studies that categorise X-ray pictures of shoulder implants using the technique of deep learning. Despite this, there are still a few research that carry out categorization using ESA models that have been pre-trained. It is essential to evaluate the findings of this study in light of those obtained from other research. In this investigation, Urban et al. made use of the data set that was used in [2]. The greatest level of classification accuracy, 0.80, was achieved by Urban et al. via their use of six distinct pre-trained ESA models in their work. However, the classification accuracy obtained in this investigation by three of the stepwise models that were fed with the YOLO detection technique was found to be greater than the accuracy reported in the study by Urban et al. This finding provides more evidence that improving classification performance by using pre-trained ESA designs that are fed with YOLO into stepwise models is beneficial. In a separate piece of research, Yi et al. [4] used the ResNet pre-trained ESA architecture to categorise five distinct types of shoulder implants. They tested five different models and found that the classification accuracy ranged from 0.90 to 0.94 to 0.95 to 0.98 to 1. Through this research, Yi and colleagues were able to improve their categorization accuracy. On the other hand, one 5-outlet ESA model can accommodate all of the numerous shoulder implants that are available.

They constructed five distinct ESA models, each of which assigns a binary rating (True-False) to each implant rather than categorising the implants themselves. This explains why they performed so much better in the categorization. Because in binary classification, the model is trained using just one implant from one manufacturer, and it responds by determining if the X-ray picture in question corresponds to the relevant implant company. In addition, the data set that they utilised is distinct from the one that was used in this research. Because of all of these factors, it would not be particularly realistic to compare studies in terms of their categorization abilities straight to one another. According to the examination of the relevant literature, there is no other research that categorises shoulder implants save these two studies. On the other hand, there aren't very many research that categorise hip and knee implants using deep learning. Yi et al. [8] used ResNet ESA architecture to classify two distinct knee implants, and their results showed a high level of accuracy. Once again, a binary classifier is used in this situation. An accuracy of 0.96 was achieved by Ghose et al. [17] in their classification of six distinct orthopaedic knee implants. During the classification process, they used eight unique pre-trained ESA topologies and found that the MobileNetV2 architecture provided the best results, earning it the highest rating. Borjali et al. [6] used DenseNet-201's pre-trained ESA architecture to classify three distinct complete hip replacement implants, achieving an accuracy of 1.

In earlier work on implants, including the shoulder implant and others, direct use of pre-trained ESA

architectures was made in every case. However, in our research, in addition to the direct use of ESA architectures, we also used it as a cascade model by combining it with YOLO. The results of this combination showed that it increased the classification performance, which will serve as a foundation for future research in this area of study. The data set itself is the source of the most significant shortcoming of our investigation. As was previously underlined, the poor and varied resolution of the X-ray pictures, the varying aspect ratios and low image contrast, and the unequal distribution according to the manufacturer had a major detrimental influence on the model performances. This was the case for a number of reasons. The cascade models that were developed in the research could display improved classification accuracy on a dataset of greater quality and one that is free of these unfavourable characteristics.

V. CONCLUSION AND FUTURE WORK

Through the use of seven distinct pre-trained ESA architectures and stepwise models in which these ESA architectures were fed with YOLO, the authors of this work were able to distinguish between four distinct manufacturers of shoulder implants using X-ray pictures. According to the findings of the research, the DenseNet201 ESA cascade model that was fed with YOLO had the greatest classification performance. This model had an accuracy rate of 84.76 percent. The accuracy of this categorization is greater than that of another research [2] that used a dataset that was comparable to this one. Additionally, it has been shown without a reasonable doubt that the classification performance of cascade models is superior to that of individual ESA structures. Instead of concentrating on the photos as a whole, the cascade models zero down on certain aspects of the X-rays of the implants taken with YOLO that stand out from the rest. The accuracy of the categorization is improved as a result. It's possible that the stepwise model that was provided in this study would be useful for future research on the implant. In addition, the classification performance of the ensemble model, which incorporates voting in addition to conventional machine learning models, was worse than that of the majority of the ESA designs. Deep learning techniques may be used to determine the manufacturer of an implant, which can be helpful to radiologists as well as orthopaedic surgeons when making judgments.

Cascading models will be created with detection algorithms such as YOLO, R-CNN, and Fast R-CNN, and all possible ESA architectures will be demonstrated on a larger data set and their classification performances will be shown. If new datasets on shoulder implants can be reached in future studies, they will allow cascading models to be created. Additionally, X-ray pictures will be rebuilt via the autoencoder deep learning approach, and graded ESA models will be trained utilising features from a variety of various layers.

REFERENCES

[1] Barrett,W.P.; Franklin, J.L.; Jackins, S.E.;Wyss, C.R.; Matsen, F.A. Total shoulder arthroplasty. J. Bone Jt. Surg. 1987, 69, 865–872.

[2] Hawkins, R.J.; Bell, R.H.; Jallay, B. Total shoulder arthroplasty. Clin. Orthop. Relat. Res. 1989, 242, 188–194.

[3] Bohsali, K.I.;Wirth, M.A.; Rockwood, C.A., Jr. Complications of total shoulder arthroplasty. J. Bone Jt. Surg. 2006, 88, 2279–2292.

[4] Cofield, R.H. Total shoulder arthroplasty with the neer prosthesis. J. Bone Jt. Surg. 1984, 66, 899–906.

[5] Lunati, M.P.;Wilson, J.M.; Farley, K.X.; Gottschalk, M.B.;Wagner, E.R. Preoperative depression is a risk factor for complication and increased health care utilization following total shoulder arthroplasty. J. Shoulder Elb. Surg. 2021, 30, 89–96.

