A Deep Learning Framework for Detection and Classification of Implant Manufacturer using X-Ray Radiographs

Attar Mahay Sheetal, K. Sreekumar*
Department of Computing Technologies, SRM Institute of Science and Technology,
Kattankulathur, 603203, Tamil Nadu, India

***Abstract*—Now-a-days, artificial prosthesis is widely used to mitigate pain in damaged shoulders and restore their movement ability. The process involves a complex surgery that attempts to fix an artificial prosthesis into a dead shoulder as a replacement for the ball and socket joints of the shoulder. Long after the surgical process, the need for revision or reoperation may arise due to some problems with the prosthesis. Identification of prosthesis manufacturer is a paramount step in the reoperation exercise. Traditional approach compares the prosthesis under consideration with prosthesis from a vast number of manufacturers. This approach is cost-efficient and requires no extra training for the physician to identify the prosthesis manufacturer. However, the method is time inefficient and is prone to mistakes. Systems based on machine learning have the potential to reduce human errors and expedite the revision process. This paper proposes a shallow 2D convolution neural network (CNN) for the classification of shoulder prosthesis to speed-up the learning process and improve the performance of the deep learning model for implant classification, this paper employed three different techniques. Firstly, a generative adversarial network (GAN) is applied to the dataset to augment the classes with fewer samples to ensure the data imbalance problem is eliminated. Secondly, the highly discriminating features are extracted using principal component analysis (PCA) and used to train the proposed model. Lastly, the model hyper-parameters are optimised to ensure optimal model performance. The model trained with extracted features with a variance of 0.99 achieved the best accuracy of 99.8%.**

Keywords—Machine learning; deep learning; convolution neural network; Adversarial Network (GAN); Principal Component Analysis (PCA); shoulder implants

I. Introduction

One of the invasive methods used to reduce pain and restore movement in injured shoulders is Total Shoulder Arthroplasty (TSA) [1]. Shoulder malfunction is generally caused by rheumatoid arthritis, abrasion, calcification, deterioration of cartilage tissue, and damage to surrounding bones [2]. Surgery on the shoulder is required in order to repair the damaged shoulder's function. The damaged, non-functional joint is surgically removed and replaced with a prosthetic joint [3–5]. Different prostheses are currently produced by a number of manufacturers. Acumed, Biomet, Cofield, Depuy, Encore, Exactec, Tornier, and Zimmer are among the most widely used manufacturers [18]. These manufacturers produce prostheses in various models according to the patient and case type [6]. In order to determine whether prosthesis is compatible with a specific issue in the shoulder, x-ray images of the implants are used. Presently, x-rays plays crucial role in the diagnosis of medical conditions like bone fracture, COVID-19, and many more [23].

After surgery, the implanted prostheses might require repairs for a specific amount of time. In addition, the prostheses might require replacement due to damage from events like accidents [1]. In this instance, the replacement requires knowledge about the prosthesis. The course of treatment is slowed down when this information is either unavailable or unknown to the patient and the doctor. The primary surgical procedure to reduce common complications and prevent treatment delays is determining the prosthesis' model and manufacturer so that it can be positioned correctly. Traditionally, the model and manufacturer are identified by a thorough inspection and visual comparison of the prosthesis's x-ray images with pictures of the prosthesis that are currently available. This method is laborious and prone to mistakes.

Several deep learning techniques have been proposed to reduce errors caused by the conventional approach for the identification of the prosthesis manufacturer and model and to expedite the treatment process. A deep CNN-based method for implant manufacturer classification was presented by the authors in study [6]. The model's accuracy cap is set at 80%. In order to predict the maker of prostheses, the researcher of [7] presents a framework that employs the Squeeze-and-Excitation (SE) network and the conventional 50 layer Residual Network (ResNet50). The suggested method reaches a 97% accuracy level at most. Additional deep learning techniques used are K-Nearest Neighbour [8], Inception, Random Forest, VGG16 [8], ResNet50 [9], and many more. Although these techniques demonstrate a high degree of performance, the proposed deep CNN model incur huge training time and cannot be generalized due to limited number of training samples. On the other hand, the pre-trained models have limited flexibility due to their specific architecture. Adapting these techniques is highly challenging especially when there is need for modification in the model to fit other form of prosthesis datasets. For these reasons, a more effective and reliable solution is still required.

To ensure more accurate and reliable implant prediction, this paper proposes a shallow 2D convolution neural network (CNN) for the classification of shoulder implants. To speed up the learning process of the proposed method and improve the performance of the deep learning method for implant classification, a generative adversarial network (GAN) is applied to the dataset to augment the classes with fewer samples to ensure the data imbalance problem is eliminated, and the highly discriminate features are extracted using principal component analysis (PCA) and used to train the proposed model. Also, the model hyper-parameters are optimised to ensure optimal model performance.

The proposed framework in this paper will remove class imbalances in the dataset, which will make the model unbiased. Also, the feature extraction significantly reduced the training features, thus reducing the model's training time and improving its performance.

The objectives of this paper are as follows:

- To develop a robust deep learning framework for shoulder implant classification with high classification accuracy
- To reduce the processing time and ensure high model performance through dimensionality reduction
- To eliminate data imbalances in the TSA dataset through data augmentation

The remaining portions of this article are organized as follows: In Section II, researchers' efforts to categorize and identify prosthetics manufacturers are examined. In Section III of this paper, the recently built deep learning framework is explained in detail. The experiment's specifics and the outcomes of the training process and the evaluation of the suggested deep learning model are presented in Section IV. Section V presents our findings discussion and Section VI wraps up this paper by outlining our plans for future research.

This template, modified in MS Word 2007 and saved as a “Word 97-2003Document” for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

II. Related Work

To solve the issue of implant manufacturer identification, the authors of [1] use conventional Convolution Neural Network (CNN) and conventional methods of machine learning. The effectiveness of CNN and conventional machine learning methods are contrasted by the authors. The CNN model was given a fresh perspective on channel selection in order to produce filter features. To identify the implant manufacturer and model, the authors use both conventional machine learning techniques and deep learning techniques.

In research [6], researchers categorize shoulder implants in X-ray photographs using a deep learning methodology. The authors assess how well deep learning algorithms perform in comparison to other machine learning classification algorithms, such as gradient boosting and random forest. The authors' findings indicate that Deep Convolutional Neural Network (DCNN) outperforms other machine learning classification algorithms, particularly when an ImageNet pre-trained model is used for classification. While other machine learning classification methods reach an optimal accuracy of 56%, the deep learning model presented in this work uses 10 fold cross validation to achieve an average accuracy of 80%.

