JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EARLY DETECTION OF KNEE ARTHRITIS USING AI: ENHANCING DIAGNOSTIC ACCURACY WITH DEEP LEARNING

Advait Kothuri
Student, Chantilly High School, Chantilly, United States

Abstract : Knee arthritis is a significant cause of chronic pain and disability for millions of people worldwide. Early detection of this disease is critical to effective intervention. However, the conventional manual diagnosis using X-ray images is usually subjective, time-consuming, and prone to human error. This work, therefore, proposes a deep learning-based automatic model to classify knee arthritis into five different grades: Normal, Doubtful, Mild, Moderate, and Severe. This study utilizes CNNs and transfer learning to retrain the models of EfficientNetB0, DenseNet121, and InceptionV3 using a custom dataset of X-ray images of knees. These were resized to a standard and normalized before initiating the training process. As a result, EfficientNetB0 showed excellent performance in predicting Moderate and Severe cases of arthritis, which is vital for early treatment, with an accuracy score of 85%. Densenet121 and InceptionV3 performed commendably, with 80% and 79% accuracy, respectively. However, all models showed some struggles in classifying Mild cases, which indicates that further refinement is needed to improve early-stage detection. Assessment metrics, including confusion matrices and classification reports, have shown a better recall for Severe cases but struggle to identify subtle early symptoms. These findings show that artificial intelligence-driven models can help decrease diagnostic variability and accelerate the diagnosis of arthritis, especially in areas where health services are inadequate. Future work will involve data diversification to extend generalizability and incorporate explainable AI methods to garner clinician buy-in. These innovations could bring about a revolution in the diagnosis of arthritis, significantly improving outcomes for patients.

IndexTerms - knee arthritis, osteoarthritis, X-ray, deep learning, CNN, transfer learning, EfficientNetB0, DenseNet121, InceptionV3

I. INTRODUCTION

Arthritis is among the most common conditions that usually result in chronic pain and loss of mobility in millions of people worldwide; the most susceptible part is knee arthritis. Treatment of arthritis, especially at early stages, requires the diagnosis to be precise (Reference). The X-ray image diagnosis of arthritis can be challenging since the differences among the stages of the disease are not easily marked. There is a need for an automated system to classify knee arthritis X-ray images into five stages: normal, doubtful, mild, moderate, and severe. Doctors might then identify conditions more rapidly and precisely using machine learning, which could imply better patient care. Equally important is the limitation of manual diagnoses, which may be very slow and full of human errors. Automating this classification would make diagnoses more consistent and speedier, especially in areas with poor access to specialized doctors. The AI-based classification system will also recognize patients who require emergency medical care, thus ensuring that those with debilitating arthritis get quality and timely treatment. The approach hastens the diagnosis and ensures patients receive proper and timely treatment. Artificial intelligence (AI) in medical imaging, particularly for musculoskeletal conditions like arthritis, has gained significant attention in recent years. Many works also target the implementation of machine learning algorithms, especially CNNs, for diagnosing and classifying arthritis from X-ray images. It was

