Architecture
Image Studies

Vol.6, Issue 3, pp.1766-1784, 2025
DOI: https://doi.org/10.62754/ais.v6i3.518

https://journals.ap2.pt/index.php/ais/index

Medical Image Classification Using Hybrid Deep Learning Methods and Advanced Preprocessing Using Wrist X-ray Images

Vipul.A.Shah[1], Hemantika Prabhatsinh Chauhan[2]

Abstract

The most common kind of crack, and one that happens rather often, is a wrist fracture. Even though X-ray medical imaging is useful for identifying wrist fractures, the depiction of these breaks may sometimes provide problems. These days, there's a lot of hope that artificial intelligence (AI) might help orthopaedic X-ray interpreters make better, faster fracture diagnoses. The purpose of this investigation is to use “deep learning” (DL) to assist physicians, particularly those who work in emergency departments, in diagnosing wrist X-ray fractures. To build a reliable and effective system that classifies wrist X-ray abnormalities, this study used a MURA (XR_WRIST) dataset from the Stanford Machine Learning Group website. To obtain higher accuracy, the data was preprocessed using various key steps, initially resizing the images and converting them to greyscale; CLAHE was implemented. To mitigate minor noise, Gaussian blur is implemented, and the "haar" wavelet is employed in the discrete wavelet transform (DWT) at level 1 to further reduce noise by zeroing high-frequency coefficients. The model's robustness is ensured by using a 3-fold cross-validation strategy. For classification purposes, the VGG19_AlexNet_AdaptiveHybrid_CNN model is subsequently implemented. Model performance has been assessed using metrics such as the confusion matrix, “Cohen's kappa”(CK), “F1-score”(f-measure), “sensitivity”(Sn), “specificity”(Sp), “recall”(recl), “accuracy”(acc), “precision”(prec), and ROC-AUC. The comparative results of existing and proposed DL models demonstrate thatproposed VGG19_AlexNet_AdaptiveHybrid_CNN model has the highest accuracy99.81%, precision99.9%, recall99.81% and f1-score99.86%. These outcomes show that a proposed model is highly reliable in wrist X-ray abnormalities detection.

Keywords: *Bone Fracture Detection, Hybrid Deep Learning, Wrist X-ray Abnormality Detection.*

Introduction

Wrist fractures and deformities comprise a considerable number of patients with musculoskeletal conditions, especially if patients experience falls or trauma to the upper limbs. This is important when it comes to treating, rehabilitating and helping the patient get back on their feet as soon as possible (Hardalaç et al., 2022). Conventional identification of wrist fractures, joint dislocations and other abnormalities use the services of a radiologist who interprets the analysis from X-ray images himself (Bauskar & Clarita, 2020)(Narayan et al., 2023). It might be difficult to analyse X-ray images to find bone fractures when specialists are not easily accessible (Bowers, 2011)(Javed et al., 2023). Outcomes may be affected if this causes missed diagnosis and postpones treatment. Due to the common occurrence of diagnostic mistakes, it is essential to promptly and accurately identify fractures (Khare et al., 2024).

AI is a new and exciting development in medical imaging (Mustafa & Alneamy, 2022), with the potential to enhance orthopaedic X-ray fracture diagnosis (Chan et al., 2020). AI systems, driven by deep learning methods, are able to look for patterns in masses of data; this allows for very accurate automated fracture detection, localisation, and classification.

In fracture diagnosis, AI is important since it may make the procedure more efficient and reliable. Radiologists may expedite and improve the accuracy of fracture identification by using AI-powered computer-assistedparaphernalia that help in analysing X-ray pictures(Sharma, 2023). In addition, AI systems might potentially provide uniform and standardised evaluations, which can decrease variability among observers and ameliorate patient outcomes (Yang et al., 2020).

[1] Instrumentation and Control Engineering Department, Faculty of Technology, Dharmsinh Desai University,Nadiad-Gujarat.India. E-mail: vashah.ic@ddu.ac.in (corresponding author).
[2] Instrumentation and Control Engineering Department, Faculty of Technology, Dharmsinh Desai University,Nadiad-Gujarat. India.

The field of medical imaging has been one of most promising areas for DL, a branch of AI (Jabbar et al., 2022)(Arora et al., 2024). It has been increasingly functional to medicinal image analysis and can be seen as a practical and effective solution to improving the discovery of variances in these images (Kuo et al., 2022). In fact, "Convolutional Neural Networks" (CNNs)(Tandon, 2023), or class of ANNs, have much higher performance in feature extraction from images; as a result(Pranav Khare, 2022)(Pranav Khare, 2023)(Khare, 2022), they may be used for uncovering of minor changes in the X-ray images that might point to possible fractures, dislocations or other pathological states. Nevertheless, X-ray wrist completed abnormality detection is still difficult to attain high precision because of various factors; low contrast to the image, variation in anatomical structures of the hand, wrist, and the fact that some of the Wrist abnormalities are difficult to detect(Tahir et al., 2024).

Motivation and Significance

The present research study's objective is to propose and deploy an efficient and robust deep neural network that can accurately diagnose multiple anomalies through analysis of X-ray images of wrist. In medical imaging, there is an urgent need for rapid and accurate diagnostic tools; this is particularly true when it derives to X-rays of the wrist, where human processing is both laborious and prone to mistakes. By leveraging advanced DL techniques and hybrid architectures, this study aims to enhance diagnostic accuracy and reliability, supporting radiologists in making faster and more precise decisions. The implication of this work lies in its efficacy to revolutionise medical diagnostics through an innovative combination of preprocessing methods, hybrid CNN models, and robust evaluation metrics, offering a scalable and effective solution for real-world clinical applications.

Novelity and Contribution

The inventiveness of this work lies in a development of a hybrid CNN, VGG19_AlexNet_AdaptiveHybrid_CNN, which uniquely integrates the feature extraction strengths of pre-trained VGG19 and AlexNet models with additional custom convolutional layers to achieve superior performance in wrist X-ray classification. The model is complemented by an advanced preprocessing pipeline, including CLAHE, Gaussian blur, and Discrete Wavelet Transform, alongside the use of SMOTE for addressing data imbalance. This innovative combination of preprocessing techniques, hybrid architecture, and feature fusion results in exceptional accuracy and robustness, making it a significant advancement in medical image diagnostics. Here are the key contributions of this study.

- This study introduces the VGG19_AlexNet_AdaptiveHybrid_CNN model, which combines the feature extraction capabilities of pre-trained VGG19 and AlexNet architectures with additional custom layers, enabling highly accurate wrist X-ray classification.
- The research implements an enhanced preprocessing technique, including CLAHE for contrast enhancement, Gaussian blur, and Discrete Wavelet Transform (DWT) for denoising, ensuring high-quality input images for model training.
- The study utilises SMOTE to resolve class imbalance in dataset, contributing to more reliable and unbiased model predictions.
- The model's potentiality to detect complex patterns in medical pictures is improved by fusing features from VGG19 and AlexNet using Global Average Pooling and then further refining them with additional convolutional layers.
- To provide ainclusive picture of model's usefulness, study assesses it using a wide range of measures, such as recl, acc, prec, Sn, Sp, ROC AUC, Cohen's Kappa(CK), and F1 score.
- Model's high accurateness and solidity make it apposite for real-world deployment in diagnostic workflows, potentially aiding radiologists in identifying wrist abnormalities more efficiently.

The following paper is organised as: Sections II provide literature review on this topic, Section III discussed the proposed methodology with each process. Discussion of comparative analysis and experimental findings are included in Section IV. Section V wraps up study by talking about possible future work and any restrictions.

Literature Review

Research using open-source picture datasets gathered from a variety of medical equipment for the purpose of fracture identification is available in the literature. In this study, the classification of X-ray imageries of wrist fractures in MURA dataset was carried out i using built deep learning models, new models thereof adapted. Therefore, main topics have been taken into account while examining the literature regarding the study.

El-Saadawy et al. (2021) suggested approach by integrating the GNG network with eight custom-built models that drew inspiration from the VGG model to maximise speed while minimising computations. The biggest publicly available collection of bone X-ray images, the MURA database, has been used in experiments. Results from the first stage averaged 99.63% specificity and 95.86% sensitivity. Stage two achieved an ideal average Sn of 92.50% and an optimum averageSp of 92.12%.

Du et al. (2022), the open mura data set's bone images are analysed using the adaptive contrast restriction approach, which enhances the visibility of the focus point and bone edge. This research presents a Densenet- and perception-based IDNet network for medical picture recognition: For classification job of upper limb bone components, IDNet achieved an accuracy of 0.99, while for the task of upper limb bone damage classification, it obtained 0.92.

Singh (2022), suggested ComDNet-512, a unique hybrid architecture based on DNN, for effective detection of bone anomalies in musculoskeletal radiographs. The ComDNet-512 pipeline consists of three stages: compression, dense neural network training, and progressive resizing. An AUC of 0.894 was attained by the model, with sensitivity being 0.941 and specificity being 0.847. With values of 0.86 prec, 0.94 recl, 0.89 f-measure, and 0.78 CK, this model outperforms all of its predecessors.

Erzen et al. (2023), provide a CAD system that can automatically identify and categorise different kinds of paediatric wrist injuries from start to finish. They use the publicly available GRAZPEDWRI-DX dataset for training and evaluating the suggested CAD system. Collecting medical X-ray data, labelling it, preprocessing it, optimising our prediction AI model, and finally evaluating it are all meticulous phases in the construction of our CAD system. Our system relies on YOLOv8, the most recent and sophisticated version of YOLO, to perform object identification and classification tasks. Experiments show that our suggested CAD system performs well in evaluations with respect to prec (77.80%), recl (54.60%), mAP@50 (59.10%), and mAP50@95 (37.20%).