To improve the accuracy of shoulder prostheses prediction based on x-ray images on the conventional SIXIC x-ray dataset, authors in study [7] proposed the X-Net framework. The Residual Network module incorporates the Squeeze and Excitation (SE) blocks as part of the suggested model. Through the process of weighing every one of the feature maps obtained using the Residual Network (ResNet) component the method enhances the efficiency of shoulder prosthesis prediction. For obtaining more pertinent features from the xray images in the dataset, both the ResNet and SE components are used. Ultimately, the ResNet and SE modules' fine-grained feature extractions are categorised into Cofield, Depur, Tornier, and Zimmer categories.

In research [8] the performance of traditional ML techniques like RF and KNN is compared with that of deep learning techniques like the 16 layer visual geometric group, Vision transformer, the 50 layer conventional residual network and Inception. The researchers apply a vast DL and ML approaches to the augmented arthroplasty dataset generated by authors of [6, 10]. The results reported by the authors indicate that data augmentation enhances the accuracy of models and lowers the likelihood of over-fitting.

In order to distinguish between the reverse and the normal Total Shoulder Arthroplasty (TSA), as well as between different prosthesis models, the authors of [9] proposed a binary classifier based on a Residual Network (ResNet) Deep Convolution Neural Network (DCNN). For every model, the authors employ five different classifiers, and they assess each model's performance. For the purpose of differentiating between TSA and RTSA and classifying the five distinct prosthesis models, the suggested DCNN achieves a higher AUC-ROC.

A classification tool was proposed by the authors of [10] to identify the manufacturer of shoulder prostheses. The authors sought to remove the obstacles that medical professionals encountered when trying to determine the prosthesis' manufacturer through visual inspection of xray images. After locating the implant using the Hough transform for circles, the authors segment the implant using the seeded region growing method. The results of the suggested software solution in this work were verified visually and by comparing the outcomes of classification with the manually segmented real-world images.

The methods proposed in the literature have achieved promising performance. However, the methods fail to address the issue of class imbalance in the dataset, which tends to learn more about the classes with large samples. These make the model bias towards the class with the larger samples. Also, the methods take longer training time due to the size of features to be used for training and the number of layers in the models. Other models used in the literature are pre-trained models with limited flexibility. Employing these models is highly challenging especially when there is need for modification in the model to fit other form of prosthesis datasets. In this paper, the limitations in the state-of-the-art methods are addressed by eliminating class imbalance using GAN network to generate artificial dataset, which eliminate model bias. The PCA is used to reduce the number of training features, which results in a model with fewer layers, training time and greater performance.

III. MATERIALS AND METHODS

The main motivation behind the development of the proposed deep learning framework is to automatically identify the manufacturer of shoulder implants before the replacement of problematic prostheses in an arthroplasty patient. The workflow of the proposed DL framework for detection and classification of implant manufacturers using X-ray radiographs is depicted in Fig. 1. The workflow consists of the following steps: dataset collection, data preprocessing phase, building and training of various transfer learning models (DenseNet201, Inseption-V3, MobileNet, and ResNet50), and the proposed 2D Convolution Neural Network (2DCNN).

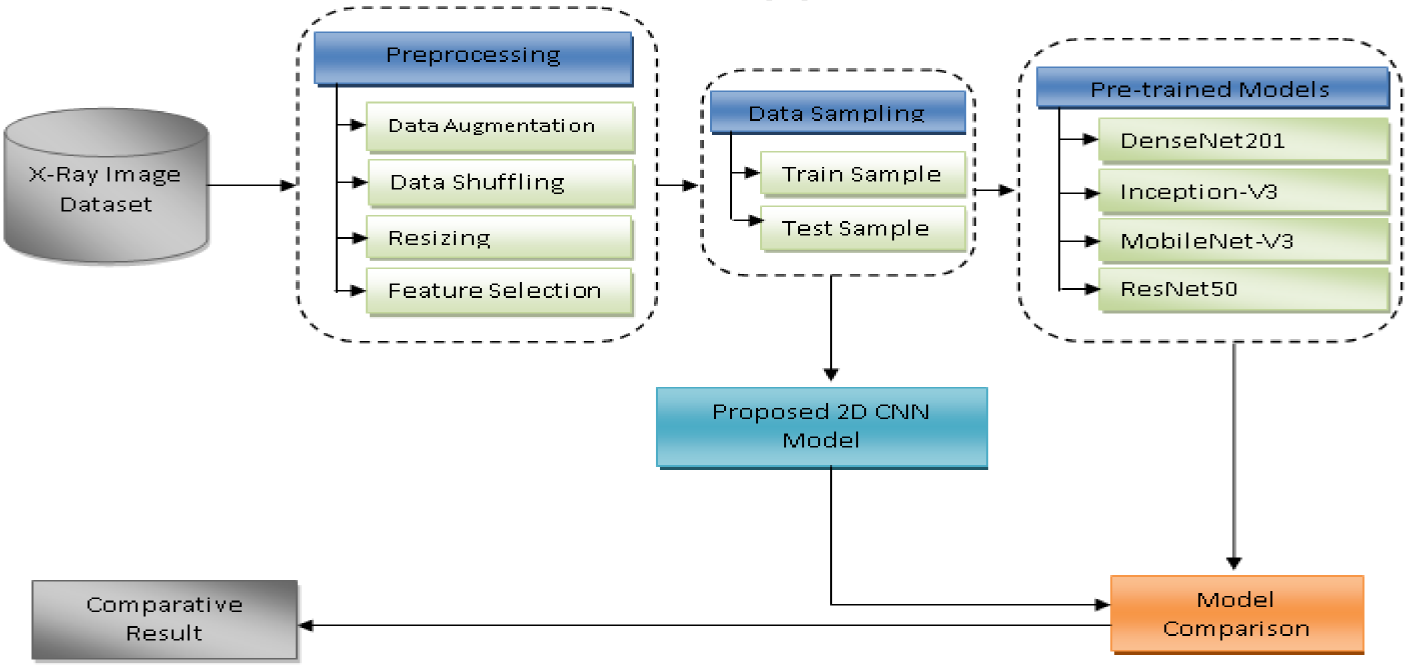


Fig. 1. Workflow of the proposed automated implant manufacturer classification method using X-ray radiograph.

