IECE

RESEARCH ARTICLE

Comparative Analysis of Automated Knee Osteoarthritis Severity Classification from X-Ray Images Using CNNs and VGG16 Architecture

Sundas Iqbal[1], Danish Ali[2,*], Sumaria Rafique[3], Kajol Bagga[4], Atif Ur Rahman[5] and Salman Khan[6]

[1] School of Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

[2] School of Computer Science, Wuhan University, Wuhan 430072, China

[3] National University of Computer and Emerging Sciences, Lahore 54770, Pakistan

[4] Department of Computer Science, Technical University of Applied Sciences Würzburg-Schweinfurt, 97070 Würzburg, Germany

[5] Department of Computer Science, Iqra National University, Swat 19200, Pakistan

[6] CodeNinja Inc., Lahore 54792, Pakistan

Abstract

Osteoarthritis (OA) is a degenerative joint disease that primarily affects the knee, causing cartilage deterioration and discomfort. Early diagnosis is crucial for effective management, as it can slow disease progression and improve the quality of life. This study proposes a deep learning approach to automatically classify knee OA severity from X-ray images using Convolutional Neural Networks (CNNs) and the VGG16 model. The models were trained on a dataset of knee X-ray images, and performance was evaluated using accuracy, precision, recall, and F1-score. The proposed CNNs model achieved 99% training accuracy and 80% testing accuracy after 50 epochs, while the VGG16 model, after fine-tuning for 20 epochs, reached 99.8% training accuracy and 91% testing accuracy. Additionally, a web app was developed using Streamlit, enabling users to upload knee X-ray images and receive real-time predictions of OA severity. The results demonstrate the effectiveness of CNN and transfer learning with VGG16 in knee OA classification, offering a promising tool for early diagnosis in clinical settings.

Keywords: knee osteoarthritis, biomedical imaging, transfer learning, knee disease.

1 Introduction

Knee osteoarthritis (OA) is a degenerative joint disease characterized by the deterioration of cartilage in the knee joint, leading to discomfort, stiffness, and reduced mobility [1]. Early diagnosis is critical for preventing disease progression and enabling timely and effective treatment. Automated detection of knee OA is essential to ensure prompt and efficient patient care

[2]. Early intervention can slow disease progression, reduce the likelihood of adverse outcomes, and significantly improve patients' quality of life by alleviating pain and enhancing mobility [3]. Moreover, early-stage treatment is generally less costly compared to advanced-stage interventions, which often require more complex and expensive therapies [3].

Compared to conventional diagnostic methods, automated identification and diagnosis of knee OA using deep learning techniques offer higher accuracy and consistency [4], thereby reducing the risk of misdiagnosis and ineffective treatment [5]. This is particularly important in cases where knee OA symptoms overlap with those of other conditions, making diagnosis challenging for clinicians. Automated systems can alleviate the burden on healthcare professionals, freeing up resources for other medical needs and reducing strain on the healthcare system [6]. Osteoarthritis affects approximately 250 million people worldwide, accounting for 4% of the global population. Risk factors include age, obesity, previous joint injuries, occupation, and genetic predisposition. Given the knee joint's critical role in daily activities, the progression of OA can significantly impair mobility and quality of life [7]. Therefore, early detection is crucial to prevent the disease from advancing to a severe stage. The most common diagnostic method for OA is joint X-ray imaging [8].

Deep learning (DL) techniques have emerged as powerful tools for medical image analysis, with significant potential for assisted living applications [9]. These techniques have shown great promise in the automated diagnosis and classification of knee OA. However, prior studies often relied on limited datasets, leading to models that perform well during training but fail to generalize to diverse real-world cases. Additionally, imbalanced distributions of OA severity levels in datasets can introduce bias, resulting in reduced model accuracy. Recent advances in deep learning, particularly convolutional neural networks (CNNs), have demonstrated high performance in addressing these challenges. CNNs automatically learn hierarchical features from complex images, which are crucial for detecting variations in knee joint degradation. Among these models, VGG16 stands out due to its deep architecture and effective use of transfer learning. By leveraging pre-trained weights from large datasets, VGG16 can achieve robust performance even with limited data, making it highly relevant for this study.

Transfer learning, a machine learning (ML) approach, enables the reuse of pre-trained models on new but related tasks without the need to train from scratch [10]. This technique is particularly beneficial for image classification tasks, where collecting large amounts of labeled data and training deep neural networks from scratch can be time-consuming and costly [11]. In the context of medical image classification, transfer learning involves adapting models pre-trained on large datasets (e.g., ImageNet) to biomedical imaging data, including MRI scans, CT scans, and X-rays [12, 13]. The primary objective of transfer learning in medical imaging is to enhance the performance of classification tasks by leveraging knowledge from non-medical domains.

This study aims to develop and evaluate an automated method for classifying knee OA severity using X-ray images, employing convolutional neural networks (CNNs) and transfer learning with the VGG16 architecture. The proposed approach categorizes knee OA into four severity levels: normal, mild, moderate, and severe. The primary research objective is to assess the performance of CNNs and VGG16 in knee OA severity classification and to explore their potential for real-world clinical applications.

The significance of this research lies in its potential to assist healthcare professionals in making faster and more accurate diagnoses, thereby improving patient outcomes. By automating the classification process, this study aims to reduce the workload of medical practitioners and minimize the risk of misdiagnosis. Furthermore, this work contributes to the broader field of medical image analysis, where deep learning techniques have demonstrated significant promise in diagnosing various conditions, including cancer, brain tumors, and cardiovascular diseases. Ultimately, this study seeks to provide a robust, reliable, and efficient tool for knee OA diagnosis, which can be integrated into clinical practice to support early detection and management of the disease. This research aligns with the growing body of work exploring the application of artificial intelligence (AI) in healthcare, particularly in automating the analysis of medical imaging data.

However, the collection of labeled data for training deep neural networks in medical imaging is often challenging due to logistical, ethical, and regulatory constraints [14]. Transfer learning addresses this issue by enabling the reuse of pre-trained models on new medical image classification tasks, thereby reducing the need for large labeled datasets. Given the high

prevalence of knee OA, there is a pressing need for novel methodologies and technologies to evaluate its development, grading, and detection [15, 16]. In this study, we propose a transfer learning-based approach for the automated identification and classification of knee OA using X-ray images with varying severity levels. The dataset includes four severity grades, with Level 0 representing a healthy knee and Level 4 indicating severe osteoarthritis. This study addresses key challenges in knee OA diagnosis, including the need for large labeled datasets, the risk of misdiagnosis, and the burden on healthcare professionals [17]. The structure of this article is as follows: Section 2 provides a comprehensive literature review, Section 3 details the proposed methodology, Section 4 presents the results and evaluation, and Section 5 concludes the study.

