Journal of Computer Science

Research Article

A Combined Deep Learning Model with Attention Mechanism for Detection of Implant Manufacturer Using X-Ray Images

Attar Mahay Sheetal and K. Sreekumar

Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India

Article history
Received: 09-07-2024
Revised: 08-10-2024
Accepted: 20-11-2024

Corresponding Author:
Attar Mahay Sheetal
Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India
Email: as9069@srmist.edu.in

Abstract: Shoulder replacement surgery is one of the invasive techniques in orthopedic disciplines that replaces dead shoulder joints with prostheses made from polyethylene and metal components. To perform the surgery, there is a need to know the implant accessories and manufacturer of the implant. Some problems arise in a situation where the patient experiences pain and shoulder malfunctions that need replacement, and the accessories and implant manufacturer are mysterious to the doctor or the patient. In such a case, the solution to the problem depends on the accuracy of the identification of the manufacturer of the prosthesis. This research study proposes a novel detection and classification approach that integrates models based on deep learning with an attention mechanism to identify the implant manufacturers prior to surgery. The ensemble deep learning model utilizes the more sophisticated Long Short Term Memory architecture (LSTM) and the traditional multi layer Convolution Neural Networks for extraction of features and predicting the implant manufacturer. The model employs an attention mechanism to focus on the critical part of the prosthesis that is crucial in the detection of the prosthesis manufacturer. The features map from the attention layer is finally fed into the LSTM for prediction by the implant manufacturer. Collection implant images of 597 from different implant manufacturers, which include 294 images generated by the Depuy manufacturer, 83 images generated by the Cofield manufacturers, 149 generated by the Zimmer manufacturer, and 71 generated by the Tornier manufacturer, are utilized as a dataset not only for training but also for testing the model. The results show that the combined Deep Learning (DL) model with attention mechanism performs better than the Convolution Neural Network model, Convolution Neural Network + Attention, and Convolution Neural Network + LSTM models. Depending on the accomplishment of the model, it is concluded that this model could become an important tool for planning the preoperative procedure and this can be implemented for identifying and classifying the implants from different manufacturers.

Keywords: Deep Learning, Arthroplasty, Convolution Neural Network (CNN), Attention Mechanism, Long Short Term Memory (LSTM)

Introduction

One of the invasive procedures deployed by surgeons to lessen the discomfort and regain the swift mobility of the damaged upper arm joint is Total Shoulder Arthroplasty (TSA) (Yılmaz, 2021). The main causes of shoulder dysfunction are typically rheumatoid arthritis, abrasion, cartilage tissue deterioration, calcification, and injury to the adjacent bones (Bohsali *et al.*, 2006). Medical procedures for the shoulder are essential to restore the injured shoulder and repair its functionality. The injured, non-functioning joint is removed during surgery, and it is restored with an artificial joint (Cofield, 1984; Sanchez-Sotelo, 2011; Lunati *et al.*, 2021). Numerous artificial joints are presently manufactured by different manufacturers. Tornier, Acumed, Depuy, Biomet, Encore, Exactec, Cofield, and Zimmer are some of the most accepted producers (Matsen, 2007). Depending on the patient's needs and the type of case, these manufacturers offer a variety of prostheses (Yi *et al.*, 2020). The compatibility of the prosthesis is assessed using x-ray scans of the prosthesis.

The implanted prosthesis could require repairs for a particular duration of time after the medical procedures. Additionally, the artificial arm might require to be updated if they could assist with the damage from incidents such as accidents (Yılmaz, 2021). In this situation, the replacement of bones will require more information about prostheses. The objective of treatment is delayed when the patient and the doctor lack access to or are unaware of this information. To fit the prosthesis correctly, recognizing the implant prototype and its designer is a paramount. This will help to minimize common difficulties and avert treatment delays. Traditionally, to determine the manufacturer and model, a detailed inspection and ocular differentiation of X-ray imaging of the artificial arm with the images of the existing arm is performed. This procedure will take a huge amount of time and is prone to error.

Different deep learning algorithms have been progressed to accelerate the treatment process and minimize the errors caused by standard methods for identifying the manufacturer and model of prostheses. A technique based on Deep CNN was suggested by the authors of (Urban *et al.*, 2020) for differentiating the implants based on their manufacturers. The utmost accuracy obtained by the model is 80%. Authors of (Vo *et al.*, 2022) present a methodology depending on the SE (Squeeze and Excitation) and ResNet50 (Residual Network) to predict the prosthetics manufacturers. The maximum accuracy for the suggested method is 97%. ResNet50 (Yi *et al.*, 2020), Random Forest, VGG16 (Cha *et al.*, 2021), Inception, K-Nearest Neighbor (Zhou and Mo, 2021), and many other deep learning techniques have also been utilized. These methods are associated with limited focus on relevant features as they tend to process all parts of the input equally, lack of temporal context handling, which can be handled by the LSTM classifier in the proposed method.

This study introduces a Deep Learning (DL) framework integrated with an attention mechanism to predict implant manufacturers. The ensemble DL model uses the more sophisticated LSTM and multi layer CNN for the extraction of features and the implant manufacturer prediction. The feature maps derived from the multiple layered CNN are input into the Squeeze and Excitation (SE) block for obtaining optimal channel attention. The features map from the attention layer is finally fed into the LSTM for prediction by the implant manufacturer. The ensemble model is trained by utilizing the dataset collected by the authors of Urban *et al.* (2020) and the recall, f1-score, accuracy, and precision, are calculated according to the introduced models effectiveness. The introduced method enhances feature extraction by incorporating an attention mechanism, enabling the model to concentrate on the most critical regions of the X-ray image, such as the distinctive characteristics of the implant. This results in more accurate implant identification, as the model pays attention to the most relevant features while minimizing irrelevant information. Also, the combination of CNN and LSTM provide a better handling of complex data and improved generalization.