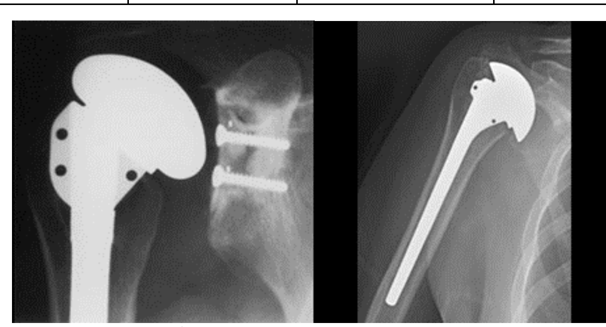
A. Dataset

The dataset used in this research was collected from various sources by the authors of [6, 10]. The initial sample collection includes 605 x-ray radiographs in the Joint Photographic Expert Group (jpeg) format with an 8-bit grey scale and variable sizes. Duplicate images from similar patients were removed from the collection, resulting in a new total of 597 samples spanning four manufacturers: Cofield with 83 sample images, Depuy with 294 sample images, Tornier with 71 sample images, and Zimmer with 149 sample images. The sources of the samples include the Feeley Lab and BIDAL lab at the Californian University and San Francisco state University. Other sources include the various websites of the implant manufacturers and the Common US Shoulder Prosthesis. Table I below shows the initial sample size and the augmented sample size used for training and validation of the pre-trained models and the proposed shallow 2D CNN model.

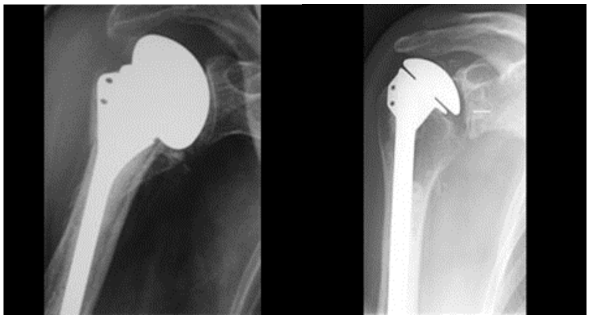
Fig. 2 shows samples of the respective classes of the dataset, which include Cofield, Depuy, Tornier and Zimmer implants.

TABLE I. DISTRIBUTION OF DATASET

Implant Class	Initial Samples Size	Augmented Samples Size	Total Samples
Cofield	83	217	300
Depuy	294	6	300
Tornier	71	229	300
Zimmer	149	151	300



(a). Cofield (b). Depuy



(c). Tornier (d). Zimmer

Fig. 2. Samples of x-ray implant images used for training and validation of the proposed framework.

B. Data Preprocessing

This phase is one of the crucial steps in the proposed workflow. It helps improve the performance of the model, reduce the training time of the model, and prevent model overfitting. The preprocessing steps in the workflow include data augmentation, shuffling, resizing, and feature selection.

1) Data augmentation: The dataset used in this work consists of a few X-ray images with some imbalance among the classes. To improve the downstream performance of the proposed model and avoid poor approximation, we augment the X-ray images to create a bigger dataset for more generalization. To eliminate the class imbalance problem, a generative adversarial network (GAN) is used. GAN is one of the common approaches used by image generation functions to create artificial image data with similar characteristics to real image data. GAN is a multi-layer perceptron neural network consisting of generator (G) and discriminator (D) elements. The generator element generates data similar to the original data during the training, while the discriminator distinguishes between the generated and actual data. The generator element G takes in as input a random noise vector z and generates synthetic data G(z); the discriminator element D takes G(z) as input and outputs a probability D(G(z)) to distinguish between synthetic data and true data from the distribution. To train the generator and discriminator, a two-player min-max game is formed where the generator attempts to generate realistic data to fool the discriminator, whereas the discriminator attempts to distinguish between synthetic and real data. The objective function to be optimised is given as follows:

$$min_G max_D \, V(D,G) = E_{X \sim P_{data(X)}}[\log D(X)] + E_{Z \sim P_{Z(Z)}}\left[\log\left(1 - D\big(G(Z)\big)\right)\right] \quad (1)$$

To prevent the discriminator D from rejecting samples from the generator G with a close confidence of 1, we trained the generator G to maximiseD(G(Z)) so that the discriminator should not be able to distinguish between the synthetic and real data. To achieve the data augmentation, both horizontal and vertical shifts and random r were used.

2) Data shuffling: In this phase of preprocessing, the X-ray images from the various manufacturers were shuffled to ensure that each class of the manufacturer was represented in every batch. It helps the proposed model learn the various patterns in each epoch and increase the speed at which the model converges.

3) Resizing and normalization: In this stage of preprocessing, the images of prostheses belonging to various manufacturers were resized to a common dimension to fit the input of the model. Since the work employs various types of CNN with different input dimension requirements, the dimension of the training set was resized to 224 by 224 by 3 to accommodate the common dimensions of the variants of CNN. To assist in stabilising the problem of gradient propagation and speed up the training of the model, the image pixels are normalised to the range of 0 and 1.

4) Feature selection: To obtain highly discriminative features from the shoulder x-ray images with the potential to enhance the performance of the proposed 2D CNN model, principal component analysis (PCA) was used to reduce the dimension of the features. The choice of the PCA was attributed to its simplicity, efficiency, and well-known multivariate approach to feature extraction. PCA is a statistical approach that employs orthogonal transformation to transform observations of correlated features into a group of linearly uncorrelated features with the highest variance, called principal components. PCA generally attempts to determine lower-dimensional surfaces in order to project higher-dimensional data. In PCA, the principal component is directly connected to the size of the information to be retained. Therefore, to reduce data dimensionality using PCA, an appropriate number of principal components should be selected.

The PCA algorithm reduces a 2D matrix of image pixel values M with dimension A x B to another smaller matrix N with dimension A x P using a linear transformation U of dimension B x P. During the linear transformation process, information from the image data is retained. The transformation process is presented in Eq. (2) below.

$$N = U^T M \quad (2)$$

where, A, B and P represents total pixels after masking, number of instances and the number of pixels such that A, P < B.The covariance C_N of the output of the transformation process N is determined as follows:

$$C_N = \frac{1}{A} N^T N \quad (3)$$

where C_N is a matrix with dimension of P x P

The obtained covariance, C_N is maximised to obtain an eigenvector with Lagrange multiplier λ , which is broken down into three matrices using matrix diagonalization. The decomposition process yields another matrix C, which is a product of the three matrices.

$$C = XDX^{-1} \quad (4)$$

where, X and D represents the matrix of eigenvector and diagonal matrix consisting of Eigen values.

