Noname manuscript No.
(will be inserted by the editor)

Knee Osteoarthritis Severity Classification

Shrunmay Shinde · Prathamesh Gadekar · Ishkumar Savaliya · Daksh Agarwal · Sanjeev Sharma

Received: date / Accepted: date

Abstract Osteoarthritis, a prevalent degenerative condition, greatly affects the elderly population, particularly in the context of knee osteoarthritis (KOA). Traditional X-ray analysis using the Kellgren and Lawrence grading system presents challenges in accurately diagnosing KOA due to its subjective nature. To address these challenges, this study introduces an innovative approach leveraging artificial intelligence, specifically deep learning techniques. By employing Convolutional Neural Networks (CNNs), transfer learning and Ensemble Techniques, the study aims to improve the accuracy of KOA detection and classification based on knee X-ray images. This method facilitates early and precise diagnosis, enabling timely interventions that may potentially slow down the disease progression. Comparative analyses with previous models demonstrate the effectiveness of the proposed approach. This reserach has achieved

Shrunmay Shinde
Indian Institute of Information Technology, Pune, India.
E-mail: 112115148@cse.iiitp.ac.in

Prathamesh Gadekar
Indian Institute of Information Technology, Pune, India.
E-mail: 112115115@cse.iiitp.ac.in

Ishkumar Savaliya
Indian Institute of Information Technology, Pune, India.
E-mail: 112115142@cse.iiitp.ac.in

Daksh Agarwal
Indian Institute of Information Technology, Pune, India.
E-mail: 112116012@cse.iiitp.ac.in

Sanjeev Sharma
Indian Institute of Information Technology, Pune, India.
E-mail: sanjeevsharma@iiitp.ac.in

84% accuracy for binary classification, 90% accuracy for tertiary classification and 71% for 5 category classification. The incorporation of deep learning in KOA diagnosis not only enhances diagnostic accuracy but also reduces the dependence on manual expertise, promising significant advancements in medical imaging and patient care in osteoarthritis management.

Keywords Knee Osteoarthritis · Deep Learning · Convolutional Neural Networks · Medical Imaging

1 Introduction

Osteoarthritis, commonly known as wear-and-tear arthritis, is a degenerative disease characterized by the progressive erosion of cartilage, the natural cushioning of joints. This erosion causes the relationship between the bones in the joint to weaken, causing the protective cartilage to weaken. For this reason, people may experience symptoms such as pain, swelling, stiffness, and decreased mobility. Sometimes this friction can lead to the formation of bone spurs that can worsen symptoms. It mainly affects the elderly and occurs in two main types: primary osteoarthritis (occurs when there is no obvious cause) and secondary osteoarthritis (which can be caused by neck injury or disease and other problems). Symptoms often worsen over time, affecting the patient. Factors include age, weight, genetic predisposition, gender, repetitive strain injuries, and certain medical conditions such as rheumatoid arthritis and metabolic syndrome. and using imaging techniques such as X-rays and MRI scans. This tool helps identify bone and cartilage damage as well as the presence of bone spurs. Early diagnosis is important, so timely intervention can be used to reduce symptoms and slow joint progression. Treatment strategies focus on controlling symptoms through pain management, lifestyle changes, daily exercise, and sometimes surgery[44]. Doctors who understand the symptoms and risks of knee osteoarthritis can provide care and support, ultimately improving the patient's quality of life.

It influences a significant section of the populace, especially the elderly, with around one in three people encountering OA at a few organize of their lives. Despite its far-reaching event, successful treatment for serious KOA remains tricky.Be that as it may, early determination is fundamental to prevent the malady from progressing to a serious stage.