demonstrated by Tiulpin et al. 2018 that CNNs have great potential in quantifying the severity of osteoarthritis. The results obtained were close to those from trained radiologists. These results point to particular possibilities of AI improving diagnostic precision and facilitating workflow, possibly reducing human errors regarding the detection and classification of arthritis severity. However, specific challenges still exist to disseminating AI for arthritis diagnosis. First, there is an issue with the datasets these models have used for training. Most reviewed studies concluded that one needs to move toward more diverse and comprehensive datasets that would allow generalization across a wide range of populations and conditions by AI models. Although deep learning models are very effective, their "black box" nature has raised concerns about interpretability and trust in clinical decision-making. Ongoing research is directed toward improving the interpretability of these models so they will smoothly fit into clinical practice, providing accuracy with explainability in real-world healthcare settings. The model is designed to automate the classification of knee arthritis X-ray images into five stages: Normal, Doubtful, Mild, Moderate, and Severe. This will help doctors to develop a better tool to identify the progression of arthritis with more consistency and precision. Early detection of the stages of arthritis is critical to ensure that patients receive the proper care at the right time, which helps to delay disease progression and improves long-term outcomes. The model aims to handle the limitations brought about by manual diagnoses, which are time-consuming, subjective, and prone to human error. One of the key objectives of the model is to reduce the diagnostic variability between radiologists. Even experienced doctors have differing opinions on the severity of arthritis when reviewing the same X-ray images, particularly in the early stages of the disease, where differences are subtle. It also has the feature to perform diagnosis consistently so that inconsistencies can be reduced by bringing a uniform, automated classification system. Such diagnosis consistency is crucial in remote or underserved areas where well-trained specialists might not be readily available. With an AI-based tool, primary care doctors or general practitioners could make more accurate assessments to ensure that patients with severe arthritis get timely referrals to specialists. Another critical objective is that it should allow doctors to follow the evolution of arthritis over time. This model will enable doctors to differentiate between the different stages of arthritis and, hence, provide better patient monitoring. Doctors could alter their treatment plans to slow further degradation if a patient progresses from mild to moderate in several months. Classifying patients' stage-wise also enlightens them about more precise functioning or facts about their condition, helping them understand how their disease progresses and what interventions may be necessary. Apart from achieving high accuracy in the classification tasks, the model seeks to address a key challenge with AI solutions in healthcare: trust and interpretability. Many current models in AI work as "black boxes" and provide results without explanations; hence, doctors are skeptical about relying on their output. This paper focuses on making the model's predictions interpretable, so doctors understand each classification's reasoning. It brings in explainable AI to help the model gain the trust of healthcare professionals for better clinical adoption. The overall aim is to develop a practical and scalable solution that will enhance speed and accuracy in diagnosing arthritis, thereby helping doctors with better guidance for patients.

II. DATASET

The dataset utilized in this study consists of images designated for classifying various stages of knee arthritis. These images are organized into categories representing different severity levels: 'Normal,' 'Doubtful,' 'Mild,' 'Moderate,' and 'Severe.' Figure 1 "Images of doubtful, mild, moderate, severe cases of arthritis" illustrates an example of these different severity levels. The grayscale images undergo preprocessing steps, including resizing and normalization. After preprocessing, each image is resized to a consistent 300x160 pixels. This standardization ensures continuity across the dataset, which is essential for practical model training.

Figure 1: Images of doubtful, mild, moderate, severe cases of arthritis

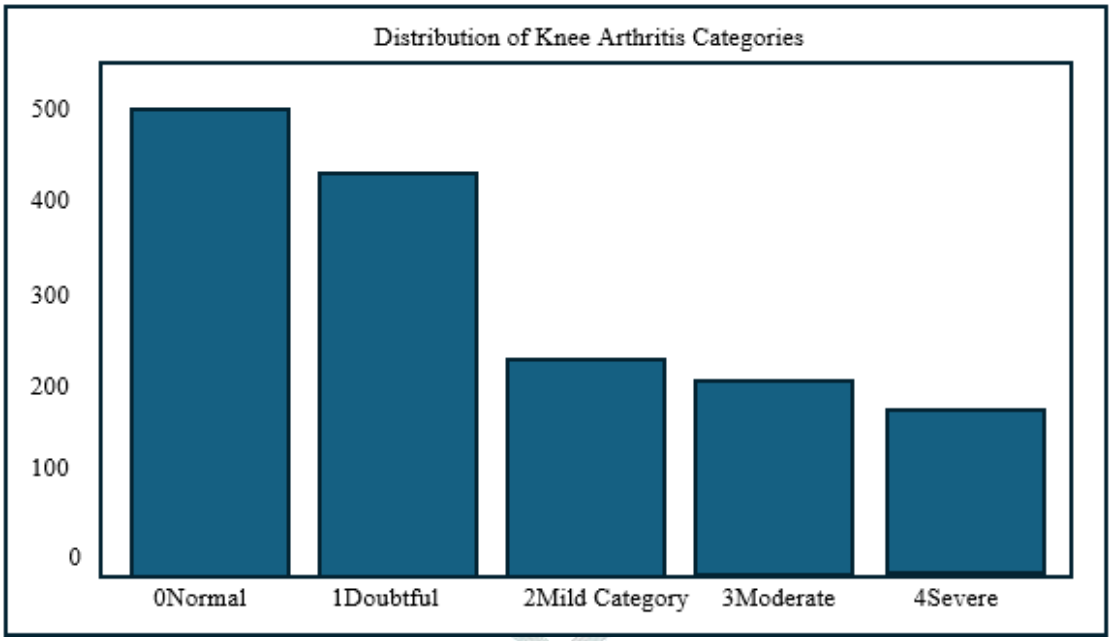

Figure 2: Count distribution of different knee arthritis categories.