The overall variance of the transformation is therefore represented as the sum of the eigenvalues in Eq. (5) below:

$$N^{Total} = \sum_{i=1}^{B} \lambda_i \tag{5}$$

The percentage information retained by the PCA is calculated as follows:

$$I^{R} = \frac{\sum_{i=1}^{P} \lambda_i}{\sum_{i=1}^{B} \lambda_i} \tag{6}$$

where, $\sum_{i=1}^{P} \lambda_i$ represent the variance retained as the top P eigenvectors data from the subsets of the B vectors

C. Data Sampling

At this phase, the x-ray datasets collected from sources are divided into train and test split. 20% of the dataset is made up of the test set, while 80% is made up of the training set. The models are trained using the training split, and they are validated using the validation split. The train and validation split is used to train and validate each of the pre-trained models and the proposed 2D CNN. The performance of the pre-trained models and the 2D CNN model is then evaluated using the test set.

D. Pre-trained Models

In this phase, we trained different transfer learning models that are already pre-trained on very large datasets. Pre-trained models have been widely used to address numerous deep learning problems caused by inadequate labelled training data, improve the performance of Deep Neural Network and address problems in computer vision. The pre-trained models used in this paper include DenseNet201, InceptionV3, MobileNet, NasNet, ResNet50 and Xception.

1) DenseNet201: DenseNet [11] is a CNN that employs dense connections between the layers of the structure to reduce layer interdependencies by reusing feature maps from various layers. The shortcut connections of variable lengths between layers provide dense and differentiated input features that minimise the gradient disappearance problem in the deep networks [12]. The features from all the layers of DenseNet are finally used to make predictions on a standard dataset with better performance using small-size models with less computation effort. DenseNet has four different variants based on the depth of the layers. In this paper, the DenseNet201 variant, consisting of 201 densely connected layers, is used.

2) InceptionV3: Inception-V3 is a CNN architecture from the Inception family that employs a number of techniques to optimise the earlier versions of the architecture. The initial version of Inception (GoogleNet) employs multiple filters of varying sizes at the same level, thus reducing the size of the deep layer to parallel layers. Inception V1 was later refined by introducing batch normalisation for Inception V2 [13]. A number of factorizations were introduced in Inception V2 to form Inception V3. Inception V3 employs level smoothing, factorised convolutions, and an auxiliary classifier to communicate the class information to other layers of the network [14].

3) MobileNet V3: MobileNet V3 is a CNN architecture from the MobileNet family that employs a number of techniques to optimise the earlier versions of the architecture. The initial version of MobileNet reduced the number of parameters by using dept-wise convolution. In the second version of MobileNet, an expansion layer was introduced to obtain expansion filtering compression. MobileNet V3 introduces a squeeze and excitation layer to the initial building block of MobileNet V2, which later goes for further treatment. The squeeze and excitation layers result in unequal weights for the various channels from the input when generating the output feature maps.

4) ResNet50: The ResNet50 model is a deep convolution neural network consisting of 50 convolution layers, introduced by Microsoft in 2015 [15]. The ResNet50 model consists of approximately 26 million parameters, with the input of the ith layer directly connected to the (i+j)th layer. The ResNet50 model establishes its deep network by stacking additional layers on the input layer. In the residual network, residuals, which are the subtraction of features learned from the input, are learned rather than learning the features [16, 17].

E. 2D CNN Model

Two-dimensional convolutional neural networks, or 2D CNNs, are a class of neural network architecture intended for the processing and analysis of two-dimensional structured data. Tasks involving grid-like data structures, like images, are especially well-suited for it. Image classification, object detection, and image segmentation are just a few of the computer vision tasks in which CNNs have demonstrated remarkable success. A 2D CNN uses convolutional and pooling layers to learn hierarchical features from input data (like images), then fully connected layers to make predictions. The architecture works well for tasks involving spatial relationships and patterns, especially in images, because it is made to automatically learn and extract pertinent features from the input.

The convolution operation is the main function of a 2D CNN. Convolutional layers are made up of filters, sometimes referred to as kernels, which move over the input data, such as an image, and multiply local regions element-wise to create feature maps.

Different characteristics or patterns found in the input data are captured by filters. For instance, deeper layers may capture complex patterns or high-level features, while earlier layers may learn basic features like edges or textures. A number of kernels are used in each convolution layer to determine the feature map tensor. Equation 7 below describes the operation of the convolution layer.

$$y_t = \mathrm{k}(x_t * w_t + b_t) \tag{7}$$

where, yt, k, xt, wt, and bt represents the output of the convolution layer operation, the activation function, the input vector, the layer weight and the bias of the filter or kernel. By using the activation function, the feature maps become more nonlinear. The Rectified Linear Unit (ReLU), which keeps the threshold input at zero, is a commonly used tool for activation computation. The operation is described as follows:

$$f(x) = \max(0, x) \tag{8}$$

The features that the convolution layer extracts have enormous dimensions. A pooling layer is added to reduce the cost of network training and solve the dimensionality issue in the convolution layer. To down sample the parameter sizes, the pooling layer uses the output of the previous convolution layer.

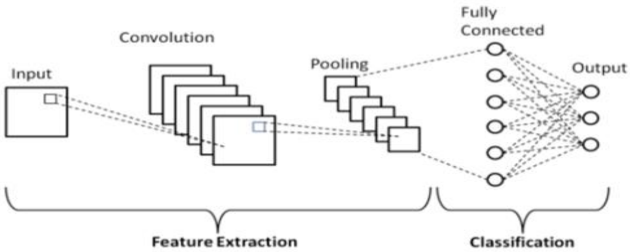


Fig. 3. Architecture of CNN [18].

The proposed 2D CNN used in the classification of implants consists of four (4) convolution layers consisting of the ReLu activation function, four (4) max pooling layers, one (1) flattening layer, a fully connected layer, and an output layer that uses Softmax as the activation function. The architecture and network topology of the proposed 2D CNN are presented in Fig. 4 and Table II, respectively.