Traditionally, KOA conclusion has depended on X-ray imaging, utilizing the Kellgren and Lawrence (KL) reviewing framework to evaluate seriousness. Be that as it may, the elucidation of X-rays, especially the qualification of KL evaluation, is subjective and dependent on the mastery of doctors. This subjectivity presents uncertainty and complicates the early discovery of KOA, driving to deferred mediation and movement to serious stages.To address the impediments of manual determination, especially the equivocalness related to X-ray interpretation, counterfeit insights (AI) methods, especially profound learning, offer promising arrangements. Later headways have driven to the improvement of calculations for knee joint acknowledgment and KL review classification, utilizing Convolutional Neural Organize (CNN) designs and machine learning methods for highlight extraction and classification. The key contributions of this paper are:

- With the help of transfer learning[9] and ensembling methods[29], this paper proposed a solution to the problem of early detection of Knee Osteoarthritis.
- Several hyperparameters are tuned to find the best parameters for each model.
- In this paper, we propose a method to classify OA symptoms based on neural systems and learning change and combine deep learning and machine learning architectures[36] to accurately classify the weight of the KL level.
- The various models comparatively has achieved 84% accuracy for binary classification, 90% accuracy for tertiary classification and 71% for 5 category classification.

Here is the rest of the paper's organization. The literature review of the related work is discussed in Section 2. The study of material and methods is covered in Section 3. The results of the experiments are presented in Section 4. Finally, we are concluding section 5 with our work.

2 Literature Review

Computer vision technology[26] has emerged as a robust tool for knee osteoarthritis (OA) diagnosis via X-ray image classification. Analogous to its application in facial recognition or object detection, computer vision scrutinizes knee X-rays to discern OA indicators and ascertain severity levels. Leveraging an extensive corpus of medical literature, encompassing scholarly works and journal publications, researchers acquire foundational insights into the historical trajectory, contemporary understanding, and persisting challenges inherent in OA diagnosis. This corpus serves as a bedrock for the development and refinement of computer vision algorithms[36] tailored for X-ray analysis, thereby augmenting efficiency and precision in knee OA classification endeavors. The integration of image processing techniques into X-ray interpretation harbors significant potential for automating facets of OA classification, heralding expedited diagnoses, refined disease progression monitoring, and elevated standards of patient care.

Cheng-Tzu Wang et al.[45] have proposed a osteoarthritis classification deep-learning model employing YOLOv4 and Conv-BN-ReLU block based model. Their research aimed to swiftly and accurately classify knee osteoarthritis with the help of osteoarthritis initiative (OAI) data set comprising of 8964 knees and clinical AP radiographs of 246 knees from FEMH with 5 KL grade classes, achieving an impressive 78% accuracy.

Paulione S. Q. Yeoh et al.[50] research works has proposed 3D assisted knee osteoarthritis classification leveraging the CNN models(ResNet, DenseNet, VGG and AlexNet). The Osteoarthritis Initiative (OAI) dataset is used which contains 4,796 image. ResNet18 achieved an highest AUC score of 94.5%.

Haresh Rajamohan et al.[35] work proposed prediction of total knee replacement with knee MRIs using a CNN architecture. This research is based on OAI dataset classified into IW-TSE, FS-IW-TSE, and DESS having 4796 images in total, leveraging 90% AUC score.

Nide Nasir et al.[23] proposed multi-modal image classification of COVID-19 images captured using computed tomography and X-rays scans using transfer learning models such as VGG16, ResNet50, InceptionResNetV2 and MobileNetV2. The research is based on the covid-xray-dataset, which binary classifies into "Covid" and "Non-Covid". The accuracy turned out to be 97.8%.

Aleksei et al.[43] works proposed Knee Osteoarthritis Diagnosis from Plain Radiographs mounted on deep learning. This research is based on a 5960 image-oriented database, Osteoarthritis Initiative dataset using Deep Siamese CNN architecture. The research claimed a quadratic Kappa coefficient of 0.83 and average multiclass accuracy of 66.71%.

Hua Wang et al.[46] research work proposed ankle fracture detection deploying EfficientNetB5, ResNet50 with the Squeeze-and-Excitation Network (SENet). In this research,CT images of ankle fractures images are collected numbering 987 , deemed apt for research purpose, segmented into 255 images depicting fractures and 732 illustrating normal ankles. The result for ResNet50 with SENet was 93%, and EfficientNetB5 was 90%.

