Shoulder Implant X-Ray Manufacturer Multiclass Image Classification

Name: Rajat Lingwal

1. Introduction

Shoulder pain is a global concern affecting numerous individuals, sometimes leading to the necessity of Total Shoulder Arthroplasty (TSA) for joint replacement. Between 1999 and 2020, there was a notable increase in the standardized incidence of primary shoulder replacement surgeries in England, rising from 2.6 to 10.4 per 100,000 population, as reported by the NHS **(Valsamis et al., 2023)**. However, with the increasing number of surgeries, arises the issue of post-operative maintenance for artificial joints. Multiple manufacturers offer various models of artificial shoulders to cater to different patient needs. Over time, implanted shoulders may require servicing or replacement, and patients often struggle to recall the specific model or manufacturer of their implant.

In our literature review, we examined various papers focusing on deep learning models for multi-image classification in the healthcare domain. Our review is divided into two sections: the first part investigates papers that explore the classification of x-ray shoulder implants, using the same dataset as our research. In the second part, we analysed papers addressing other image classification challenges in the healthcare sector.

2. Literature Review

2.1. Shoulder Implant X-Ray Dataset

Urban et al. (2020) employed deep learning techniques, particularly deep convolutional neural networks (CNNs), comparing their performance with other classifiers like random forests and gradient boosting. Utilizing the similar shoulder implants dataset from 4 manufacturers and 16 different models, the research aimed to classify implants accurately. Results showed that pre-trained CNNs, especially those trained on ImageNet data, outperformed other classifiers, achieving approximately 80% accuracy in identifying manufacturers. They emphasized the importance of data augmentation in improving model performance. Future plans include addressing class imbalance and potentially extending the classification task to include both manufacturer and model identification, requiring more data collection.

(Yi et al. 2020) utilized 482 radiography studies from publicly available repositories, focusing on native shoulders, reverse TSA (RTSA) implants, and five TSA models. Utilizing a combination of CNN architectures, particularly ResNet-152 models, they achieved an impressive accuracy of 82.12% in identifying individual TSA models. However, the approach proved to be less scalable, requiring separate models for each TSA model and resulting in a high number of optimizable parameters, particularly due to the complexity of ResNet-152 architecture. Future efforts could focus on developing a unified architecture capable of identifying all implant models, potentially reducing the computational complexity and enhancing scalability for broader applications in orthopaedic implant identification

Sultan et al. (2021) proposed a deep learning-based framework utilizing CNNs to classify shoulder implants from X-ray images. They enhanced the training dataset through rotational invariant augmentation and employed a dense residual ensemble-network (DRE-Net) combining modified ResNet and DenseNet models. Their evaluation on an X-ray dataset demonstrated superior performance of DRE-Net compared to existing methods, achieving high accuracy of 85.92%. Their study also explored the use of class activation maps to visualize the network's decision-making process. Future research aims to expand the dataset, address class imbalance, and further optimize the proposed technique for clinical implementation, reflecting the potential of computer-based algorithms to enhance shoulder arthroplasty outcomes.

Sivari et al. (2022) proposed an automated expert system based on hybrid machine learning models to classify shoulder implant manufacturers using X-ray images. Their system incorporated ten hybrid models combining deep learning and machine learning algorithms, with the DenseNet201 + Logistic Regression model achieving the highest accuracy of 95.07%. Future directions include expanding the dataset to include more implant manufacturers, optimizing feature extraction algorithms, and exploring applications of the proposed method in computer-aided diagnosis systems.

Kanakatte et al. (2023) proposed an encoder-decoder based classifier with a supervised contrastive loss function, capable of identifying implant manufacturers with increased accuracy of 92% using X-ray images. They employed techniques such as image pre-processing, adaptive learning rates, and utilize TensorFlow and OpenCV for implementation. Comparative analysis with other methods on the same dataset reveals the superiority of deep learning approaches over traditional methods, with their proposed method outperforming others with higher precision, recall, and F1-scores. They highlighted the efficiency and robustness of their approach, emphasizing its potential to handle class imbalance and reduce computational complexity compared to existing methods like DenseNet, NASNet, X-Net, and DRE-Net. Future research directions include extending classification to different implant models from various manufacturers.

