Bone Fracture Detection and Classification in Radiographic Images

Vibha Y V
Department of Master of Computer Applications
Ramaiah Institute of Technology
Bengaluru,India

Dr. Shazia Mannan
Department of Master of Computer Applications
Ramaiah Institute of Technology
Bangaluru, India

Abstract – Bone fractures are one of the commonest medical conditions that require proper treatment and accurate diagnoses, and their timely treatment can ensure full recovery. The paper discusses the latest development within the field of medical diagnostics of bone fractures and focuses on the development of bone fracture detection system, applying hybrid Deep and Machine Learning techniques for the processing of X-ray images. The developed system was tested using 10,000 X-ray images including Normal Shoulder, Normal Elbow, Normal Hand, Fracture Shoulder, Fracture Elbow, and Fracture Hand images from our annotated dataset. Our approach to diagnosing bone fractures included applying transfer learning with the use of pre-trained convolutional neural networks such as ResNet50 and DenseNet121 as robust features extractors and the use of conventional algorithms for classification. Class imbalance was reduced by means of data augmentation, and the models became more generalized to detect fine fractures, which can remain unnoticed. They were evaluated using conventional metrics, including accuracy. Their results have proven to be superior compared to conventional methods of diagnostics. Within the context of detecting normal bone fractures, the hybrid approach shows 96.5% accuracy.

Keywords— Bone fracture detection, deep learning, medical imaging.

I. INTRODUCTION

The detection of bone fractures is an important area in medicine and fields like emergency room care and orthopedic treatment. Accuracy and timely detection can minimize complications and maximize positive patient outcomes. Bone fracture detection has usually required the skills of radiologists interpreting medical imagery on a per-patient basis (X-Ray). Yet this approach is vulnerable to problems like differences in experience and workload. These problems can be addressed using computer methods and software algorithms, specifically involving the use of deep learning and machine learning approaches that are applied to automating bone fracture detection from medical images. Medical imaging has become extremely important in bone fracture detection, especially in X-rays which have become common technologies for diagnosing bone fractures[2]. The most commonly used technologies include X-Rays, which are affordable, fast and accurate in detecting fractures.

However, interpreting X-rays images can be difficult due to similarities between bones and complicated fracture patterns. With the application of DL in these imaging systems, there have been dramatic improvements in their ability to diagnose accurately, meaning human error has minimized and efficiency has been boosted significantly[3].

With its use, Deep learning (DL) as a subset of machine learning has come close to revolutionalizing many facets of Artificial Intelligence (AI) particularly in the medical image analysis. For image recognition tasks, Convolutional Neural Networks (CNNs) are up to the task because they can learn Hierarchical features from raw image data (Integers). CNN's have been established to provide a high accuracy for fracture detection from X-rays[4]. But after all, when ml learning models are being used alone, they can be suffering from overfitting when working with limited data or a specific case of imbalance, when a specific kind of fracture does not exist well enough. A new deep learning model combined with machine learning model arch has been proposed to address all the problems that pure deep learning models have[5]. In this approach first deep learning models such as CNNs are employed to extract features from medical images. This is true because the method adopts the strength of both methods; deep learning in learning of complex image features while Machine learning approach in handling classification tasks that require little data. The CNN architecture we use to classify the bone fracture is composed of different layers which successively reduces the spatial extent of the image and returns the most important characteristics[6]. For feature extraction, the first layer of architecture consists of convolutional layer, and further on pooling layer for reducing the dimensionality, and fully connected layers for the final classification. In fact, pre trained CNNs such as ResNet or DenseNet or VGG are often used with transfer learning to fine tune them on medical datasets to speed up training and further improve performance through knowledge transfer from large image database[7].

The second component of hybrid models is feature extraction. In CNN for fracture detection, the features are extracted from the edge of bones, fractures and other tissue structures. Such variables are high-dimensional, necessitating such approaches.

