***Abstract No: (for official use only) Type your Theme here.***

**ENSEMBLE DEEP LEARNING APPROACHES FOR MULTICLASS CLASSIFICATION OF HIP REGION FRACTURES IN X-RAY IMAGES**

# Introduction

# Fracture of the hip, such as pelvic, femoral neck, intertrochanteric, and subtrochanteric, is a serious cause, particularly in elderly patients. Such injuries severely affect mobility and may result in severe conditions of deep vein thrombosis, pulmonary embolism, and permanent inactivity. Diagnosis should play a crucial role in initiating proper medical or surgical care and achieving a better patient outcome.

# The X-ray imaging is the most popular diagnostic method because it is inexpensive, quick, and accessible. Nonetheless, interpretation of X-rays by the sight is qualitative and prone to human error, especially in poor countries where there are few competent radiologists. There is a potential solution to automate and enhance the process of detecting fractures through the use of artificial intelligence (AI), specifically its deep learning and computer vision capabilities.

# A more recent invention in deep learning has been ensemble learning, wherein several models are emulated and then merged in order to provide a single and ultimate forecasting. Such learning has been observed to be considerably more accurate and stable than single-model deep learning. The work under consideration attempts to design and test ensemble deep learning models to perform multiclass classification of hip region fractures based on X-ray images.

# Materials and Methods

# 

Figure 1: High level architecture

# Dataset Collection

# The anonymized data was in the form of a dataset consisting of 1000 hip X-rays-based images that were received in Sri Lankan hospitals during 2022 and 2023. The images were categorized into five classes:

# Non-fracture

# Femoral neck fracture

# Fracture intertrochanteric

# Subtrochanteric fracture

# Combinatory fracture

# Data Preprocessing and Data Augmentation

# All the images were resized similarly, normalized and turned to grayscale when required to maintain consistency. Data augmentation was used to provide additional diversity to dataset and control class imbalance.

# Deep Learning Models

# The data was split into 70% on training, 15 % validation, 15% testing using deep learning models such as RestNet101, RestNet50, EfficientNetB0, and EfficientNetV2. Ensemble learning was used to fine-tune these models on the dataset of hip fractures. The training was implemented on Adam optimizer and a loss of categorical cross-entropy.

Table 1: Deep learning model layers

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Trainable Parameters | Key Convolutional Base Layers | Pooling Layer | Dense Layers | Dropout | Output Layer |
| Efficientnetv2 | Base frozen | EfficientNetv2S  (pretrained) | GlobalAveragePooling2D | Dense (512, ReLU) | 0.5 | Dense (Softmax) |
| EfficientNetB0 | Base frozen | EfficientNetB0 (pretrained) | GlobalAveragePooling2D | Dense (512, ReLU) | 0.5 | Dense (Softmax) |
| ResNet50 | Base frozen | ResNet50 (pretrained) | GlobalAveragePooling2D | Dense (512, ReLU) | 0.5 | Dense (Softmax) |
| ResNet101 | Base frozen | ResNet101 (pretrained) | GlobalAveragePooling2D | Dense (512, ReLU) | 0.5 | Dense (Softmax) |

# Results and Discussion

Table 2: Performance evaluation of deep learning models

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Training accuracy** | **Validation accuracy** | **Testing accuracy** | **Training time (s)** | **Precision** | **Recall** | **F1** |
| RestNet50 | 0.8919 | 0.7194 | 0.7786 | 1237.22 | 0.7298 | 0.7785 | 0.7364 |
| RestNet101 | 0.8735 | 0.7338 | 0.8000 | 1817.57 | 0.8179 | 0.8000 | 0.7690 |
| EfficientNetB0 | 0.8750 | 0.7482 | 0.7286 | 404.34 | 0.6613 | 0.7286 | 0.6769 |
| EfficientNetV2 | 0.8858 | 0.7338 | 0.7500 | 1330.01 | 0.6678 | 0.7500 | 0.6912 |

