Osteochondroma Identification Through Transfer Learning and Convolutional Neural Networks

Ayesha Afridi[1], Muhammad Kamran Abid[1*], Naeem Aslam[1], Shumaila Khan[2], Arsalan Khan[2], Muhammad Ahmad Nawaz ul Ghani[3]

[1]NFC Institute of Engineering and Technology Multan, 60000, Pakistan.

[2]Department of Computer Science, University of Science & Technology, Bannu, Pakistan.

[3]School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China.

*Correspondence: kamranabidhiraj@gmail.com

Accurate and timely diagnosis of musculoskeletal conditions like osteochondroma is pivotal in ensuring effective treatment and improved patient outcomes. However, traditional diagnostic methods relying on manual interpretation of medical images can be susceptible to human errors, potentially leading to misdiagnosis or delayed detection. Previous studies have explored Deep Learning (DL) techniques for automated disease detection, but they often face challenges such as limited dataset availability and generalization capabilities across diverse imaging modalities. This research addresses these gaps by proposing a robust Convolutional Neural Network (CNN) framework for osteochondroma identification, leveraging transfer learning and data augmentation techniques. The ResNet-50 architecture, pre-trained on a large dataset, is fine-tuned with dense layers and an output layer for binary classification. Extensive data pre-processing and offline augmentation strategies enhance model performance and generalizability. The proposed model achieves an impressive 97.67% accuracy on the test dataset, demonstrating its effectiveness in distinguishing between normal and osteochondroma cases. Furthermore, its generalizability is validated by training and testing on the publicly available Potato Leaf Disease dataset, showcasing consistent performance in multi-class classification scenarios. While the model exhibits promising results, future work could explore integrating more extensive and diverse datasets and investigating advanced architectures for improved accuracy and computational efficiency. The implications of this research extend to empowering medical practitioners with accurate and swift osteochondroma diagnostics, ultimately contributing to enhanced patient care in orthopaedics.

Keywords: Osteochondroma; Deep Learning; Convolutional Neural Networks; Transfer Learning; Medical Imaging, Image Classification

Introduction:

The osteochondroma is a reasonably common benign bone tumour typically affecting long bones like the femur and tibia. While normally non-cancerous, it can cause difficulties and discomfort, making it critical to understand its causes, treatment options, and potential consequences for afflicted people. The cause of osteochondroma is unknown, but it is widely thought to result from aberrant bone growth during childhood and adolescence. Osteochondromas arise in the growth plates of the bones, where new bone tissue is generated. A genetic mutation known as EXT1 or EXT2 has increased the risk of developing osteochondromas. This mutation impairs the normal action of the genes that drive bone growth, producing these benign tumors. Most osteochondromas are solitary but can also be associated with a genetic disorder known as Multiple Hereditary Exostosis (MHE). MHE causes many osteochondromas throughout the skeleton, potentially leading to severe health problems [1]. Osteochondromas are often asymptomatic and may be discovered incidentally during imaging studies for other reasons. When symptoms do appear, they may include discomfort, edema, and a restricted range of motion in the afflicted joint. Rarely, nerve compression or vascular compromise might occur, resulting in serious complications. Osteochondroma is usually diagnosed using a combination of clinical and imaging investigations. X-rays, CT scans, and MRIs are frequently utilized to assess the tumor's size, location, and influence on adjacent tissues. Genetic testing may also be recommended to discover the underlying genetic abnormalities, particularly in situations of numerous osteochondromas [2].

In traditional osteochondroma diagnostics, health practitioners heavily rely on X-ray and CT scan images, encountering challenges such as interpretive bias, potential oversight of minor abnormalities, and time-consuming analysis [3]. DL diagnostics emerge as a transformative alternative, providing a more objective, efficient, and precise approach to osteochondroma detection. By leveraging advanced algorithms, DL can significantly enhance diagnostic accuracy and streamline operations, ultimately improving patient care [4]. The integration of DL into medical diagnostics, particularly in musculoskeletal imaging, represents a paradigm shift, offering new avenues for enhanced accuracy, efficiency, and early detection of osteochondromas. Despite the effectiveness of conventional imaging methods, challenges persist in distinguishing between benign and malignant osteochondromas, underscoring the need for nuanced expertise. DL, mainly through Convolutional Neural Networks (CNNs), has shown remarkable potential in addressing these challenges by recognizing specific features indicative of osteochondromas and facilitating automated image segmentation and classification for precise measurements [5]. Federated learning, as discussed in 'A federated learning approach for anomaly detection in high-performance computing', complements this DL framework by pooling anonymized data from diverse healthcare settings. This approach mitigates common data biases and enhances the diagnostic models' ability to discern subtle osteochondroma characteristics across varied populations [6].

The study aimed to collect a large dataset to develop a lightweight CNN model for identifying osteochondroma diseases from X-ray images. A CNN model leveraging transfer learning was developed for classifying X-ray images as either normal or indicative of osteochondromas. The model incorporates ResNet-50 as the backbone for feature extraction, followed by two Fully Connected (FC) layers and a final output layer for binary classification—the proposed model produced comparatively good results of 97.67% on the test dataset. This high accuracy suggests its potential practical application by healthcare practitioners and officials, underscoring its reliability in aiding accurate diagnoses. Further refinements and advancements could pave the way to achieving even higher accuracies for enhanced clinical utility. The main contributions of the proposed study are:

- We are curating a substantial dataset comprising 1500 normal and 1500 osteochondroma X-ray images, addressing the scarcity of publicly available datasets for this specific condition.
- We employed strategic data pre-processing techniques to enhance model robustness and generalization capabilities, including resizing, normalization, and augmentation through rotations, flipping, and zooming.
- We leveraged transfer learning by fine-tuning the pre-trained ResNet-50 architecture, a powerful deep-learning model, to accurately classify osteochondroma cases, achieving an impressive 97.67% testing accuracy.
- The proposed model's versatility and robustness can be analyzed by evaluating its performance on a publicly available multi-class dataset, validating its potential for broader applications in medical image analysis.