TABLE II. NETWORK TOPOLOGY OF THE PROPOSED 2D CNN

Layer	Type	Kernel Size	Stride	Activation Function	Dropout
Convolution Layer	Conv2D	3 x 3	1	ReLu	0
Convolution Layer	Conv2D	3 x 3	1	ReLu	0.25
Pooling Layer	Max Pooling	2 x 2	2	-	-
Convolution Layer	Conv2D	5 x 5	1	ReLu	0
Convolution Layer	Conv2D	5 x 5	1	ReLu	0.5
Pooling Layer	Max Pooling	2 x 2	2	-	-
Fully Connected Layer	-	-	-	-	-
Output Layer	-	-	-	Softmax	-

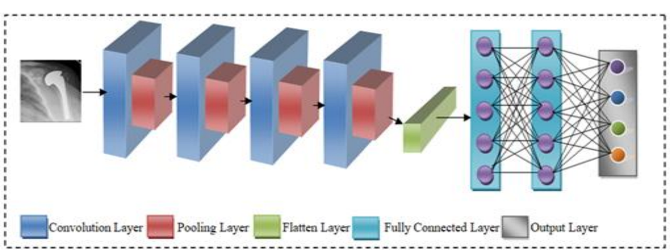


Fig. 4. Architecture of the proposed 2D CNN.

Fig. 4 shows the architecture of the proposed 2D Convolution Neural Network (CNN), consisting of the input image of a shoulder implant, the convolution layers, the pooling layers, the flattening layer, the fully connected layer, and the output layer, which classifies the implant based on the output obtained from the layers that precede it.

F. Performance Evaluation Metrics

The performance indicators of the proposed 2D CNN and the pre-trained models are presented in equation 9 through 12 below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

$$Precision = \frac{TP}{TP+TN} \tag{10}$$

$$Recall = \frac{TP}{FN+TP} \tag{11}$$

$$F1\ Score = \frac{2(Sensitivity*Precision)}{Sensitivity+Precision} \tag{12}$$

IV. EXPERIMENTAL RESULTS ANALYSIS

In this section, the experimental details and the results obtained are presented and discussed.

A. Experimental Setup

In the experiment, the Total Shoulder Arthroplasty (TSA) dataset was split into training and testing sets in the ratio of 80:20. The proposed 2D Convolution Neural Network (CNN) consists of four (4) convolution layers with pooling layers, as described in Fig. 3 and Fig. 2, respectively. The 2D CNN network was implemented using Python version 3.10 and the Keras version 3.0.2 library with Tensorflow version 2.15 on a machine with an Intel (R) processor (Core (TM) i7 CPU @ 2.30 GHz) and 16GB of RAM. In addition, the proposed 2D CNN model was built using Jupyter Notebook and trained using the Graphic Processing Unit (GPU) of the Intel GTX 1050 Ti.

At the initial phase of the implementation, a Generative Adversarial Network (GAN) was built to augment the Total Shoulder Arthroplasty (TSA) dataset. The GAN augmentation approach is used to create artificial image data with similar characteristics to the real image data that resolve the data imbalance in the dataset. The parameter settings of the GAN model are presented in Table III.

TABLE III. PARAMETER SETTINGS FOR GAN NETWORK

Parameter	Parameter Value
Convolution Layer	2
Filter Size	5 x 5
Activation Function	ReLu
Learning Rate	0.0001
Dropout	0.25
Optimizer	Adam
Loss Function	Binary Cross-entropy
Batch Size	4

The real image data and the augmented data are combined to form a total of 1200 images belonging to 4 classes. To enhance the performance of the proposed 2D CNN, the Principal Components Analysis (PCA) method is applied to the compiled dataset. The PCA method extracts the features with the most important information that the model learns. At this stage, top features with a variance between 1 and 0.95 are considered for training the proposed 2D CNN.

The features extracted from the PCA are used to individually train the proposed 2D CNN and the pre-trained model, and the performance of the models is monitored for each feature. For each feature used to train the 2D CNN and the pre-trained model, different values were used for the hyper-parameters, and the optimal values were chosen. Table IV shows the various hyper-parameter configurations used and the chosen values.

TABLE IV. HYPER-PARAMETER CONFIGURATIONS

Parameter	Options	Chosen
Input Shape	64, 128, 224, 512	224
Batch Size	16, 32, 64, 128	64
Learning Rate	0.1, 0.01, 0.001, 0.0001	0.001
Optimizers	SDG, RMSprop, Adam	Adam
Epoch	10, 25, 50, 100	50
Activation function	Sigmoid, Softmax, tanh	Softmax

Based on the performance portrayed by the model using different hyper-parameter values, the final model was trained with an input shape of 224x224x3, a batch size of 64, a learning rate of 0.0001, and an Adam optimizer for 50 epochs.

B. Results Analysis

Tables V through VII show the performance of the proposed 2D Convolution Neural Network (CNN) for individual classes corresponding to the various PCA sets. Based on the results in Tables V to VII, it can be observed that training the proposed 2D CNN with a dataset of the extracted features with a variance of 0.99 achieves greater performance than other features with a different variance. The model trained with extracted features with a variance of 0.99 achieved overall precision, recall, and a f measure of 99.2%.

Based on the performance results in Tables V, VI, and VII, the model trained with extracted features with a variance of 0.99 is better than models trained with extracted features with a variance of 0.95, 0.96, 0.97, 0.98, and 1, thus the model is considered for comparison with pre-trained models trained with extracted features with a variance of 0.99.