III. METHODS AND MODELS

CNN (Convolutional Neural Networks): A CNN is a class of deep learning models visualized to operate on data presented in grid-like structures, for instance, images, videos, and time-series data. The most crucial reason why CNNs outperform other algorithms in computer vision tasks is that they automatically learn and detect spatial hierarchies of features, ranging from low-level details, such as edges, to high-level patterns, such as objects, without explicit manual feature engineering. The key word is that CNNs efficiently preprocess data with local dependencies using convolution operations. This is quite an improvement compared to traditional neural networks, where each input pattern is considered independently. Moreover, CNNs emulate the hierarchical processing of a visual signal in the human brain, where neurons in the visual cortex are responsible for feature recognition at different spatial scales. These bio-inspired architectures improve generalization capabilities in models for a wide range of visual contexts and tasks. CNN architecture consists of three major components: convolutional, pooling, and fully connected layers. Convolutional Layers: In convolutional layers, filters or kernels slide over the input data, capturing features at a specific spatial location. Typically, pooling layers follow convolutional layers to down-sample data to decrease computational load while making the

network invariant or robust to spatial transformations like translation and rotation. Fully connected layers serve to map the learned features to what is usually the ultimate output, typically a classification decision. Finally, an activation function, such as ReLU, follows each convolutional layer to introduce nonlinearities into the network to enrich its capabilities of capturing complicated patterns. Training the CNN involves updating weights through backpropagation and optimization algorithms, such as stochastic gradient descent, through iteratively minimizing the loss function over many iterations of data. CNNs have revolutionized several areas besides computer vision, including medical imaging, speech recognition, and natural language processing. In medical applications, CNNs are employed to analyze radiological images to assist in anomaly detection, such as tumors or bone fractures, with a very high degree of accuracy. In this line, CNNs are also crucial in object detection in real-time for autonomous driving systems to find pedestrians, traffic signs, and other vehicles. More recently, innovation in CNN architectures such as AlexNet, VGG, ResNet, and Inception has shown impressive performance improvements, establishing new benchmarks on accuracy in image classification and object detection tasks. This has helped increase not only the functionalities but also the application areas for the deep learning models. InceptionV3: InceptionV3 is a very advanced, deep convolutional neural network that tries to achieve an optimal balance between computational efficiency and image classification accuracy. It was developed as part of the Google Inception family, which introduced the notion of "Inception modules," which simultaneously perform convolutions at multiple scales. Using parallel convolutions with different filter sizes (1x1, 3x3, and 5x5) and aggregating the outputs, InceptionV3 extracts fine-grained and large-scale image features, making it particularly effective for complex classification tasks. This work employed a pre-trained InceptionV3 model, fine-tuned for classifying knee arthritis stages. Pre-trained layers of the model were trained on the ImageNet dataset and were initially frozen to leverage their ability to detect basic visual patterns. Then, a custom classification head was added consisting of a GlobalAveragePooling2D layer, a dense layer with 512 neurons and ReLU activation, a dropout layer to mitigate overfitting, and a softmax output layer to classify the five arthritis stages. After training the added layers, fine-tuning was done by unfreezing the last few layers of the InceptionV3 base model and training these with a reduced learning rate, adapting the model to the specifics of the dataset. EfficientNet: EfficientNet is a state-of-the-art CNN architecture that achieves remarkable performance with minimal computational cost. Unlike traditional CNNs, where the depth or width of the network is often scaled up arbitrarily, EfficientNet uses a principled compound scaling method to systematically scale up all three dimensions of the network: depth, width, and resolution. For this study, EfficientNetB0, the base variant of the EfficientNet family, was fine-tuned for arthritis classification. The pre-trained model was frozen to retain the generic visual features learned from the ImageNet dataset. A custom classification head was added, consisting of a GlobalAveragePooling2D layer, a dense layer with 512 neurons using ReLU activation, a dropout layer to prevent overfitting, and a softmax layer for multi-class classification. The model was then fine-tuned by unfreezing the last few layers and training them with a lower learning rate to adapt the pre-trained weights to the arthritis dataset. EfficientNet's compound scaling and squeeze-and-excitation modules Enhanced the model's ability to focus on the most significant regions of the X-ray images while reducing irrelevant noise. This efficiency enabled EfficientNet to excel in medical imaging tasks where computer resources were sometimes in short supply. Combining EfficientNet's scalability with data augmentation and techniques like the ReduceLROnPlateau callback, the model achieved balanced performance across all arthritis stages, even for the underrepresented classes. DenseNet: DenseNet is a modern CNN architecture that maximizes the information flow between layers by connecting each layer to every subsequent layer in a feed-forward fashion. Such dense connections help reduce the vanishing gradient problem, diminish redundancy, and improve feature reuse, making DenseNet highly effective and efficient for medical imaging tasks. In this study, the authors used a 121-layer model of DenseNet and fine-tuned it for the classification of arthritis. The pre-trained weights from the model were frozen, which had been trained on ImageNet. A custom classification head was added, which consisted of a GlobalAveragePooling2D layer, a dense layer of 512 neurons using ReLU activation, a dropout layer to handle overfitting, and finally, a softmax layer for the classification of five stages of arthritis and fine-tuning involved unfreezing the last few dense blocks and retraining them with a reduced learning rate. Densely connected layers in DenseNet allowed the model to grab features from shallow to deep, making it very proficient in describing minute differences among stages in arthritis. Data augmentation techniques and the Adam optimizer were also employed to enhance the model's performance. This efficient use of parameters decreases the chances of overfitting while maintaining the high classification accuracy in DenseNet; hence, it was selected as a robust model for this work.