[6] Urban, G.; Porhemmat, S.; Stark, M.; Feeley, B.; Okada, K.; Baldi, P. Classifying shoulder implants in X-ray images using deep learning. Comput. Struct. Biotechnol. J. 2020, 18, 967–972.

[7] UCI Machine Learning Repository, Shoulder Implant X-ray Manufacturer Classification Dataset. Available online: https://archive.ics.uci.edu/mL/datasets/Shoulder+Implant+X-ray+Manufacturer+Classification (accessed on 20 February 2021).

[8] Stark, M.B.C.G. Automatic Detection and Segmentation of Shoulder Implants in X-ray

Images. Master's Thesis, San Francisco State University, San Francisco, CA, USA, 2018.

[9] Lindsey, R.; Daluiski, A.; Chopra, S.; Lachapelle, A.; Mozer, M.; Sicular, S.; Hanel, D.; Gardner, M.; Gupta, A.; Hotchkiss, R. Deep neural network improves fracture detection by clinicians. Proc. Natl. Acad. Sci. USA 2018, 115, 11591–11596. [PubMed]

[10] Bredow, J.; Wenk, B.; Westphal, R.; Wahl, F.; Budde, S.; Eysel, P.; Oppermann, J. Software-based matching of X-ray images and 3d models of knee prostheses. Technol. Health Care 2014, 22, 895–900. [PubMed]

[11] Nachimuthu, U.; Geethalakshmi, S.N. Multiple classification system for fracture detection in human bone X-ray images. In Proceedings of the Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), Karur, India, 26–28 July 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–8.

[12] Takahashi, T.; Nozaki, K.; Gonda, T.; Mameno, T.;Wada, M.; Ikebe, K. Identification of dental implants using deep learning—pilot study. Int. J. Implant Dent. 2020, 6, 53. [PubMed]

[13] Kim, J.E.; Nam, N.E.; Shim, J.S.; Jung, Y.H.; Cho, B.H.; Hwang, J.J. Transfer learning via deep neural networks for implant fixture system classification using periapical radiographs. J. Clin. Med. 2020, 9, 1117. [PubMed]

[14] Sukegawa, S.; Yoshii, K.; Hara, T.; Yamashita, K.; Nakano, K.; Yamamato, N.; Nagatsuka, H.; Furuki, Y. Deep neural networks for dental implant system classification. Biomolecules 2020, 10, 984. [PubMed]

[15] Sharma, M.; Malhotra, H.; Kumar, N.; Yadav, J. Deep learning with convolutional neural network for controlling human arm prosthesis using sEMG signals. In Proceedings of the IEEE International Conference on Computing, Power and Communication Technologies (GUCON), New Delhi, India, 27–28 September 2019.

[16] Yi, P.Y.; Kim, T.K.; Wei, J.W.; Li, X.; Hager, G.H.; Sair, H.I.; Fritz, J. Automated detection and classification of shoulder arthroplasty models using deep learning. Skelet. Radiol. 2020, 49, 1623–1632. [PubMed]

[17] Gowd, A.K.; Agarwalla, A.; Amin, N.H.; Nicholson, G.P.; Verma, N.N.; Liu, J.N. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. J. Shoulder Elbow Surg. 2019, 28, 410–421. [PubMed]

[18] Borjali, A.; Chen, A.F.; Bedair, H.S.; Melnic, C.M.; Muratoglu, O.K.; Morid, M.A.; Varadarajan, K.M. Comparing performance of deep convolutional neural network with orthopaedic surgeons on identification of total hip prosthesis design from plain radiographs. MedRxiv 2020.

[19] The University of Washington Shoulder Site. Available online: http://faculty.washington.edu/alexbert/Shoulder/ (accessed on 11 March 2021).

[20] Yılmaz, A. YapayZeka, 7th ed.; Kodlab Press: ˙Istanbul, Turkey, 2017.

[21] Yılmaz, A.; Kaya, U. DerinÖ˘grenme, 2nd ed.; Kodlab Press: ˙Istanbul, Turkey, 2019.

[22] Çelik, A.; Arıca, N. Enhancing face pose normalization with deep learning. Turk. J. Electr. Eng. Comput. Sci. 2019, 27, 3699–3712.

[23] Yan, Z.; Xu, Z.; Dai, J. The big data analysis on the camera-based face image in surveillance cameras. Autosoft J. Intell. Autom. Soft Comput. 2019, 24, 123–132.

[24] Adem, K.; Közkurt, C. Defect detection of seals in multilayer aseptic packages using deep learning. Turk. J. Electr. Eng. Comput. Sci. 2019, 27, 4220–4230.

[25] Alqudah, A.M.; HiamAlquraan, H.; Qasmieh, I.A.; Alqudah, A.; Al-Sharu,W. Brain tumor classification using deep learning technique—A comparison between cropped, uncropped, and segmented lesion images with different sizes. Int. J. Adv. Trends Comput. Sci. Eng. 2019, 8, 3684–3691.

[26] Jinsakul, N.; Tsai, C.; Tsai, C.; Wu, P. Enhancement of deep learning in image classification performance using xception with the swish activation function for colorectal polyp preliminary screening. Mathematics 2019, 7, 1170.

[27] Yılmaz, A. Çokkanallıcnnmimarisiile X-ray görüntülerinden Covid-19 tanısı. GaziÜniversitesiMühendislikMimar. FakültesiDerg. 2021. (In press)