Beri et al. (2024), a dataset consisting of 4906 X-ray images is utilised for study. These images are partitioned into two groups: fractured and non-fractured, respectively. The Sequential CNN model is utilised in this algorithm. The CNN model is able to attain an astounding acc of 98% in fracture classification by utilising a training set of 4099 photographs, a testing set of 401 photographs, and an additional 406 photographs for validation.

Tahir et al. (2024), advanced a multi-layered model for X-ray fracture detection. The model's preprocessing layers are ResNet50, InceptionV3, Histogram Equalisation, Vgg16, and MobileNetV2, while the feature extraction layer is a Global Average Pooling layer. Utilising a single training-validation split, an analysis is performed utilising the whole humerus from publically accessible Mura-v1.1 dataset. The suggested model scored F1 scores of 92.14%, recall of 91.62%, and accuracy of 92.96%, edging out the modified deep-learning models.

Table 1 provides a comprehensive summary of these related studies discussed below. This table summarises the methodologies, data sources, key results, research gaps, and potential future directions for each study. The research gaps primarily highlight areas where model performance could be further improved, and the future work column insinuate directions for extending the research into broader or more complex problems.

Table 1. Summary of Related Work on Wrist X-ray Abnormality Detection

Study	Methodology	Data	Results	Research Gaps	Future Work
El-Saadawy et al. [2021]	GNG network combined with 8 models inspired by VGG for bone x-ray classification	MURA dataset (bone X-ray images)	Stage 1: Sn= 95.86%, Sp = 99.63% Stage 2: Sn = 92.50%, Sp = 92.12%	Limited to upper bone areas; computational complexity of models	Explore multi-stage optimisation for reducing model size without sacrificing accuracy, and apply to other bone types

Du et al. [2022]	Adaptive contrast restriction algorithm + IDNet (perception &Densenet)	MURA dataset (bone X-ray images)	Upper limb bone injury classification: acc = 0.92 Upper limb bone part classification: Accuracy = 0.99	Focused on bone injury and parts classification only; limited exploration of edge cases in medical imaging	Extend to other anatomical regions and diseases; evaluate generalizability across diverse datasets
Singh [2022]	Hybrid deep neural network (ComDNet-512): Compression, dense network training, progressive resizing	Finger radiograph	AUC = 0.894 Sn = 0.941, Sp = 0.847 prec= 0.86, recl= 0.94, f-measure= 0.89	Limited to binary classification (normal/abnormal); could improve feature representation.	Expand to multi-class classification for various bone abnormalities and include more diverse datasets.
Erzen et al. [2023]	End-to-end CAD system using YOLOv8 for pediatric wrist injury detection	GRAZPED WRI-DX dataset (pediatric wrist X-rays)	Prec= 77.80%, recl = 54.60%, mAP@50 = 59.10%, mAP50@95 = 37.20%	Lower recall and mAP50@95 results indicate potential limitations in detecting subtle injuries	Improve recall through advanced training methods; include 3D data to enhance injury detection and classification
Beri et al. [2024]	Sequential CNN model for fracture detection	X-ray images (4906 images, fractured/non-fractured)	Acc = 98%	Focus on binary classification model performance for different bone regions not explored.	Extend to multi-class classification for different types of fractures and other musculoskeletal disorders.
Tahir et al. [2024]	Ensemble model combining MobileNetV2, VGG16, InceptionV3, and ResNet50 for fracture detection	MURA-v1.1 dataset (humerus X-rays)	acc = 92.96%, recl = 91.62%, f-measure = 92.14%	Lack of detailed performance metrics for other anatomical regions; non-fracture bone abnormalities not considered	Investigate the use of more diverse bone datasets and improve ensemble models for better performance in rare cases.

Methodology

The methodology for detecting wrist X-ray abnormalities that have been proposed commences with the acquisition of the MURA (XR_WRIST) dataset from Stanford's ML group. The dataset underwent preprocessing, which involved resizing each image to 224x224 pixels and converting it to greyscale. Then, in order to improve the finer details, CLAHE is implemented, and the pixel values are normalised to the [0, 1] range. To mitigate minor noise, Gaussian blur is implemented, and the "haar" wavelet is employed in the discrete wavelet transform (DWT) at level 1 to further reduce noise by zeroing high-frequency coefficients. SMOTE is implemented to mitigate class imbalances, and images are converted to RGB format for model compatibility following denoising. With 80% of data set set aside for training model and 20% for testing, the processed data is partitioned into an 80:20 train-to-test ratio. The model's robustness is ensured by using a 3-fold cross-validation strategy. For classification purposes, the VGG19_AlexNet_AdaptiveHybrid_CNN model is subsequently implemented. Model performance has been assessed using metrics such as the confusion matrix, CK, f-measure, Sn, Sp, recl, acc, prec, and ROC-AUC. Ultimately, the production of classification results for detection of wrist X-ray anomalies validates the model's effectiveness. The flowchart of the suggested technique for wrist X-ray abnormality identification is displayed in **Figure 1** below.

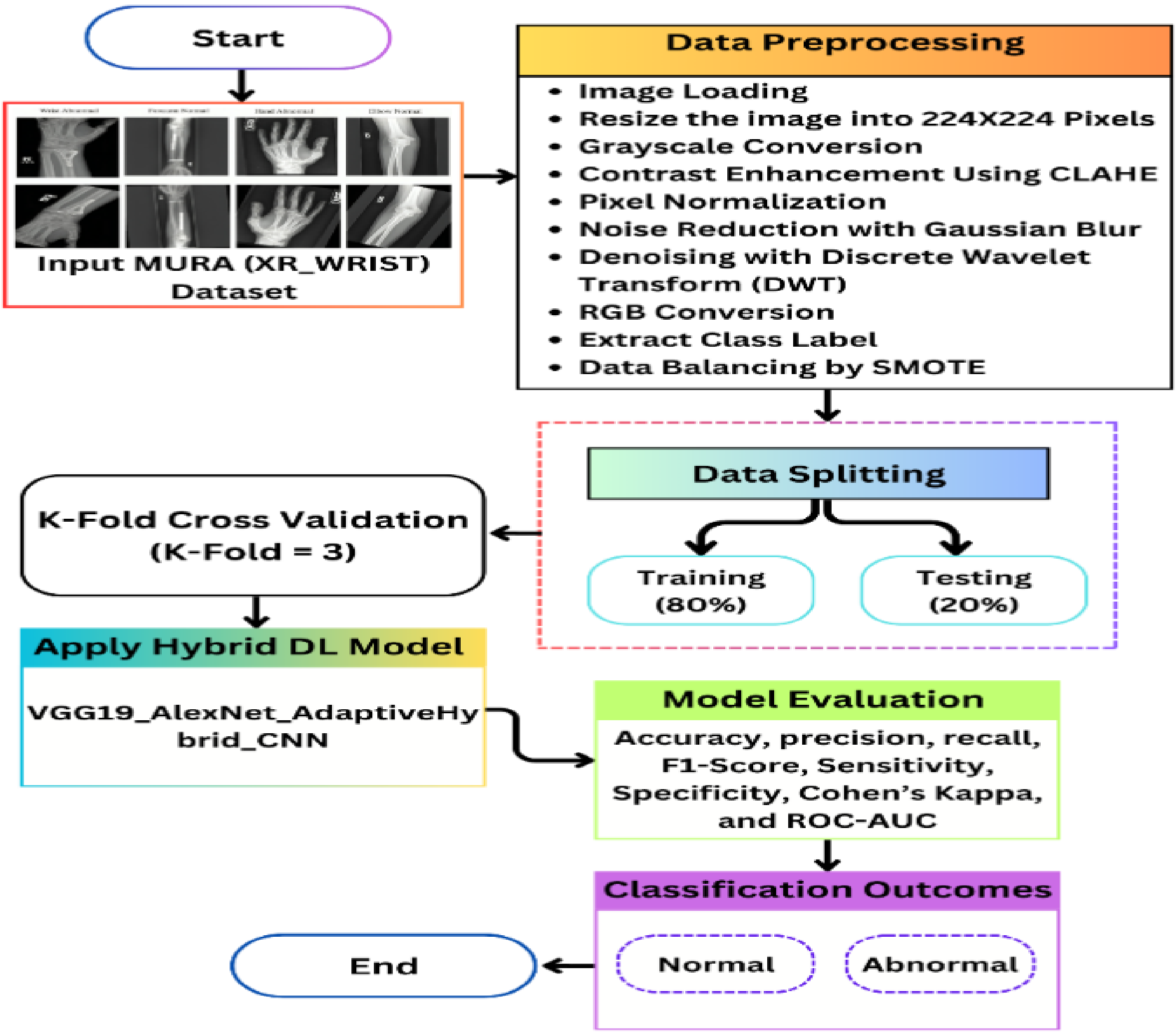

Figure 1. Flowchart of proposed methodology for Wrist X-ray Detection using DL models

Each technique of this work are discussed below;

Data Pre-Processing

XR_WRIST dataset preprocessing procedure consists of many important phases meant to improve picture quality and lower noise levels. To ensure consistency throughout the collection, each picture loads and resizes to 224x224 pixels first. The scaled photos are greyscale transformed, and CLAHE is used to boost contrast so that finer details may be seen. Pixel values are normalised to a [0, 1] range after contrast change. Then a Gaussian blur is used to lower some little noise. After that, discrete wavelet transforms (DWT) using the "haar" wavelet at level 1 denoise by setting high-frequency coefficients to zero, hence reducing noise without sacrificing the necessary picture structure. After this, balance the dataset with SMOTE techniques. At last, the denoised greyscale photos are restored to RGB format for consistency, and the treated images are kept for eventual study.