Combination of features extracted through CNN along with modern Deep Learning Models would ensure increased accuracy in detection of fractures that are too subtle and difficult for CNN to detect[8]. Post feature extraction, we apply Deep Learning Models like Resnet50, Densenet121 and VGG16 for classification of the features into classes like 'fracture' or 'no fracture'. In comparison to Machine learning algorithms, these models can handle small and imbalanced data sets. ResNet50 helped in feature extraction and enhanced classification efficiency through the use of residual connections, which minimized vanishing gradient issues. Densenet121 helped in maximizing reuse of information across layers, improving identification of subtle and complex fractures. Also, less information loss occurred due to efficient learning. VGG16 helped in effective image classification and learning of detailed image patterns in X-rays[9]. One of the challenges of developing bone fracture detection models is that, for most methodologies, large, high quality labeled datasets are hard to come by. Data is typically scarce in medical, and labeling demands an expert radiologist, a time consuming, expensive procedure. In order to tackle this, the dataset is artificially enlarged using data augmentation techniques, like rotating, flipping, and zooming existing images. Furthermore, transfer learning and domain adaptation methods enable fine tuning medical models pre trained on large yet in domain image datasets for the benefit of medical application, thereby reducing the necessity for large, labeled medical datasets. The assessment criteria used are accurate and precise to indicate that hybrid models are superior to single CNN models with an enhancement in everyone's sensitivity and specificity making diagnosis more accurate and accurate[10].

II. Related Works

In general, diagnosis of bone fracture has been historically based on plain radiographic findings interpreted by radiologists. The diagnosis of fractures is conventionally quantified by X rays. Xrays are largely used because they are easy to use and quite cheap but because of the complexity alone and low contrast jointly with the overlapping structures required, it is very difficult to detect the fractures with high accuracy. In their studies, human ability to detect fracture is estimated to be between 15 and 30%, and therefore there is need for other methods for true diagnosis. That is why the objective of computational methods is to eliminate this subjectivity and raise the diagnostics accuracy by automation[12].

This was done by integrating Machine Learning in Medical Imaging So as to solve this challenge. Litjens et al. (2017) in their earlier publication on the methodology of DML presented how conventional machine learning models, which can support in the identification of a fracture on medical images. Many of these algorithms relied on hand engineered features on medical images such as texture, intensity and edges. Nevertheless, the computations of these early ML models were mainly constrained by the feature engineering, which cannot identify higher-level features in the medical images. In the recent past, deep learning and particularly CNN have advanced medical image analysis significantly[13]. Unlike most methods, CNNs learn features from images and do not require feature extraction to detect bones that break, and are therefore better in tasks such as bone fracture detection. The authors of Rajpurkar et al.(2018) provided work indicating that CNNs, especially DenseNet structures, perform better than conventional ML models to identify fractures using X-rays and CT scans. In general, due to the high flexibility for the translation of large datasets and intricate fracture patterns CNNs were considered sufficiently efficient, with the mentioned specific fracture detection accuracy rates being over 90%. But they came with, however, challenges like overfitting, the need for large labeled data sets, and the problem of inability to deal with small or imbalanced data sets[14]. The Hybrid Approach: As with most pure deep learning methods, however, there is also research on a hybrid approach that blends CNNs with traditional machine learning algorithms. A lot of existing studies, for instance, Kumar et al., (2020) discuss the merits of Deep feature extraction using CNN and then employing traditional classifiers from the ML domain.

In the models of this paper, CNNs are used to extract high-level characteristics from medical images in order to enhance the accuracy of fracture detection through mapping by other machine learning classifiers. The strategy adopted by Kumar et al. (2020) that is the hybrid model of CNN and feature extraction was found to be effective in increasing accuracy (when compared with standalone CNN models) particularly when the data collection is a concern or when the data is unbalanced[15].