2 Related Work

Recently, convolutional neural networks (CNNs) and attention-based approaches have demonstrated remarkable performance in analyzing visual data [18–20]. Researchers have developed and evaluated various CNNs architectures for classifying the severity of knee osteoarthritis (OA) from X-ray images. Transfer learning approaches have also been extensively explored, leveraging models pre-trained on natural image datasets to enhance the efficiency of medical image analysis [21, 22]. Previous studies have reported high accuracy in classifying different grades of knee OA using deep learning (DL) models [23]. The field continues to advance rapidly, with new architectures, larger datasets, and multimodal approaches further improving performance.

Sharma et al. [24] proposed a deep learning model trained on the Knee Osteoarthritis Severity Grading dataset from Kaggle to detect OA in its early stages. The dataset consists of X-ray images of knees labeled as healthy, moderate, or severe. The model was optimized using the Adamax and Adam optimizers, and its performance was evaluated based on recall, accuracy, F1-score, and precision. Adam outperformed Adamax, achieving 93.84% accuracy, 87.80% precision, 85.70% recall, and an F1-score of 0.862, compared to 93.24% accuracy for Adamax. The model successfully classified 1,656 test images with only 102 errors, demonstrating its potential for assessing bone health. Further optimization of this model is recommended to improve early OA detection.

Thomson et al. [25] employed form and surface analysis combined with a random forest classifier to achieve high accuracy in knee OA diagnosis using radiographic images. Cueva et al. [17] proposed a method that integrates two deep neural networks to classify knee joints based on the Kellgren-Lawrence (K.L.) scale, achieving an accuracy of 69.7%. This performance was comparable to that of ResNet and DenseNet, while utilizing an enhanced VGG-19 model. Almhdie-Imjabbar et al. [10] introduced an automated technique for OA diagnosis and prediction. Their approach identifies regions of interest (ROIs) and locates cortical bone plates in radiographs using morphological methods and active shape models. This technique was specifically designed to recognize ROIs in the tibial trabecular bone regions of the knee joint.

Using radiographs, Shamir et al. [26] proposed a template-based method for knee joint identification. Their approach employs a sliding window and pre-specified joint center images as templates, computing Euclidean distances across each patch in an X-ray image. The patch with the smallest distance is identified as the center of the knee joint [11]. Once the center is located, a 700×500-pixel image segment representing the knee joint region is extracted. This method utilizes X-ray images from the BLSA dataset and achieves an accuracy of 83.3%. Antony et al. developed the CaffeNet and VGG16 architectures for knee OA classification. Tiulpin et al. [27] introduced attention maps to highlight radiological features, providing medical practitioners with insights into the decision-making process of deep learning models.

Rehman et al. [28] developed a model for the early identification of osteoarthritis using knee X-ray images. Osteoarthritis, a degenerative joint condition, is characterized by cartilage wear, bone-on-bone contact, stiffness, pain, and loss of function. Early diagnosis can significantly improve patients' quality of life. Their study compared traditional machine learning methods with advanced CNNs, proposing a novel feature engineering approach called CRK (CNNs Random Forest K-neighbors). This method leverages transfer learning, initially using a 2D CNNs to extract spatial features from X-ray images. These features are then fed into random forest and k-neighbors models to generate a probabilistic feature set, which is used to build the final machine learning model. Experiments demonstrated that the CRK model achieves 99% accuracy in osteoarthritis prediction. The model's performance was validated using k-fold cross-validation and hyperparameter tuning. This innovative approach has the potential to revolutionize osteoarthritis prediction by intelligently extracting features through transfer learning.

Table 1. Comparative analysis of related works.

Methods	Model Used / Dataset	Accuracy
Sharma et al. (2023) [24]	Deep Learning (Adam, Adamax) / Knee Osteoarthritis Severity Grading (Kaggle)	93.84% (Training), Not reported (Testing)
Cueva et al. (2022) [17]	ResNet, DenseNet, VGG-19 / K.L. Scale Dataset	Not reported (Training), 69.7% (Testing)
Rehman et al. (2023) [28]	CNNs, Random Forest, K-Neighbors / Not specified	Not specified (Training), 99% (Testing)
Ahmed et al. (2022) [5]	CNNs, Transfer Learning (VGG16, VGG19, ResNet50) / X-ray images dataset	Not reported (Training), 91.51% (Testing, ResNet50)
Ours	CNNs, VGG16 (Transfer Learning) / Osteoarthritis Initiative (OAI) X-ray images	99% (Training, CNNs), 80% (Testing, CNNs), 91% (Testing, VGG16)

Ahmed et al. [5] proposed a CNNs-based method for the identification and classification of knee diseases using X-ray images. Their study combined deep learning methods, such as CNNs, with transfer learning and fine-tuning. The authors evaluated three pre-trained models—VGG16, VGG19, and ResNet50—for the classification task. Among these, ResNet50 outperformed the others, achieving a validation accuracy of 91.51%.

A comparative analysis of the aforementioned studies is presented in Table 1, highlighting the models, datasets, and accuracy achieved by each approach. This comparison underscores the advancements in knee OA classification and provides a foundation for evaluating the performance of our proposed method.

2.1 Gaps in Literature

While many studies have demonstrated the effectiveness of deep learning techniques for knee osteoarthritis (OA) classification, several challenges persist in the field that limit the applicability and reliability of these models. These gaps include:

- **Small and Imbalanced Datasets:** A common limitation in medical image analysis is the lack of large, well-balanced datasets. Small datasets can lead to overfitting, where models perform well on training data but fail to generalize to new, unseen data. Many prior studies on knee OA classification are limited by the availability of labeled X-ray images, which results in models trained on small sample sizes and limited diversity in the data.

- **Class Imbalance:** In many studies, the distribution of severity grades is skewed, with certain classes (e.g., severe OA) having fewer samples than others. This imbalance negatively impacts model performance, as the model becomes biased toward predicting the majority class, leading to reduced accuracy in identifying less frequent severity levels.

