JOURNAL OF DATA SCIENCE | Vol.2024:43
eISSN:2805-5160

Convolutional Neural Network Model for Bone Fracture Detection and Classification in X-Ray Images

M. Fariz Fadillah Mardianto[1*], Elly Pusporani[1], Fatiha Nadia Salsabila[1], and Alfi Nur Nitasari[1]

[1]Statistic Study Program, Faculty of Science and Technology, Universitas Airlangga, Indonesia

[*]**Email:** m.fariz.fadillah.m@fst.unair.ac.id

Abstract

Bone fractures are one of the most common medical conditions worldwide. Proper and rapid diagnosis of fractures is essential to ensure effective treatment and reduce the risk of further complications. This study uses a Convolutional Neural Network (CNN) for fracture classification on X-ray images, which aims for the clinical implementation of CNN models in supporting the diagnostic process in the orthopedic field to minimize misdiagnosis due to human error. The analysis results show that fracture classification using CNN has accuracy, precision, recall, and F1-score reaching 99%, indicating highly accurate classification performance. This research aligns with the 3rd SDG's goal of good health and well-being: to ensure a healthy life and support well-being. The results of this research are expected to significantly contribute to the medical world, especially in improving the accuracy and efficiency of fracture diagnosis and become a foundation for developing more innovative diagnostic technologies to support more equitable and quality health services globally.

Keywords

Bone Fracture, Classification, Convolution Neural Network, SDGs, X-Ray Images

Introduction

Fractures are frequent traumatic injuries that affect many individuals (Almigdad et al., 2022). According to World Health Organization (WHO) data, 2019 there were 178 million new fracture cases, an increase of 33.4% since 1990, primarily due to population growth and ageing. In the same year, there were 455 million acute or chronic fracture symptoms, up 70.1% compared to 1990 prevalence. Globally, fractures caused 25.8 million years of life lived with disability (YLD), an increase of 65.3% since 1990 (WHO, 2024). The fracture healing process is complex, involving various biological and mechanical factors to restore the bone to its normal state (Marongiu et al., 2020). However, not all fractures heal entirely, which can lead to complications such as impeded healing or failure of the bone to fuse (Gillman & Jayasuriya, 2021).

Over the past few years, Artificial Intelligence (AI)-based approaches have shown great potential in improving the efficiency and accuracy of medical diagnosis (Kaur et al., 2020). Convolutional Neural Network (CNN) is one of the deep learning models that has undergone significant development, including the introduction of various types of convolutions such as 1-D,

Submission: 24 July 2024; **Acceptance**: 28 October 2024

2-D, and multidimensional CNN is one type of artificial neural network specifically designed for image processing that has been successfully applied in various medical applications (Abdou, 2022). Previous research has been conducted by Xie et al. (2022) regarding the classification of brain tumours using CNN. Other research was also conducted by Nanni et al. (2021) related to bioimage classification to show strong discriminatory power and good generalization across various bioimage and medical datasets. Other related research on classification using CNN was also conducted by Mardianto et al. (2024) to classify pneumonia from X-ray images. In addition, another study using CNN was also conducted for breast cancer classification (Ratna et al., 2024). Based on several previous studies, CNN has proven to be an essential tool in various classification applications, ranging from medical to hyperspectral images. Various strategies have been developed to improve classification performance, such as ensemble, multi-view fusion, and integration with graph convolutional networks.

This study focuses on designing an optimized CNN model for classifying fractures. The performance of the proposed model will be assessed through metrics like accuracy, sensitivity, specificity, and F1 score to evaluate the reliability of its predictions (Naidu et al., 2023). Furthermore, this study investigates the potential clinical applications of CNN models in assisting the orthopaedic diagnostic process, especially in healthcare facilities with limited resources. The novelty of this research is the classification of fractures using CNN. This research is anticipated to promote AI-based technology adoption for fracture diagnosis significantly. Moreover, the findings can potentially enhance healthcare efficiency by offering fast, accurate, and accessible fracture diagnostic solutions. Aligning with the Sustainable Development Goals (SDGs) Goal 3, this study aims to ensure healthy lives and promote well-being for all ages by improving access to technology-driven diagnostic tools. Such advancements can save lives and prevent severe complications caused by delayed diagnoses.