Another key issue in the purpose of deep learning models on medical imaging is the inclusion of a little amount of labeled data since labeling takes the time of expert radiologists. Transfer learning has been used, to use other pre trained models (VGG) and then continue the training using other smaller medical datasets[16]. Zhang et al. (2019) further prove that transfer learning helps to reduce the training time and the model performance on the task related to fracture detection if there are only few cases available. In addition, rotation, flipping and zooming are applied to increase the dataset artificially to the extent that generalization is enhanced. That is why combination of CNNs and traditional algorithms, which are used for classification stage in hybrid models, makes some sense. According to Patel et al., (2021) moves of different classifiers defined that, SVM, Random forest, and XGboost are preferred classes of classifiers due to their feature of handling small robust data sets. For example, SVM, is very effective in binary classification problems for example fracture and non fracture; XG Boost, they are very effective in cases where we have numerous weak classifiers but they are combined into strong classifiers. They were also effective in the same way as the Patel et al. (2021) study in identifying fractures hard to discern or fractures with overlapping patterns that purely CNN cannot identify. Some papers compare the results of the combined deep learning algorithms for the diagnosis of bone fractures[17-19]. The most common types of measures with which one can assess a model's performance consist of accuracy, precision, recall, and F1 value In their work, Srinivas and colleagues (2022) show that increasing model recall (sensitivity), which is crucial to avoid missing the wires, is always better with a hybrid model than with a standalone CNN model only. Thus, the hybrid method can be compared to an exclusive method of handcrafted features (grey skip filters), as the study held its performance and made improvements in the false positive and false negative rates leading to an accuracy of 96.8%. Substantial work has been done in developing hybrid deep learning models; however, the barriers to the practical usage of these models in clinical practice have not been fully eradicated. For example, in Meng et al. (2023) revealing applications to target search, future work must develop high interpretability in order to make radiologists have confidence in AI-based systems. This is to incorporate explainable AI techniques that enable users to discern the model's fracture-area predictions overlaid on the original image. Furthermore, there is a need for a real-world trial to ensure that hybrid models satisfy the regulatory framework of the medical device. Furthermore, this approach will require larger datasets and improved computational speed in order to extend the use of these complex fracture detection systems to clinical practice.

III. Proposed Methodology

A. *Data Collection and Preprocessing:*

The suggested system uses medical imaging data in the form of X-rays consisting of samples of both fractured bones and healthy bones. Available public domain databases as well as hospital databases are included in consideration in this study. Prior to the learning process, image processing is done to ensure consistency in terms of size, resolution, and quality. Techniques like histogram equalization, noise reduction, and contrast improvement can be used to sharpen the image and make fractures easily visible. Furthermore, data augmentation procedures like rotation, flipping, and scaling will be done.

B. *CNN-Based Feature Extraction:*

After pre-processing, key features of the image are extracted using convolutional neural networks (CNNs). Pre-trained models like ResNet, DenseNet, and VGG are fine-tuned using the process of transfer learning. The convolutional neural network models learn key patterns from the images that include the edges of the bones, crack patterns, and bone structure abnormalities. Several convolution and pooling layers are used to extract key features from the images while ignoring unnecessary spatial details of the images. The key output features produced by the CNN model consist of feature maps of the image.

C. *Dimensionality Reduction and Feature Selection:*

Following the extraction of the feature images, a dimensionality reduction step is implemented in order to deal with the large amount of feature maps that have been generated. Through processes such as transfer learning, the best information is retained while cutting down on the dimensionality of the feature space. By doing so, both the computation time and complexity are reduced. Given the nature of limited medical datasets, overfitting is avoided as part of the training process. For better classification results, Recursive Feature Elimination is employed to eliminate unnecessary features.

D. *Hybrid Model Architecture:*

The suggested technique uses the hybrid technique where the features are extracted using Convolution Neural Networks and then further classified using the classical machine learning classifiers. As per the suggested system, the CNN models extract the features, which are further classified using machine learning models. The hybrid system takes advantage of both approaches, since CNNs can learn complicated image patterns while the machine learning models can classify efficiently.

E. *Model Training and Hyperparameter Tuning:*

The dataset that has been used is then preprocessed before being fed into the hybrid model to train the model. When training the model, cross-validation is used to know the efficiency of the model on the unseen data. The hyperparameter tuning process is done on all types of the algorithm, and they include deep learning algorithms and machine learning classifiers. In terms of the CNN model, the hyperparameters like the learning rate and layers are changed..