- **Overfitting:** Deep learning models, particularly those with a large number of parameters like CNNs, are prone to overfitting when trained on small or imbalanced datasets. This results in high training accuracy but poor generalization on testing data, which is a major concern in deploying these models in real-world clinical settings.

2.2 Proposed Solutions

In this study, we address these challenges by implementing the following strategies:

- **Larger and More Diverse Dataset:** We use the Osteoarthritis Initiative (OAI) dataset, which contains over 12,000 labeled knee X-ray images, to provide a more robust basis for model training.

- **Data Augmentation Techniques:** To mitigate class imbalance, we generate synthetic samples to ensure better representation of all severity levels, improving the model's ability to generalize.

- **Regularization Methods:** Incorporating dropout and other regularization techniques reduces overfitting, allowing the models to generalize effectively to unseen data.

By addressing these limitations, our approach aims to develop a more robust and reliable model for knee OA classification, enabling effective deployment in clinical settings.

3 Methodology

The CNNs and VGG16 models are used to detect and classify Knee Osteoarthritis (OA), which combines the power of convolutional neural networks. The methodology involves training the model on a large dataset of knee joint X-ray images that are annotated with corresponding OA labels. The first step of the proposed method is data gathering.
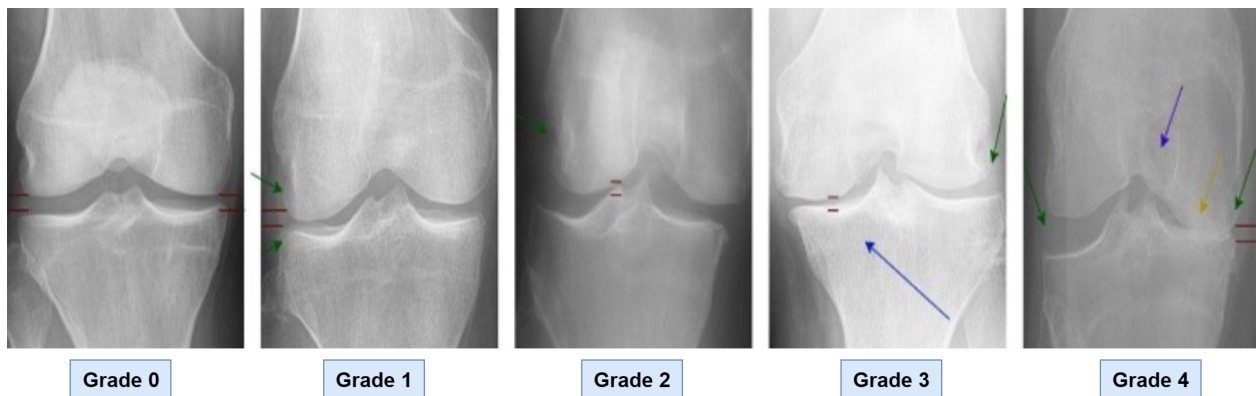
Figure 1. Illustration of knee X-ray grading: Grade 0 shows a normal knee, Grade 1 indicates early osteoarthritis, Grade 2 represents mid-level, Grade 3 is moderate, and Grade 4 signifies severe osteoarthritis.

3.1 Dataset

As shown in Figure 1, the data set of knee X-ray images labeled in accordance with the Kellgren Lawrence (K.L.) scale was used in this study. The dataset was formed by using the data from the Osteoarthritis Initiative (OAI) and also mentioned by Abedin et al. [3]. Data was used to train and validate the suggested model. There are 5947 people in this dataset, with ages ranging from 45 to 79. There are around 12,000 photos in the dataset. Following the loading of the images, the data was divided into 70% and 30% for the model's training and testing, respectively. Grade 0 illustrates the normal knee x-ray, whereas Grade 1 displays a knee that may have osteoarthritis. Knee osteoarthritis is classified as grade 2 for mid-level and grade 3 for moderate. The ultimate grade of 4, or severe osteoarthritis, is used.

3.2 Data preparation

Because the number of images varies with the K.L. scale, the dataset has a problem with class imbalance. The effect of class imbalance on categorization has a detrimental impact on the model's performance [29]. To create a balanced dataset and improve deep learning model performance for small datasets, we use data augmentation. New training data were made from old data by using data augmentation methods. Picture data augmentation is the process of converting identically categorized copies of the source picture. Three image augmentation techniques, rotation, distortion, and flipping, are used to produce the visuals. During the rotation process for picture enhancement, the images are rotated both clockwise and counterclockwise by a maximum of 10 degrees. After that, we randomly deform the image. Finally, we flip the photos vertically and horizontally with a probability of 0.5. We transform them to an array

dataset and resize them to 224 × 224. Finally, we encode every label once. Due to class imbalance in the dataset, where certain categories had fewer images, data augmentation techniques were applied to balance the dataset and improve model performance. The data augmentation techniques used were rotation, flipping, and distortion. After augmentation, the total number of images increased from 12,000 to 36,000, with each image being modified to generate new training samples. This increased dataset size helped the model better generalize and reduced the risk of overfitting.

3.3 Model Selection

Convolutional Neural Networks (CNNs): CNNs are the cornerstone of modern image classification tasks due to their ability to automatically learn relevant features from raw image data without the need for manual feature extraction [30–33]. For knee osteoarthritis (OA) classification, CNNs are particularly well-suited because they can effectively capture spatial hierarchies in X-ray images, enabling the learning of discriminative features at multiple scales, such as bone texture and joint spaces. Prior studies, including those by Jena et al. [23] and Rehman et al. [28], have demonstrated the effectiveness of CNNs in diagnosing knee OA from radiographs, making CNNs a natural choice for this task.

VGG16: VGG16 is a widely used deep learning architecture known for its simplicity and depth. It consists of 16 layers, including 13 convolutional layers, and has achieved impressive results in various image classification tasks. VGG16 has been successfully applied in several medical imaging domains, such as brain tumor detection and breast cancer classification, demonstrating its ability to generalize well to medical images. The architecture of VGG16, characterized by small 3×3 filters and deep convolutional layers,

enables it to learn fine-grained patterns in complex images, such as knee X-rays, which are crucial for accurate OA severity classification. Furthermore, transfer learning with VGG16 allows the utilization of pre-trained weights from large datasets like ImageNet, enhancing model performance even with relatively smaller datasets of knee X-rays.