TABLE V. PRECISION OF PROPOSED FRAMEWORK WITH VARIANCE BETWEEN 0.95 AND 1.0

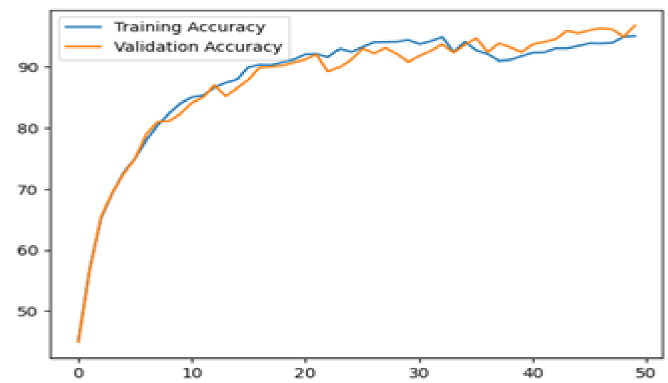
Variance	Cofield	Depuy	Tornier	Zimmer	Overall
1.0	0.987	0.972	0.978	0.977	0.979
0.99	0.992	0.992	0.992	0.992	0.992
0.98	0.987	0.972	0.978	0.977	0.979
0.97	0.987	0.972	0.978	0.977	0.979
0.96	0.987	0.972	0.978	0.977	0.979
0.95	0.987	0.972	0.978	0.977	0.979

TABLE VI. RECALL OF PROPOSED FRAMEWORK WITH VARIANCE BETWEEN 0.95 AND 1.0

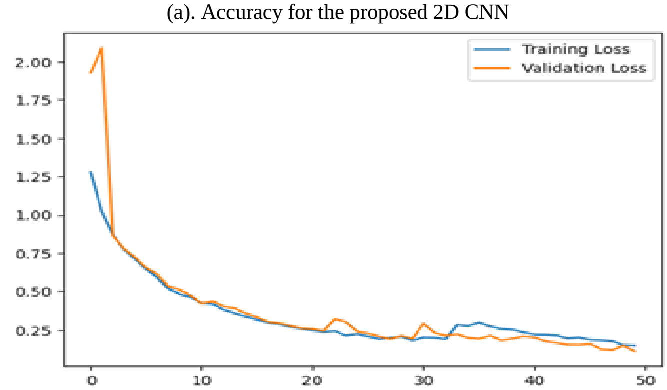
Variance	Cofield	Depuy	Tornier	Zimmer	Overall
1.0	0.984	0.975	0.981	0.979	0.980
0.99	0.992	0.992	0.992	0.992	0.992
0.98	0.984	0.975	0.981	0.979	0.980
0.97	0.984	0.975	0.981	0.979	0.980
0.96	0.984	0.975	0.981	0.979	0.980
0.95	0.984	0.975	0.981	0.979	0.980

TABLE VII. F MEASURE OF PROPOSED FRAMEWORK WITH VARIANCE BETWEEN 0.95 AND 1.0

Variance	Cofield	Depuy	Tornier	Zimmer	Overall
1.0	0.989	0.975	0.980	0.982	0.982
0.99	0.992	0.992	0.992	0.992	0.992
0.98	0.989	0.975	0.980	0.982	0.982
0.97	0.989	0.975	0.980	0.982	0.982
0.96	0.989	0.975	0.980	0.982	0.982
0.95	0.989	0.975	0.980	0.982	0.982



(a). Accuracy for the proposed 2D CNN



(b). Loss for the proposed 2D CNN

Fig. 5. Accuracy and loss for the proposed 2D CNN trained with extracted features with variance 0.99.

Fig. 5 shows the training and validation accuracies and training and validation loss of the proposed 2D CNN trained with extracted features with a variance of 0.99.

Table VIII shows the performance of the proposed 2D CNN and the pre-trained models trained with extracted features of variance 0.99. Based on the performance results in the table, the proposed 2D CNN achieved a 99.79% recall and F1 score and 99.8% accuracy and precision. When compared with the pre-trained models, the 2D CNN recorded the best performance in terms of precision, recall, f1 score, and accuracy when trained with extracted features with a variance of 0.99 for 50 epochs.

TABLE VIII. PERFORMANCE COMPARISON OF PRE-TRAINED MODELS AND PROPOSED 2D CNN TRAINED WITH EXTRACTED FEATURES OF VARIANCE 0.99

Metrics	DenseNet-201	Inception-V3	MobileNet-V3	ResNet50	Proposed 2D CNN
Precision	97.96	98.1	98.2	91.2	99.8
Recall	97.99	97.9	97.2	91.4	99.79
F Meassure	97.65	98.1	98.2	91.2	99.79
Accuracy	97.4	97.8	98.2	92	99.8
Training time (E-4)	53.6	25.2	22.1	38.7	16
Model Size (MB)	80	92	15.3	98	4

Fig. 6 shows the graphical representation of the performance of the proposed 2D CNN and the pre-trained models trained using extracted features with a variance of 0.99. According to the performance comparison, it can be seen that the proposed 2D CNN recorded the minimum training time and memory size as compared to the pre-trained models. The 2D CNN achieves this due to the limited number of convolution layers in the model. On the contrary, the pre-train models have more depth than the 2D CNN, which results in large feature extraction activity at the convolution layers and more memory space to store the model.

Fig. 7(a) through e show the confusion matrix for the proposed 2D CNNs, MobileNetV3, InceptionV3, DenseNet201, and ResNet50, trained with extracted features with a variance of 0.99. Based on the figures, the 2D CNN recorded the least misclassification, with two images belonging to Zimmer misclassified as Tornier and another 2 misclassified as Depuy. Also, one image belonging to Cofield is misclassified as Zimmer, while all test samples belonging to Depuy and Tornier are correctly classified. MobileNetV3 became the second to the proposed 2D CNN, with 7 images belonging to Cofield and Zimmer misclassified as Tornier and Depuy. Also, InceptionV3, which achieves an overall accuracy close to MobileNetV3, recorded seven misclassifications for the Cofield, Tornier, and Zimmer classes, while all test samples from Depuy were correctly classified. The DenseNet201 model recorded 8 misclassifications, with 4 samples from Cofield misclassified as Depuy, Tornier, and Zimmer and the other 4 samples from Zimmer misclassified as Cofield. ResNet50 recorded a total of 9 misclassifications, with 4 images from Zimmer misclassified as Tornier, 2 from Tornier misclassified as Depuy, and 3 from Cofield misclassified as Depuy and Zimmer.

Table IX shows the performance comparison of the proposed 2D CNN with the state-of-the-art methods.