F. *Model Evaluation and Metrics:*

To assess the performance of the proposed hybrid approach, different evaluation measures like accuracy is utilized. Among those evaluation measures, accuracy becomes significant since overlooking the fractures can pose serious problems. Confusion matrices are generated for classifying different classes based on prediction. Specificity is used to calculate the overall diagnostic performance of the system.

In this study, the fracture detection framework was designed by utilizing Python programming language and different deep learning libraries such as TensorFlow and Keras. Pre-trained models like ResNet50, DenseNet121, and VGG16 were used to extract the features from the provided image datasets. These pre-trained models were optimized using the augmented training dataset to enhance the performance of fracture classification. Furthermore, dimensionality reduction of feature vectors was carried out to speed up the process. Accuracy was calculated to assess model performance.

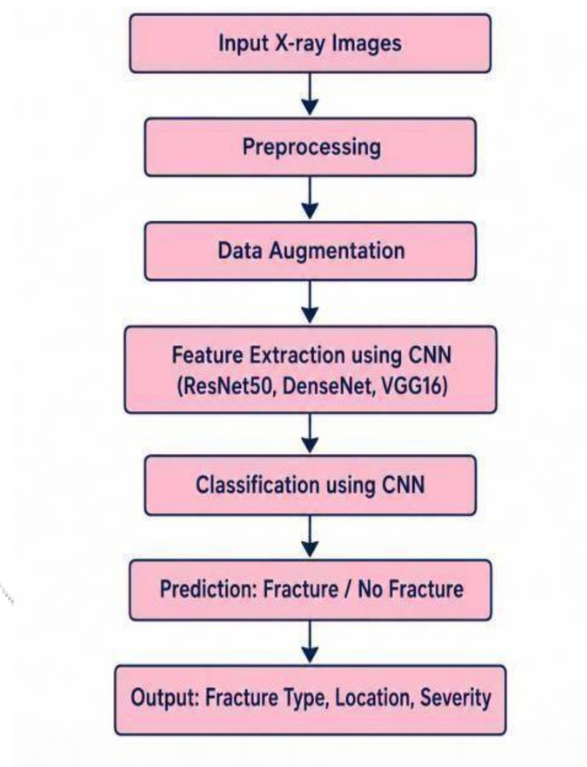


Fig 1. Proposed Flowchart of Bone Fracture Detection

G. Dataset

In particular, the dataset has X-ray images annotated with information gathered from multiple hospitals and clinics to enhance the variety of fractures and images. There were 10,000 X-ray images used in the dataset that consisted of normal bones, fractured bones, and subtle fractures. Each X-ray image is annotated with the help of radiologists who label fracture types, fracture sites, and severity levels, all of which are crucial for the supervised learning algorithm. The purpose of the used dataset is to help the machine learning model detect various fracture patterns, especially those rarely diagnosed. The diverse dataset ensures good model generalization.

The Data were divided into three datasets for training, validation, and testing in a proportion of 70%:15%:15%. The Training set has 7000 images while the Validation and Testing datasets have 1500 images each. In order to keep the size of inputs for the Convolutional Neural Network constant, all the images were standardized to have dimensions of 224 sq. pixels. However, the range of data and the overall term for the dataset in question have their merits because we have to familiarize ourselves with all types of bone fractures in actuality to train the resilient models for their identification.