3.4 VGG16 Architecture

The Visual Geometry Group (VGG) at Oxford University created the deep convolutional neural network architecture known as VGG16. The architecture of VGG16 is illustrated in Figure 2. Its sixteen-weight layers contain three fully connected layers and thirteen convolutional layers. The convolutional layers enable the network to effectively learn hierarchical feature representations from pictures by using max pooling and tiny 3x3 filters for downsampling, as shown in 2. For knee osteoarthritis detection, VGG16 leverages transfer learning. The model is first pre-trained on a large dataset like ImageNet to learn general image features. We then fine-tune this pre-trained model by retraining the top classification layers on our dataset of knee X-ray images labeled with osteoarthritis severity grades. Fine-tuning allows VGG16 to adapt to recognizing features specific to knee OA from the X-rays. In our experiments, we trained two versions of VGG16: one for ten epochs and another for 20 epochs. In order for the model to be able to generate correct predictions on the training data, the model's parameters must be adjusted during the training phase, which is a crucial stage in the machine learning or deep learning process. In our instance, we use X-ray pictures of knee OA to train the suggested model. Once trained, the model may be used to predict the presence or absence of knee OA, if present, the disease's stage.

The additional epochs allow VGG16 to learn better representations for identifying different grades of knee osteoarthritis. The hierarchical feature learning capabilities and transfer learning approach enable VGG16 to classify knee osteoarthritis severity from X-ray images effectively. The model utilizes multiple modules, which consist of parallel convolutional layers of varying kernel sizes, to capture multi-scale features from the input images. The extracted features are then aggregated and fed into fully connected layers for classification. The activation function is used for the output layer. We utilized the proposed model weights to train it on the input data because it is a pre-trained model.

Backpropagation will only update the thick layer, not any other layers. Our data's mean and variance might not match the pre-trained weights, therefore we didn't freeze the batch normalization layers to improve performance. Consequently, we applied the auto-tuning approach to these layers. The suggested model uses the Stochastic Gradient Descent (SGD) as its optimization function. SGD was chosen because it performs better for multiclass classification issues. The performance of the model is assessed using the loss function Categorical Cross Entropy. Through this methodology, the Inception v4 model achieves high accuracy and robustness in detecting and classifying Knee OA, contributing to improved diagnosis and treatment of this debilitating condition.

The VGG16 model uses a series of convolutional layers to extract features from the input images. The output of each convolutional layer is passed through a rectified linear unit (ReLU) activation function, which is defined as:

$$f(x) = \max(0, x) \qquad (1)$$

The convolutional layers are followed by max-pooling layers, which perform downsampling of the feature maps. The max-pooling operation is defined as:

$$y_{i,j}^k = \max_{(m,n) \in R_{i,j}} x_{m,n}^k \qquad (2)$$

where $R_{i,j}$ is the pooling region centered at (i, j), and $x_{m,n}^k$ and $y_{i,j}^k$ are the input and output feature maps, respectively. The VGG16 model also includes fully connected layers, which combine the features extracted by the convolutional layers to produce the final classification output. The output of the fully connected layers is passed through a softmax function to obtain the class probabilities:

$$P(y = j|x) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \qquad (3)$$

where z_j is the input to the softmax function for class j, and K is the total number of classes.

4 Results and Analysis

The performance results demonstrate the potential of deep convolutional neural networks for automated knee osteoarthritis severity classification from X-ray images. Both the custom CNNs architecture and transfer learning with VGG16 yielded high training accuracy and good testing accuracy in predicting K.L. grades, indicating osteoarthritis severity.
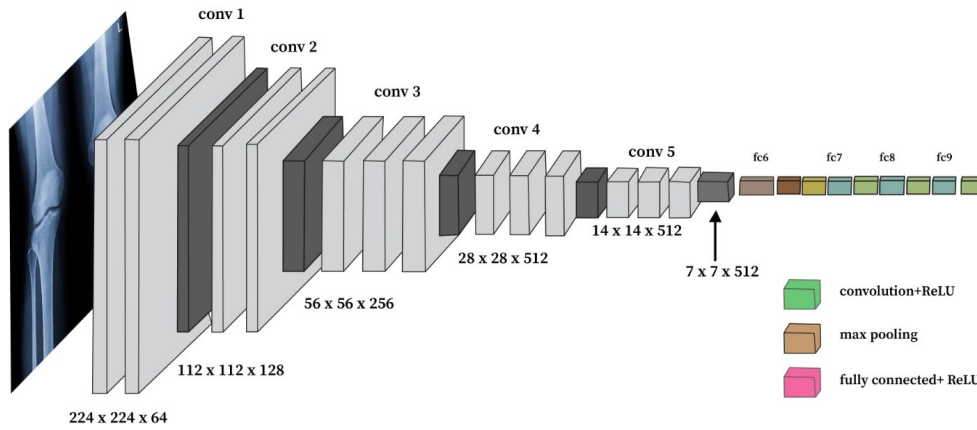
Figure 2. Layer architecture of the VGG16 model.

4.1 Performance of Proposed CNNs

The proposed CNNs model achieved strong performance, attaining 99% training accuracy and 80% testing accuracy after 50 epochs of training as shown in Figures 3 and 4. The high training accuracy indicates that the CNNs model was able to effectively learn the features and patterns in X-ray images associated with different grades of knee osteoarthritis severity. The gap between training and testing performance implies some overfitting, but the competitive 80% testing accuracy is promising for real-world applications. As a tailored architecture designed specifically for this task, the CNNs leverages domain-specific features crafted through an optimized combination of convolutional, pooling, and fully connected layers. The use of dropout regularization and data augmentation also helped improve generalization and testing performance. The proposed CNNs demonstrates that specially engineered deep networks can deliver excellent results for knee OA classification from X-rays [34].