IV. EVALUATION METRICS

A confusion matrix is a simple evaluation device that helps compare predicted and actual labels. This confusion matrix tabulates the actual classes on the rows and predicted classes on the columns to give a complete breakup of model performance across categories. It shows four key outcomes: true positives, indicating correct predictions for positive instances; true negatives, indicating correct predictions for negative instances; false positives, meaning incorrectly predicted positive instances; and false negatives, which are incorrectly predicted negative cases. That would allow for nuancing in the model's mistakes to identify better the specific weaknesses, such as over-predicting some classes or under-predicting others. For example, a high rate of false negatives could be critical in medical diagnosis because it would mean that your model is missing the disease cases. While the confusion matrix gives some insight into the raw results, the classification report goes into more detail for quantitative analysis of the model's performance with metrics such as precision, recall, F1-score, and support. On the other hand, precision is defined as the ratio of accurate optimistic predictions to all positive predictions; this is very useful in applications where the cost of false positives is prohibitive, such as in spam detection. Recall or sensitivity is the measure that informs about how many relevant instances in the dataset the model can detect; therefore, it is a valuable metric when there are false negatives one would not want to see. The F1-score expresses precision and recall in a single metric by calculating their harmonic mean. Therefore, this gives a good balance between the two when evaluating the accuracy of a model, especially in those cases where precision and recall are at odds. It is the support metric that will essentially provide the number of actual instances per class, thus helping in the evaluation of model performance within the context of a class distribution, making sure example performance is not only for the majority classes but also for the minority ones.

V. MODEL TRAINING AND TRANSFER LEARNING

In this study, several transfer learning models were fine-tuned to classify knee arthritis stages more accurately and efficiently. Transfer learning involves leveraging pre-trained models that have already learned general features from large datasets and adapting them to the specific task. We want to improve these models by reducing training time, improving their work with small medical datasets, and helping the model perform better. We selected three top CNN architectures for this study: EfficientNetB0, InceptionV3, and DenseNet121. Each model was modified in some ways to improve its performance on the knee arthritis dataset. The base layers of these models were initially frozen to retain the pre-trained knowledge of general image features, such as edges and textures. Subsequently, new layers were added to the models to adapt them for classifying arthritis stages. After training the new layers, fine-tuning was performed by unfreezing select layers from the base models to improve performance by allowing the models to learn more task-specific features from the dataset. Layer Modifications For each model, the number of layers unfrozen varied depending on the architecture’s depth and complexity. In the case of EfficientNetB0, the final 20 layers of the base model were unfrozen to allow the network to learn task-specific features from the arthritis images. Similarly, for InceptionV3, the last 50 layers were unfrozen, while for DenseNet121, the final dense blocks were unfrozen to refine the model’s performance on the target dataset. After unfreezing these layers, we added custom classification heads tailored to the arthritis classification task. The new layers included a GlobalAveragePooling2D layer to reduce the dimensionality of the feature maps and retain the most critical features. This was followed by a fully connected Dense layer with 512 neurons and ReLU activation, which introduced non-linearity to enhance the model's learning capacity. A Dropout layer with a dropout rate 0.5 was added to mitigate overfitting by randomly deactivating 50% of neurons during each training iteration. Finally, a Dense output layer with five neurons and a softmax activation function was added to classify the images into the five categories of arthritis severity: Normal, Doubtful, Mild, Moderate, and Severe. Compilation and Optimization Each modified model was compiled using the Adam optimizer, widely used for its adaptive learning rate and efficient handling of sparse gradients. The loss function selected was sparse categorical cross entropy, appropriate for multi-class classification problems where the target labels are integers. The model's performance was evaluated using accuracy as the primary metric, ensuring the model's ability to classify images into the correct arthritis stages. A learning rate scheduler (ReduceLROnPlateau) was employed to optimize training further. This callback function monitored the validation loss during training and reduced the learning rate by 0.5 when the validation loss plateaued for five consecutive epochs. Dynamical learning rate adaptation helped to enhance the convergence of the model by allowing it to take more significant steps during its earlier training while taking smaller steps when it gets closer to finding the optimal solution. Training Procedure The training process