“"Contrast Limited Adaptive Histogram Equalisation (CLAHE)”

It is possible to ameliorate an image's contrast without amplifying noise using the CLAHE preprocessing approach. The CLAHE method(Bhan & Patel, 2017) is used for real-color images in this research. Individual equalisation is applied to each of the three RGB colour spaces. The color-equalized image is created by combining these equalised RGB components. The initial purpose of CLAHE was to effectively improve low-contrast medical photographs. While conventional histogram equalisation processes the whole image, CLAHE works on smaller image tiles, allowing for better preservation of localised contrast. The "contrast limiting" prevents excessive contrast in homogeneous areas, making it particularly suitable for medical images like X-rays, where subtle details are critical. This method improves the visibility of bones and abnormalities in wrist X-rays without distorting the overall image quality.

“Discrete Wavelet Transform (DWT)”

The filter-based DWT breaks down a signal into its component wavelets. DWT has a leg up on Fourier transform since it can handle both frequency and time location (Zhang, 2023). This transformation finds widespread use in several applications, including picture compression, signal processing, and noise reduction. The low-low (LL), low-high (LH), high-low (HL), and high-high (HH) sub-bands are created by DWT from the host image transform. The approximation portion is located in the LL sub-band, but the other three sub-bands hold specific information like the image's borders and textures. A multi-level decomposition may be achieved by further subdividing each sub-band into four sub-bands. In general, the LL sub-band typically has bigger DWT coefficient magnitudes than the other sub-bands. We have used a four-level DWT on the LL sub-band to deconstruct the host picture in our suggested technique. In this work, DWT aids in removing noise while preserving essential structural details in wrist X-rays, ensuring that the hybrid CNN model processes clean and high-quality images for feature extraction. This step is crucial for accurate medical diagnostics.

“SMOTE (synthetic minority oversampling technique)”

A data augmentation approach called SMOTE was developed to fix datasets that have an imbalance of classes. Using interpolation amidst existing minority samples, it creates artificial samples for the marginal class, bringing the dataset into proportion (Dablain et al., 2023). This prevents the ML model from becoming skewed towards the majority class, which is critical in medical applications where the minority class (e.g., abnormal cases) carries significant importance. By using SMOTE, the model learns more effectively and achieves higher accuracy and generalizability in classifying wrist X-rays.

A 3-fold cross-validation

The statistical technique of cross-validation involves partitioning dataset into many subsets (folds) after assess a performance of a model. The dataset is split in half lengthwise in 3-fold cross-validation; half is utilised for training the model and the other half is utilised for validation. Three iterations of this procedure are carried out, using one fold for validation. By testing the model throughout the whole dataset, this minimises overfitting and yields a more accurate performance estimate(Arlot et al., 2010). The 3-fold approach balances computational efficiency with robust performance evaluation, making it suitable for training the hybrid CNN on the MURA dataset.

Deep Learning-Based Classification Models

DL is a branch of ML that attempts to imitatorinvolvedverdict-making process of human brain via use of multilayered neural networks (Sinha, 2024). This section provides a synopsis of the classification models built using deep learning that have been utilised in a creation of hybrid models for the detection of anomalies in wrist X-rays.

VGG19

VGG19 is a CNN architecture. It is an upgrade to the VGG16 architecture with 19 layers(Medaramatla et al., 2024). The VGG19 architecture consists of nineteen layers, plus convolutional, pooling, and three dense layers. Employment of modest 3 × 3 filters in each convolutional layer increases feature representation acquisition in the architectural design. One of the main aims of the network is to decrease a spatial dimensionality of an input data and it can be achieved by including max-pooling layers to perform feature map-down sampling(Kandel et al., 2020). An input image may be mapped to a probability distribution across a predetermined range of classes using the VGG19 architecture, which can be formally defined as a sequence of nonlinear transformations.

AlexNet

In AI, AlexNet is a CNN construction that is considered unique(Thaarakaraam et al., 2024). The eight-layer architecture consists of three fully linked layers and five convolutional layers. It works by applying non-linearity with ReLU, reducing with max-pooling layers, and extracting features from input pictures using convolutional layers. Fully connected layers perform the processing of the attributes for classification after their extraction. To put an end to overfitting, washout is used. Expected class probabilities are provided by the final SoftMax layer(Noureen et al., 2023). The 2012 ImageNet competition, which AlexNet won, was a watershed moment because it proved deep learning could solve difficulties with picture recognition. Success has led to further advancements in CNN designs, which in turn have spurred research into computer vision and deep learning. The incorporation of deep architectures and ReLU into AlexNet's design has had a significant influence on subsequent CNN models and has played a pivotal role in the field's progress.

Proposed VGG19_AlexNet_AdaptiveHybrid_CNN

This paper presents the "VGG19_AlexNet_AdaptiveCNN_Hybrid_Model," a hybrid CNN architecture developed for accurate identification of wrist X-ray abnormalities. It takes the robust feature extraction capabilities of VGG19 and AlexNet, using pre-trained models without their top layers to preserve learnt features while circumventing retraining. The weights of VGG19 and AlexNet are fixed, with each model extracting unique characteristics from the input pictures, succeeded by Global Average Pooling to summarise spatial information. The combined features from both models are concatenated and processed via specialised convolutional layers that enhance feature extraction for improved flexibility. This is succeeded by fully linked layers, which include dropout layers for regularisation to mitigate overfitting. The output layer is configured for binary classification using a softmax activation function, facilitating anomaly detection. The model is fine-tuned using the Adam optimiser, running for five epochs with a batch size of fifty, using binary cross-entropy loss and learning rate of 0.0001. This method attains improved detection efficacy by integrating several feature representations and a resilient design. The internal architecture of suggested model is illustrated in Figure 2 below.

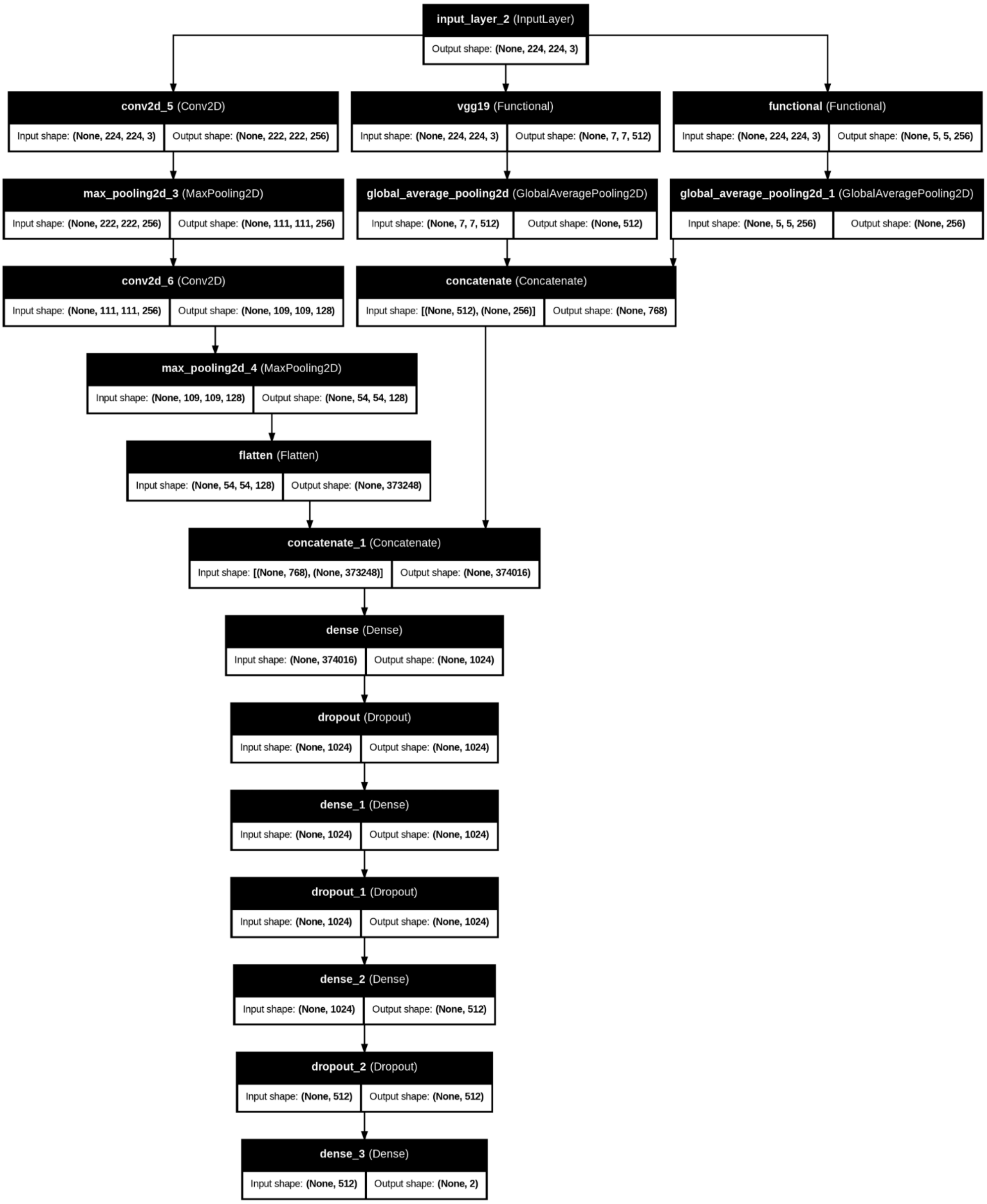

Figure 2. Building of Proposed VGG19_AlexNet_AdaptiveHybrid_CNN Model for Wrist X-ray Abnormality Detection

```
80134624/80134624 ━━━━━━━━━━━━━━━━━━━━ 0s 0us/step
Model: "VGG19_AlexNet_AdaptiveCNN_Hybrid_Model"
```