Related Works:

The rise of bioinformatics has revolutionized medical imaging and agricultural analysis, driving significant advancements in automation. The rapid progress in image classification has significantly bolstered automatic diagnostics in medical imaging. This innovative approach ensures swift and accurate results, expediting the diagnostic process [7]. Additionally, its user-friendly interface facilitates easy accessibility, marking a pivotal advancement in medical diagnostics for improved patient care. The research [8] proposed a fine-tuned CNN model for multi-class lung disease classification using frontal chest X-ray images, addressing diseases like COVID-19, Pneumonia, and TB. The study involved a dataset gathered from various sources, which was pre-processed, balanced, and augmented and achieved 98.89% accuracy, surpassing other pre-trained models. Auto-encoder was used for image denoising, and CLAHE enhancement and filtering techniques were applied. The study [9] introduced a novel deep learning model named Tumor-ResNet for accurate MRI scan-based brain tumor diagnosis. The model's architecture contains twenty convolution layers with Leaky ReLU activation and three dense layers for feature extraction. The model identifies 99.33% of brain tumors based on binary class. The article [10] examined the use of neural networks to identify breast cancer and proposed an enhanced CNN approach.

Malik H. et al. [11] proposed a novel fusion model combining hand-crafted features with deep CNNs for classifying ten chest diseases, including COVID-19, using chest X-rays. The approach involves Info-MGAN for lung image segmentation and a pipeline incorporating essential point extraction methods (ORB and SURF), DCNNs (VGG-19), and ML models for disease classification. The model achieves a high accuracy of 98.20%, demonstrating the potential for automated diagnosis of chest diseases, including COVID-19. Aytaç et al. [12] proposed a novel adaptive momentum optimization algorithm for accelerating the convergence and stability of CNNs in medical image classification. The adaptive momentum rate was dynamically adjusted based on changes in the error over epochs, eliminating the need for hyperparameter tuning. It was tested on three medical datasets: REMBRANDT Brain Cancer, NIH Chest X-ray, and COVID-19 CT scan. Results demonstrated improved performance over Stochastic Gradient Descent (SGD) and other optimizers, achieving lower classification errors (5.44%) and higher accuracy. The proposed approach was applied to various CNN architectures, outperforming SOTA models with 95% accuracy.

Eweje, F. R. et al. [13] developed a DL model using 1,060 histologically confirmed bone lesions from routine MRI scans, based on the EfficientNet-B0 architecture, combined image-based features with patient demographics for classification. The ensemble achieved similar accuracy, sensitivity, and specificity compared to expert radiologists. External validation demonstrated a robust ROC AUC of 0.79, highlighting the algorithm's effectiveness in distinguishing benign and malignant bone lesions. Sampath K. et al. [14] focused on classifying

normal and cancerous bone images using image processing and CNNs. Utilizing a dataset of 1,141 CT scan images, the Alex-Net model emerged as the most effective, achieving impressive training, validation, and testing accuracies of 98%, 98%, and 100%, respectively. The pre-processing involved grayscale conversion and median filtering, followed by K-means clustering and canny edge detection for image segmentation. Comparative analysis with other CNN models highlighted Alex-Net's superior performance, offering accurate classification and promising potential for early detection of bone cancer, with the added benefit of a shorter computational processing time. The study [15] aimed to enhance the classification accuracy of diverse bone tumors using an optimized DL algorithm. The proposed VGG16-ViT fusion model, combining the strengths of VGG-16 and Vision Transformer (ViT), demonstrated a notable classification accuracy of 97.6% on a dataset of 786 Computed Tomography (CT) images. The innovative approach involved selecting 27 features from the third layer of VGG-16, integrated into the ViT model for comprehensive training. The model showcased an 8% improvement in sensitivity and specificity optimization, reducing training time and offering robust classification performance for different bone tumor types.

Proposed Methodology:

The proposed framework introduced a robust pipeline for osteochondroma identification, harnessing the prowess of Convolutional Neural Networks (CNNs) augmented by strategic transfer learning techniques. The workflow commences with meticulous data pre-processing, wherein images are resized to uniform dimensions and normalized to optimize neural network performance. Data augmentation strategies, including rotations, flips, and zooming, are judiciously applied to enhance the model's generalization capabilities from the training data to unseen instances. The framework's core revolves around adapting a pre-trained ResNet-50 model, a robust CNN architecture, to the specific task of binary classification – distinguishing between normal and osteochondroma-affected images. This adaptation entails replacing the output layer and fine-tuning the network with the target dataset, iterating through numerous epochs of forward and backward propagation while dynamically adjusting learning rates based on validation performance feedback. Figure 1 encapsulates the comprehensive workflow, depicting the intricate interplay between data pre-processing, model architecture, training dynamics, and performance evaluation components, orchestrating a seamless end-to-end pipeline for accurate osteochondroma identification.
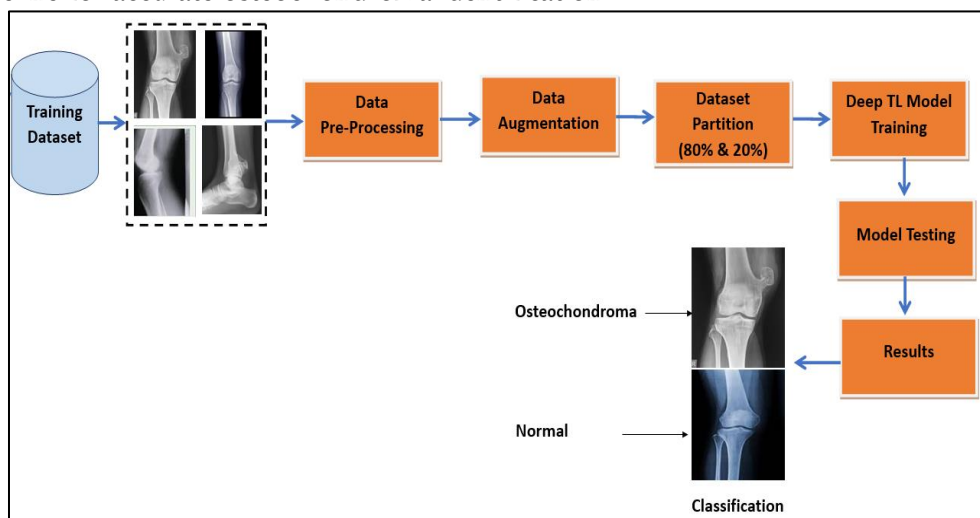

Figure 1: Comprehensive workflow of the proposed transfer learning-based convolutional neural network model for osteochondroma identification, detailing the stages of data pre-processing, model architecture, training dynamics, and performance evaluation.

The proposed algorithm presents a logical and technically sound approach to training a Convolutional Neural Network (CNN) for osteochondroma identification using transfer

learning. It encompasses crucial steps such as data pre-processing (resizing, normalization), augmentation (rotations, flipping, zooming), and partitioning into training, validation, and test sets. The model setup involves loading the pre-trained ResNet-50 architecture and modifying the top layer for binary classification. The training loop employs cross-entropy loss computation, expressed as $L = -(1/|B|) \Sigma_{i=1}^{|B|} [Y\_bi \log(\hat{Y}\_bi) + (1 - Y\_bi) \log(1 - \hat{Y}\_bi)]$, where Y_bi and $\hat{Y}\_bi$ represent true labels and predicted probabilities, respectively. Backpropagation updates weights using an optimizer and specified learning rate. Periodic validation evaluates performance and adjusts the learning rate accordingly. Finally, the algorithm computes performance metrics (accuracy, precision, recall, F1-score) on the test set, ensuring a comprehensive and technically robust approach to osteochondroma identification.

Data Collection:

In DL, the essence of model training for a specific domain problem lies in the quality and quantity of the training data. The significance of high-quality data cannot be overstated, as it directly impacts the model's performance. Training a DL model in medical imaging poses a substantial challenge due to the requirement for many images, especially for less common conditions like osteochondroma, which may lack public datasets. A proactive approach was taken, involving the collection of images from various sources, including the internet and a local health institute.

A total of 1200 healthy images and 1500 osteochondroma images were meticulously gathered. In addition, 1100 images were collected from the internet, and 1600 images were collected from a local health institute. They provided hard copies of X-ray images, and we converted them using a smartphone. Offline augmentation techniques were applied to address the dataset imbalance, resulting in a balanced dataset of 1,500 healthy and 1,500 osteochondroma images. The final dataset comprised 3000 images, with 300 reserved for testing and 2700 for model training. This dedicated effort underscored the commitment to obtaining comprehensive and diverse data and emphasized the ethical considerations in acquiring and utilizing medical images. The success of the DL model hinges on this well-curated dataset, ensuring its ability to robustly discern and classify osteochondroma, ultimately contributing to advancements in medical image-based diagnostics. Sample images illustrating the dataset are depicted in Figure 2, providing a glimpse into the variety and complexity of the acquired data.

Figure 2: Training dataset samples

Pre-Processing:

Pre-processing plays a pivotal role in the success of CNN image classification, acting as a crucial step in refining input data quality. Each image undergoes manual scrutiny involving domain experts for accurate labeling according to predefined classes to ensure adequacy. In this meticulous process, all images were standardized to a uniform size, and pixel values were rescaled to a standardized range, typically from 0 to 1 in this study. Such uniformity in data presentation optimizes the model's ability to discern relevant features and patterns during training. The deliberate efforts in pre-processing, including manual checks and expert labeling,

contributed to a robust and reliable dataset that enhanced the overall performance of the CNN, resulting in more precise and reliable image classification outcomes.

Data Augmentation:

Data augmentation is a powerful tool in mitigating overfitting and enhancing the generalization capacity of models to previously unseen images. This technique involves expanding the dataset by introducing variations to existing images, encompassing transformations such as rotations, adjustments in brightness, shearing, flipping, and shifting. Figure 3 illustrates various image modifications, like rotations and flips, to expand the training dataset and improve the model's generalization capabilities.
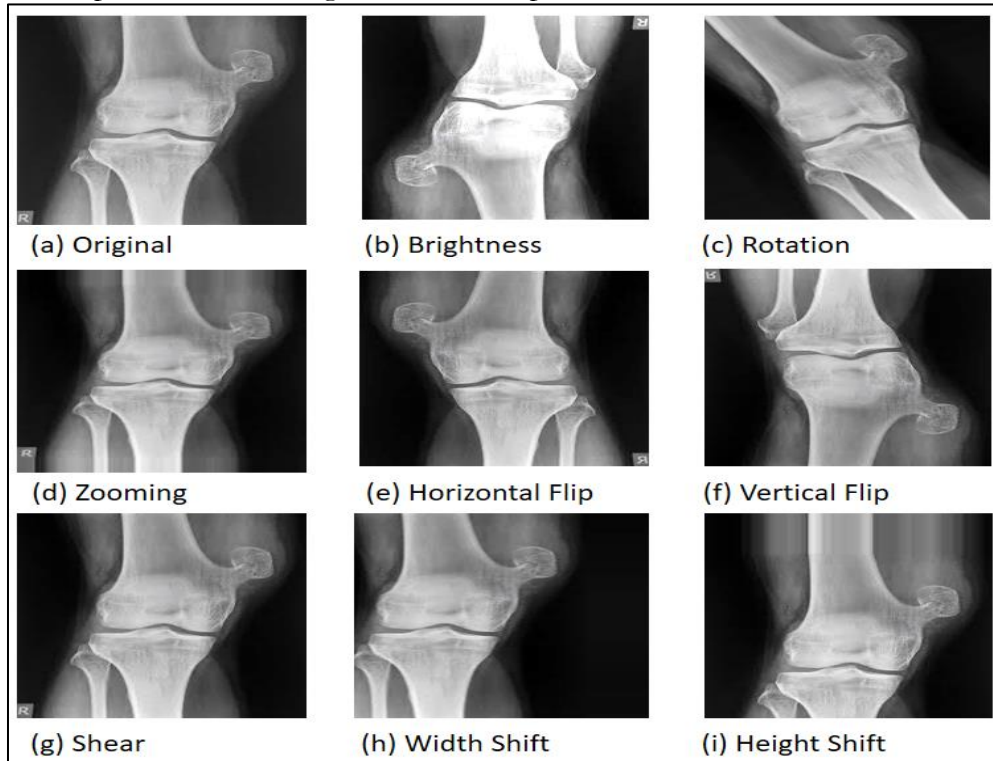

Figure 3: Image augmentation Techniques used in this study.

Exposing the model to a diverse range of augmented images during training allows it to discern essential features amidst variability, improving its ability to generalize well on novel data. This helps prevent overfitting, where the model memorizes the training set, struggles with new data, and fosters resilience to real-world variations. The versatility introduced by data augmentation reinforces the model's adaptability, resulting in a more robust and effective image classification system with broader applicability.

Convolutional Neural Networks (CNNs):

A CNN is a DL model optimized for processing structured grid data, such as images. It excels at image recognition and classification because it can automatically learn hierarchical feature representations from raw pixel data. It consists of three layers: convolutional, dense, and output. A convolutional layer in CNNs is an essential component for analyzing patterns of space in input data, especially in computer vision tasks. It performs convolution operations on input feature maps using learnable filters or kernels. A typical CNN model contains multiple convolutional layers, depending on the domain task. Mathematically, the convolution operation is expressed as:

$$(f * g)(s) = \sum_{a=1}^{m}\sum_{b=1}^{n} f(a,b).(s-a,t-b) \qquad (1)$$

Here, f represents the input feature map, g is the learnable filter/kernel, and (s,t) is the spatial coordinates. The outcome is a feature map that depicts local patterns and spatial hierarchy. Convolutional layers extract hierarchical characteristics while preserving spatial linkages and reducing the number of parameters compared to dense layers, making them useful for image identification and other spatial data problems.

Dense layers, commonly known as fully connected (FC) in CNNs, connect each neuron to every neuron in the preceding layer. They capture high-level abstractions and make final predictions after convolutional and pooling layers. FC layers learn complex correlations from previous layers' feature representations, mapping hierarchical features to output classes during training. In classification tasks, the final layer often includes a SoftMax (for multi-class) or sigmoid (for binary class) activation function, converting the raw output into class probabilities. This study employed a sigmoid activation for binary osteochondroma classification, expressing the input's probability of belonging to the positive class within the 0 to 1 range.

Transfer Learning (TL):

TL is a DL technique that uses knowledge acquired via training on one task to improve performance on another related activity. One significant advantage of transfer learning is its potential to expedite training and improve model performance, mainly when dealing with limited labeled data for the intended job. TL allows the transfer of valuable features and representations from large and diverse datasets, allowing the model to generalize to new tasks more effectively. This approach is especially effective in computer vision and natural language processing. TL increases efficiency, eliminates the need for extensive computational resources, and allows DL to be applied to a broader range of tasks with practical benefits.

Proposed Model:

In this study, a CNN utilizing TL was trained to identify osteochondromas in X-ray images. TL was chosen to address the limited dataset size, leveraging pre-trained ResNet-50 as the backbone. Using a sigmoid activation function, the proposed model incorporated two FC layers with a final layer for binary classification. The FC layers comprised 512 and 256 neurons, followed by a dropout layer with a 0.3 dropout ratio. ResNet-50, introduced by Microsoft in 2015, is a deep CNN architecture designed to address training challenges in intense neural networks [16]. Its main innovation is using residual blocks with skip connections that enable direct learning of residual functions and facilitate gradient flow, allowing the training of deeper networks. Comprising 50 layers of convolutional, batch normalization, and fully connected layers, ResNet-50 excels at image classification tasks. Its depth and skip connections aid in capturing complex hierarchical features, making it a popular choice for transfer learning in computer vision applications and a suitable backbone for the proposed osteochondroma identification model. The skip connection can be represented mathematically as:

$$Y = f(x) + x \qquad (2)$$

Where Y is the output of the skip connection, f (x) represents the transformation applied by a block of layers in the network, which can be any function representing the computations within the block. x is the input to the skip connection, often the output from an earlier layer in the network. Adding (+) combines the original input x with the block's transformed output f(x), allowing the network to preserve information and gradients that might be lost through multiple layers.