consisted of two stages: initial training with frozen base layers and subsequent fine-tuning with unfrozen layers. Only the newly added layers were trained during the initial training phase, allowing the model to adapt the high-level features learned from the base model to the arthritis dataset. The fine-tuning phase began once the new layers reached satisfactory performance. In the fine-tuning phase, select layers from the pre-trained base models were unfrozen, and the entire model was trained with a reduced learning rate. This step allowed the models to adjust their pre-trained features better to capture the specific patterns in knee arthritis X-ray images. The models were trained for 100 epochs, with early stopping criteria to prevent overfitting. Data augmentation techniques were applied during training to improve the model's robustness by introducing variations in the input images, such as rotations, zooms, and flips. Adding previous knowledge, tailor-made classification heads, and fine-tuning improved the performance measures of all three models, thus yielding balanced accuracy scores for the different stages of arthritis, including even the rarer classes like Mild and Severe.

VI. RESULTS AND DISCUSSION

Base CNN Model:

EfficientNetB80 CNN implementation Model: The classification report for the EfficientNetB0 model (shown in Figure 9) The model achieved an overall accuracy of 85%, indicating that it correctly classified 85% of the test images across all categories. The macro average F1 Score and the weighted average F1 Score were both 0.85, demonstrating consistent performance across all five categories.