Layer (type)	Output Shape	Param #	Connected to
input_layer_2 (InputLayer)	(None, 224, 224, 3)	0	-
conv2d_5 (Conv2D)	(None, 222, 222, 256)	7,168	input_layer_2[0][0]
max_pooling2d_3 (MaxPooling2D)	(None, 111, 111, 256)	0	conv2d_5[0][0]
vgg19 (Functional)	(None, 7, 7, 512)	20,024,384	input_layer_2[0][0]
functional (Functional)	(None, 5, 5, 256)	3,747,200	input_layer_2[0][0]
conv2d_6 (Conv2D)	(None, 109, 109, 128)	295,040	max_pooling2d_3[0][0]
global_average_pooling2d (GlobalAveragePooling2D)	(None, 512)	0	vgg19[0][0]
global_average_pooling2d… (GlobalAveragePooling2D)	(None, 256)	0	functional[0][0]
max_pooling2d_4 (MaxPooling2D)	(None, 54, 54, 128)	0	conv2d_6[0][0]
concatenate (Concatenate)	(None, 768)	0	global_average_poolin… global_average_poolin…
flatten (Flatten)	(None, 373248)	0	max_pooling2d_4[0][0]
concatenate_1 (Concatenate)	(None, 374016)	0	concatenate[0][0], flatten[0][0]
dense (Dense)	(None, 1024)	382,993,408	concatenate_1[0][0]

Figure 3. Summary of Proposed VGG19_AlexNet_AdaptiveHybrid_CNN Model for Wrist X-ray Abnormality Detection

To identify abnormalities in wrist X-rays, a hybrid CNN model called "VGG19_AlexNet_AdaptiveCNN_Hybrid_Model" has been suggested (see Figure 3 for a general overview). The model incorporates an architecture that integrates layers from both VGG19 and AlexNet. Input layers handle 224x224 RGB pictures, and then feature extraction and dimensionality reduction are handled by convolutional, pooling, and functional layers. In order to combine characteristics from several phases, important layers use methods such as global average pooling and concatenation. A last output layer and regularisation dropout layers round out the model. Dense (completely linked) layers are used for this. The 406,245,954 parameters in this structure enable efficient and adaptive feature extraction.

Model Evaluation

An essential part of any ML or DL project is assessing model's efficacy on various metrics. This study used several classification parameters, namely, sensitivity, specificity, recl, f-measure, acc, prec, confusion matrix, ROC-AUC, and Cohen's Kappa to evaluate the proposed VGG19_AlexNet_AdaptiveHybrid_CNN model performance in classifying wrist X-ray abnormalities.

Confusion Matrix: An effective method for gauging a classifier model's potentiality is the confusion matrix. Certain datasets, particularly those with imbalances, might make accuracy seem deceptive. A bone fracture detection scenario is shown in the image, along with the several prediction types: "true positive", "true negative", "false positive", and "false negative". Figure 4 below provides a graphic depiction of the confusion matrix.

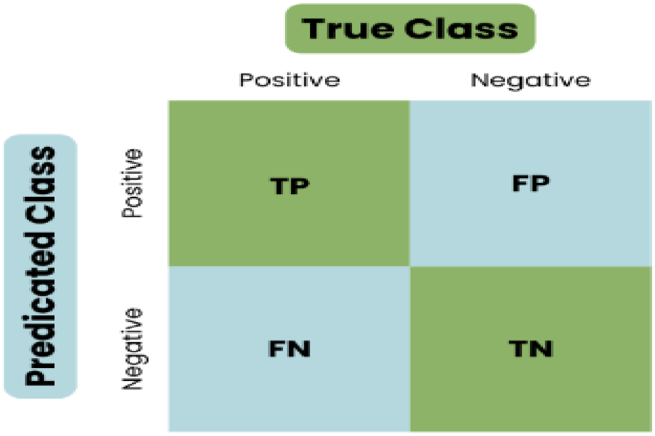

Figure 4. "Confusion Matrix"

Accuracy: It is calculated using proportion of all successfully predicted photos to all test images. Equation (1) may be used to determine accuracy.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: By separating actual findings by total amount of true positives, one may determine the precision. It can be calculated using Equation (2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall or Sensitivity: This statistic's formula is the total of all forecasts divided by sum of all positive samples. With equation (3), recall can be computed.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Specificity: A test's specificity, often known as its true negative rate, measures its ability to detect false negatives (Monaghan et al., 2021). Another way of putting it is that specificity is the percentage of those who get a negative result who were really assigned a negative outcome. Equation (4) may be used to determine it.

$$Specificity = \frac{TP}{TN+FP} \quad (4)$$

F1-Score: The f-measureis quantitative indicator of model's harmonic mean efficacy. Eq. (5) allows for its calculation.

$$F1 - Score = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (5)$$

Cohen's Kappa: The degree of agreement among two judges who assign things to separate categories is measured by Cohen's Kappa Statistic (Sabharwal, 2021). Kappa examines the likelihood of predicted agreement among two raters relative to the frequency of actual agreement to establish the independence of the ratings. The degree to which two raters agree on the application of a criteria to the determination of the validity of a condition is determined by the kappa statistic. It can be calculated using Equation (6).

$$k = \frac{p_o - p_e}{1 - p_e} \quad (6)$$

ROC-AUC: A ROC curve displays a performance of a classifier in two dimensions. It may be useful to simplify ROC performance into a single scalar number that represents predicted performance in order to compare classifiers (Fawcett, 2006). A popular approach is to determine the AUC, or area under the ROC curve. Its value will consistently fall somewhere amidst zero and one and a half since it represents a fraction of the unit square. No practical classifier, however, should have an AUC lower than 0.5 as random guessing generates the 0.5-area crosswise line amongst (0, 0) and (1, 1).

“Results Analysis and Discussion”

This segment evaluates similarities and differences among studied subjects, highlighting key findings and implications. The results of this study are implemented on a Dell PC that was running on Windows 11 operating system. The system has 32GB of RAM, 500GB of SSD and 16 GB NVIDIA GPU for fast processing. The research study also used Jupyter Notebook and the Python language for better understanding and code reusability. The following sections provide the dataset description with visualisation results, and provide investigational results of DL models in standings of performance measures, and then provide the comparative analysis amidst base and proposed models with discussion.

Dataset Description and visualisation

The MURA (XR_WRIST) dataset, published by the Stanford Machine Learning Group, comprises an extensive collection of bone X-rays intended for the development and assessment of algorithms that categorise X-ray studies as normal or pathological. Musculoskeletal diseases, impacting over 1.7 billion individuals worldwide and serving as a primary source of significant pain and disability, highlight critical role of this dataset in enhancing medical imaging technology, especially in areas with restricted access to proficient radiologists. MURA comprises 40,561 multi-view radiographic pictures derived from 14,863 investigations involving 12,173 patients, each classified into one of 7 upper limit radiographic categories: shoulder, elbow, finger, forearm, humerus, hand, and wrist. Expert radiologists at Stanford Hospital meticulously categorised each research as normal or abnormal according to clinical radiography evaluations performed between 2001 and 2012. The MURA (XR_WRIST) dataset's sample pictures are shown in Figure 5.

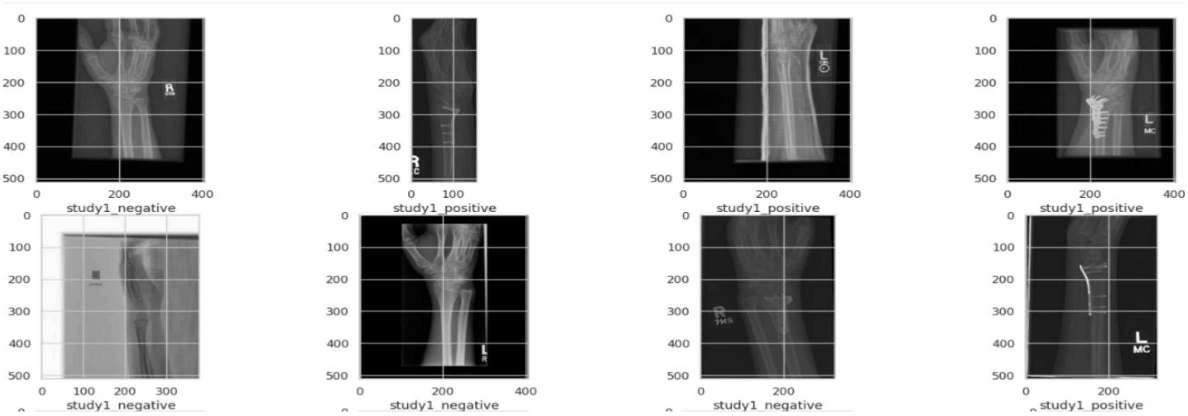

Figure 5. MURA (XR_WRIST) dataset's sample images

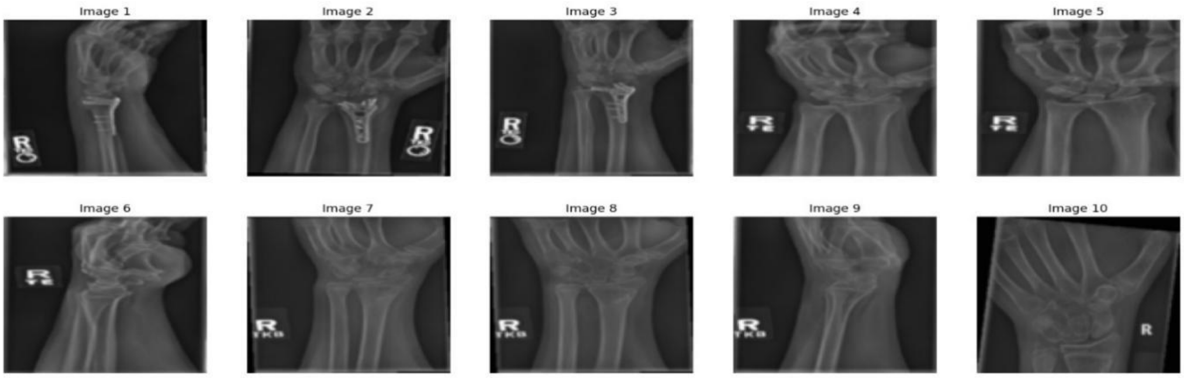

Figure 6. Sample Images of MURA (XR_WRIST) Dataset After Preprocessing Technique

Figure 6 displays the preprocessed wrist X-ray pictures of MURA (XR_WRIST) dataset. After preprocessing, images have been standardised in size and contrast-enhanced to improve the visibility of bone structures. Each image shows a different angle or section of the wrist, aiding in consistent evaluation for diagnostic purposes. The right ("R") marker in each image indicates the anatomical side.