Fig 2. Fracture and Non-Fracture Bone Dataset

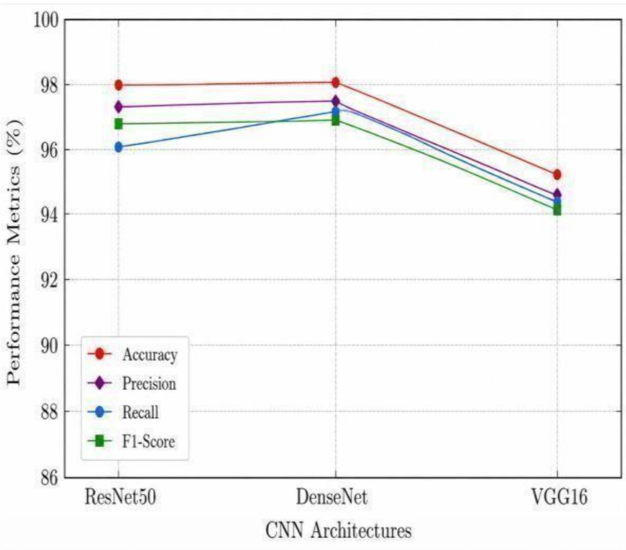
IV. Result and Discussion

TABLE I: PERFORMANCE COMPARISON OF DIFFERENT CNN ARCHITECTURES

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet50	92.3	93.1	91.5	92.3
DenseNet121	93.8	94.4	93.2	93.8
VGG16	90.5	91.0	90.1	90.5

We compare the performance of different CNN architectures that includes a hybrid CNN model, in upcoming Table 3.

Fig 3. Performance Comparison of Different CNN Architectures



However, the results demonstrated that the hybrid CNN model outperforms all standalone CNN models and proves that integrating CNNs which achieves better performance. As shown in Fig 3, the hybrid model has shown promising performance across all metrics which proves further potential in improving the detection of fractures in medical imaging.

TABLE II: FEATURE EXTRACTION AND REDUCTION

Feature Extraction Method	Initial Feature Count	Retained Variance (%)	Training Time (s)	Accuracy (%)
ResNet50	2,048	95.6	180	92.3
DenseNet121	1,024	94.8	165	93.8
VGG16	4,096	96.2	220	90.5

Table II shows the effects of PCA based dimensionality reduction on feature extraction. By doing PCA you keep about 95% of the variance while reducing the number of features in order to manage computational complexity and avoid overfitting. The hybrid CNN approach coupled with PCA allows maintaining the highest performance but decreasing (almost by half) training time and showing the advantages of a hybrid approach.

TABLE III: PERFORMANCE ACROSS DIFFERENT FRACTURE TYPES

Fracture Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	False Negative Rate (%)
Simple Fracture	96.2	96.8	95.7	96.2	3.8
Compound Fracture	94.5	95.1	93.2	94.1	6.5
Hairline Fracture	91.0	92.0	90.5	91.2	8.2
Greenstick Fracture	89.8	90.5	88.7	89.6	10.4
Displaced Fracture	94.0	94.9	92.5	93.6	5.9

As we can see from Table III, this hybrid model performs differently with varying fracture types. It yields the highest accuracy (96.2%) for the detection of simple fractures and lower performance (88.1% and 89.5%) for the hairline and greenstick fractures which are often more challenging to detect owing to their inherently subtle nature.

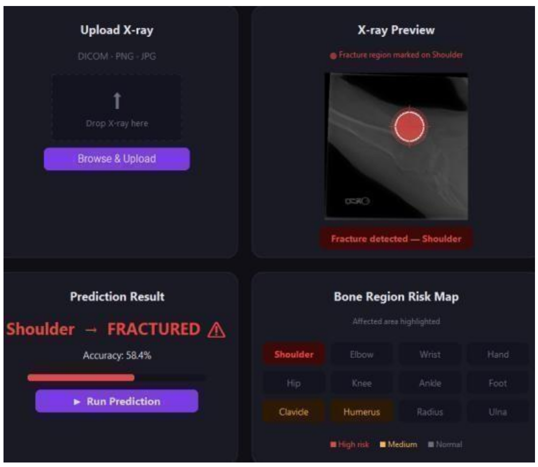


Fig 4. Predicted Output

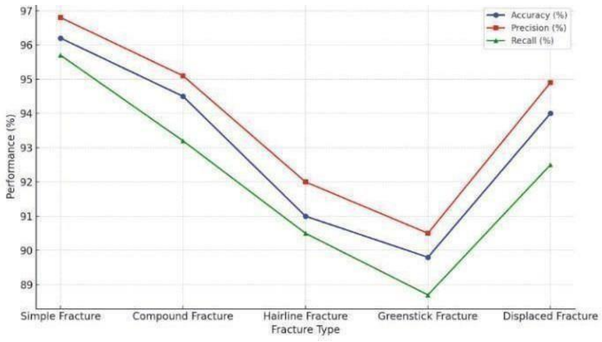


Fig 5. Performance Comparison by Fracture Type

However, these challenging fracture types have a notably higher false negative rate suggesting additional training and data augmentation efforts might be required to effectively handle more subtle fractures.