Methodology

Convolutional Neural Network (CNN)

One component of the feed-forward network that has shown efficacy for image or image processing applications is the CNN. CNN is an advancement of the Multilayer Perceptron (MLP), which was first introduced by Yann Lecun in 1980 and intended to process two-dimensional input. Images are processed by filters in CNNs to extract important features like corners and edges. Convolution is used to apply the filter to the image repeatedly (Witten & Frank, 2017). In object image classification, CNN has proven to outperform other machine learning methods, including Support Vector Machines (SVM) (Jozdani et al., 2019). CNNs are widely employed for object detection, image recognition, and analysis (Kumara et al., 2024). Their architecture consists of multiple interconnected layers, often referred to as multi-building blocks. These include convolutional layers, pooling layers, activation functions, fully connected layers, loss functions, regularization techniques, and optimizers.

Data Collection

In this study, each model will be trained using datasets classified into two classes, not fractured and fractured obtained from the Kaggle repository of open access datasets with the details shown in Table 1 (Ashraf, 2023). Before classification analysis, the balance of data distribution for each class or category in the dataset is first identified. This study comprised 2,445 fractured data and 2,445 non fractured data. By considering the number that is not much different between the categories, it can be concluded that the research dataset has been distributed evenly for each category (balance) so that it can be continued for the analysis stage. Furthermore, the image is divided into data training, validation, and testing. This division has been available in the dataset source format with a ratio of 84:8:8 or 4,097 training data, 399 validation data, and 404 testing data.

Table 1. Comprehensive Overview of Datasets

Category	Number of Images		
	Training	Validation	Testing
Fractured	2,078	199	168
Not Fractured	2,019	200	236
Total	4,097	399	404

Data Pre-Processing

The research data was thoroughly pre-processed before entering the analysis stage. In this study, the images were normalized by rescaling the pixel values to between 0 and 1. In addition, to ensure that the image was compatible with the model architecture, it was resized to 224 x 224 pixels and converted into RGB format.

Architecture of The Model

The model architecture utilized in this study is a sequential CNN designed explicitly for binary classification tasks, with the output defined by class_num=1 and a sigmoid activation function in the final layer. The input to the model comprises RGB images resized to 224 x 224 pixels. The architecture includes three convolutional blocks, each consisting of Conv2D, BatchNormalization, and MaxPooling2D layers. The number of filters in the Conv2D layers increases with depth: 32 in the first block, 64 in the second, and 128 in the third. All Conv2D layers employ a 3 x 3 kernel and ReLU activation function. Following the convolutional blocks, a Flatten layer converts the feature maps into a one-dimensional vector, passed through two Dense layers with 256 and 128 units using ReLU activation. The output layer consists of a single Dense unit with a sigmoid activation function to produce class probabilities. BatchNormalization and a Dropout rate of 0.3 are incorporated to enhance the model's performance and generalization capabilities.

Evaluation Metric

Various evaluation metrics can be employed to assess the performance of a classification model. This study assesses the model's predictions using metrics such as accuracy, precision, recall, and F1-score. Higher values of these metrics indicate superior model performance. The evaluation is conducted through a Confusion Matrix, which is based on four parameters: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Validation metrics are computed through mathematical formulas, as follows (Kamil, 2021).

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{4}$$

Analysis Steps

The analysis steps in this study include data collection, data preprocessing, model training, model evaluation, and performance measurement. Figure 1 provides an overview of the analysis steps adopted for this study.