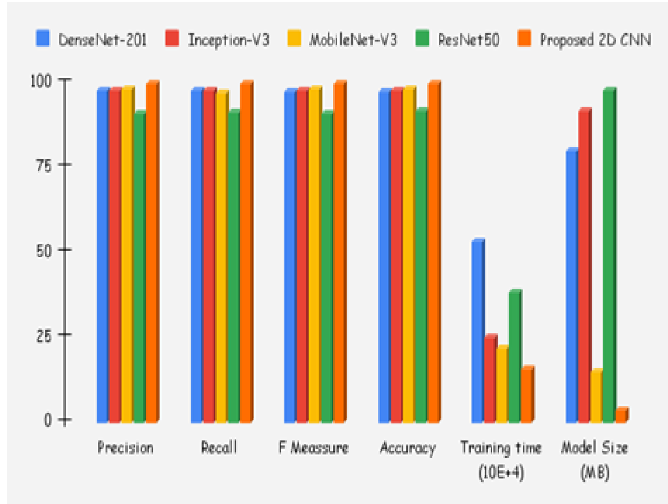
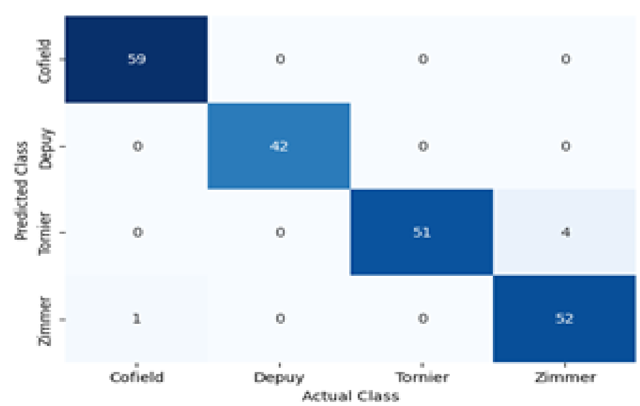


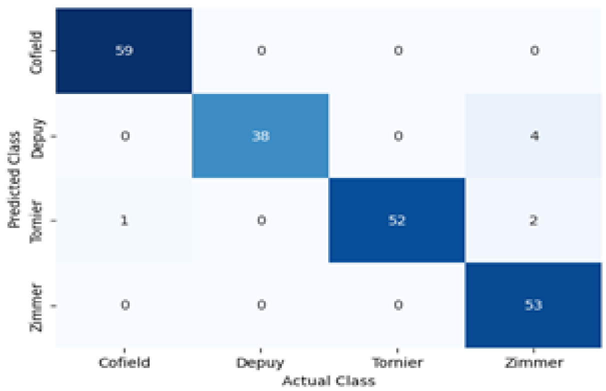
Fig. 6. Graphical representation of the performance of 2D CNN and pre-trained models.

TABLE IX. COMPARISON OF PROPOSED 2D CNN WITH STATE-OF-THE-ART METHODS

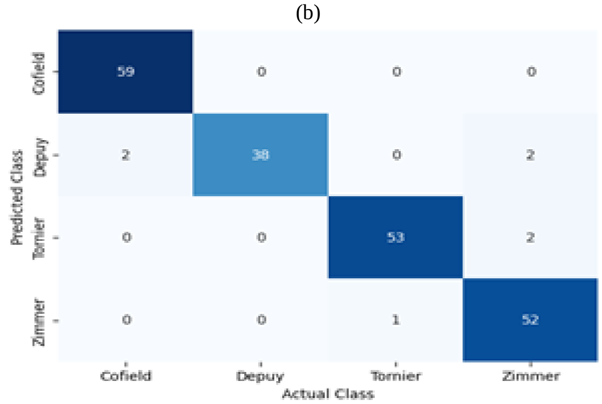
Authors	Method	No of Images	Accuracy (%)
Proposed	2D CNN	900	99.8
Yılmaz, A. [1]	CNN with channel selection	597	97.2
Sivari, E. et al. [19]	Hybrid DL & ML based on DenseNet201+Logistic Regression	597	95.07
Geng, E. et al [20]	CNN	696	93.9
Sultan, H. et al [21]	DRE-Net	597	85.92
Uysal, F. et al [22]	Ensemble Learning Models (E1 & E2) based on ResNet, ResNeXt, DenseNet, VGG, Inception, MobileNet	8942	85
Vo, M. T et al. [7]	X-Net (ResNet + Squeeze & Excitation(SE) block)	597	82
Yi, P. H. et al. [9]	DCNN	482	-
Urban, G. et. Al [6]	Custom CNN	597	80
Zhou, M. et al [8]	Random Forest, KNN, VGG16, ResNet50, InceptionV3, Vision Transformer	597	77 (ResNet50)



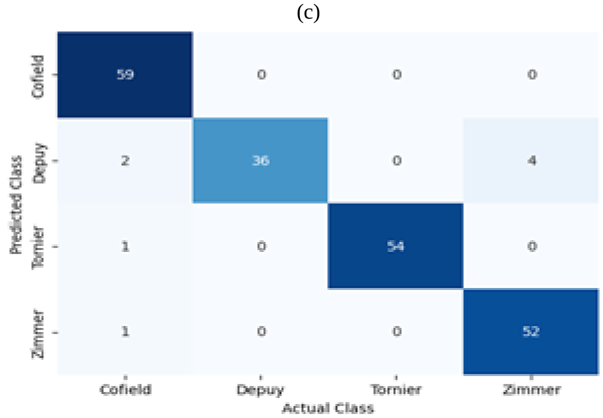
(a)



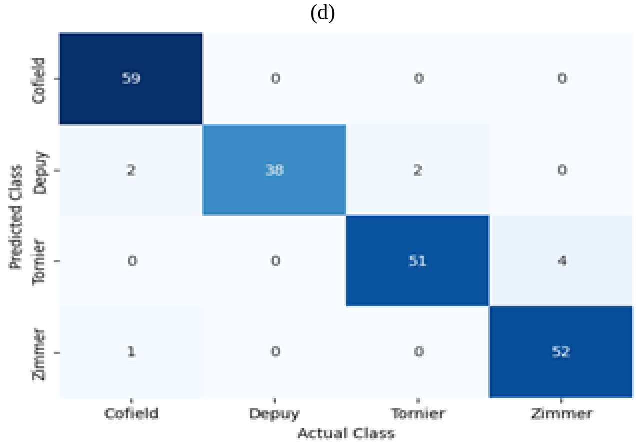
(b)



(c)



(d)



(e)

Fig. 7. Confusion matrices for proposed 2D CNN and pre-trained models trained with extracted features with variance of 0.99.