Outline of the contributions for the paper is granted as follows:

- To create a combined Deep Learning (DL) model with an attention mechanism for accurate implant manufacturer prediction during shoulder bone replacements
- To reduce the number of errors when using the manual method of implant manufacturer identification
- To develop an attention-based method that focuses on the relevant features that determine the class of implants

Literature Review

For improving the prediction of the shoulder artificial arm by utilizing an X-ray image from the SIXIC dataset, researchers in Vo *et al.* (2022) have suggested the X-Net framework. Squeeze and Excitation (SE) blocks are incorporated into the Residual Network (ResNet) module for the suggested model. By utilizing the ResNet (Residual Network) module the distinct feature map weights are extracted, this technique boosts the prediction effectiveness of the upper arm joint prosthesis. Both the ResNet and SE modules are leveraged to derive more important traits from the upper arm joint image dataset. In the end, the traits from the four classes of prosthesis were applied to the fine-grained characteristics retrieved by the ResNet and SE modules.

The authors in Urban *et al.* (2020) utilized Deep Learning (DL) to classify shoulder implants in X-ray images. The researchers evaluate the effectiveness of deep learning models in comparison to several machine learning algorithms, like gradient boosting and random forest. According to the author's research, when using a pre-trained model like ImageNet for classification, the DCNN (Deep Convolutional Neural Network) excel beyond other ML (machine learning) methods. By utilizing 10-fold cross-validation, 80% of an average accuracy is achieved by utilizing the DL model in this study while other machine learning classifiers top out at 56%.

The authors of Stark (2018) propose a categorization strategy for determining the implant manufacturer of the shoulder arm prostheses. Authors try to address the difficulties encountered when medical practitioners attempt to determine the manufacturer of the prosthesis by visually examining x-ray pictures. The implant is located by the authors utilizing the cripple transform for circles, and the artificial limb is subsequently segmented by utilizing a seeded region growth method. The classification outcomes were compared with the manual segmentation ground truth imagery, and the findings were visually examined to assess the validity of the outcome of the suggested software solution in this study.

The working of deep learning techniques like Vision Transformer, VGG16, Inception, and ResNet50 (Residual Network) is compared to that of more developed ML solution such as KNN and RF in Yi *et al.* (2020). The authors apply different DL and ML methods for increasing the Total Shoulder Arthroplasty (TSA) dataset that they received from the Irvine machine learning repository, University of California. According to the authors' findings, data augmentation enhances model performance and lowers the risk of over-fitting.

The research presented in Zhou and Mo (2021) introduces a Deep Convolutional Neural Network based on the ResNet architecture, aimed at detecting shoulder arthroplasty implants in X-ray images. The proposed model is designed to differentiate between various types of shoulder replacements, including Total Shoulder Arthroplasty (TSA) and Reverse Total Shoulder Arthroplasty (RTSA). Furthermore, it classifies different prosthesis models associated with these procedures. Utilizing a binary classification method, the network effectively distinguishes implant types, enabling automated recognition. This technique improves diagnostic precision and aids healthcare professionals in making informed clinical decisions by providing an advanced tool for analyzing orthopedic implants in radiographic images. The performance of models is assessed by the authors using five various classifiers for each model. The proposed DCNN provides a better AUC-ROC for classifying the five various prosthetic models and differentiating between TSA and RTSA.

The authors in Sultan *et al.* (2021) present an innovative approach to identifying shoulder implants using AI technologies. The investigation emphasis on building a deep learning system in accordance with the traditional dense residual ensemble prototype to automatically recognize various types of implants of the upper arm joint out of X-ray images, aimed at enhancing personalized medicine practices. This method helps automate the identification process, reducing the need for manual inspections by clinicians, which might require significant time and vulnerable to errors. Their model outperforms traditional approaches by achieving high recognition accuracy and reliable implant classification. This research is positioned as an important step toward improving orthopedic care by allowing quick and accurate identification of implants, which is crucial for providing individualized treatment plans and avoiding complications in revision surgeries. Additionally, the research underscores the importance of intelligent systems in advancing medical image analysis and personalized medicine.

In Geng *et al.* (2023), authors introduce a machine learning-based solution for classifying implants of the upper arm joint. The focus of the investigation is on differentiating between TSA and RSA upper arm joint implants, two common types of shoulder joint replacements, using a large dataset of clinical images. The investigators develop a ML based algorithm specifically trained to recognize these two implant types, aiming to assist clinicians in the accurate identification of implants for follow-up care and surgical planning. The prototype demonstrates an elevated level of accuracy in classifying TSA and RSA implants, providing reliable automated support for clinicians in orthopedic settings. This advancement can reduce diagnostic errors and improve decision-making in revision surgeries by quickly and accurately identifying the type of implant present in the patient. The paper underscores the growing function of systems with intelligence in orthopedic surgery and highlights its ability to advance efficiency and consequences in clinical workflows by offering fast, consistent implant identification from X-ray images.