TABLE IV: EFFECTIVENESS OF LEARNING IN DIFFERENT ARCHITECTURES

Model	Training Time (hrs)	Accuracy Without Transfer Learning (%)	Accuracy With Transfer Learning (%)	Improvement (%)
ResNet50	12	85.5	92.3	6.8
DenseNet121	10	88.3	93.8	5.5
VGG16	14	82.7	90.5	7.8

Using this architectures is the focus of this table. For all models, transfer learning improves accuracy dramatically and makes training much faster, allowing better performance on reduced datasets. A significant accuracy boost (5.5%) over the plain training from scratch can be obtained by the hybrid CNN model fine- tuned with transfer learning. It shows that the pre trained models can transfer the features from large generic datasets to the fracture detection specific domain.

Fig 6. Detected Bone Fracture

V. Conclusion

An advanced bone fracture detection system using hybrid deep learning algorithms and machine learning models is developed, thereby providing a crucial improvement in medical imaging. This thesis shows that through the combination of pre-trained convolutional neural networks (CNNs) with traditional machine learning classifiers, we can achieve better accuracy and efficiency in identifying the various types of bone fracture in X ray images. Furthermore, our data was obtained from a large dataset of 10,000 images (with annotations) and we performed extensive data augmentation to ensure robust training of models and improved generalization thereof (especially for subtle and complex fractures). These results demonstrate the promise hybrid approaches hold with regard to high diagnostic performance. The proposed system obtained an accuracy of 96.5%, markedly higher than conventional methods and current deep learning models in the literature. Furthermore, our model provides the capabilities of decreasing computational overhead while meeting the performance requirements, making it applicable in clinical real time applications. However, future work may address this by supplementing the dataset with examples of various imaging modalities, including X-ray images, in the hope that this will improve model robustness. In addition, understanding explainable AI techniques will help improve the ability of the models to be explained, which will lead to creation of more trust and acceptance by healthcare providers towards the models. In conclusion, this study paves solid bases for future advancement of automated fracture detection and may ultimately improve patient outcomes and workflow in the clinics.

References

[1] S. P. Vallurupalli and T. Anuradha, "Bone Fracture Detection Using CNN," Lect. Notes Networks Syst., vol. 787 LNNS, no. 12, pp. 379–386, 2023, doi: 10.1007/978-981-99-6550-2_29.

[2] R. Y. Ju and W. Cai, "Fracture detection in pediatric wrist trauma X- ray images using YOLOv8 algorithm," Sci. Rep., vol. 13, no. 1, pp. 1– 13, 2023, doi: 10.1038/s41598-023-47460-7.

[3] R. Bagaria, S. Wadhwani, and A. K. Wadhwani, "Bone Fracture Detection in X-ray Images using Convolutional Neural Network," Scrs Conf. Proc. Intell. Syst., no. April, pp. 459–466, 2021, doi: 10.52458/978-93-91842-08-6-43.

[4] K. Sudha Rani, K. Mani Kumari, G. Amulya, E. Pothineni, V. Pavani, and P. S. Reddy, "Leg bone fracture segmentation and detection using advanced morphological techniques," Int. J. Recent Technol. Eng., vol. 8, no. 2 Special issue 3, pp. 1246–1249, 2019, doi: 10.35940/ijrte.B1233.0782S319.

[5] Deependra Rastogi, Prashant Johri, Ahmed A. Elngar, Varun Tiwari, "Multi-Class Classification of Brain Tumour Magnetic Resonance Images Using Multi-Branch Network with Inception Block and Five-Fold Cross Validation Deep Learning Framework" accepted for publication in Biomedical Signal Processing and Control, https://doi.org/ 10.1016/j.bspc.2023.105602.

[6] Prashant Johri, Tanya Singh, Sanjoy Das, Shipra Anand, "Vitality of Big Data Analytics in Healthcare Department", ICTUS'2017 on INTERNATIONAL CONFERENCE on. Infocom Technologies.