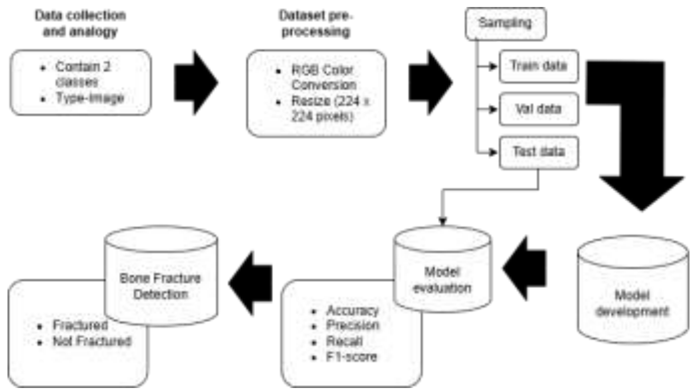
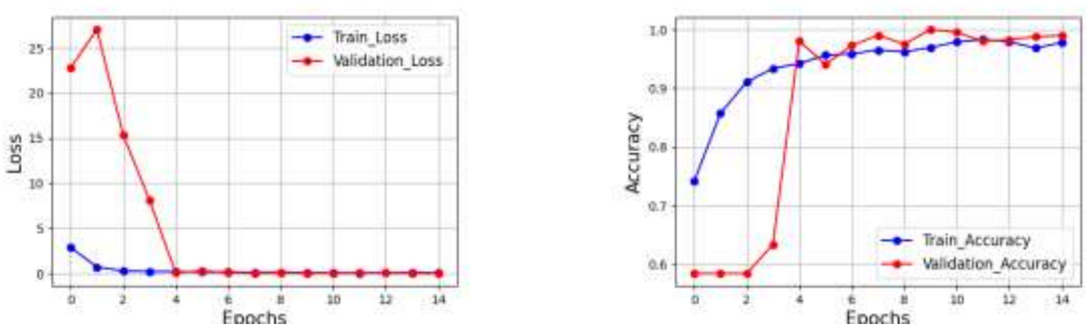


Figure 1. Research Flowchart

Results and Discussion

The model in this study is configured using Adam optimizer and binary cross-entropy as the loss function, which is suitable for binary classification tasks. The training process is performed with a maximum of 15 epochs using the training dataset and validated with the validation dataset. Two callbacks are used, namely ModelCheckpoint, to save the best weights in "model.keras" format and EarlyStopping, with the provision of stopping training if there is no significant improvement in 5 consecutive epochs. The training results are visualized in Figure 2 as follows.



(a) (b)

Figure 2. Graph (a) Loss Function and (b) Accuracy Rate

After completing the training process, the model was evaluated using the testing data to assess its performance on unseen data. Figure 3 illustrates the confusion matrix generated by the model in this study. This matrix provides a tabular representation of the comparison between actual labels and the labels predicted by the model. The diagonal values, highlighted in dark blue, represent the correctly classified instances for each class, whereas the off-diagonal values indicate misclassifications or prediction errors.

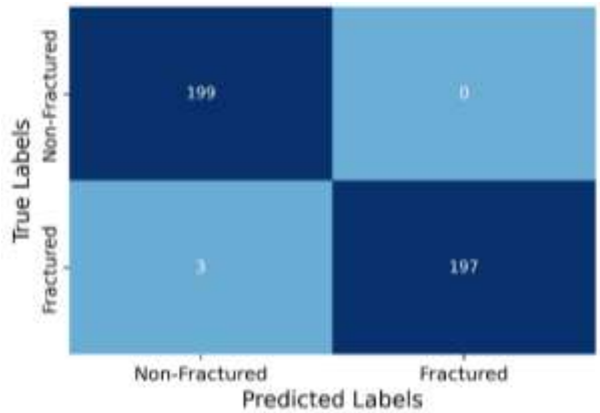


Figure 3. Confusion Matrix

According to the confusion matrix in Figure 3, the performance of the model is based on various evaluation metrics including accuracy, precision, recall, and F1-score. The model in this study produces excellent performance for each evaluation metric, reaching 99%. This shows that the model is very capable of distinguishing between fractured and non-fractured categories.