The confusion matrix demonstrates that the proposed CNNs model performs very well in correctly classifying the different grades of knee osteoarthritis severity. The strong diagonal from the top left to the bottom right shows that the majority of samples are classified correctly, with 116, 109, 119, and 115 correct predictions for each of the four classes, as shown in Figure 4. The high values along this diagonal compared to the off-diagonal elements indicate that the model is able to effectively distinguish between the different disease grades based on the X-ray image features. There is some confusion between neighboring classes; for example, grade 1 is misclassified as grade 0, but very few severe
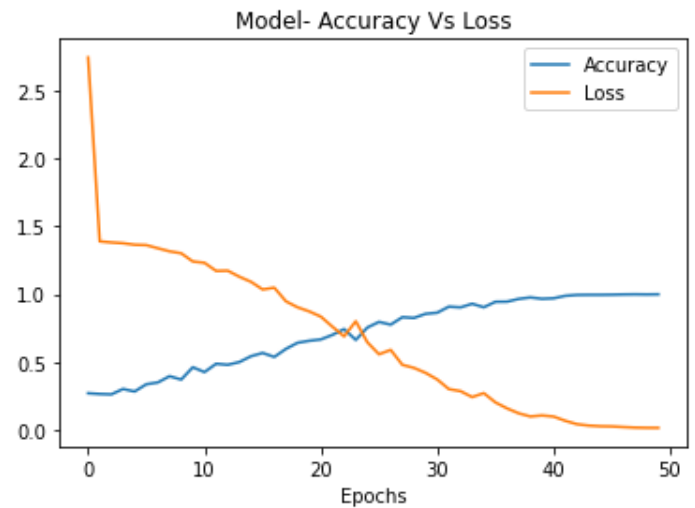

Figure 3. CNNs accuracy and loss curve.

misclassifications. The confusion matrix reflects the high accuracy and robust performance of the CNNs architecture for this classification task.

The accuracy of the model is then calculated as:

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

The precision, which measures the fraction of correct positive predictions, is given by:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

The recall, which measures the fraction of actual positives that are correctly identified, is defined as:
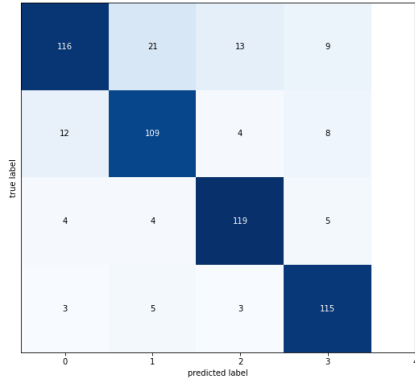
$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Figure 4. CNNs confusion matrix.

Finally, the F1 score, which is the harmonic mean of precision and recall, is calculated as:

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \qquad (7)$$

4.2 VGG16 Model with 10 epochs

The VGG16 model trained for ten epochs also achieved good performance, reaching 98% training accuracy and 89% testing accuracy, as shown in Figures 5 and 6. The high training accuracy shows that VGG16 was quickly able to adapt its pre-trained feature hierarchies to the knee OA classification task. The smaller gap between training and testing performance indicates less overfitting compared to the CNNs model. This can be attributed to the regularization effect of transfer learning, where the foundations of the model were pre-trained. The competitive testing accuracy demonstrates the power of fine-tuning VGG16's pre-trained representations on a modest-sized dataset of just a few thousand X-ray images. Even with just ten epochs, VGG16 is able to learn meaningful features for discriminating between different grades of knee osteoarthritis.

The confusion matrix for the 10-epoch VGG16 model also shows strong performance in correctly classifying the knee osteoarthritis severity grades. The diagonal values of 143, 109, 133, and 105 indicate that the majority of samples are predicted correctly for each class, as shown in Figure 6. There is some confusion between the moderate and severe grades, with a few mild cases misclassified as severe and vice versa. This suggests the model may need more training to distinguish better the more damaged cartilage features between these advanced disease stages. The high diagonal values compared to the off diagonals demonstrate that fine-tuning VGG16 for just ten epochs can already achieve good classification accuracy for
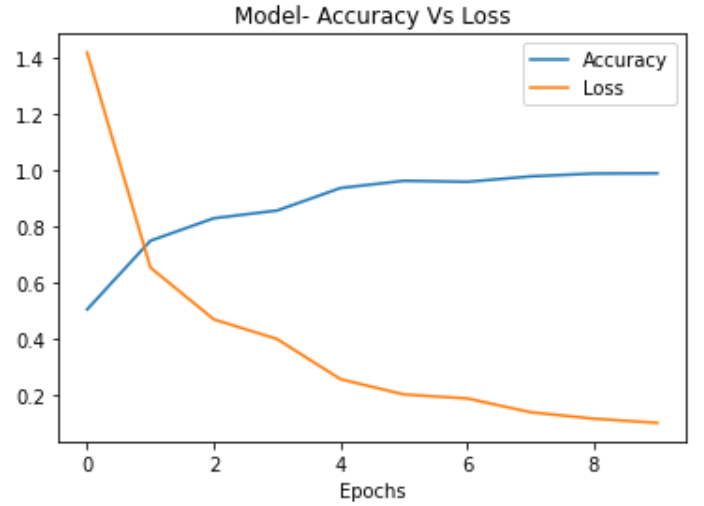

Figure 5. Shows accuracy and loss matrix of the (VGG16 with 10 epochs).
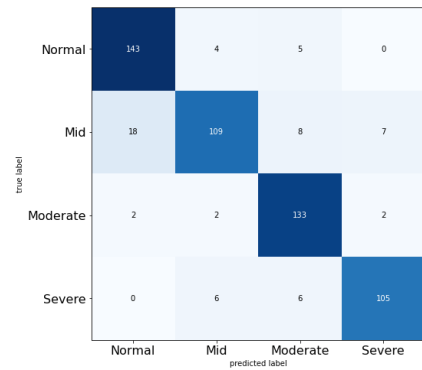

Figure 6. VGG16 (10 Epochs) confusion matrix.

the four grades of knee osteoarthritis severity from the X-ray images.

The categorical cross-entropy loss, which is commonly used for multi-class classification problems, is defined as:

$$L = -\sum_{i=1}^{N} \sum_{j=1}^{K} y_{i,j} \log(p_{i,j}) \qquad (8)$$

where N is the number of samples, K is the number of classes, $y_{i,j}$ is a binary indicator (0 or 1) indicating whether the i-th sample belongs to class j, and $p_{i,j}$ is the predicted probability that the i-th sample belongs to class j. During the training process, the model's parameters are optimized to minimize this loss function using an optimization algorithm, such as stochastic gradient descent (SGD). The update rule for the model's weights w and biases b using SGD is given by:

$$w \leftarrow w - \alpha \frac{\partial L}{\partial w} \, b, \qquad \leftarrow b - \alpha \frac{\partial L}{\partial b} \qquad (9)$$

where α is the learning rate, and $\frac{\partial L}{\partial w}$ and $\frac{\partial L}{\partial b}$ are the gradients of the loss function with respect to the weights and biases, respectively.