The proposed model used TL with ResNet-50 as its backbone for extracting features from the X-ray images. Due to the relatively small size of the training dataset, TL was used. The ResNet-50 model was fine-tuned by adjusting two FC layers with 512 and 256 neurons attached to classify the features extracted by the backbone of the model. A dropout layer followed each FC layer, and a sigmoid activation function was used as the output layer for binary classification. Figure 4 details the CNN architecture, showcasing layers and pathways critical for learning distinctive features in osteochondroma images.
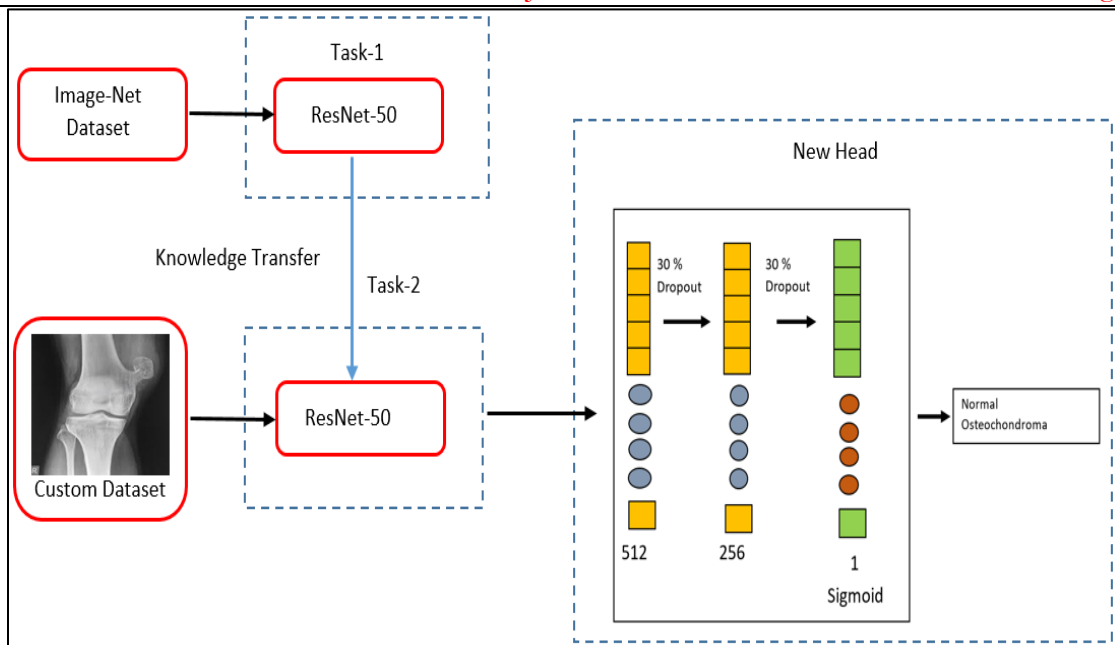
Figure 4: CNN Model architecture of the proposed model

Experimental Setup:

To train a DL model, robust hardware with sufficient memory and CPUs is required. When dealing with image-based problems in deep learning, GPU-based systems become crucial for expeditious model training. Without a robust GPU, the training session may extend into entire days or span several days. To reduce this inconvenience, Google Colab, a cloud-based computing environment, was used, taking advantage of its high-performance GPUs and extensive memory resources. This strategic decision enabled fast and timely model training, overcoming the limitations of previous setups and ensuring optimal use of computational resources.

Performance Measurement Metrics:

Diverse mathematical metrics play a pivotal role in assessing the model's performance. This study's evaluation criterion encompasses accuracy, precision, recall, and F1 score, each calculated using the mathematical formulas outlined below. These metrics comprehensively understand the model's effectiveness in classification tasks. They should discuss the results and how they can be interpreted from the perspective of previous studies and the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \qquad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (5)$$

$$F1 - \text{Score} = 2 \cdot \frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}} \qquad (6)$$

(Where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives)

Accuracy: Accuracy gauges the correctness of the model's predictions across all classes.
Precision: Precision focuses on the accuracy of positive predictions, indicating the model's ability to avoid false positives.

Recall (Sensitivity or True Positive Rate): Recall measures the model's ability to identify positive instances out of all actual positives correctly.

F1 Score: The F1 score is the harmonic mean of precision and recall, providing a balanced metric for model evaluation.

In this study, a CNN model was trained using TL. Training CNN models from scratch typically demands many images, which were unavailable in this case. To circumvent this limitation, Transfer Learning was employed as an alternative technique to train a model for binary classification, distinguishing between normal and osteochondroma images. The model was trained for 98 epochs, using an early stopping mechanism to avoid overfitting. Figure 5 shows the accuracy and loss graphs, demonstrating the model's performance throughout training. Table 1 outlines the architecture of the convolutional neural network model, incorporating the ResNet-50 framework for feature extraction. It details each layer's structure, including filter sizes and kernel dimensions, emphasizing the model's systematic construction for enhancing image classification accuracy.