Results of proposed hybrid model for wrist image classification

The following section provide the expeirmental results of proposed VGG19_AlexNet_AdaptiveCNN_Hybrid_Model for wrist image classification.

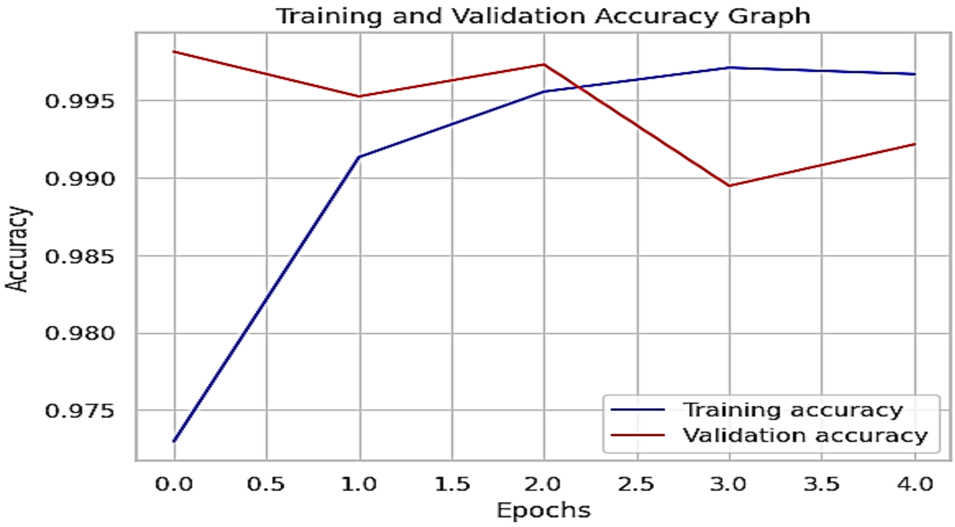

Figure 7. Plotting accuracy curve of hybrid method on Training and Validation datasets

The suggested hybrid model's acc in training and validation is shown in Figure 7, a line graph. An x-axis of a graph shows amountof epochs, between 0 and 4, while a y-axis shows a combined training and validation accuracy. To differentiate training and authentication accuracy graph utilised two different colour lines: blue for training accuracy and red for validation acc. The graph clearly demonstrates that the proposed VGG19_AlexNet_AdaptiveHybrid_CNN model achieves training acc of 99.83% and validationacc of 99.81%.

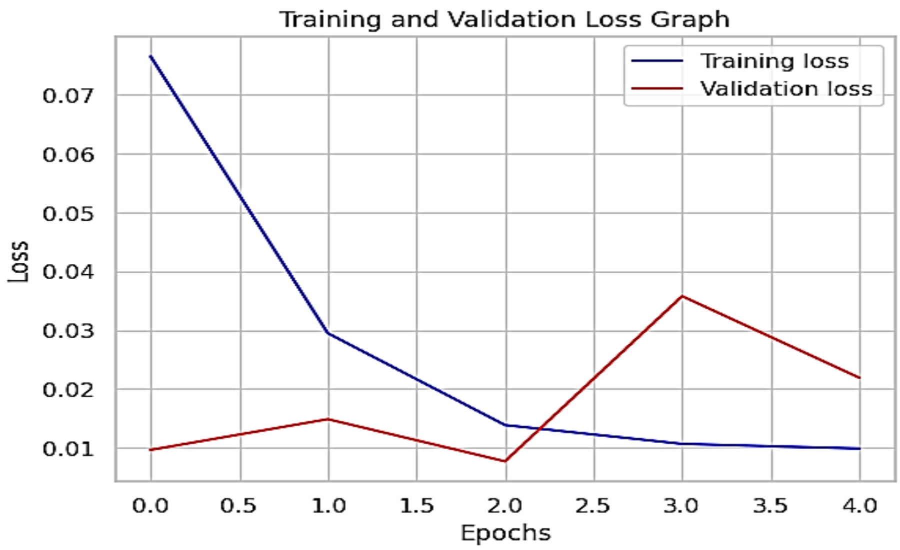

Figure 8. Plotting loss curve of hybrid model on Training and Validation datasets

A line graph showing suggested hybrid model's failure in training and validation is displayed in Figure 8. An x-axis of a graph shows amountof epochs, between 0 and 4, while a y-axis shows a total loss throughout training and validation. To differentiate train and validation loss graph utilised two different colour lines: blue for training loss and red for validation loss. The graph clearly demonstrates that the proposed VGG19_AlexNet_AdaptiveHybrid_CNN modelaccomplishes a training loss of 0.0048 and a validation acc of 0.0095.

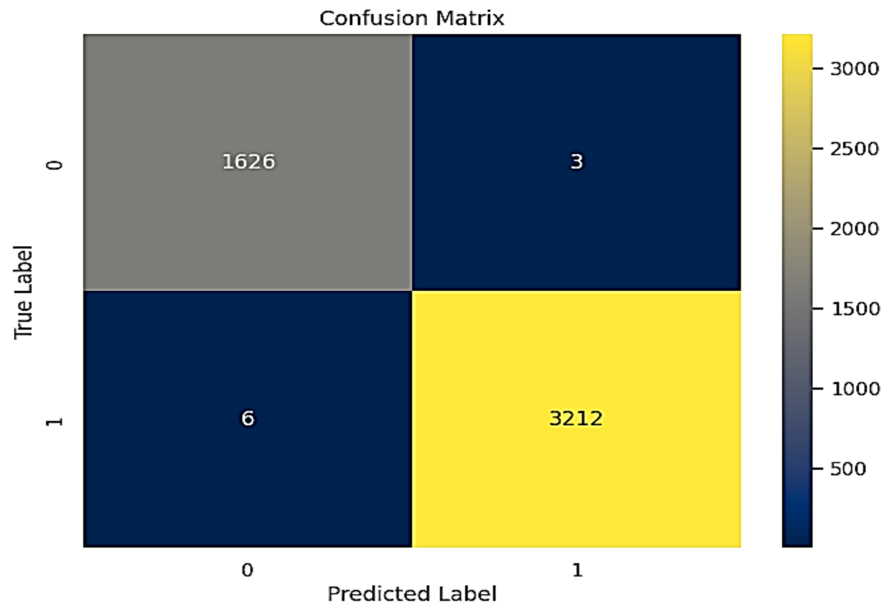

Figure 9. Confusion Matrix for VGG19_AlexNet_AdaptiveHybrid_CNN Model

The suggested VGG19_AlexNet_AdaptiveHybrid_CNN model is shown in Figure 9 as a confusion matrix to illustrate its performance in wrist X-ray abnormality classification. The x-league of confusion matrix illustrates predicted label, while the y-axis shows real label. During the testing phase, the model accurately predicted 1626 cases for the normal (0) class and 3212 instances for the abnormal class, according to the data in the confusion matrix.

Table 2. Proposed Model's performance of Classification Metrics on Wrist Facture Images dataset

Parameters	Performance of Hybrid model
Accuracy	99.81
Precision	99.9
Recall	99.81
F1-Score	99.86
ROC-AUC	99.99
Cohen's Kappa	99.58
Sensitivity	99.81
Specificity	99.81

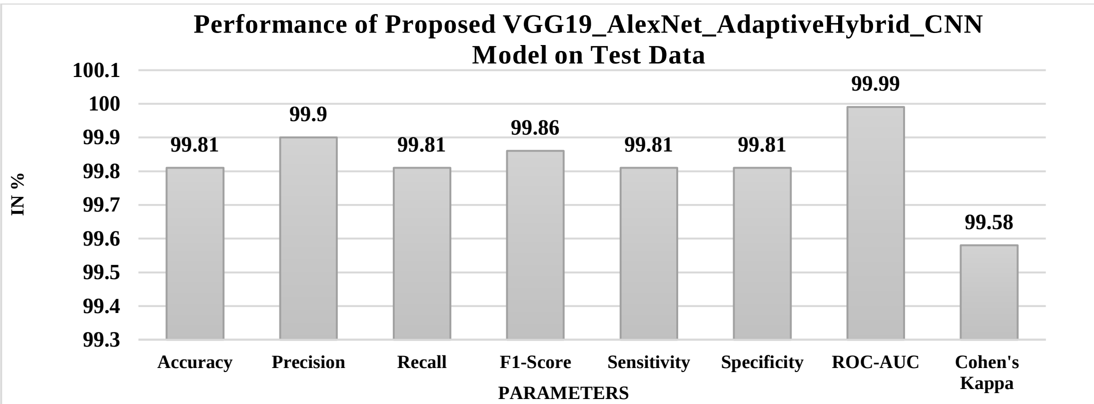

Figure 10. Performance Graph of the Proposed VGG19_AlexNet_AdaptiveHybrid_CNN Model's Classification Metrics

To demonstrate efficacy of a proposed model in wrist X-ray abnormality classification, Figure 10 and Table 2 provides potentiality of classification VGG19_AlexNet_AdaptiveHybrid_CNN model. The graph's y-axis shows percentage score for each classification parameter, while x-axis shows parameters. The bar graph clearly depicts that proposed model gets an accuracy99.81%, a precision99.9%, a recall99.81%, an f1-score99.86%, a sensitivity99.81%, a specificity99.81%, ROC-AUC99.99% and Cohen's kappa99.58% throughout the testing phase. These outcomes highlight a model's performance in effectively wrist X-ray abnormality classification.