V. Discussion

By analysing the results obtained, it is demonstrated that feature extraction and data augmentation have a significant effect on the performance of the proposed model. The performance of the proposed model varies with variations in the extracted feature variance. This indicates that features with high information can be selected for model training to obtain the best performance. Aside from the increase in performance of the model when features with high information are used, the model has a lower training time and occupies less memory space as compared to the pre-trained models. This occurs as a result of the few network layers present in the proposed model, which is attributed to the reduced training time and memory space. Based on the performance presented in the result analysis section, the proposed system could differentiate between the manufacturers of the four shoulder implants with high accuracy. A comparison between the proposed 2D CNN and the state-of-the-art methods in terms of accuracy is shown in Table IX. From Table IX, it can be seen that some of the state-of-the-art methods [14–16, 18, 19] obtain slightly lower accuracy compared to the methods in [11–13] and the proposed method. Despite the promising performance displayed by these methods, the proposed 2D CNN recorded a significant improvement over the state-of-the-art methods, as depicted in Table V.

VI. Conclusion

The number of shoulder replacements performed has spiked dramatically over the last few decades. Replacement surgery is typically required if an implanted prosthesis is inadvertently damaged or if specific problems arise with the operating shoulder. Knowledge about the prosthesis is necessary for its replacement process. In many instances, the surgeon closely inspects the prosthesis' x-ray image and visually compares it to existing images of prostheses from various manufacturers in order to determine the prosthesis' manufacturer. This method takes a lot of time and is highly susceptible to errors. This work proposes a shallow 2D Convolution Neural Network (CNN) for implant classification in order to prevent delays, lessen errors and complications in the conventional method, and guarantee robust, dependable, and time-efficient implant classification. To address the class imbalance in the dataset and improve the performance of the proposed 2D CNN, a generative adversarial network (GAN) is used to augment the implant images in the classes with low samples. Principal Component Analysis (PCA) is then applied to the initial and augmented datasets to extract highly discriminating features for model training. The GAN augmentation and PCA feature extraction have a significant impact on the model performance and training speed, as presented in the results section. The proposed 2D CNN recorded an overall accuracy of 99.8%, a training time of 16x104ms and 4MB of memory space, which outperformed the pre-trained models and other state-of-the-art deep learning models. The model in this work is limited to only four classes of prosthesis manufacturers, therefore cannot be generalized. In this feature, the model generalisation will be tested when more datasets are available and the results be compared with that of experts.

REFERENCES

[1] Yılmaz, A. (2021). Shoulder implant manufacturer detection by using Deep Learning: Proposed Channel Selection Layer. Coatings, 11(3), 346. doi:10.3390/coatings11030346.

[2] Bohsali, K. I., Wirth, M. A., & Rockwood Jr, C. A. (2006). Complications of total shoulder arthroplasty. JBJS, 88(10), 2279-2292.

[3] Cofield, R. H. (1984). Total shoulder arthroplasty with the Neer prosthesis. The Journal of Bone and Joint surgery. American Volume, 66(6), 899-906.

[4] Sanchez-Sotelo, J. (2011). Total shoulder arthroplasty. The open orthopaedics journal, 5, 106.

[5] Lunati, M. P., Wilson, J. M., Farley, K. X., Gottschalk, M. B., & Wagner, E. R. (2021). Preoperative depression is a risk factor for complication and increased health care utilization following total shoulder arthroplasty. Journal of Shoulder and Elbow Surgery, 30(1), 89-96. doi:10.1016/j.jse.2020.04.015.

[6] Urban, G., Porhemmat, S., Stark, M., Feeley, B., Okada, K., & Baldi, P. (2020). Classifying shoulder implants in X-ray images using Deep Learning. Computational and Structural Biotechnology Journal, 18, 967-972. doi:10.1016/j.csbj.2020.04.005

[7] Vo, M. T., Vo, A. H., & Le, T. (2021). A robust framework for shoulder implant X-ray image classification. Data Technologies and Applications.

[8] Zhou, M., & Mo, S. (2021). Shoulder implant X-ray manufacturer classification: exploring with vision transformer. arXiv preprint arXiv:2104.07667.

[9] Yi, P. H., Kim, T. K., Wei, J., Li, X., Hager, G. D., Sair, H. I., & Fritz, J. (2020). Automated detection and classification of shoulder arthroplasty models using deep learning. Skeletal radiology, 49(10), 1623-1632.

[10] Stark, M. B. C. G. (2018). Automatic detection and segmentation of shoulder implants in X-ray images (Doctoral dissertation, San Francisco State University).

[11] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[12] Zhou, T., Ye, X., Lu, H., Zheng, X., Qiu, S., & Liu, Y. (2022). Dense convolutional network and its application in medical image analysis. BioMed Research International, 2022.

[13] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1).

[14] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[15] Ikechukwu, A. V., Murali, S., Deepu, R., & Shivamurthy, R. C. (2021). ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. Global Transitions Proceedings, 2(2), 375-381.

[16] Kamal, K., & Hamid, E. Z. (2023). A comparison between the VGG16, VGG19 and ResNet50 architecture frameworks for classification of normal and CLAHE processed medical images.

[17] Mascarenhas, S., & Agarwal, M. (2021, November). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In 2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON) (Vol. 1, pp. 96-99). IEEE.

[18] Islam, M. Z., Islam, M. M., & Asraf, A. (2020). A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Informatics in medicine unlocked, 20, 100412.

[19] Sivari, E., Güzel, M. S., Bostanci, E., & Mishra, A. (2022, March). A novel hybrid machine learning based system to classify shoulder implant manufacturers. In Healthcare (Vol. 10, No. 3, p. 580). MDPI.

[20] Geng, E. A., Cho, B. H., Valliani, A. A., Arvind, V., Patel, A. V., Cho, S. K., ... & Cagle, P. J. (2023). Development of a machine learning algorithm to identify total and reverse shoulder arthroplasty implants from X-ray images. Journal of Orthopaedics, 35, 74-78.

[21] Sultan, H., Owais, M., Park, C., Mahmood, T., Haider, A., & Park, K. R. (2021). Artificial intelligence-based recognition of different types of shoulder implants in X-ray scans based on dense residual ensemble-network for personalized medicine. Journal of Personalized Medicine, 11(6), 482.

[22] Uysal, F., Hardalaç, F., Peker, O., Tolunay, T., & Tokgöz, N. (2021). Classification of shoulder x-ray images with deep learning ensemble models. Applied Sciences, 11(6), 2723.

[23] Meedeniya, D., Kumarasinghe, H., Kolonne, S., Fernando, C., De la Torre Díez, I., & Marques, G. (2022). Chest X-ray analysis empowered with deep learning: A systematic review. *Applied Soft Computing*, *126*, 109319.