R. V. Manjunath et al.[20] research work proposed a detection and classification model of liver disease using CT images based on a deep learning model called modified Unet 60 . A public dataset 3Dircadb is utilized containing 864 images, 360 images belongs to Metastasis cancer and leftover 360 images affiliated to Cholangiocarcinoma. This research attained an accuracy of 98.61% and a dice score of 98.59%.

Qingji Guan et al.[14] proposed the Chest X-ray Image Classification with Noisy Labels using Heteroscedastic Modeling Method. This research embraces an ChestX-ray2017 dataset containing 5856 anterior-posterior Chest X-ray images including 14 pathologies: Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, PT, and Hernia. This model achieved an average AUC score of 14 pathologies as 81.2%.

Ahmed Khalid et al.[17] works proposed to predict Knee Osteoarthritis Grades based on fusion features FFNN (CNN and handcrafted). This research is based on the OAI and RCU datasets containing 5 classes of severity grading Healthy, Doubtful, Minimal, Moderate, Severe, having total of 11,436 imagest.

It gained 99.1% recognition accuracy.

Dalila Say et al.[38] research work proposed Categorization of Multi-class Welding defects using the X-ray Image utilizing image augumentation and CNNs on 4479 X-ray images belonging to six groups: cavity, cracks, inclusion slag, lack of fusion, shape defects, and normal defects. This research attained 92% accuracy.

Mitra Daneshmand et al.[10] works outlined a deep learning-based model for the detection of osteophytes in radiographs and magnetic resonance imagings of the knee using 2D and 3D morphology. This research utilized Osteoarthritis Initiative dataset, deploying 3 models for 3 different tasks such as a U-Net-based model28 for X-ray segmentation, 2D ResNet-18 purposed for X-ray-based mask analysis and 3D ResNet-10 for MRI-based mask analysis . An overall accuracy of 80% is obtained.

Goram M. M. Alshmrani et al.[2] work proposed to classify multi class lung diseases with deep learning architectures based on X-ray(CXR) images. The database has 5 categories namely Pneumonia, Lung Cancer, tuberculosis (TB), Lung Opacity and COVID-19 with CXR images of 3615 COVID-19, 6012 Lung opacity, 5870 Pneumonia, 20,000 lung cancer, 1400 tuberculosis, and 10,192 normal images. The research attained an average accuracy of 96.48% with VGG19 and CNN architecture.

Suganya Athisayamani et al.[5]introduced a study on MRI Brain Tumor Classification. The primary methodology employed entails the utilization of a residual deep CNN (ResNet-152) and Optimized Feature Dimension Reduction for classification task. The synthesized dataset of Figshare, BRATS 2019 and MICCAI BRATS BRATS 2019 is used having various classes such as meningiomas, pituitary tumors, and gliomas of various tumors .The study reports a classification accuracy of 98.85%.

Fahad Ahmed et al. [1] research proposed a transfer learning approach empowered with explainable artificial intelligence for Identification of kidney stones in KUB X-ray images. The classification problem is addressed by implementing VGG16 using manually fetched dataset. The model demonstrated the highest classification accuracy, reaching 97.41%.

Sun-Woo et al. [30] proposed a method for automated osteoarthritis grading in knee X-ray images through ensemble deep learning, specifically targeting five prevalent severities: Healthy, Doubtful, Minimal, Moderate, Severe. Utilizing a Osteoarthritis Initiative dataset comprising 8260 images and employing the 8 optimised model ensemble network as the foundational network, the study achieved enhanced accuracy, registering 76% in the test set.

Soumya Ranjan Nayak et al. [24] proposed a proficient method for automated diagnosis of COVID-19 from Chest X-ray Images through a Lightweight Deep Convolutional Neural Network. The model employed for classifying key chest diseases, including normal, COVID-19, and pneumonia, demonstrated a remarkable average accuracy of 98.33% on the test set. This inventive approach, utilizing the authors' Figshare dataset, containing 750 samples for each class.

Ming Ni et al. [25] outlined an innovative deep-learning methodology for classifying calcaneofibular ligament injuries in the ankle joint, utilizing Mask R-CNN models. The study integrates image augmentation and hybrid contrast stretching techniques to enhance the dataset and improve visual quality along with randomly adding Gaussian noise to the data. The proposed LeNet-5 technique attains a notable classification of classes 0–2 with accuracy of 93%, surpassing recent methodologies.