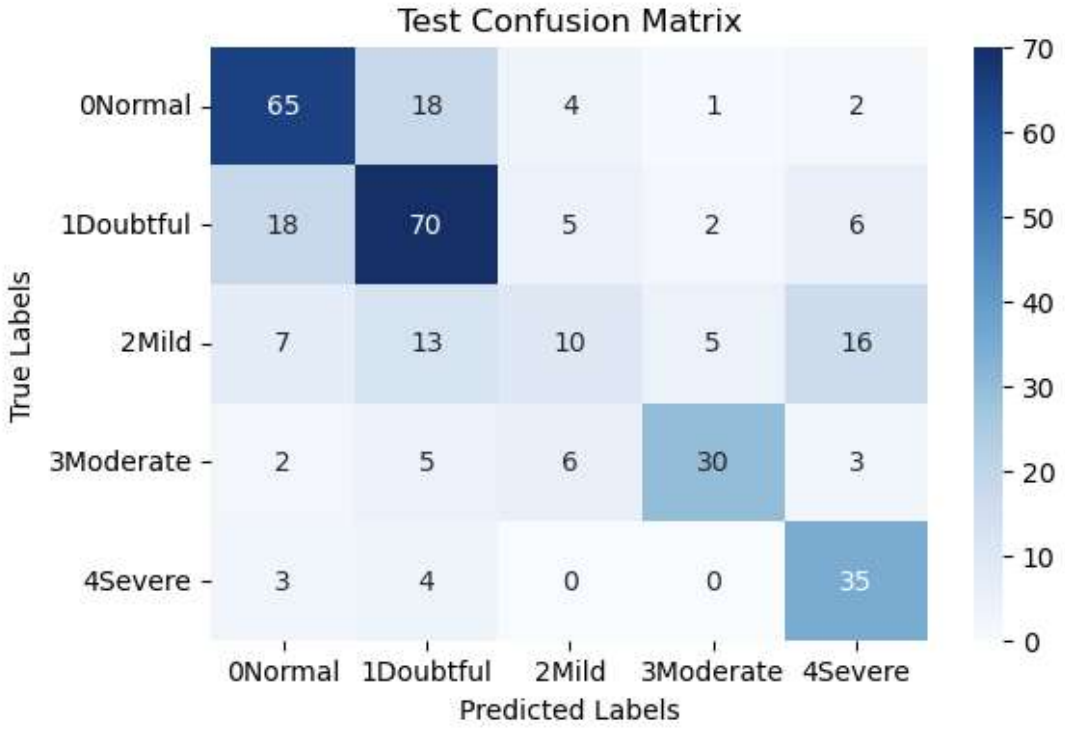

Figure 4: Testing Data Confusion Matrix.

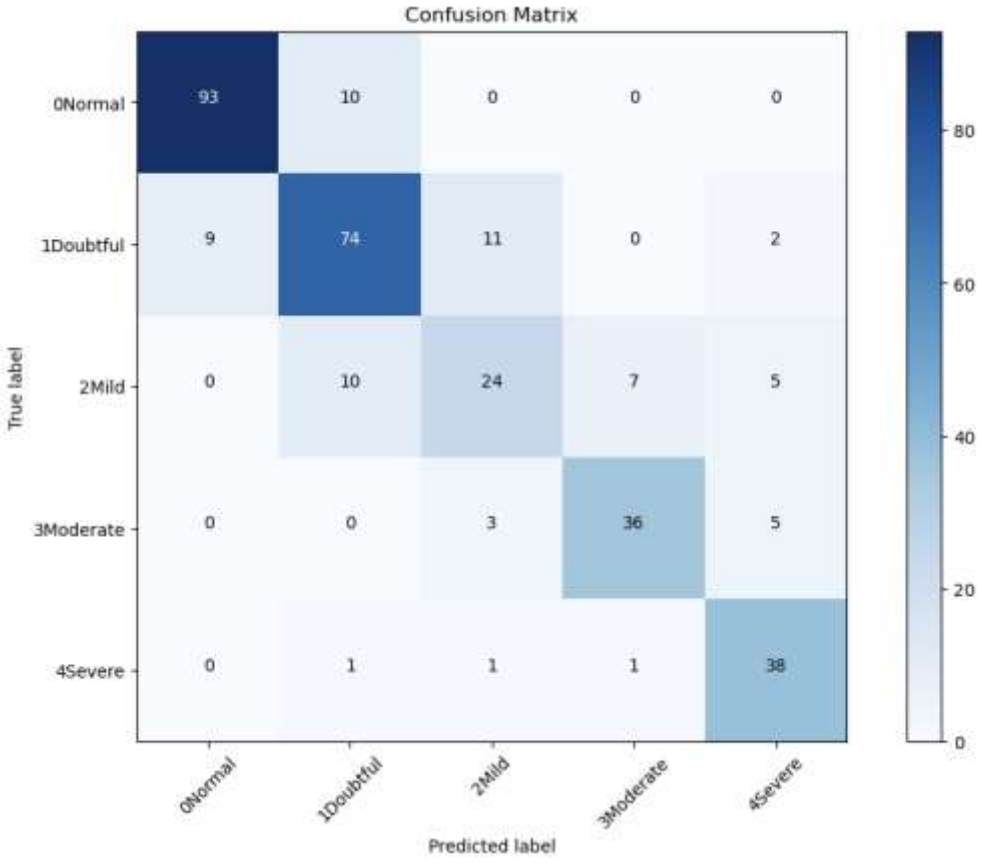

Figure 6: Testing Data Confusion Matrix (DenseNet121).