Table 1: Detailed Architecture of the Proposed Transfer Learning Model Utilizing ResNet-50

S No.	Layer	Filters/Neurons	Kernal Size
1	Input (224x224x3)	-	-
2	Backbone (ResNet-50) for Transfer Learning	-	-
3	Convolutional layer	64	3x3
4-14	Residual Blocks	Varied	3x3
15	Average pool	-	Global
16	Flatten layer	-	-
17	Dense (Dropout (0.3)	512	-
18	Dense (Dropout (0.3)	256	-
19	Dense (Sigmoid)	1	-

Table 2 summarizes the key parameters and settings employed during the training phase of the CNN model. This includes learning rates, batch sizes, and other optimization techniques, which are crucial for understanding the model's learning environment and procedural setup.

Table 2: Training Parameters and Settings for CNN Model Optimization

Serial No	Parameter	Value
1	Optimizer	SGDM
2	Loss Function	Binary Cross Entropy
3	Batch Size	32
4	Train Set	0.8
5	Validation Set	0.2
6	Learning rate	0.001
7	Max Epochs	120
8	E-Stopping	98
9	Shuffle	Every Epoch

Results of the Proposed Model on the Test Dataset:

Table 3 presents the performance metrics of the CNN model, including accuracy, precision, recall, and F1 score, calculated from the test dataset. It quantitatively evaluates the model's effectiveness in classifying X-ray images as indicative of osteochondromas.

Table 3: Performance measurement of the proposed model on test data

S No.	Metric	Value (%)
1	Accuracy	97.67
2	Precision	98.64
3	Recall	96.67
4	F1-Score	97.63

These results underscore the model's efficacy in bone disease classification. The

potential integration of this model into real-world environments holds promise for enhancing the diagnostic systems of orthopedic departments. By providing practitioners with early and prompt diagnostics, this model contributes to advancing orthopedic diagnostics, offering valuable support for timely medical interventions.
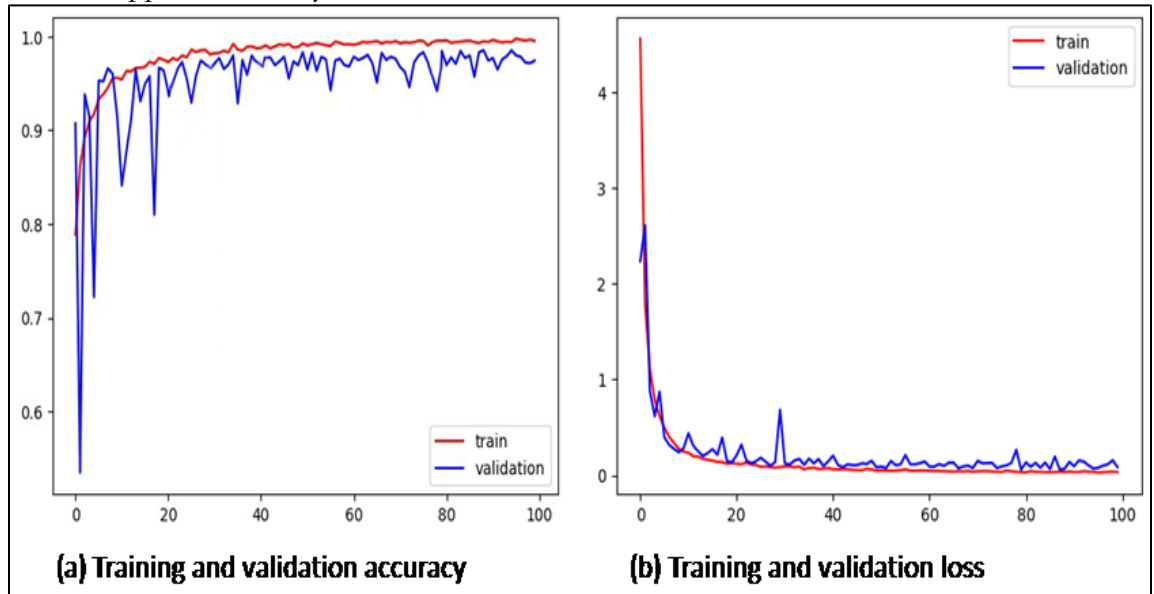

Figure 5: Training progress of the proposed model

In the iterative course of the experimental process, diverse models underwent training and fine-tuning. We systematically adjusted the layered structure, modulated parameters, and iteratively manipulated the model architecture by introducing and removing layers—such as batch normalization, dropout, and substituting pre-trained models. After rigorous evaluation, a model featuring ResNet-50 as its backbone, coupled with two fully connected layers and the SGD optimizer, demonstrated superior performance and was selected as the proposed model. Table 4 presents a comprehensive comparison between our proposed model and other famous pre-trained models, highlighting the comparatively good performance of our model in terms of various metrics. This meticulous model refinement process ensures the selection of an optimal architecture for effective bone disease classification, setting our proposed model apart in performance and potential real-world applicability.

Table 4: Performance comparison of the proposed model with other SOTA pre-trained models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EfficientNetB0	96.35	97.00	95.80	96.23
MobileNet-V2	95.00	96.50	94.40	95.80
Proposed Model	97.67	98.64	96.67	97.63

**Performance of the Proposed Model on the Public Dataset:**

To benchmark the model's performance, it is advised that the model should be trained and tested using datasets other than the original one that are publicly available. In our study, we trained and tested the suggested model using the Potato Leaf Disease (PLD) dataset, which is publicly available on Kaggle and contains 4072 images. It was divided into three classes: early blight, late blight, and normal. While our first problem focused on binary classification, applying the proposed methodology to a multi-class setting revealed flexibility. The results in Table 5 support the model's performance in multi-class classification scenarios, demonstrating its adaptability and reliability across varied datasets. This more significant review emphasizes the model's potential for applications beyond the setting of bone disease classification.