Authors in Uysal *et al.* (2021) explore the application of ensemble DL prototypes to classify upper arm joint prosthesis. The authors combine multiple DL models to enhance the accuracy of classifying different conditions and abnormalities seen in shoulder X-rays. The ensemble method capitalizes on the advantages of each individual model, leading to improved robustness and performance compared to single-model methods. Their model achieves high classification efficacy, showcasing the proficiency of using ensemble strategies in biomedical image evaluation. The investigation highlights the prospect of deep learning in automating diagnostic tasks, offering valuable support to radiologists by speeding up and improving diagnostic precision. The results show that ensemble deep learning models can handle the complexity of medical images better than traditional methods, contributing to advancements in computer-aided diagnosis systems.

In Sivari *et al.* (2022), the author's presents a composite ML system designed to categorize the different designers of upper arm joint from X-ray imagery. The authors combine numerous ML techniques, to boost the reliability of implant identification. This composite approach harnesses the advantages of various algorithms, improving general effectiveness compared to using single models. Their system achieves high classification accuracy, offering a valuable tool for automating implant identification, which is crucial for follow-up care and revision surgeries. The paper emphasizes the importance of accurate manufacturer identification in orthopedic practices, particularly when dealing with a variety of implant types. The proposed hybrid system present the prospects of combining traditional and DL methods in medical imaging analysis, offering reliable and fast solutions for clinical applications.

The authors of Ahmed *et al.* (2021) focus on developing a method for diagnosing COVID-19 pneumonia using X-ray lung images. The paper proposes

DOI: 10.3844/jcssp.2025.1322.1331

a composite approach that fuses deep feature extraction methods with traditional ML to boost the effectiveness of diagnosis. Initially, DL prototype extract relevant traits out of the image data, which are then fed into machine learning classifiers to distinguish between COVID-19 pneumonia and other conditions. The investigation reveals that the proposed solution attains superior precision in classifying lung images, providing a rapid and effective tool for supporting healthcare professionals in the diagnosis of COVID-19. By automating the diagnostic process, the system aims to reduce the burden on medical staff and facilitate timely treatment. The study underscore the capabilities of integrating deep learning and machine learning techniques in clinical imaging, particularly in the context of the ongoing pandemic. This research contributes to the broader efforts of using artificial intelligence to improve healthcare outcomes during critical times.

Materials and Methods

In this section, a combined deep learning model with an attention mechanism is proposed. The deep learning model involves both long short-term memory and convolution neural network. The convolution neural networkis utilized for extracting features from the implant radiographs and the long short-term memory is utilized for classifying the implants based on their manufacturer. The main goal of attention mechanism is to make the model concentrate on features relevant to the classification. The complete system architecture of the proposed implant manufacturer prediction system is depicted in Figure (1).

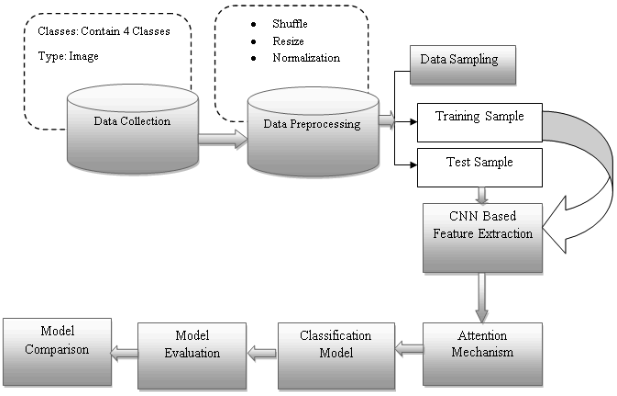


Fig. 1: The overall system architecture of the proposed system for implant manufacturer prediction

Data Collection

For training and testing the combined deep learning model, the data collected by Urban et a. (2020); Stark (2018) from San Francisco State University's BIDAL Lab; and Feeley Lab from the University of California, San Francisco; Common US Shoulder Prosthesis; as well as the websites of various manufacturers were used. This dataset incorporates 605 X-ray pictures in a jpeg file format. Eight images from the initial set were eliminated because they appeared to be the work of the same photographers. The dataset contains graphics from the following manufacturers: Tornier (71), Cofield (83 pictures), Zimmer (149) and Depuy (294).

Figure (2) illustrates the distribution of X-ray images across different implant manufacturers in the dataset. This visualization underscores the need for data preprocessing and sampling techniques to mitigate potential biases during model training. The figure provides a clear overview of the dataset composition, which is critical for understanding the experimental setup and results.

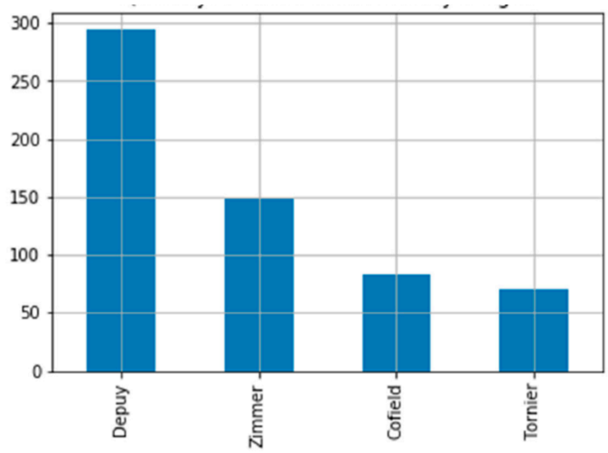


Fig. 2: Count of manufacturers X-Ray images

Data Preprocessing

In this phase, the dataset is preprocessed to enhance the features in the image and suppress some irrelevant features. The implant images are first resized to 224×224 to reduce the number of pixels, which reduces the amount of processing at the training phase. Secondly, we normalize the dataset into a specific range to improve the models performance and reliability. Lastly, we shuffle the implant images to minimize the variance and make the model more general.