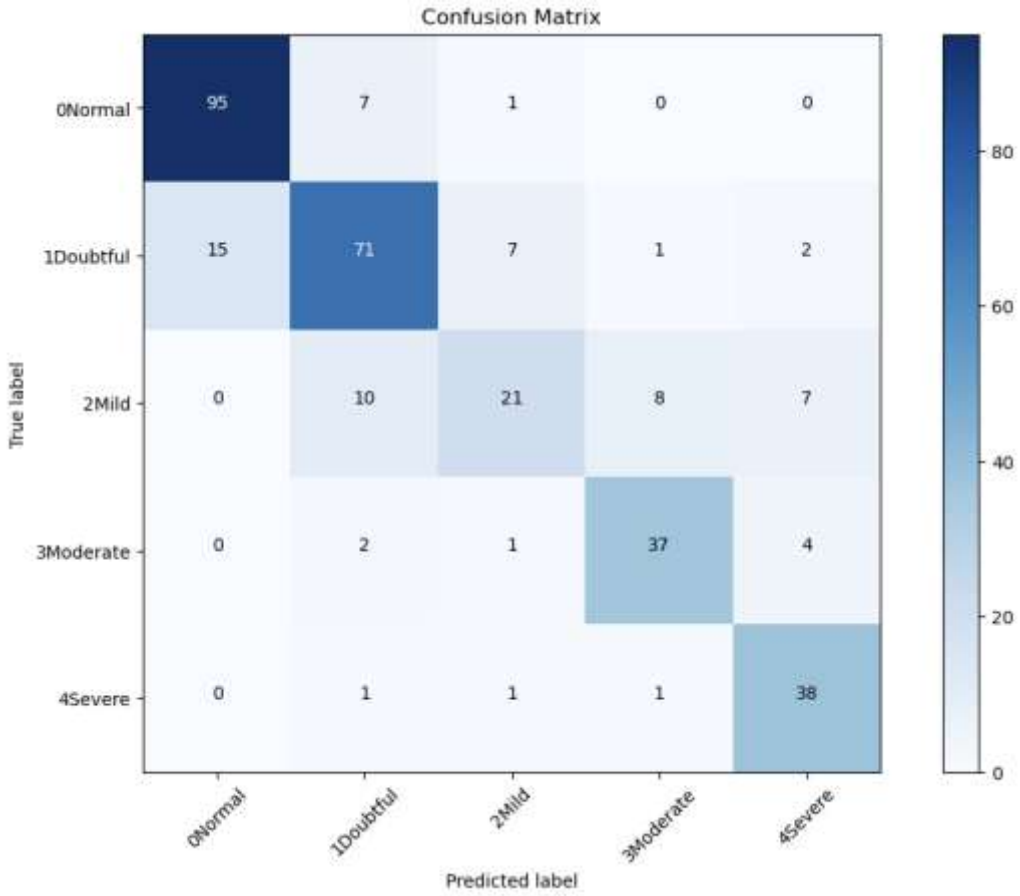

Figure 8: Testing Data Confusion Matrix (InceptionV3).

VII. CONCLUSION

The main goal of this research was to develop and optimize convolutional neural network models for automatically classifying knee arthritis stages based on X-ray images. This study attempted to overcome the drawbacks of manual diagnosis., which can be subjective, time-consuming, and prone to human error. Among the models tested, EfficientNetB0 outperformed the others, achieving the best balance between precision, recall, and scores across all categories. The EfficientNetB0 model identified Normal, Moderate, and severe arthritis cases; it achieved a high F1 Score for the moderately affected group and good recall for cases belonging to the severe class, an essential factor for early medical intervention. However, it struggled to classify cases with mild arthritis, indicating that more refinement is required to diagnose the early stages of arthritis. The DenseNet121 and InceptionV3 architectures achieved similar performances, with overall accuracies of 80% and 79%, respectively. All models distinguished between the regular and severe classes well but performed significantly poorly in correctly classifying cases with mild effects. The baseline convolutional neural network performed the worst, further ascertaining the benefits of using pre-trained transfer learning models in medical imaging tasks.

VIII. SIGNIFICANCE AND IMPACT

The findings of this study have important implications for the medical community. Deep learning can help automate the classification of stages of knee arthritis, decreasing diagnostic variability and increasing diagnostic accuracy. This can speed up the detection of arthritis, mainly in resource-poor areas. Eventually, this will lead to earlier treatment interventions that have the potential to slow disease progression and thereby reduce the need for expensive imaging techniques like MRIs. The paper presents how artificial intelligence can help radiologists manage their workload and engage with patients. An automated framework can also be used as a decision-support system that consistently and reliably offers classifications, improving patients' outcomes. Additionally, these models integrated into mobile applications or cloud-based platforms bear enormous potential for enhancing accessibility to diagnostics in remote and under-resourced locations.