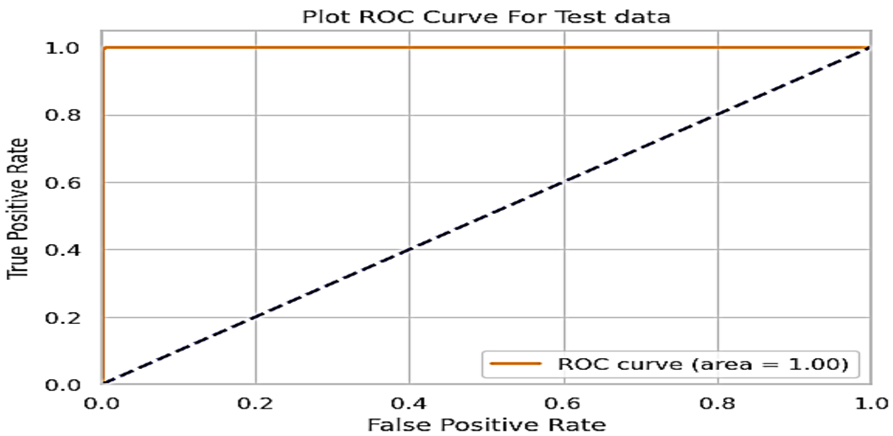

Figure 11. ROC Curve for the VGG19_AlexNet_AdaptiveHybrid_CNN Model

To illustrate an effectiveness of a proposed model in a classification of wrist X-ray abnormalities Figure 11 shows the ROC curve for the VGG19_AlexNet_AdaptiveHybrid_CNN model. In the graph, TPR is shown on y-axis and FPR is shown on x-axis. The model achieves perfect score during testing, as shown by the ROC curve.

Table 3. Comparison Analysis of different DL models for Wrist X-ray Abnormality Detection Using Classification Metrics

Parameters	**Base Model**			**Propose Model**
	DenceNet-201(Chada, 2019)	**VGG16**(Ashkani-Esfahani et al., 2022)	**MobileNetV2**(Sandler et al., 2018)	**VGG19_AlexNet_AdaptiveHybrid_CNN**
Accuracy	88.00	78.00	88.00	99.81
Precision	84.00	77.00	86.00	99.9
Recall	84.00	77.00	88.00	99.81
F1-Score	88.00	78.00	88.00	99.86

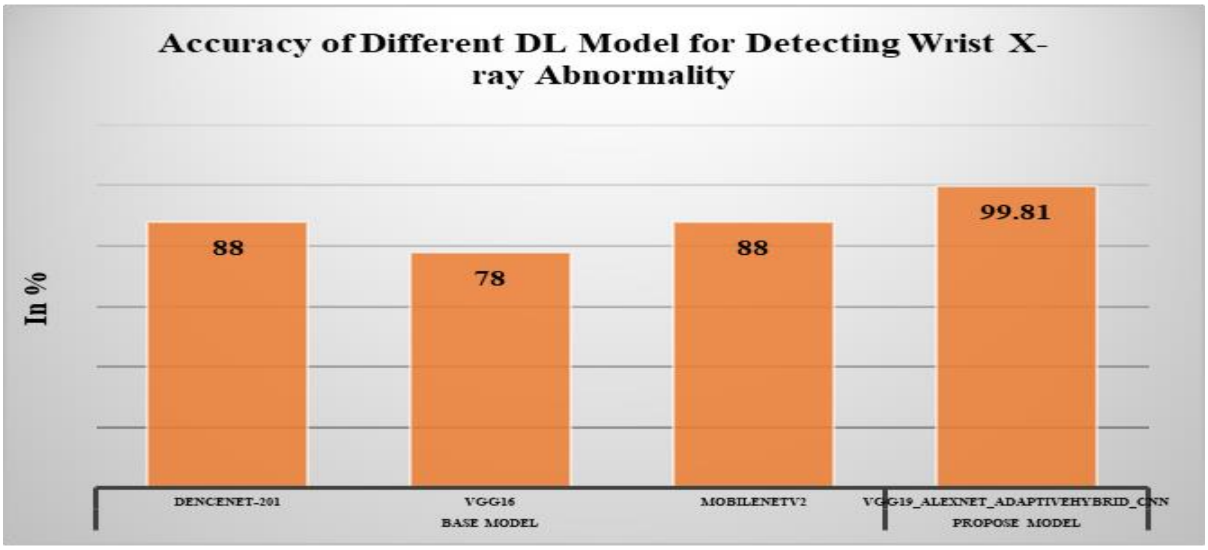

Figure 12. Comparison of Accuracy Measures for Different DL Models in Wrist X-ray Abnormality Detection

Table 3 and Figure 12 illustrate the comparison of accuracy measures for various DL models in the detection of wrist X-ray abnormalities. In the figure, graph's x-axis indicates the existing (DenceNet-201, VGG16, and MobileNetV2) and proposed (VGG19_AlexNet_AdaptiveHybrid_CNN) model, while a y-axis represesnts an accuracy score of every model as a percentage. The bar graph clearly demonstrates that DenceNet-201 model has an accuracy88%, the VGG16 model has an accuracy78%, the MobileNetV2 model has an accuracy88%, and the proposed VGG19_AlexNet_AdaptiveHybrid_CNN model has an accuracy of 99.81%. These findings demonstrate that, in comparison to other models already in use, the suggested model has the best accuracy. As a result, the proposed model outperforms the others when recognising abnormalities in wrist X-rays.

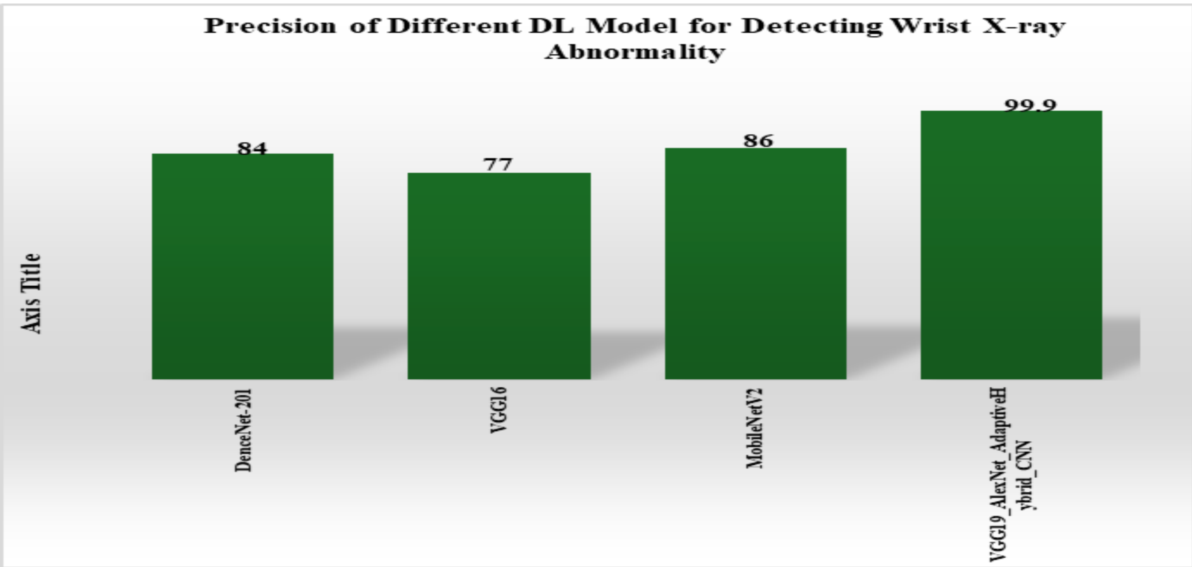

Figure 13. Comparison of Precision Measures for Different DL Models in Wrist X-ray Abnormality Detection

Table 3 and Figure 13 illustrate the comparison of precision measures for various DL models in detection of wrist X-ray abnormalities. In figure, graph's x-axis indicates the existing (DenceNet-201, VGG16, and MobileNetV2) and proposed (VGG19_AlexNet_AdaptiveHybrid_CNN) model, while a y-axis displays precision score of every model as a proportion. The bar graph clearly demonstrates that DenceNet-201 model has a precision of 84%, the VGG16 model has a precision of 77%, the MobileNetV2 model has a precision of 86%, and the proposed VGG19_AlexNet_AdaptiveHybrid_CNN model has a precision of 99.9%. These findings demonstrate that, in comparison to other models currently in use, the suggested model has the best precision. So, a proposed model is an optimal model for classifying wrist X-ray abnormalities.