4.3 VGG16 Model with 20 epochs

When taken to 20 epochs of training, the performance of VGG16 improved further, achieving 99.8% training accuracy and 91% testing accuracy see Figure 7. The additional ten epochs allow the model to optimize its weights better to precisely recognize features associated with each OA grade from the X-ray images. The training accuracy is now almost perfect, while the testing accuracy sees a worthwhile boost as well. The model is able to generalize even better, likely due to the depth of representations learned from the convolutional layers pre-trained on a diverse dataset. The results for the 20-epoch VGG16 model demonstrate the benefits of increased training time and highlight the potential of transfer learning for knee OA detection. Competitive performance can be reached rapidly by fine-tuning a robust pre-trained model like VGG16.
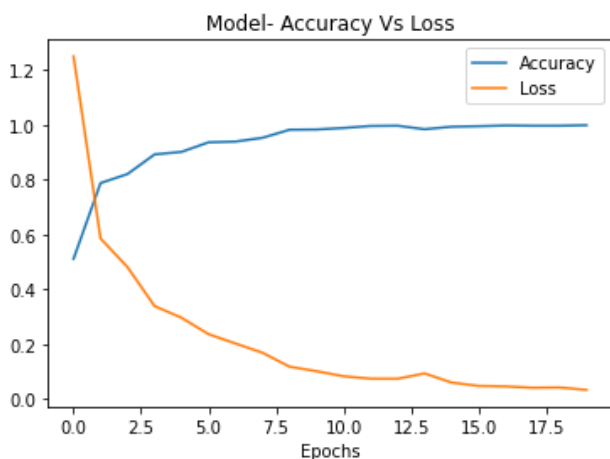

Figure 7. VGG16 (20 Epochs) accuracy and loss curve.

Table 2 summarizes the performance evaluation metrics of the proposed CNN model and the VGG16 model across different training epochs. The proposed CNN achieved high accuracy (99%) with a faster training time, while the VGG16 model showed slightly improved recall and F1-Score with increased training epochs. These results highlight the trade-off between performance and computational efficiency.

4.4 Comparison of CNNs and VGG16

The proposed CNNs model, tailored specifically for knee osteoarthritis classification, achieved strong performance but required the most training time. Trained for 50 epochs, it reached 99% training accuracy

and 80% testing accuracy but took 161 seconds per step, totaling over 2 hours to converge. The extensive training enables robust learning of domain-specific features through its optimized convolutional and dense layer architecture. In contrast, transfer learning with VGG16 leverages broad pre-trained feature representations, allowing quicker training. The 10-epoch VGG16 attained 98% training accuracy and 89% on the test set in just 528 seconds total. It was extending VGG16 to 20 epochs further enhanced performance, achieving near-perfect 99.8% training accuracy and 91% testing accuracy in 240 seconds per step as shown in the Table 2 and Figure 8.

The additional ten epochs enable better adaptation of VGG16's hierarchical feature extractors to the target task. The proposed CNNs reaches marginally higher testing accuracy, but VGG16 offers a superior speed accuracy trade-off by leveraging pre-trained representations. After just 20 epochs, it matches the CNNs's performance at 3x faster training. For real world applications, the rapid training of transfer learning is preferred over extensive optimization from scratch. VGG16's learned general features transfer well for classifying osteoarthritis severity, reaching competitive results with the CCNNsNN in significantly less training time. Its established architecture also provides more predictable convergence compared to individually designing and tuning a custom CNNs.
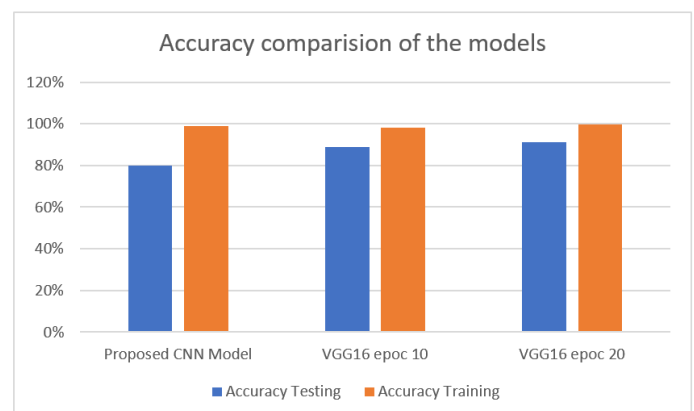

Figure 8. Comparative analysis of the proposed models.

5 Conclusion

The conclusion of this study highlights the potential of deep learning for automated knee osteoarthritis (OA) classification from X-ray images, offering significant promise for enhancing diagnostic accuracy and reducing clinician workload. The proposed models, CNNs and VGG16, both demonstrated strong

Table 2. Performance across multiple evaluation metrics.

Model	Accuracy	Precision	Recall	F1-Score	Training Time (s)
Modified CNN	99%	97%	78%	0.85	161/step
VGG16 (10 Epochs)	98%	96%	83%	0.89	528/step
VGG16 (20 Epochs)	99.8%	98%	90%	0.94	240/step

performance, with the CNNs achieving 99% training accuracy and 80% testing accuracy, while VGG16 reached 99.8% training accuracy and 91% testing accuracy after 20 epochs. These results underscore the effectiveness of using deep learning to automate knee OA classification, providing clinicians with a reliable tool to assist in diagnosing and managing the disease. The implementation of such models could facilitate earlier detection, improving treatment outcomes and reducing healthcare burdens. For clinical deployment, integrating these models into existing radiology workflows could allow for real-time OA severity classification directly from X-ray images, offering immediate support for clinical decision-making. The proposed models can help streamline diagnostic processes and improve patient care by providing clinicians with automated results, thus saving time and enhancing diagnostic precision. However, further research and validation are necessary to ensure that these models can generalize across different clinical environments and patient demographics. Future work should focus on testing these models on larger, more diverse datasets and incorporating multimodal data (e.g., MRI scans, patient history) to improve the accuracy of OA classification. Looking ahead, deployment in clinical settings requires addressing several practical challenges, such as integration with Electronic Health Records (EHR) and Picture Archiving and Communication Systems (PACS). Additionally, models should undergo continuous monitoring and refinement based on real-world feedback to ensure their continued effectiveness. The future of this research lies in improving the scalability and robustness of these AI tools, ensuring they can be seamlessly adopted in healthcare systems globally to improve the early diagnosis and management of knee osteoarthritis.