IX. LIMITATIONS

While these findings are encouraging, the present study has several limitations: the dataset used to test the approach included a relatively small number of examples; this may be insufficient to generalize the model to more extensive and more diverse populations. Additionally, the dataset could include biases tied to specific demographics that may lead to less accurate generalization when applying the model to patients belonging to other demographics. The model also struggled to classify Mild cases accurately, highlighting the need for further improvements in feature extraction to capture subtle differences in early-stage arthritis.

X. REFERENCES

1. Tiulpin, A., Thevenot, J., Rahtu, E., Lehenkari, P., & Saarakkala, S. (2018).Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. Scientific Reports, 8(1), 1727. https://doi.org/10.1038/s41598-018-20132-7

2. Antony, J., McGuinness, K., O'Connor, N. E., & Moran, K. (2016). Automaticdetection of knee osteoarthritis from X-ray images using deep learning. InternationalSymposium on Biomedical Imaging. https://doi.org/10.1109/ISBI.2016.7493411

3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification withdeep convolutional neural networks. Advances in Neural Information ProcessingSystems, 25, 1097–1105. https://doi.org/10.1145/3065386

4. Zhang, X., Wang, Z., Liu, D., & Li, Y. (2020). EfficientNet: Rethinking modelscaling for convolutional neural networks. Proceedings of the 36th InternationalConference on Machine Learning. https://arxiv.org/abs/1905.11946

5. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., & Ghafoorian, M.(2017). A survey on deep learning in medical image analysis. Medical ImageAnalysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

6. Wang, S., Yin, Y., Cao, G., Wei, B., Zheng, Y., & Yang, G. (2021). Interpretabilityof deep learning in healthcare: A technical overview. IEEE Access, 9, 103931–103946. https://doi.org/10.1109/ACCESS.2021.3089824

7. Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., & Mehta, H. (2018). Deeplearning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXtalgorithm to practicing radiologists. PLOS Medicine, 15(11), e1002686.https://doi.org/10.1371/journal.pmed.1002686

8. Shamir, L., Ling, S. M., Scott, W. C., & Ferrucci, L. (2009). Knee X-ray imageanalysis to detect osteoarthritis. BMC Medical Imaging, 9(1), 1–9.https://doi.org/10.1186/1471-2342-9-10

9. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., & Blau, H. M. (2017).Dermatologist-level classification of skin cancer with deep neural networks. Nature,542(7639), 115–118. https://doi.org/10.1038/nature21056

10. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L.(2018). Artificial intelligence in radiology. Nature Reviews Cancer, 18(8), 500–510.https://doi.org/10.1038/s41568-018-0016-5

11. Zhang, Z., Yang, L., & Zheng, Y. (2021). Classification of osteoarthritis severityusing deep learning on X-ray images. Journal of Medical Imaging, 8(4), 041206.https://doi.org/10.1117/1.JMI.8.4.041206

12. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention. https://arxiv.org/abs/1505.04597

13. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436– 444. https://doi.org/10.1038/nature14539

14. Ghorbani, A., Ouyang, D., Abid, A., He, B., & Chen, P. H. C. (2019). Deep learning interpretation of echocardiograms. Nature Medicine, 26(5), 886–891.
https://doi.org/10.1038/s41591-019-0447-x

15. Thomas, D. J., & Young, S. W. (2020). The role of AI in predicting osteoarthritisprogression. Orthopaedic Journal of Sports Medicine, 8(1), 2325967120903084.
https://doi.org/10.1177/2325967120903084

16. Ma, D., He, X., & Zhang, Y. (2019). Explainable AI: Developing trust in medical AIpplications. Computers in Biology and Medicine, 114, 103491.
https://doi.org/10.1016/j.compbiomed.2019.103491

17. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., & Anguelov, D. (2015). Goingdeeper with convolutions. Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2015.7298594

18. Liu, F., Zhou, Z., Samsonov, A., Blankenbaker, D., Larison, W., & Kanarek, A.(2018). Deep learning for knee MRI analysis. Radiology, 289(1), 160–169.
https://doi.org/10.1148/radiol.2018171843

19. Chesbrough, H. W., & Appleyard, M. M. (2007). Open innovation and arthritisresearch. California Management Review, 50(1), 57–76.
https://doi.org/10.2307/41166416

20. World Health Organization. (2023). Global burden of arthritis. World HealthOrganization. https://www.who.int/news-room/fact-sheets/detail/arthritis