Conclusion

Based on the analysis results, the CNN model in this study shows excellent performance as shown by the evaluation metrics on testing data reaching 99%. Therefore, the model used can identify X-ray results as fractured and non-fractured so that it can be used to assist medical personnel in providing diagnoses to patients to minimise errors due to human error. Suggestions that can be used for future research are the model in this study can be trained on a more extensive and diverse dataset by applying the data augmentation process and tunning parameters to be more optimal.

References

Abdou, M. A. (2022). Literature review: Efficient deep neural networks techniques for medical image analysis. *Neural Computing and Applications*, *34*(8), 5791–5812.

Almigdad, A., Mustafa, A., Alazaydeh, S., Alshawish, M., Mustafa, M. B., & Alfukaha, H. (2022). Bone Fracture Patterns and Distributions according to Trauma Energy. *Advances in Orthopedics*, *2022*(1), 8695916. https://doi.org/10.1155/2022/8695916

Ashraf, A. (2023). *Bone Fracture Classification Using CNN*. https://www.kaggle.com/code/ahmedashrafahmed/bone-fracture-classification-using-cnn

Gillman, C. E., & Jayasuriya, A. C. (2021). FDA-approved bone grafts and bone graft substitute devices in bone regeneration. *Materials Science and Engineering: C*, *130*, 112466. https://doi.org/10.1016/j.msec.2021.112466

Jozdani, S. E., Johnson, B. A., & Chen, D. (2019). Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine Algorithms for Object-Based Urban Land Use/Land Cover Classification. *Remote Sens*, *11*(14), 1713. https://doi.org/10.3390/rs11141713

Kamil, M. Y. (2021). A deep learning framework to detect Covid-19 disease via chest X-ray and CT scan images. *International Journal of Electrical and Computer Engineering*, *11*(1), 844–850. https://doi.org/10.11591/ijece.v11i1.pp844-850

Kaur, S., Singla, J., Nkenyereye, L., Jha, S., Prashar, D., & Joshi, G. P. (2020). Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives. *IEEE Access*, *8*, 228049–228069.

Kumara, B., Koti, M. S., & Azami, N. I. A. (2024). Facial Recognition Using Convolutional Neural Network Using Real-Time Data. *Journal of Data Science2024*, *10*, 1–7.

Mardianto, M. F. F., Yoani, A., Soewignjo, S., Putra, I. K. P. A., & Dewi, D. A. (2024). Classification of Pneumonia from Chest X-ray images using Support Vector Machine and Convolutional Neural Network. *International Journal of Advanced Computer Science & Applications*, *15*(6), 1015. https://doi.org/10.14569/ijacsa.2024.01506104

Marongiu, G., Dolci, A., Verona, M., & Capone, A. (2020). The biology and treatment of acute long-bones diaphyseal fractures: Overview of the current options for bone healing enhancement. *Bone Report*, *12*, 100249. https://doi.org/10.1016/j.bonr.2020.100249

Naidu, G., Zuva, T., & Sibanda, E. M. (2023). A review of evaluation metrics in machine learning algorithms. *Computer Science On-line Conference*, 15–25. https://doi.org/10.1007/978-3-031-35314-7_2

Nanni, L., Ghidoni, S., & Brahnam, S. (2021). Ensemble of convolutional neural networks for bioimage classification. *Applied Computing and Informatics*, *17*(1), 19–35.

Ratna, S., Muflih, M., Budiman, H., Syapotro, U., & Hamdani, M. (2024). Comparison of ELM, LSTM, and CNN Models in Breast Cancer Classification. *Journal of Data Science*, *2024*(55).

WHO. (2024). *Fragility fractures*. https://www.who.int/news-room/fact-sheets/detail/fragility-fractures

Witten, I. H., & Frank, E. (2017). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. Elsevier Inc.

Xie, Y., Zaccagna, F., Rundo, L., Testa, C., Agati, R., Lodi, R., Manners, D. N., & Tonon, C. (2022). Convolutional neural network techniques for brain tumor classification (from 2015 to 2022): Review, challenges, and future perspectives. *Diagnostics*, *12*(8), 1850. https://doi.org/10.3390/diagnostics12081850