Data Availability Statement

The knee X-ray images used in this study were obtained from the publicly available Osteoarthritis Initiative (OAI) dataset. The dataset can be accessed at https://nda.nih.gov/oai. All data used were fully anonymized and utilized in accordance with the OAI data usage agreement.

Conflicts of Interest

Salman Khan is an employee of CodeNinja Inc., Lahore 54792, Pakistan.

Ethical Approval and Consent to Participate

This work uses publicly available, anonymized X-ray images from the Osteoarthritis Initiative (OAI) dataset. All data were fully de-identified prior to use. Therefore, ethical approval and informed consent were not required, as per the terms of use of the OAI dataset.

References

[1] Abdullah, S. S., & Rajasekaran, M. P. (2022). Automatic detection and classification of knee osteoarthritis using deep learning approach. *La radiologia medica*, 127(4), 398-406. [CrossRef]

[2] Sikkandar, M. Y., Begum, S. S., Alkathiry, A. A., Alotaibi, M. S. N., & Manzar, M. D. (2022). Automatic Detection and Classification of Human Knee Osteoarthritis Using Convolutional Neural Networks. *Computers, Materials & Continua*, 70(3). [CrossRef]

[3] Abedin, J., McGuinness, K., O'Connor, N. E., Moran, K., Rebholz-Schuhmann, D., & Newell, J. (2019). Predicting knee osteoarthritis severity: Comparative modeling based on patient's data and plain X-ray images. *Scientific Reports*, 9(1), 5761. [CrossRef]

[4] Olsson, S., Akbarian, E., Lind, A., Razavian, A. S., & Gordon, M. (2021). Automating classification of osteoarthritis according to Kellgren-Lawrence in the knee using deep learning in an unfiltered adult population. *BMC musculoskeletal disorders*, 22, 1-8. [CrossRef]

[5] Ahmed, H. A., & Mohammed, E. A. (2022). Detection and classification of the osteoarthritis in knee joint using transfer learning with convolutional neural networks (CNNs). *Iraqi Journal of Science*, 5058-5071.

[6] Mendon, A., Patil, M., Gupta, Y., Kadakia, V., & Doshi, H. (2023). Automated healthcare system using AI

based chatbot. In *Intelligent Computing and Networking: Proceedings of IC-ICN 2022* (pp. 191-205). Singapore: Springer Nature Singapore. [CrossRef]

[7] Alhusaini, A. A., Aldosari, R., Aly, E., & Alkhalifah, A. (2022). Thermal imaging for clinical evaluation of knee joint disorders: A study on osteoarthritis patients and control subjects. *Bioengineering*, 9(6), 232. [CrossRef]

[8] Brahim, A., Jennane, R., Riad, R., Janvier, T., Khedher, L., Toumi, H., & Lespessailles, E. (2019). A decision support tool for early detection of knee OsteoArthritis using X-ray imaging and machine learning: Data from the OsteoArthritis Initiative. *Computerized Medical Imaging and Graphics*, 73, 11-18. [CrossRef]

[9] Qureshi, M. A., Qureshi, K. N., Jeon, G., & Piccialli, F. (2022). Deep learning-based ambient assisted living for self-management of cardiovascular conditions. *Neural Computing and Applications*, 1-19. [CrossRef]

[10] Almhdie-Imjabbar, A. (2021). Trabecular bone texture analysis of conventional radiographs in the assessment of knee osteoarthritis: Review and viewpoint. *Arthritis Research & Therapy*, 23(1), 1-13. [CrossRef]

[11] Antony, J., McGuinness, K., O'Connor, N. E., & Moran, K. (2016). Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 1195-1200). IEEE. [CrossRef]

[12] Rawat, A., Dhyani, A., Panwa, D., Bisht, G., Vats, S., & Sharma, V. (2024, July). Knee Osteoarthritis Prediction using Transfer Learning. In *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)* (pp. 1001-1007). IEEE. [CrossRef]

[13] Mahfus, M., Tosun, M., & Göker, H. (2025). Transfer Learning-Based Classification of Knee Osteoarthritis Severity from X-Ray Images. *Mühendislik Bilimleri ve Tasarım Dergisi*, 13(1), 325-339. [CrossRef]

[14] Zhong, B., Xing, X., Luo, H., Zhou, Q., Li, H., Rose, T., & Fang, W. (2020). Deep learning-based extraction of construction procedural constraints from construction regulations. *Advanced Engineering Informatics*, 43, 101003. [CrossRef]

[15] Chaugule, S., & Malemath, V. S. (2022). An extreme learning machine-based autoencoder (ELM-AE) for denoising knee X-ray images and grading knee osteoarthritis severity. In *International Conference on Recent Trends in Image Processing and Pattern Recognition* (pp. 164-175). Springer Nature Switzerland. [CrossRef]

[16] Chuang, C. H., Kuo, C. C., Chiang, Y. F., Lee, P. Y., Wang, F. H., Hsieh, C. Y., & Lin, C. L. (2023). Enriched peripheral blood-derived mononuclear cells for treating knee osteoarthritis. *Cell Transplantation*, 32, 9636897221149445. [CrossRef]

[17] Cueva, J. H., Castillo, D., Espinós-Morató, H., Durán, D., Díaz, P., & Lakshminarayanan, V. (2022). Detection and classification of knee osteoarthritis. *Diagnostics*,

12(10), 2362. [CrossRef]

[18] Khan, H., Usman, M. T., Rida, I., & Koo, J. (2024). Attention enhanced machine instinctive vision with human-inspired saliency detection. *Image and Vision Computing*, 152, 105308. [CrossRef]

[19] Wang, X., Wang, T., & Su, Z. (2025). Attention-enhanced deep learning and machine learning framework for knee osteoarthritis severity detection in football players using X-ray images. *Journal of Radiation Research and Applied Sciences*, 18(2), 101428. [CrossRef]

[20] Ding, Y., Ma, Z., Wen, S., Xie, J., Chang, D., Si, Z., ... & Ling, H. (2021). AP-CNN: Weakly supervised attention pyramid convolutional neural network for fine-grained visual classification. *IEEE Transactions on Image Processing*, 30, 2826-2836. [CrossRef]