Table 5: Performance of the proposed model on a publically available dataset

Dataset	Source	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
PLD	Kaggle	97.36	96.50	96.35	96.44

Discussion:

The role of the dataset in training CNNs is paramount, influencing the model's generalization, robustness, and overall performance. A large and diverse dataset ensures that the CNN learns representative features and patterns, making it more adaptable to unseen images. The proposed study used a custom dataset to train the model. Figure 6 contrasts the original dataset with the augmented images, illustrating the significant role of augmentation in improving dataset diversity and balance, which is crucial for the robust training of the model.
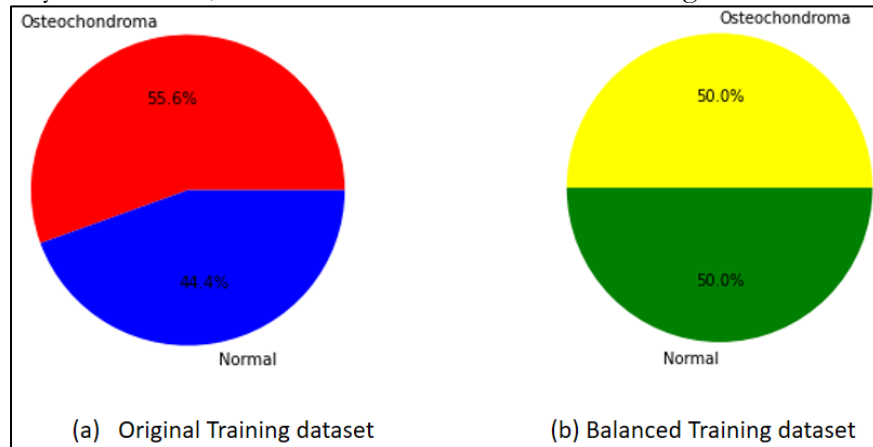
Figure 6: shows the original and final training datasets after offline augmentation.

Data augmentation is pivotal in model generalization, but its extended usage may cause problems like overfitting, which goes against model generalization. In the proposed study, offline augmentation was used only on a small portion of the training dataset; therefore, online data augmentation was also used with pre-processing techniques to enhance model performance by standardizing input data and introducing variations during training. The model training process was done iteratively from shallow to deep architectures, but due to the small size of the training dataset, it was not easy to finalize a CNN model with training from scratch. Therefore, TL was an alternative to training the model on a limited set of training datasets. Various models were trained and tested in this chain of model training processes, but ResNet-50, as the backbone for TL with two FC layers and a sigmoid activation function, performed well compared to other pre-trained backbones shown in Table 4.

The incorporation of early stopping is a strategic measure to counter overfitting. By monitoring the model's performance on a validation set and halting training when improvement plateaus, early stopping prevents the model from memorizing the training data and promotes better generalization to new, unseen images. Various optimizers, including Adam, RMSprop, and SGD with momentum, were employed to fine-tune the model's learning process. In this context, SGD with momentum stood out, demonstrating effectiveness in reaching a desirable solution for the identification of osteochondroma. SGD is generally slower than Adam optimization but exhibits smoother convergence, which can be advantageous for model training due to reduced oscillations, enhancing stability during optimization. Oscillations refer to rapid changes in parameter values during training, potentially hindering convergence towards the optimal solution. In conclusion, the success of the CNN model for osteochondroma identification can be attributed to a combination of factors, including dataset size, diversity, balanced distribution, pre-processing techniques, online and offline augmentation, early stopping, and strategic use of transfer learning and optimizers. This holistic approach results in a robust and generalizable model capable of accurate identification in real-world scenarios.

Theoretical Implications:

This study's successful integration of transfer learning and convolutional neural networks contributes to advancing the theoretical understanding of automated medical image analysis. The proposed framework demonstrates the potential of leveraging pre-trained models and fine-tuning them for specific diagnostic tasks, paving the way for more efficient and effective deep-learning applications in the medical domain.

Practical Implications:

From a practical standpoint, this research holds significant implications for enhancing the diagnostic capabilities of healthcare professionals in orthopaedics. The accurate and efficient identification of osteochondroma through the proposed model can streamline diagnostic processes, reducing the risk of misdiagnosis and enabling timely medical interventions, ultimately improving patient outcomes and overall quality of care.

Ablation Study:

The ablation study systematically evaluated the contribution of each component by removing or modifying them individually and observing the impact on model performance. The findings validated the effectiveness of data augmentation, transfer learning with the ResNet-50 backbone, pre-processing techniques like resizing and normalization, choice of ResNet-50 architecture, and inclusion of dropout layers. The study quantified how the absence of these components adversely affected accuracy, highlighting their significance in achieving robust and accurate osteochondroma identification. This comprehensive analysis reinforced the proposed methodology's technical robustness and provided insights into the interplay of various elements for optimal performance.