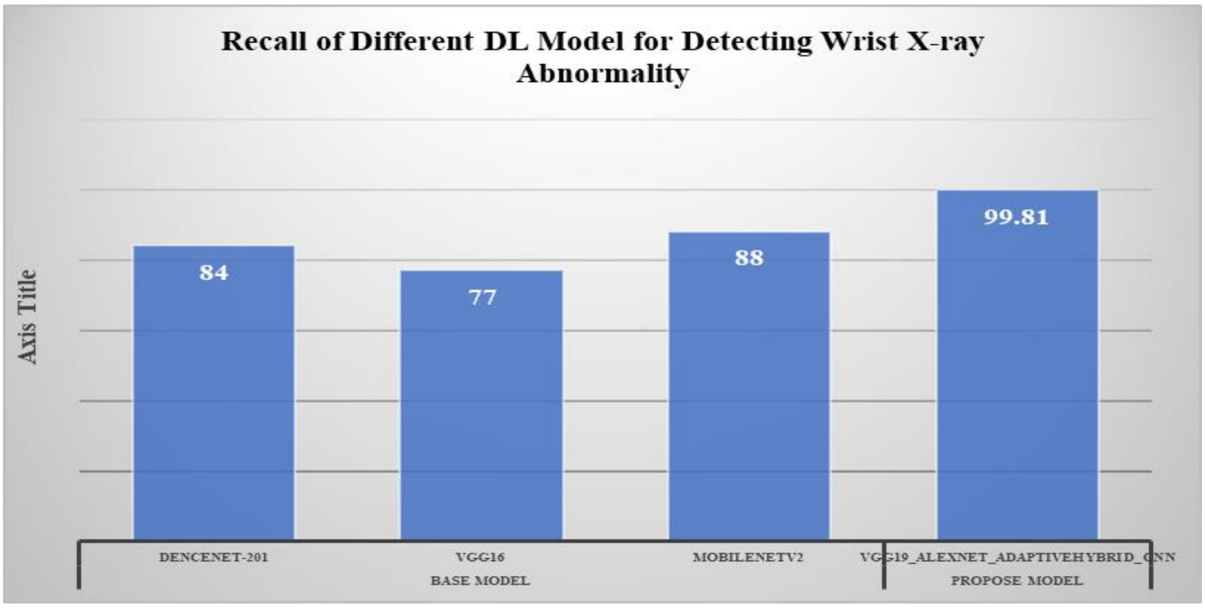

Figure 14. Comparison of Recall Measures for Different DL Models in Wrist X-ray Abnormality Detection

Table 3 and Figure 14 illustrate the comparison of recall measures for various DL models in detection of wrist X-ray abnormalities. In the figure, graph's x-axis indicates the existing (DenceNet-201, VGG16, and MobileNetV2) and proposed (VGG19_AlexNet_AdaptiveHybrid_CNN) model, while a y-axis displays recall score of every model as fraction. The bar graph clearly demonstrates that DenceNet-201 model has a recl of 84%, VGG16 model has a recl of 77%, the MobileNetV2 model has a recall of 88%, and the proposed VGG19_AlexNet_AdaptiveHybrid_CNN model has a recall of 99.81%. These results demonstrate that, in comparison to other models in existence, the suggested model has the greatest recall. So, a proposed model is an optimal model for classifying wrist X-ray abnormalities.

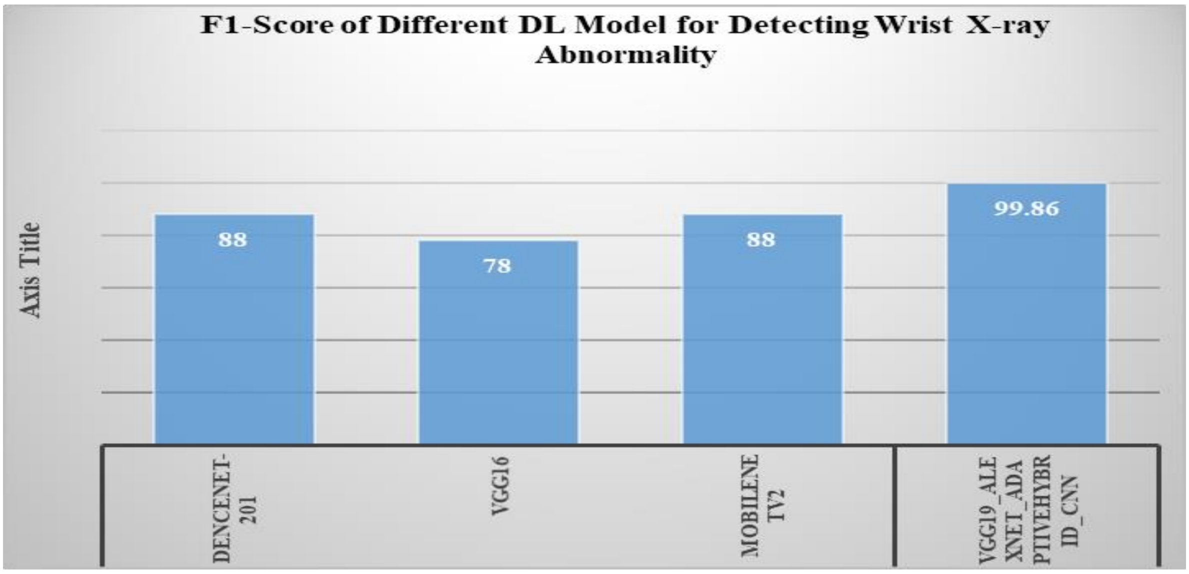

Figure 15. Comparison of F1-Score Measures for Different DL Models in Wrist X-ray Abnormality Detection

Table 3 and Figure 15 illustrate the comparison of F1 measures for various DL models in the detection of wrist X-ray abnormalities. In the figure, graph's x-axis indicates the existing (DenceNet-201, VGG16, and MobileNetV2) and proposed (VGG19_AlexNet_AdaptiveHybrid_CNN) model, while a y-axis displayed a f1-score of every model as a proportion. The bar graph clearly demonstrates that DenceNet-201 model has a f1-score88%, the VGG16 model has a f1-score78%, the MobileNetV2 model has a f-measure of 88%, and the proposed VGG19_AlexNet_AdaptiveHybrid_CNN model has f-measure of 99.86%. These findings demonstrate that, when compared with other models already in use, the suggested model achieves the greatest f1-score. As a result, the suggested model is the best one for identifying abnormalities in wrist X-rays.

Figure 16. Prediction of Proposed VGG19_AlexNet_AdaptiveHybrid_CNN Model for Wrist X-ray Abnormality

Figure 16 illustrates the prediction of the proposed VGG19_AlexNet_AdaptiveHybrid_CNN model for classification of the wrist X-ray abnormality. In figure, X-ray picture on the left, with the caption "Predicted: Positive," indicates anomalies, as the model had predicted. Some abnormalities in the bone structure are shown in the X-ray, which might indicate a medical problem. On the other hand, a wrist X-ray showing no obvious abnormalities is shown in the picture on the right, labelled "Predicted: Negative," which causes the model to forecast a negative result. This prediction shows the model's reliability and effectiveness in classifying wrist X-ray abnormalities.

Discussion with Limitations

The proposed VGG19_AlexNet_AdaptiveHybrid_CNN model significantly outperforms existing models like DenseNet-201, VGG16, and MobileNetV2 in wrist X-ray abnormality classification. Achieving an accuracy99.81%, a precision99.9%, a recall99.81%, and an F1-score99.86%, it demonstrates superior efficacy across all evaluation metrics compared to DenseNet-201 and MobileNetV2 (both with 88% accuracy) and VGG16 (78% accuracy). Additionally, the proposed model excels in precision and recall, addressing the critical need for reliable detection of abnormalities in medical imaging. With its advanced hybrid architecture, integrating features from VGG19 and AlexNet, and enhanced preprocessing and data handling techniques like CLAHE and SMOTE, the model ensures robust generalisation and effective handling of class imbalances, establishing itself as an optimal choice for wrist X-ray abnormality classification.

Despite the remarkable performance of the proposed VGG19_AlexNet_AdaptiveHybrid_CNN model, there are a few limitations to consider. Firstly, the model's evaluation is limited to a specific dataset of wrist X-rays, which might not generalise well to other types of medical imaging or datasets with different imaging modalities or demographic variations. Secondly, while achieving high accuracy, the computational complexity and training time are relatively high due to the hybrid architecture, which may limit its deployment on resource-constrained devices. Lastly, the model focuses primarily on binary classification and may require further enhancements to handle multi-class problems or subtle abnormalities in X-ray images. Addressing these limitations could ameliorate model's adaptability and clinical utility in real-world scenarios.

Conclusion and Future Scope

Wrist injuries are common, and fractures of the wrist, therefore, need to be well-diagnosed through the X-ray technique for adequate management and healing. This paper developed a new DL model known as the VGG19_AlexNet_AdaptiveHybrid_CNN for differentiating between normal and abnormal wrist X-rays with high accuracy. Compared to models such as DenseNet 201, VGG16, and MobileNetV2, this model outperformed them all with a 99.81% acc, 99.9% recl, and 99.86 f-measure on the test set. These outcomes demonstrate model's ability to provide high classification accuracy in terms of wrist X-ray abnormalities and its potential for improving diagnostic efficacy in aiding clinical decisions for early, effective treatment. Such an approach may open new possibilities for developing subsequent techniques for carrying out the analysis of medical imaging. Future work could be to come up with different models to detect other types of fractures and to try the model with different sets of patients. Furthermore, the incorporation of explainable AI methods would also enable clinicians to understand the model outputs better, making the model applied to clinical practice easier and more reasonable for early diagnosis and enhanced care of wrist fractures a reality.