Jindal & Singh (2024) proposed a transfer learning-based method to detect implant manufacturers. The method involved modifying network architecture using techniques like batch normalization, dropout, and fully convolutional layers, along with cyclical learning rates and weighted loss calculation to address class imbalance. They performed the evaluation using seven pre-trained CNN models, with the DenseNet-201 variant achieving the highest accuracy of 89.5%. Their study emphasized the importance of techniques such as data augmentation and transfer learning in improving model performance. Future research includes exploring advanced data augmentation methods, investigating different transfer learning strategies, and addressing class imbalance issues to enhance classification accuracy further.

2.2. Other Healthcare Image Classification Dataset

Borjali et al. (2021) introduced a novel approach by utilizing deep convolutional neural networks (CNNs) to automatically identify total hip replacement (THR) implant designs from plain radiographs, achieving 100% accuracy in classifying three commonly used designs. Their method significantly reduces identification time, potentially enhancing patient outcomes and reducing healthcare expenses. By employing transfer learning, they demonstrated the effectiveness of their approach compared to pretrained CNNs, which failed to learn to categorize X-rays accurately. Additionally, they highlight the importance of dataset size and the potential of online data augmentation to enhance CNN performance. While acknowledging limitations such as dataset size and the need for additional radiographic views, they proposed future research directions to expand their method's applicability to various implant designs, ultimately aiming to aid revision arthroplasty surgeons in accurately identifying failed implant components.

Naeem et al. (2022) proposed a novel framework, SCDNet, for multiclassifying skin cancer types using deep learning algorithms, specifically Vgg16 combined with CNN. The proposed model's accuracy is compared with four state-of-the-art pre-trained classifiers, Resnet 50, Inception v3, AlexNet, and Vgg19, using the ISIC 2019 dataset. SCDNet achieves an accuracy of 96.91% for skin cancer multiclassification, outperforming Resnet 50 (95.21%), AlexNet (93.14%), Vgg19 (94.25%), and Inception-v3 (92.54%). They emphasized on the importance of early skin cancer diagnosis and highlights the potential of deep learning-based diagnostic systems in improving detection accuracy. Future research will explore the use of pre-trained models with other publicly available datasets and address the limitation of not using image datasets of dark-skinned individuals.

Joseph et al. (2022) focused on improving multi-classification of BC histopathological images using handcrafted features and DNN classifiers. The BreakHis dataset was utilized for this purpose. Handcrafted techniques like Hu moment, Haralick textures, and color histogram were employed to extract features, which were then used to train DNN classifiers with data augmentation to tackle overfitting. Their proposed method achieved impressive accuracy scores of 97.87% for 40x, 97.60% for 100x, 96.10% for 200x, and 96.84% for 400x magnification-dependent histopathological images classification. They also highlighted the importance of data augmentation for further improving classification accuracy. Future research may explore enhancements in data augmentation techniques and investigate the model's efficiency compared to other existing methods.

Hu et al. (2022) aimed to enhance the detection of COVID-19 and non-COVID-19 pneumonia using chest x-ray images by integrating radiomics analysis into a deep learning model. Utilizing a 2D sliding kernel for radiomics, three deep neural network architectures (VGG-16, VGG-19, and DenseNet-121) were initially trained on x-ray images alone and then integrated with selected radiomic feature maps (RFMs). Notably, VGG-16 exhibited significant improvement in COVID-19 classification, while DenseNet-121 showed the most enhancement in healthy individual classification. The highest accuracies achieved were 0.973 (COVID-19), 0.936 (non-COVID-19 pneumonia), and 0.933 (healthy individuals) using VGG-19. The findings suggest that integrating radiomics analysis into deep learning models enhances accuracy and robustness in COVID-19 and pneumonia detection, with potential clinical implications during the pandemic.

Ge et al. (2023) researched about the viability of deep learning (DL) models in multiclassifying reflux esophagitis (RE) endoscopic images based on the Los Angeles (LA) classification system. They utilized images from the HyperKvasir dataset and other hospitals data for training and validation, and independent test set. They utilized DL models, including MobileNet, ResNet, Xception, EfficientNet, ViT, and ConvMixer, via transfer learning using Keras. The EfficientNet model demonstrated superior performance in both the validation and test sets, outperforming other models and endoscopists in terms of accuracy (96.2% and 95.7%). Additionally, Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize the models and highlight target lesions on original images. The findings highlight DL's potential in RE diagnosis and suggest exploring multimodal fusion for improved classification.

