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

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

# Introduction

# Hip region fractures, including pelvic, femoral neck, intertrochanteric, and subtrochanteric fractures, are critical medical conditions, especially among the elderly population. These injuries significantly impair mobility and can lead to serious complications such as deep vein thrombosis, pulmonary embolism, and prolonged immobility. Timely and accurate diagnosis is essential to initiate appropriate medical or surgical interventions and improve patient outcomes.

# X-ray imaging remains the most common diagnostic tool due to its accessibility, speed, and cost-effectiveness. However, the visual interpretation of X-rays can be subjective and prone to human error, particularly in resource-limited settings with a shortage of expert radiologists. Advances in artificial intelligence (AI), particularly deep learning and computer vision, offer a promising solution to automate and improve fracture detection.

# Recent developments in ensemble deep learning, where multiple models are combined to make a final prediction, have shown superior accuracy and robustness compared to single-model approaches. This study aims to develop and evaluate ensemble deep learning methods for multiclass classification of hip region fractures using X-ray images.

# Materials and Methods

# Dataset Collection

# A dataset of 1000 anonymized hip X-ray images were collected from Sri Lankan hospitals between 2022 and 2023. The images were categorized into five classes:

# Non-fracture

# Femoral neck fracture

# Intertrochanteric fracture

# Subtrochanteric fracture

# Combined fracture

# Data Preprocessing and Augmentation

# To ensure consistency, all images were resized, normalized, and converted to grayscale where necessary. Data augmentation techniques were applied to expand dataset diversity and address class imbalance. These techniques included:

# Random rotation

# Horizontal and vertical flipping

# Brightness and contrast adjustment

# Noise injection

# Dataset Splitting

# The dataset was split as follows:

# 70% for training

# 15% for validation

# 15% for testing

# Deep Learning Models

# Four convolutional neural networks (CNNs) were selected for comparison:

# ResNet101

# ResNet50

# EfficientNetB0

# EfficientNetV2

# These models were pre-trained on ImageNet and fine-tuned on the hip fracture dataset using transfer learning. Training was conducted using the Adam optimizer with a categorical cross-entropy loss function.

# Results and Discussion

Each model was evaluated using accuracy as the primary performance metric. The test accuracy achieved by individual models was as follows:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **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 |

The ensemble model, combining all four architectures, outperformed individual models in overall classification accuracy. This result confirms the benefit of ensemble learning in handling complex multiclass classification tasks, particularly in medical image analysis where different models may capture different fracture patterns and features.

The improved accuracy of the ensemble model is particularly important in clinical contexts where accurate fracture differentiation can influence treatment decisions. Moreover, the ensemble approach proved more robust in identifying subtle or combined fracture types that may be missed by individual models.

Although the results are promising, limitations such as dataset size and class imbalance remain. Addressing these issues in future work will be critical for improving generalizability and clinical applicability.

# Conclusion

# This study demonstrates the potential of ensemble deep learning approaches for the multiclass classification of hip region fractures using X-ray images. By combining multiple state-of-the-art CNN models, the proposed ensemble method achieved superior classification performance compared to individual models. This approach holds promise for clinical deployment, offering automated decision support to radiologists and orthopedic surgeons, especially in environments with limited access to expert interpretation.

# Future work will focus on expanding the dataset, incorporating multi-modal clinical data, and applying advanced data augmentation methods such as Generative Adversarial Networks (GANs) to balance the dataset. Further hyperparameter tuning and exploration of advanced ensemble strategies (e.g., stacking, boosting) may also enhance diagnostic performance and support integration into real-world 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.