[21] Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., ... & Acharya, U. R. (2022). Transfer learning techniques for medical image analysis: A review. *Biocybernetics and biomedical engineering*, 42(1), 79-107. [CrossRef]

[22] Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: a literature review. *BMC medical imaging*, 22(1), 69. [CrossRef]

[23] Jena, A., Goyal, N., Rana, P., Taneja, S., Vaish, A., Botchu, R., et al. (2023). Qualitative and quantitative evaluation of morpho-metabolic changes in bone cartilage complex of knee joint in osteoarthritis using simultaneous 18F-NaF PET/MRI—A pilot study. *Indian Journal of Radiology and Imaging*, 33(2), 173-182. [CrossRef]

[24] Sharma, G., Anand, V., & Kumar, V. (2023). Classification of osteo-arthritis with the help of deep learning and transfer learning. In *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 446-452). IEEE. [CrossRef]

[25] Thomson, J., O'Neill, T., Felson, D., & Cootes, T. (2016). Detecting osteophytes in radiographs of the knee to diagnose osteoarthritis. In *Machine Learning in Medical Imaging: 7th International Workshop, MLMI 2016, Held in Conjunction with MICCAI 2016* (pp. 45-52). Springer International Publishing. [CrossRef]

[26] Shamir, L., Ling, S. M., Scott, W., Hochberg, M., Ferrucci, L., & Goldberg, I. G. (2009). Early detection of radiographic knee osteoarthritis using computer-aided analysis. *Osteoarthritis and Cartilage*, 17(10), 1307-1312. [CrossRef]

[27] Tiulpin, A., & Saarakkala, S. (2020). Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks. *Diagnostics*, 10(11), 932. [CrossRef]

[28] Rehman, A., Raza, A., Alamri, F. S., Alghofaily, B., & Saba, T. (2023). Transfer learning-based smart features engineering for osteoarthritis diagnosis from knee

X-ray images. *IEEE Access*. [CrossRef]

[29] Chen, H., Li, T., Fan, X., & Luo, C. (2019). Feature selection for imbalanced data based on neighborhood rough sets. *Information Sciences*, 483, 1-20. [CrossRef]

[30] Guida, C., Zhang, M., & Shan, J. (2021). Knee osteoarthritis classification using 3D CNN and MRI. *Applied Sciences*, 11(11), 5196. [CrossRef]

[31] Arunkumar, M., Aarthi, R., & Selvagananathv, J. (2024, December). Deep Learning with 3D CNN for Knee Osteoarthritis Detection and Classification. In *2024 International Conference on System, Computation, Automation and Networking (ICSCAN)* (pp. 1-6). IEEE. [CrossRef]

[32] Mahum, R., Rehman, S. U., Meraj, T., Rauf, H. T., Irtaza, A., El-Sherbeeny, A. M., & El-Meligy, M. A. (2021). A novel hybrid approach based on deep cnn features to detect knee osteoarthritis. *Sensors*, 21(18), 6189. [CrossRef]

[33] Yoon, J. S., Yon, C. J., Lee, D., Lee, J. J., Kang, C. H., Kang, S. B., ... & Chang, C. B. (2023). Assessment of a novel deep learning-based software developed for automatic feature extraction and grading of radiographic knee osteoarthritis. *BMC musculoskeletal disorders*, 24(1), 869. [CrossRef]

[34] Khalid, A., Senan, E. M., Al-Wagih, K., Ali Al-Azzam, M. M., & Alkhraisha, Z. M. (2023). Hybrid techniques of X-ray analysis to predict knee osteoarthritis grades based on fusion features of CNN and handcrafted. *Diagnostics*, 13(9), 1609. [CrossRef]

Sundas Iqbal received the MS degree in Computer Science from National college of Business administration and Economics, Pakistan in 2018. Her research interests include Biomedical image analysis, Machine Learning, Image Processing. Currently she is PhD student at Nanjing University of Information Science and Technology (NUIST) with a research focus on domain adaptation, transfer learning and novel AI techniques for various fields.(Email: sundasiqbal058@gmail.com)

Danish Ali is a highly motivated computer researcher from Pakistan, with a strong educational background and expertise in various fields of computer science. He received the bachelor's degree in computer science from the University of Haripur, Haripur, Pakistan, in 2023. Currently, he is a MS Scholar at Wuhan University China. His research interests are in machine learning, biomedical imaging, natural language processing, and the deep learning. (Email: danishalikhan545@gmail.com)

Sumaira Rafique received the MS degree in Data Science from National University of Computer and Emerging Sciences, Lahore, Pakistan in 2024. Her research interests include Natural Language Processing, Generative AI, Deep Learning, Machine Learning, Image Processing. Currently she is working as a Machine Learning Engineer at Reallytics.ai, Lahore. (Email: rsumaira80@gmail.com)

Kajol Bagga Kajol Bagga received her Diploma and Bachelor's degree in Computer Science from Sri Sai College of Engineering, Punjab Technical University, India, and her Master's degree in Artificial Intelligence from Technische Hochschule Würzburg-Schweinfurt (THWS), Germany. Her research interests include step length estimation using neural networks, machine learning, deep learning, and artificial intelligence. (Email: kajol.bagga@study.thws.de)

Atif Ur Rahman Born on February 1, 1996, in District Malakand, Pakistan, Atif ur Rahman is a driven and ambitious individual excelling in the field of computer science. He completed his BS Computer Science, Hazara University Mansehra (2019) and now Pursuing MS Computer Science, Iqra National University Swat (ongoing). His research area includes medical image processing, Deep learning, Artificial learning. (Email: atifurrahman.inu@gmail.com)

Salman Khan is an experienced IT professional currently serving as an IT Engineer at CodeNinja Inc., Lahore. He specializes in network administration, IT support, and technical solutions, with expertise in Microsoft 365 Administration, Azure Active Directory, network security, and endpoint protection. With a proven track record in troubleshooting and optimizing IT systems, Salman is committed to leveraging technology to enhance operational efficiency and security.His research interests include Artificial Intelligence under the broad umbrella of Machine Learning and Deep Learning with a specific focus on Visual Intelligence. Passionate about leveraging cutting-edge technology, Salman focuses on delivering innovative IT solutions to enhance operational efficiency and security. (Email: salman.khan@codeninjaconsulting.com)