Table 2 is the comparative study of four deep learning models on hip and pelvic fracture classification X-ray images. ResNet101 had the most testing accuracy (0.8000), precision (0.8179), and F1-score (0.7690) although it took longer during training (1817.57s). ResNet50 was close behind and had a rather high testing accuracy of 0.7786 with a reduced training time (1237.22s). EfficientNetV2 had the highest testing accuracy (0.7500) whereas the lowest-performing model (test accuracy: 0.7286, precision: 0.6613) was EfficientNetB0 whose training took the shortest (404.34s) time. ResNet101 can perform the best to classify the fractures accurately with satisfactory performance and lack of severe computational requirements.

# Conclusion

The present research shows that it can be possible to perform multiclass classification of hip region injuries on the X-ray image to predict fracture using ensemble deep learning techniques. The proposed ensemble method where several state-of-the-art deep learning models were combined, worked better than single models in classification tasks. This method can be potentially deployed in clinics to provide automated decision support to radiologists and orthopedic surgeons in settings where it may be difficult to get expert interpretation.

The future work will be working with augmented sets of data, including multi-modal clinical data, and using sophisticated approaches to data augmentation, such as Generative Adversarial Networks (GANs) to normalize the set. There is a potential to further tune hyperparameters and explore more advanced ensemble tactics (stacking, boosting) to increase diagnostic performance and make it possible to integrate into the healthcare systems.

# References

* Sharrab, Y.O., Alsmira, M., Dwekat, Z., Alsmadi, I., and Al-Khasawneh, A.: ‘Performance comparison of several deep learning-based object detection algorithms utilizing thermal images’, in Editor (Ed.)^(Eds.): ‘Book Performance comparison of several deep learning-based object detection algorithms utilizing thermal images’ (IEEE, 2021, edn.), pp. 16-22
* fractures’, Journal of medical imaging and radiation oncology, 2019, 63, (1), pp. 27-32
* Bae, J., Yu, S., Oh, J., Kim, T.H., Chung, J.H., Byun, H., Yoon, M.S., Ahn, C., and Lee, D.K.: ‘External validation of deep learning algorithm for detecting and visualizing femoral neck fracture including displaced and non-displaced fracture on plain X-ray’, Journal of digital imaging, 2021, 34, (5), pp. 1099-1109
* Hashmi, H., Dwivedi, R.K., and Kumar, A.: ‘Comparative Analysis of CNN-Based Smart Pre-Trained Models for Object Detection on Dota’, Journal of Automation, Mobile Robotics and Intelligent Systems, 2024, pp. 31-45
* Kitamura, G.: ‘Deep learning evaluation of pelvic radiographs for position, hardware presence, and fracture detection’, European journal of radiology, 2020, 130, pp. 109139
* Kuo, C.-W., and Kira, Z.: ‘Beyond a pre-trained object detector: Cross-modal textual and visual context for image captioning’, in Editor (Ed.)^(Eds.): ‘Book Beyond a pre-trained object detector: Cross-modal textual and visual context for image captioning’ (2022, edn.), pp. 17969-17979
* Sanchez, S., Romero, H., and Morales, A.: ‘A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework’, in Editor (Ed.)^(Eds.): ‘Book A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework’ (IOP Publishing, 2020, edn.), pp. 012024
* Babu, B., and Khan, R.H.M.: ‘OBJECT DETECTION–A Comparison Between Pre-trained and Custom Model’, 2023
* Yadav, D.P., Sharma, A., Athithan, S., Bhola, A., Sharma, B., and Dhaou, I.B.: ‘Hybrid SFNet model for bone fracture detection and classification using ML/DL’, Sensors, 2022, 22, (15), pp. 5823
* Cheng, C.-T., Wang, Y., Chen, H.-W., Hsiao, P.-M., Yeh, C.-N., Hsieh, C.-H., Miao, S., Xiao, J., Liao, C.-H., and Lu, L.: ‘A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs’, Nature communications, 2021, 12, (1), pp. 1066.