Conclusion and Future Work:

This study presents a robust approach to osteochondroma identification by integrating convolutional neural networks (CNNs) and transfer learning techniques. Leveraging the pre-trained ResNet-50 architecture as a backbone, fine-tuned with dense layers and an output layer for binary classification, the proposed model remarkably distinguishes between normal and osteochondroma-affected X-ray images. Strategic incorporation of data pre-processing, augmentation techniques, and a balanced dataset enhanced the model's generalization capabilities and mitigated overfitting. The 97.67% testing accuracy and consistent performance on the public Potato Leaf Disease dataset underscore the model's efficacy and versatility across diverse imaging domains. While promising, future endeavours could explore extending the model's capabilities to the multi-class classification of various bone diseases and integrating larger, more diverse datasets to refine robustness and real-world adaptability. This research lays a foundation for integrating automated diagnostic systems in orthopaedic facilities, empowering practitioners with accurate and efficient osteochondroma identification, contributing to musculoskeletal diagnostics and improving patient care.

Conflicts of Interest: "The authors declare no conflicts of interest."

References:

[1] K. Tepelenis et al., "Osteochondromas: An Updated Review of Epidemiology, Pathogenesis, Clinical Presentation, Radiological Features and Treatment Options," In Vivo (Brooklyn)., vol. 35, no. 2, pp. 681–691, Mar. 2021, doi: 10.21873/INVIVO.12308.

[2] M. D. Murphey, J. J. Choi, M. J. Kransdorf, D. J. Flemming, and F. H. Gannon, "Imaging of Osteochondroma: Variants and Complications with Radiologic-Pathologic Correlation1," https://doi.org/10.1148/radiographics.20.5.g00se171407, vol. 20, no. 5, pp. 1407–1434, Sep. 2000, doi: 10.1148/RADIOGRAPHICS.20.5.G00SE171407.

[3] S. M. M. Doctors Marco Cañete P, Elena Fontoira M, Begoña Gutiérrez San José,

"Osteochondroma: radiological diagnosis, complications and variants," 2013, [Online]. Available: https://www.webcir.org/revistavirtual/articulos/septiembre13/chile/ch_ing.pdf

[4]   E. Farooq, M. A. Nawaz Ui Ghani, Z. Naseer, and S. Iqbal, "Privacy Policies' Readability Analysis of Contemporary Free Healthcare Apps," 2020 14th Int. Conf. Open Source Syst. Technol. ICOSST 2020 - Proc., Dec. 2020, doi: 10.1109/ICOSST51357.2020.9332991.

[5]   T. Zhao and H. Zhao, "Computed Tomographic Image Processing and Reconstruction in the Diagnosis of Rare Osteochondroma," Comput. Math. Methods Med., vol. 2021, 2021, doi: 10.1155/2021/2827556.

[6]   E. Farooq and A. Borghesi, "A Federated Learning Approach for Anomaly Detection in High Performance Computing," Proc. - Int. Conf. Tools with Artif. Intell. ICTAI, pp. 496–500, 2023, doi: 10.1109/ICTAI59109.2023.00079.

[7]   L. Pinto-Coelho, "How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications," Bioeng. 2023, Vol. 10, Page 1435, vol. 10, no. 12, p. 1435, Dec. 2023, doi: 10.3390/BIOENGINEERING10121435.

[8]   F. M. Javed Mehedi Shamrat et al., "LungNet22: A Fine-Tuned Model for Multiclass Classification and Prediction of Lung Disease Using X-ray Images," J. Pers. Med., vol. 12, no. 5, p. 680, May 2022, doi: 10.3390/JPM12050680/S1.

[9]   N. Ullah, M. S. Khan, J. A. Khan, A. Choi, and M. S. Anwar, "A Robust End-to-End Deep Learning-Based Approach for Effective and Reliable BTD Using MR Images," Sensors 2022, Vol. 22, Page 7575, vol. 22, no. 19, p. 7575, Oct. 2022, doi: 10.3390/S22197575.

[10]   M. Thilagaraj, N. Arunkumar, and P. Govindan, "Classification of Breast Cancer Images by Implementing Improved DCNN with Artificial Fish School Model," Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/6785707.

[11]   H. Malik et al., "A Novel Fusion Model of Hand-Crafted Features With Deep Convolutional Neural Networks for Classification of Several Chest Diseases Using X-Ray Images," IEEE Access, vol. 11, pp. 39243–39268, 2023, doi: 10.1109/ACCESS.2023.3267492.

[12]   U. C. Aytaç, A. Güneş, and N. Ajlouni, "A novel adaptive momentum method for medical image classification using convolutional neural network," BMC Med. Imaging, vol. 22, no. 1, pp. 1–12, Dec. 2022, doi: 10.1186/S12880-022-00755-Z/FIGURES/5.

[13]   F. R. Eweje et al., "Deep Learning for Classification of Bone Lesions on Routine MRI," EBioMedicine, vol. 68, p. 103402, Jun. 2021, doi: 10.1016/j.ebiom.2021.103402.

[14]   K. Sampath, S. Rajagopal, and A. Chintanpalli, "A comparative analysis of CNN-based deep learning architectures for early diagnosis of bone cancer using CT images," Sci. Reports 2024 141, vol. 14, no. 1, pp. 1–9, Jan. 2024, doi: 10.1038/s41598-024-52719-8.

[15]   W. Chen et al., "A fusion of VGG-16 and ViT models for improving bone tumor classification in computed tomography," J. Bone Oncol., vol. 43, p. 100508, Dec. 2023, doi: 10.1016/J.JBO.2023.100508.

[16]   R. Anand, S. V. Lakshmi, D. Pandey, and B. K. Pandey, "An enhanced ResNet-50 deep learning model for arrhythmia detection using electrocardiogram biomedical indicators," Evol. Syst., vol. 15, no. 1, pp. 83–97, Feb. 2024, doi: 10.1007/S12530-023-09559-0/METRICS.