Data Sampling

Here, the dataset is divided into training samples and testing samples. The training dataset sample contains 70% of the entire dataset, and the testing dataset sample constitutes the remaining 30%. The training set is further portioned into a training split, which have 70% of the training sample, and a validation split, which have the remaining 30% of the training sample. The 70% of training split is utilized for training the model, and the 30% of validation split of dataset is utilized for validating the model. Once the training phase is finished, the testing sample is utilized to test the performance of the proposed combined deep learning model. Table (1) below shows the distribution of the training, and testing split as well as the validation.

Table 1: Data distribution for training, testing, and validation

Dataset	Cofield	Depuy	Tornier	Zimmer
Training	51	176	43	89
Validation	16	59	14	30
Testing	16	59	14	30

Feature Extraction

Here, the features are extricated from the implant radiographs. The extraction of features is achieved by utilizing a convolution neural network (CNN). A typical CNN comprises a convolution layer comprising of a set of kernels that determine a tensor of feature maps; a pooling layer that down-samples the input dimension to minimize the number of parameters; and the fully connected layer which claims a linear transformation for the input vector through the weight matrix. The kernel in the convolution layer is convolved through the input using strides, thus making the dimension of the output volume integer. The operation of the convolution is presented in Eq. (1):

$$f(i,j) = (I * K)(i,j) = \sum\sum I(i+m, j+n)\, K(m,n) \quad (1)$$

Once the convolution operation is done, the dimension of input volume is minimized. Therefore, the input volume is padded with zeros to maintain the dimension of the input volume with minimized features. A Rectified Linear Unit (ReLU) is assigned to improve the non-linearity in the feature map.

The down-sampling of features by the pooling layer can be achieved using different methods. Max pooling, average pooling, and global pooling are methods of down-sampling in CNN. Max pooling is commonly used and the operation of the pooling layer is presented in Eq. (2):

$$f(x) = \max(0, x) \quad (2)$$

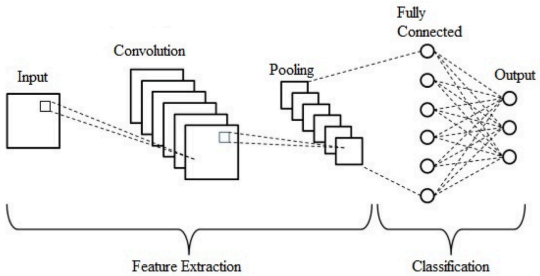


Fig. 3: A typical architecture of the CNN (Lu *et al.*, 2020)

Figure (3) depicts the typical architecture of a Convolutional Neural Network (CNN), which forms the backbone of the feature extraction process in the proposed model. The diagram includes key components such as convolution layers, pooling layers, and fully connected layers, illustrating how the CNN processes input images to extract hierarchical features. This figure is essential for understanding the foundational deep learning techniques employed in the study and their role in improving implant manufacturer identification.

Attention Mechanism

To obtain optimal channel attention, the Squeeze and Excitation Network (SENet) is used. In SENet, the correlation among the channels is utilized for improving the quality of the features created by the Convolution Neural Network (CNN). The features obtained by the CNN are calibrated with importance per channel and selectively adjust the weights of the convolution neural network channels. After the convolution operation, a SE block is attached to the CNN model to enhance its performance. By the squeeze operation, the general information is compressed to implant the global information. Excitation operation calibrates the squeezed important information and scales the importance of individual feature maps. The combination of the Squeeze and Excitation steps forms the SE block. The basic step is a conventional convolution behaviour which transforms the dimensions of X(H' × W' × C') to U(H × W × C). It creates (1×1) sized C feature maps from two-dimensional (H × W) feature maps of C channels. GAP (Global Average Pooling) method is used to form a single value by doing an average of every two dimensional feature map. Thus, the whole data through each channel is compressed. This operation is presented in Eq. (3) below:

$$Z_c = \frac{1}{HxW} \sum_{i=0}^{H} \sum_{j=0}^{W} u_c(i,j) \quad (3)$$

where, $u_c(i,j)$ is the result of convolution operation ($H' \times W' \times C'$) with filter c.

By modifying the non-linear function and the fully connected layer, the excitation operation calculates the channel-wise correlations. In order to decrease the quantity of computation, the dimensionality reduction is carried out in the middle. By applying the non-linear function at the final stage of fully connected layer's outputs. The excitation operation is presented in Eq. (4):

$$s = Sigmoid(W_2 ReLU(W_1)) \quad (4)$$

where, W_1 and W_2 represent two fully connected layers

Finally, every C feature map is multiplied as given in Eq. (5):

$$\tilde{x_c} = s_c \cdot u_c \quad (5)$$

As the output, the feature map is developed where each and every value is scaled by the channel's relevancy, with values between 0-1.