Acknowledgment

Authors' Note

It is the authors' declaration that there is no conflict of interest associated with the publication of this article. The paper's authors verified that it was free of plagiarism

References

1. Arlot, S., Celisse, A., Arlot, S., Celisse, A., Celisse, A., & Painlev, L. P. (2010). A survey of cross-validation procedures for model selection To cite this version : HAL Id : hal-00407906 A survey of cross-validation procedures for model selection.
2. Arora, R., Kumar, A., & Soni, A. (2024). Deep Learning Approaches for Enhanced Kidney Segmentation: Evaluating U-Net and Attention U-Net with Cross-Entropy and Focal Loss Functions. https://doi.org/10.20944/preprints202408.1816.v1
3. Ashkani-Esfahani, S., Mojahed Yazdi, R., Bhimani, R., Kerkhoffs, G. M., Maas, M., DiGiovanni, C. W., Lubberts, B., & Guss, D. (2022). Detection of ankle fractures using deep learning algorithms. Foot and Ankle Surgery, 28(8), 1259–1265. https://doi.org/https://doi.org/10.1016/j.fas.2022.05.005
4. Bauskar, S. R., & Clarita, S. (2020). Evaluation of Deep Learning for the Diagnosis of Leukemia Blood Cancer. International Journal of Advanced Research in Engineering and Technology (IJARET), 11(3), 661–672. https://doi.org/https://iaeme.com/Home/issue/IJARET?Volume=11&Issue=3
5. Beri, M., Singh Gill, K., Upadhyay, D., & Devliyal, S. (2024). Fracture Forecasting through Deep Learning's Role in Bone Injury Detection and Classification. 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC), 969–974. https://doi.org/10.1109/AIC61668.2024.10731022
6. Bhan, B., & Patel, S. (2017). Efficient Medical Image Enhancement using CLAHE Enhancement and Wavelet Fusion. International Journal of Computer Applications. https://doi.org/10.5120/ijca2017913277
7. Bowers, E. J. (2011). Growth in children with clefts: Serial hand-wrist X-ray evidence. Cleft Palate-Craniofacial Journal. https://doi.org/10.1597/09-129
8. Chada, G. (2019). Machine Learning Models for Abnormality Detection in Musculoskeletal Radiographs. Reports. https://doi.org/10.3390/reports2040026
9. Chan, H. P., Samala, R. K., Hadjiiski, L. M., & Zhou, C. (2020). Deep Learning in Medical Image Analysis. Advances in Experimental Medicine and Biology, 1213, 3–21. https://doi.org/10.1007/978-3-030-33128-3_1
10. Dablain, D., Krawczyk, B., & Chawla, N. V. (2023). DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data. IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/TNNLS.2021.3136503
11. Du, H., Zhao, Y., Wang, H., Yang, C., Liu, Y., Wang, L., Xu, M., Gui, J., Tan, T., & Wang, X. (2022). Research on Key Technologies of Human Upper Limb Bone Diagnosis based on Convolutional Neural Network. Proceedings - 2022 9th International Conference on Dependable Systems and Their Applications, DSA 2022. https://doi.org/10.1109/DSA56465.2022.00128
12. El-Saadawy, H., Tantawi, M., Shedeed, H. A., & Tolba, M. F. (2021). A Hybrid Two-Stage GNG-Modified VGG Method for Bone X-Rays Classification and Abnormality Detection. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3081915
13. Erzen, E. M., Bütün, E., Al-Antari, M. A., Saleh, R. A. A., & Addo, D. (2023). Artificial Intelligence Computer-Aided Diagnosis to automatically predict the Pediatric Wrist Trauma using Medical X-ray Images. ISAS 2023 - 7th International Symposium on Innovative Approaches in Smart Technologies, Proceedings. https://doi.org/10.1109/ISAS60782.2023.10391582
14. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters. https://doi.org/10.1016/j.patrec.2005.10.010
15. Hardalaç, F., Uysal, F., Peker, O., Çiçeklidağ, M., Tolunay, T., Tokgöz, N., Kutbay, U., Demirciler, B., & Mert, F. (2022). Fracture Detection in Wrist X-ray Images Using Deep Learning-Based Object Detection Models. Sensors. https://doi.org/10.3390/s22031285
16. Jabbar, J., Hussain, M., Malik, H., Gani, A., Khan, A. H., & Shiraz, M. (2022). Deep Learning Based Classification of Wrist Cracks from X-ray Imaging. Computers, Materials and Continua. https://doi.org/10.32604/cmc.2022.024965

17. Javed, R., Abbas, T., Janjua, J. I., Muhammad, M. A., Ramay, S. A., & Basit, M. K. (2023). Wrist Fracture Prediction Using Transfer Learning, a Case Study. Journal of Population and Therapeutics and Clinical Pharmacology, 30(18), 1050–1062. https://doi.org/10.53555/jptcp.v30i18.3161
18. Kandel, I., Castelli, M., & Popovič, A. (2020). Musculoskeletal images classification for detection of fractures using transfer learning. Journal of Imaging, 6(11). https://doi.org/10.3390/jimaging6110127
19. Khare, P. (2022). Signature-Based Biometric Authentication: A Deep Dive Into Deep Learning Approaches. International Research Journal of Modernization in Engineering Technology and Science, 04(08), 2414–2424.
20. Khare, P., Arora, S., & Gupta, S. (2024). Recognition of Fingerprint Biometric Verification System Using Deep Learning Model. 2024 International Conference on Data Science and Network Security (ICDSNS), 1–7. https://doi.org/10.1109/ICDSNS62112.2024.10691020
21. Kuo, R. Y. L., Harrison, C., Curran, T. A., Jones, B., Freethy, A., Cussons, D., Stewart, M., Collins, G. S., & Furniss, D. (2022). Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. Radiology, 304(1), 50–62. https://doi.org/10.1148/radiol.211785
22. Medaramatla, S. C., Samhitha, C. V., Pande, S. D., & Vinta, S. R. (2024). Detection of Hand Bone Fractures in X-Ray Images Using Hybrid YOLO NAS. IEEE Access, 12(February), 57661–57673. https://doi.org/10.1109/ACCESS.2024.3379760
23. Monaghan, T. F., Rahman, S. N., Agudelo, C. W., Wein, A. J., Lazar, J. M., Everaert, K., & Dmochowski, R. R. (2021). Foundational statistical principles in medical research : Predictive value. Medicina, 57(5), 503.
24. Mustafa, T., & Alneamy, J. (2022). Medical Images Classification Using Artificial Intelligence. Journal of Education and Science. https://doi.org/10.33899/edusj.2022.133358.1224
25. Narayan, V., Mall, P. K., Alkhayyat, A., Abhishek, K., Kumar, S., & Pandey, P. (2023). Enhance-Net: An Approach to Boost the Performance of Deep Learning Model Based on Real-Time Medical Images. In Journal of Sensors. https://doi.org/10.1155/2023/8276738
26. Noureen, A., Zia, M. A., Adnan, A., & Hashim, M. (2023). Analysis and Classification of Bone Fractures Using Machine Learning Techniques. E3S Web of Conferences, 409, 1–9. https://doi.org/10.1051/e3sconf/202340902015
27. Pranav Khare, S. S. (2022). AI-Driven Palm Print Authentication: A comprehensive Analysis of Deep Learning Approaches for Efficient Biometrics. International Journal of Science and Research Archive, 6(1), 318–327.
28. Pranav Khare, S. S. (2023). Enhancing Biometric Authentication: Deep Learning Models For Human Iris Recognition. International Journal of Creative Research Thoughts (IJCRT), 11(8), h148–h153.
29. Sabharwal, C. L. (2021). Cohen's Kappa Statistic and newKappaStatistic for Measuring and Interpreting Inter-Rater Agreement. International Journal of Research in Engineering and Science (IJRES) ISSN, 9(7), 23–28.
30. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2018.00474
31. Sharma, S. (2023). Artificial intelligence for fracture diagnosis in orthopedic X-rays: current developments and future potential. Sicot-J, 9. https://doi.org/10.1051/sicotj/2023018
32. Singh, G. (2022). Hybrid Deep Learning Approach for Automatic Detection in Musculoskeletal Radiographs Gurpreet.
33. Sinha, H. (2024). Advanced Deep Learning Techniques for Image Classification of Plant Leaf Disease. Journal of Emerging Technologies and Innovative Research (JETIR) Www.Jetir.Org, 11(9), b107–b113.
34. Tahir, A., Saadia, A., Khan, K., Gul, A., Qahmash, A., & Akram, R. N. (2024). Enhancing diagnosis: ensemble deep-learning model for fracture detection using X-ray images. Clinical Radiology, 79(11), e1394–e1402. https://doi.org/10.1016/j.crad.2024.08.006
35. Tandon, R. (2023). Face mask detection model based on deep CNN techniques using AWS. International Journal of Engineering Research and Applications, 13(5), 12–19.
36. Thaarakaraam, M. A., Reddy, M. M., & Vijayakumar, V. (2024). Cnn Based Bone Fracture Detection for Medical Imaging Using Resnet-50. International Journal of Technical Research & Science, 9(Spl), 27–35. https://doi.org/10.30780/specialissue-iset-2024/038
37. Yang, F., Wei, G., Cao, H., Xing, M., Liu, S., & Liu, J. (2020). Computer-Assisted Bone Fractures Detection Based on Depth Feature. IOP Conference Series: Materials Science and Engineering. https://doi.org/10.1088/1757-899X/782/2/022114
38. Zhang, Z. (2023). The Improvement of the Discrete Wavelet Transform. Mathematics. https://doi.org/10.3390/math11081770