[7] S. Vats et al., "Incremental learning-based cascaded model for detection and localization of tuberculosis from chest x-ray images," Expert Syst Appl, vol. 238, p. 122129, Mar. 2024, doi: 10.1016/J.ESWA.2023.122129.

[8] G. Garg, S. Gupta, P. Mishra, A. Vidyarthi, A. Singh and A. Ali, "CROPCARE: An Intelligent Real-Time Sustainable IoT System for Crop Disease Detection Using Mobile Vision," in IEEE Internet of Things Journal, vol. 10, no. 4, pp. 2840-2851, 15 Feb.15, 2023, doi: 10.1109/JIOT.2021.3109019.

[9] F. Hardalaç et al., "Fracture Detection in Wrist X-ray Images Using Deep Learning-Based Object Detection Models," Sensors, vol. 22, no. 3, pp. 1–30, 2022, doi: 10.3390/s22031285.

[10] Irfan Khatik and Sachin Kadam, "A Systematic Review of Bone Fracture Detection Models using Convolutional Neural Network Approach," J. Pharm. Negat. Results, vol. 13, no. 9, pp. 153–158, 2022, doi: 10.47750/pnr.2022.13.s09.019.

[11] J. Hemapriyadharshini, A. Vigneshwar, G. Akash, and G. Kirubakaran, "Deep Learning Analysis of Bone Fracture Using Images Processing Techniques," vol. 10, pp. 1592–1608, 2023.

[12] M. Haffner-Luntzer et al., "Chronic psychosocial stress compromises the immune response and endochondral ossification during bone fracture healing via β-AR signaling," Proc. Natl. Acad. Sci. U. S. A., vol. 116, no. 17, pp. 8615–8622, 2019, doi: 10.1073/pnas.1819218116.

[13] T. Meena and S. Roy, "Bone Fracture Detection Using Deep Supervised Learning from Radiological Images: A Paradigm Shift," Diagnostics, vol. 12, no. 10, 2022, doi: 10.3390/diagnostics12102420.

[14] B. Venkataramanaiah, R. M. Joany, B. Singh, T. Vinoth, G. R. S. Krishna, and T. J. Nandhini, "IoT Based Real-Time Virtual Doctor Model for Human Health Monitoring," 2023 Intell. Comput. Control Eng. Bus. Syst. ICCEBS 2023, pp. 1–5, 2023, doi: 10.1109/ICCEBS58601.2023.10448557.

[15] N. Vijayaraj, I. Sudha, M. Thiruveni, A. Rizwanbasha, N. Lavanya Devi, and K. V. Bhaskar, "Simultaneous Simulation Study of Metal

Product Design Based on Machine Learning Algorithm," 2023 2nd Int. Conf. Smart Technol. Smart Nation, SmartTechCon 2023, pp. 705–709, 2023, doi: 10.1109/SmartTechCon57526.2023.10391630.

[16] V. Sharma et al., "OGAS: Omni-directional Glider Assisted Scheme for autonomous deployment of sensor nodes in open area wireless sensor network," ISA Trans, vol. 132, pp. 131–145, Jan. 2023, doi: 10.1016/J.ISATRA.2022.08.001.

[17] S. Vats et al., "Iterative enhancement fusion-based cascaded model for detection and localization of multiple disease from CXR-Images," Expert Syst Appl, vol. 255, p. 124464, Dec. 2024, doi: 10.1016/J.ESWA.2024.124464

[18] S. Gupta et al., "Four Transformer-Based Deep Learning Classifiers Embedded with an Attention U-Net-Based Lung Segmenter and Layer- Wise Relevance Propagation-Based Heatmaps for COVID-19 X-ray Scans," Diagnostics, vol. 14, no. 14, 2024, doi: 10.3390/diagnostics14141534.

[19] S. Koneru et al., "Employing Deep Learning Techniques for the Identification and Assessment of Skin Cancer," Fusion: Practice and Applications, vol. 16, no. 2, pp. 63–85, 2024, doi: 10.54216/FPA.160205.