Figure (4) illustrates the overall process of the Squeeze and Excitation Network (SENet), which forms the attention mechanism in the proposed model. The figure demonstrates how the SENet block operates by first compressing global spatial information through a squeeze operation (Global Average Pooling), followed by an excitation step that recalibrates channel-wise feature responses. This process allows the model to dynamically emphasize informative features while suppressing less relevant ones, enhancing the discriminative power of the extracted features. The visualization underscores the critical role of the attention mechanism in improving the model's ability to focus on distinctive implant characteristics, ultimately contributing to higher classification accuracy.

Classification Model

In this phase, the classification model is evolved. The LSTM deep learning model is utilized for predicting the manufacturer of the implant. The attention layers feature map is fed into the classifier, and the class of each implant is predicted. The training dataset is utilized to train the LSTM model and validated by utilizing the validation set.

Long Short Term Memoryis an improvisation of conventional RNN designed to show case the long-standing problems of vanishing and gradient explosion (Islam *et al.*, 2020; Ta *et al.*, 2020; Zarrad *et al.*, 2019). Based on the memory in LSTM and their capability for making relatively accurate forecasts, it shows encouraging performance in applications such as emotional analysis, speech recognition, and text analysis, (Hochreiter, 1998; Tomislav *et al.*, 2018; Gupta and Jalal, 2020). The major difference among the conventional LSTM and RNN is the cell state utilized for saving the long-term state. The LSTM cell consists of three gates namely the input gate, the output gate, and the forget gate, as shown in Figure (4).

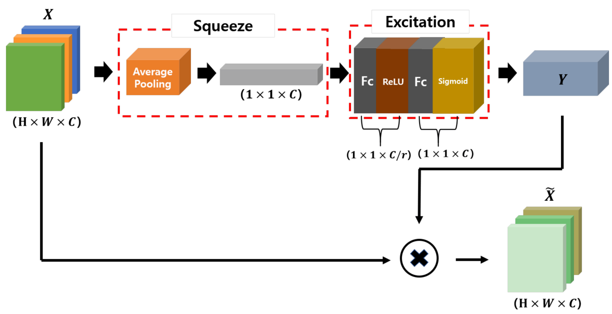


Fig. 4: SENets Overall Process (Cha *et al.*, 2021)

From Figure (5), x_t denotes the present input, C_t and C_{t-1}represent the present cell states and prior cell states. H_t and h_{t-1}represent the present output and prior output.

The procedure of LSTM is as follows:

Initially, the output of previous step $h_{t-1}$and the input of the present step $x_t$is given through the tanh layer (input gate). The current step data, $\tilde{C}_t$,and the output value of input gate are shown as follows:

$$i_t = \sigma\,(W_i.\,|h_{t-1}, x_t| + b_i) \tag{6}$$

$$\tilde{C}_t = tanh\,(W_i.\,|h_{t-1}, x_t| + b_i) \tag{7}$$

where, $i_t \in (0, 1)$, $b_{i,}$ and W_i represent bias and weight.

The present cell is then enhanced as follows:

$$C_t = fC_{t-1} + i\tilde{C}_t \tag{8}$$

With the specific probability, the option for forgetting or keeping the necessary information from previous cell state is made as follows:

$$f_t = \sigma\,(W_f.\,|h_{t-1}, x_t| + b_f) \tag{9}$$

The output O_t of the output gate is shown as follows:

$$O_t = \sigma\,(W_o.\,|h_{t-1}, x_t| + b_o) \tag{10}$$

The final output as well as the state decision vectors are multiplied to tanh layer:

$$h_t = O_t \tanh\,(C_t) \tag{11}$$

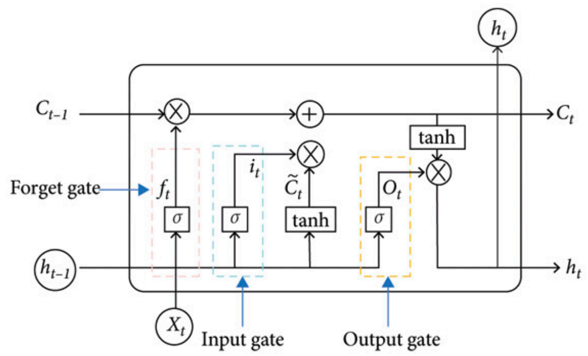


Fig. 5: The LSTM structure (Lu *et al.*, 2020)

Model Evaluation

To evaluate the efficiency of the combined deep learning model with attention mechanism for prediction of implant manufacturer, the following evaluation metrics were used:

$$Accuracy = \frac{TP+TN}{TN+FP+TP+FN} \tag{12}$$

$$Precision = \frac{TP}{TP+FP} \tag{13}$$

$$Recal = \frac{TP}{TP+TN} \tag{14}$$

$$F1_{Score} = 2 * \frac{Precission*Recall}{Precission+Recall} \tag{15}$$

where, *TP, TN, FP* and *FN* represents true positive, true negative, false positive and false positive respectively.

Model Comparison

In this phase, we compare the performance of the proposed combined deep learning model with attention mechanism with the performance of the Convolution Neural Network, CNN with attention mechanism, and combined CNN+LSTM.

Results and Discussions

The experimental setting used to implement the combined deep learning model with the attention mechanism is presented in this section.