Naeem & Anees (2024) aimed to develop a deep learning-based method, DVFNet, for skin cancer detection from dermoscopy images. Anisotropic diffusion methods are used to preprocess images, enhancing quality by removing artifacts and noise. The VGG19 architecture and Histogram of Oriented Gradients (HOG) are combined for discriminative feature extraction. The problem of imbalanced images in the ISIC 2019 dataset is addressed using SMOTE Tomek. Segmentation helps pinpoint areas of damaged skin cells, and a feature vector map is created by combining HOG and VGG19 features. Multiclassification is conducted using CNN with feature vector maps, achieving an accuracy of 98.32%. Future research will explore federated learning to improve results.

3. Conclusion

In summary, the application of various pretrained models in multiclass image classification, particularly in healthcare advancements and surgical procedures like TSA, holds significant promise. Leveraging models such as VGG16, ResNet, Inception, and DenseNet enables healthcare practitioners to effectively categorize medical images, facilitating diagnoses and treatment strategies. Furthermore, integrating data augmentation techniques strengthens model resilience by generating diverse training datasets, thereby enhancing model generalization and effectiveness. Future investigations should explore the combined potential of different pretrained models and advanced data augmentation methods to further elevate the accuracy and dependability of multiclass image classification systems in healthcare applications.

References

- Borjali, A., Chen, A. F., Muratoglu, O. K., Morid, M. A., & Varadarajan, K. M. (2020). Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. Journal of Orthopaedic Research, 38(7), 1465-1471. https://doi.org/10.1002/jor.24617
- Ge, H., Zhou, X., Wang, Y., Xu, J., Mo, F., Chao, C., Zhu, J., & Yu, W. (2023). Development and validation of deep learning models for the multiclassification of reflux esophagitis based on

the los angeles classification. Journal of Healthcare Engineering, 2023, 7023731-8. https://doi.org/10.1155/2023/7023731

- Hu, Z., Yang, Z., Lafata, K. J., Yin, F., & Wang, C. (2022). A radiomics ‐ boosted deep learning model for COVID ‐ 19 and non ‐ COVID ‐ 19 pneumonia classification using chest x ‐ ray images. Medical Physics (Lancaster), 49(5), 3213-3222. https://doi.org/10.1002/mp.15582
- Jindal, M., & Singh, B. (2024). Transfer learning-based class imbalance-aware shoulder implant classification from X-ray images. Journal of Bionics Engineering, 21(2), 892-912. https://doi.org/10.1007/s42235-023-00477-0
- Joseph, A. A., Abdullahi, M., Junaidu, S. B., Ibrahim, H. H., & Chiroma, H. (2022). Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer). Intelligent Systems with Applications, 14, 200066. Improved multi-classification of breast cancer histopathological images
- Kanakatte, A., Bhatia, D., & Ghose, A. (2023). Automated shoulder implant manufacturer detection using encoder decoder based classifier from X-ray images. Paper presented at the, 2023 1-4. https://doi.org/10.1109/EMBC40787.2023.10340429
- Naeem, A., & Anees, T. (2024). DVFNet: A deep feature fusion-based model for the multiclassification of skin cancer utilizing dermoscopy images. PloS One, 19(3), e0297667-e0297667. https://doi.org/10.1371/journal.pone.0297667
- Naeem, A., Anees, T., Fiza, M., Naqvi, R. A., & Lee, S. (2022). SCDNet: A deep learning-based framework for the multiclassification of skin cancer using dermoscopy images. Sensors (Basel, Switzerland), 22(15), 5652. https://doi.org/10.3390/s22155652
- Sivari, E., Güzel, M. S., Bostanci, E., & Mishra, A. (2022). A novel hybrid machine learning based system to classify shoulder implant manufacturers. Healthcare (Basel), 10(3), 580. https://doi.org/10.3390/healthcare10030580
- 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
- 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
- Valsamis, E. M., Pinedo-Villanueva, R., Sayers, A., Collins, G. S., & Rees, J. L. (2023). Shoulder replacement surgery's rising demand, inequality of provision, and variation in outcomes: cohort study using Hospital Episode Statistics for England. *BMC medicine*, *21*(1), 406. Shoulder replacement surgery's rising demand
- 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
- Zhou, M., & Mo, S. (2021). Shoulder implant X-ray manufacturer classification: exploring with vision transformer. *arXiv preprint arXiv:2104.07667*. https://arxiv.org/abs/2104.07667