Experimental Setup

The benchmarked models and the proposed models were implemented on the Google Colab platform with Tensorflow 2.2.0 and Python 3.7. The training of the model was done by utilizing the Colab Tesla K80 GPU, a 166GB disc, and 12GB of RAM. The CNN, CNN+LSTM, CNN with Attention, and CNN+LSTM with Attention models were trained for 125 epochs by utilizing the initial-stop callbacks with a minimized patience of 5 epochs. Dataset is randomly split into 5 folds for the cross-validation and these models are trained by utilizing the RMSprop optimization technique with a batch size of 64 for the 150 epochs and an initial learning rate of 0.0001. The best models for CNN, CNN with Attention,CNN+LSTM, and CNN+LSTM with Attention are saved in.h5 files.

Figure (6) presents the learning curves for four models: CNN, CNN with Attention, CNN+LSTM, and CNN+LSTM with Attention. Subfigures (a-d) show the training and validation accuracy and loss across epochs, demonstrating the performance improvements achieved by incorporating attention mechanisms and LSTM layers. The curves reveal that the CNN+LSTM with Attention model achieves the highest accuracy and lowest loss, validating the effectiveness of the proposed hybrid approach.

Table (2) shows the performance of the CNN, CNN+LSTM,CNN with attention mechanism, and CNN+LSTM with attention model. As seen from the table, the CNN model recorded an average of accuracy 96%, 95.4% of precision, 95.6% of recall, and f1-score of 96.2%, respectively. On introducing the attention mechanism to the same CNN model, the performance of the model increased by 1.1% of accuracy, 0.2% of precision, 0.9% of recall, and a 0.4% increase in recall value. Based on this result, it is evident that the attention mechanism plays an important role in increasing the performance of the CNN model. Under the same settings, the combined CNN+LSTM model recorded an average accuracy of 97.6%, precision of 96.2%, recall of 95.7, and f1-score of 96.2%. When comparing the combined CNN+LSTM model with the CNN and CNN with attention model, the combined CNN+LSTM model achieves better performance of the evaluation metrics. Alternatively, a proposed combined CNN+LSTM with attention model records an average of 98.7% accuracy, 98.2% of precision, 98.1% of recall, and 97.8%. of an f1-score depending on the performance of proposed combined CNN+LSTM with attention model, an increase of 1.1% accuracy, 2.0% precision, 2.4% recall, and 1.6% f1-score was recorded. From the differences among the performance of the CNN+LSTM and the CNN+LSTM with attention model, it is clear that attention mechanism has played an important role in the performance enhancement of the proposed model.

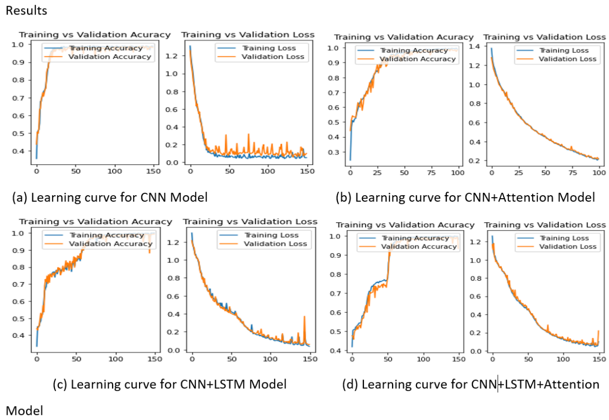


Fig. 6: Learning Curves for CNN, CNN with Attention, CNN+LSTM, and CNN+LSTM with Attention

Table 2: Performance of CNN, CNN with Attention, CNN+LSTM and CNN+LSTM with attention model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	96.0	95.4	95.6	96.2
CNN with Attention	97.1	95.6	96.5	96.8
CNN+LSTM	97.6	96.2	95.7	96.2
CNN+LSTM with Attention	98.7	98.2	98.1	97.8

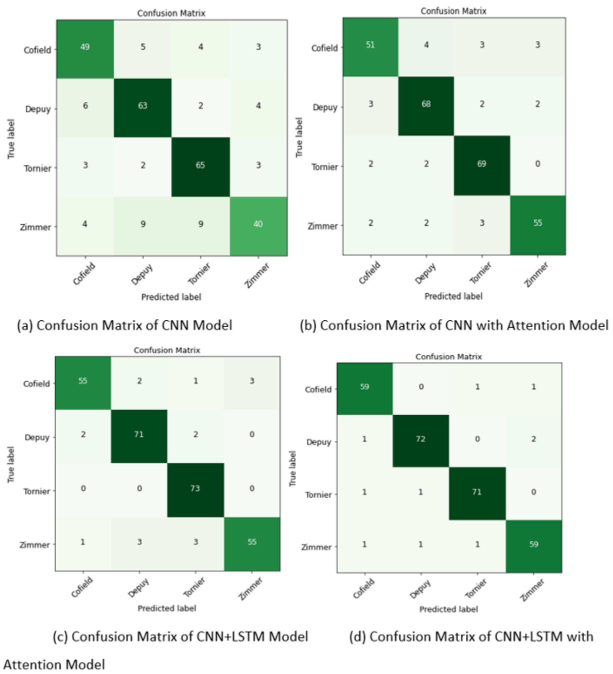
Figure (7) displays the confusion matrices for the four models, comparing their classification performance. The matrices highlight the number of correct and incorrect predictions for each manufacturer, with the CNN+LSTM with Attention model showing the fewest misclassifications. This visualization reinforces the superiority of the proposed model in accurately identifying implant manufacturers, as evidenced by the reduced false positives and negatives. The figure provides a tangible measure of the model's precision and reliability.

The Table (3) provides a performance comparison of several models based on four key metrics: accuracy, precision, recall, and F1-score. The model in Urban *et al.* (2020) has the lowest accuracy (80%), and no additional performance metrics (precision, recall, or F1-score) are reported, making it difficult to assess its overall quality. Moving up, proposed model in Vo *et al.* (2022) shows an improvement in accuracy 82% but presents an imbalance between precision 82% and recall 77%, leading to an F1-score of 79%. The model in Sivari *et al.* (2022) further improves with 84.72% accuracy, with balanced precision, recall, and F1-score all at 84%, reflecting a more consistent performance.

More advanced models, such as in Sultan *et al.* (2021), achieve higher accuracy at 85.92% and maintain

DOI: 10.3844/jcssp.2025.1322.1331

a good balance across precision 85.33%, recall 84.11%, and F1-score 84.69%. The proposed model in Geng *et al.* (2023) stands out with an impressive 93.9% accuracy and an F1-score of 94%, though the precision and recall metrics are not provided. Reference Uysal *et al.* (2021) pushes accuracy to 95.07%, and significantly improves precision to 96.77% and recall to 91.64%, leading to a high F1-score of 93.94, demonstrating strong overall performance with few false positives and negatives.



(a) Confusion Matrix of CNN Model (b) Confusion Matrix of CNN with Attention Model

(c) Confusion Matrix of CNN+LSTM Model (d) Confusion Matrix of CNN+LSTM with Attention Model

Fig. 7: Confusion Matrices of CNN, CNN+Attention, CNN+LSTM and CNN+LSTM with Attention

Table 3: Comparison with State-of-the-art

Reference	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Urban *et al.* (2020)	80.0	-	-	-
Vo *et al.* (2022)	82.0	82.0	77.0	79.0
Uysal *et al.* (2021)	84.72	84.0	84.0	84.0
Sultan *et al.* (2021)	85.92	85.33	84.11	84.69
Geng *et al.* (2023)	93.9	-	-	94.0
Sivari *et al.* (2022)	95.07	96.77	91.64	93.94
Yılmaz (2021)	96.3	97.4	97.0	97.6
CNN+LSTM with Attention	98.7	98.2	98.1	97.8

The top performers are in Yılmaz (2021) and the CNN+LSTM with attention model. Model in Yılmaz (2021) achieves 96.3% accuracy, with high precision 97.4%, recall 97.0%, and an F1-score of 97.6%. However, the CNN+LSTM with attention model surpasses all others, with the highest accuracy at 98.7%, nearly perfect precision (98.2%), recall 98.1%, and an F1-score of 97.8%. This indicates that combining convolutional networks with LSTMs and attention mechanisms significantly enhances classification performance, especially in handling complex data.

Limitations and Future work

The number of x-ray images used in the training and evaluation of the proposed model is limited. The dataset is small and lack variety in implant types and manufacturer, thus the model cannot be generalized on other implant types and manufacturers. Also, the dataset contains an imbalance in the number of samples per manufacturer, thus the model might be biased towards the more frequently occurring manufacturers. This could result in poorer performance for less common manufacturers, leading to an uneven level of accuracy across different classes. In future, In the future, we will generate new data using Generative AI techniques and develop a model depending on transfer learning to address the current problem of shoulder prosthesis data scarcity.

Conclusion

Recently, the number of shoulder replacement surgeries has increased drastically, raising the need for prostheses made from polyethylene and metal components. Different prosthesis manufacturers produce prostheses with different structures to fit different situations. With the variations in the manufacturers and structures of prostheses, accurate identification of the prosthesis manufacturer before replacement is paramount. In situations where the information is unknown to either the patient or the surgeon, the prostheses are rigorously examined and compared with prostheses from different manufacturers. This approach can result in a large error in identification, thus leading to more complications. In this study, we employ a deep learning approach to predict the manufacturer of the prosthesis utilizing the x-ray images. The proposed method here combines the CNN and LSTM models and employs an attention mechanism to focus on the critical part of the prosthesis that is crucial in the detection of the prosthesis manufacturer. We trained the CNN model, CNN with attention mechanism, combined CNN+LSTM, and combined CNN+LSTM with attention mechanism under the same settings and compared their performance. Based on the experimental results, the CNN with attention model recorded higher performance compared to the CNN model. Also, combined CNN+LSTM with attention mechanism recorded higher performance compared to the combined CNN+LSTM model. With the performance shown by the models with an attention mechanism, it is concluded that including the attention mechanism to the models will improve prediction accuracy of the models used in the identification of implant manufacturers.

Acknowledgement

The authors would like to express their sincere gratitude to the Department of Computing Technologies, SRM Institute of Science and Technology, Chennai, India, for providing the necessary infrastructure, computational resources, and academic support that facilitated this research. We are also thankful to our colleagues and research collaborators for their valuable insights and constructive feedback during the development of this study. Additionally, we acknowledge the publicly available datasets and prior research that served as the foundation for our experiments. This study would not have been possible without the institutional backing and scholarly environment fostered by SRM IST.

Funding Information

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors. However, the authors acknowledge the institutional support provided by the Department of Computing Technologies, SRM Institute of Science and Technology, Chennai, India, which facilitated access to computational resources and research infrastructure essential for this study. Any additional assistance or resources utilized were self-funded by the authors as part of their academic research endeavors.

Authors Contribution

Attar Mahay Sheetal: Conceived the research idea, designed the deep learning architecture, implemented the models, and conducted the experiments.

K. Sreekumar: Supervised the research, provided critical guidance on methodology, and contributed to manuscript preparation and revisions.

Both authors analyzed the results, interpreted the findings, and collaboratively wrote the manuscript. Additionally, both authors reviewed and approved the final version of the paper.

Ethics

This research adheres to ethical guidelines for medical image analysis, utilizing only publicly available and anonymized shoulder implant X-ray datasets from prior studies (Urban *et al.*, 2020; Stark, 2018), ensuring no patient privacy violations occurred. No human participants were directly involved, and all data were pre-processed to remove identifiable information. Institutional approval was not required as the work involved retrospective analysis of existing datasets. The authors declare no conflicts of interest and affirm that the research was conducted objectively without external influence.

References

Ahmed, M. A., Al-qaysi, Z. T., Shuwandy, M. L., Salih, M. M., & Ali, M. H. (2021). Automatic COVID-19 Pneumonia Diagnosis From x-ray Lung Image: A Deep Feature and Machine Learning Solution. *Journal of Physics: Conference Series*, *1963*(1), 012099. https://doi.org/10.1088/1742-6596/1963/1/012099

Bohsali, K. I., Wirth, M. A., & Rockwood, C. A. (2006). Complications of Total Shoulder Arthroplasty. *The Journal of Bone and Joint Surgery-American Volume*, *88*(10), 2279-2292. https://doi.org/10.2106/00004623-200610000-00024

Cha, S.-M., Lee, S.-S., & Ko, B. (2021). Attention-Based Transfer Learning for Efficient Pneumonia Detection in Chest X-ray Images. *Applied Sciences*, *11*(3), 1242. https://doi.org/10.3390/app11031242

Cofield, R. H. (1984). Total Shoulder Arthroplasty with the Neer Prosthesis. *The Journal of Bone & Joint Surgery*, *66*(6), 899-906. https://doi.org/10.2106/00004623-198466060-00010

Geng, E. A., Cho, B. H., Valliani, A. A., Arvind, V., Patel, A. V., Cho, S. K., Kim, J. S., & 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. https://doi.org/10.1016/j.jor.2022.11.004

Gupta, N., & Jalal, A. S. (2020). Integration of Textual Cues for Fine-Grained Image Captioning Using Deep CNN and LSTM. *Neural Computing and Applications*, *32*(24), 17899-17908. https://doi.org/10.1007/s00521-019-04515-z

Hochreiter, S. (1998). The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *06*(02), 107-116. https://doi.org/10.1142/s0218488598000094

Islam, Md. Z., Islam, Md. 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. https://doi.org/10.1016/j.imu.2020.100412

Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-Based Model to Forecast Stock Prices. *Complexity*, *2020*, 1-10. https://doi.org/10.1155/2020/6622927

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. https://doi.org/10.1016/j.jse.2020.04.015

Matsen, F. A. (2007). Common US Shoulder Prostheses. *The UW Shoulder Site*. http://faculty.washington.edu/alexbert/Shoulder/CommonUSShoulderProstheses.htm

Sanchez-Sotelo, J. (2011). Total Shoulder Arthroplasty. *The Open Orthopaedics Journal*, *5*(1), 106-114. https://doi.org/10.2174/1874325001105010106

Sivari, E., Güzel, M. S., Bostanci, E., & Mishra, A. (2022). A Novel Hybrid Machine Learning Based System to Classify Shoulder Implant Manufacturers. *Healthcare*, *10*(3), 580. https://doi.org/10.3390/healthcare10030580

Stark, M. B. C. G. (2018). *Automatic Detection and Segmentation of Shoulder Implants in X-Ray Images*.

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. https://doi.org/10.3390/jpm11060482

Ta, V.-D., Liu, C.-M., & Tadesse, D. A. (2020). Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. In *Applied Sciences* (Vol. 10, Issue 2, p. 437). https://doi.org/10.3390/app10020437

Tomislav , Š., Goran, Š., Đorđe, V., Katica, Š., & Roberto, L. (2018). Estimation of CNC Grinding Process Parameters Using Different Neural Networks. *Tehnicki Vjesnik - Technical Gazette*, *25*(6). https://doi.org/10.17559/tv-20180419095119

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. https://doi.org/10.1016/j.csbj.2020.04.005

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. https://doi.org/10.3390/app11062723

Vo, M. T., Vo, A. H., & Le, T. (2022). A Robust Framework for Shoulder Implant X-Ray Image Classification. *Data Technologies and Applications*, *56*(3), 447-460. https://doi.org/10.1108/dta-08-2021-0210

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. https://doi.org/10.1007/s00256-020-03463-3

Yılmaz, A. (2021). Shoulder Implant Manufacturer Detection by Using Deep Learning: Proposed Channel Selection Layer. *Coatings*, *11*(3), 346. https://doi.org/10.3390/coatings11030346

Zarrad, O., Hajjaji, M. A., & Mansouri, M. N. (2019). Hardware Implementation of Hybrid Wind-Solar Energy System for Pumping Water Based on Artificial Neural Network Controller. *Studies in Informatics and Control*, *28*(1), 35-44. https://doi.org/10.24846/v28i1y201904

Zhou, M., & Mo, S. (2021). Shoulder implant X-ray manufacturer classification: exploring with vision transformer. *ArXiv:2104.07667*.