Norio Nakata et al. [22] proposed Ensemble Learning of multiple models for multiclass classification of Ultrasound images of Hepatic Masses. The research classifies into 4 categories namely benign liver tumor (BLT), liver cyst (LCY), metastatic liver cancer (MLC), primary liver cancer (PLC). The US dataset with 26,200 images of liver masses. The results highlight ResNeXt101 as the most effective model, achieving an accuracy of 91.74% after training carried out with 16 different CNN models. This research highlights the importance of pruning in ensemble learning in classification techniques

Xin Xing et al. [48] proposed an efficient approach for 3D PET Brain Image Classification (Alzheimer's Disease Diagnosis) using learnable weighted pooling. Focused on addressing severe health-related problems affecting humans Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset is used. The research utilizes ResNet34 for feature extraction and Learnable Weighted Pooling (LWP) for converting a 3D brain image to a 2D fused image. The results indicate a significant overall accuracy of 88%.

Laith Alzubaidi et al. [4] proposed a methodology for the detection of abnormalities in shoulder images employing a deep convolutional neural network. The MURA dataset used comprised images seven skeletal bones, namely elbow, finger, forearm, hand, humerus, shoulder, and wrist. The deep feature fusion combining extracted features along with seven model such as NasNetLarge, ResNet101, DenseNet201, EfficientNet, MobilNetV2, InceptionResNetV2 and Xception, demonstrated an outstanding overall accuracy of 99.2%.

Rohit Kumar Jain et al.[16] employed multi-scale deep convolutional neural network namely High-Resolution Network (HRNet) for Knee osteoarthritis severity prediction. It integrates OsteoHRNet into radiographs, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) scans images, achieving an 71.74% classification accuracy on the OAI dataset based on the

K: grading system containing 5 classes Healthy, Doubtful, Minimal, Moderate, Severe.

Yanling Xie et al.[47] proposed a deep learning based model for enhanced multi-site fracture classification and its diagnosis, addressing challenges in obtaining sufficient training data for diverse fractures, especially fractures of the hip, shoulder, wrist, and ankle. The dataset fetched manually contains 26,098 images. The model combines features of Resnet 50 and ROI pooling, leveraging their Feature Pyramid Network (FPN). Evaluated on the FROC-AUC , 86.5% accuracy is achieved.

Suam Kim et al.[18] introduced an automated approach for ankle radiographs multiclass classification using neural network. The dataset is synthesized manually from an institution with 1493 images with 4 classes based on severity. Leveraging computer vision, image processing, and a Inceptionv3 models layers, the system attains a commendable AUC score of 91.45%.

Xiaoging Ying et al.[51] proposed a deep learning-based model for Covid-19 chest image classification in the presence of noisy labels. The approach employed Bayesian Classifier , SLIPR algorithm and CNN layers. Experimental results conducted on three datasets of Tawsifur, Skytella, CXRs, achieving a notable classification accuracy of 99.26%.

Fabi Prezja et al.[32] proposed a deep learning technique for Radiograph Classification of Knee Joint Osteoarthritis based on data augumentation techniques. The research's dataset is Osteoarthritis Initiative (OAI) consisting of Grade0-4 images classified on KL Grading and utilized EfficientNetv2 model. It attained an accuracy of 65.5% obtained by baseline rotation, while other operations are baseline, ROI, Horizontal split,Flip etc, along with different amount of Noise.

Table 1 and table 2 showcase the summary of these literature reviews in a tabular, concise format.

Table 1 Literature Review Table - Entries 1 to 5

Sr. No.	**Ref No.**	**Author**	**Accuracy**	**Problem Addressed**
1	[45]	Cheng-Tzu Wang et al.	78%	Knee Osteorthritis
2	[50]	Paulione S. Q. Yeoh	94.5% AUC	Knee Osteorthritis
3	[35]	Haresh Rajamohan et al.	90.00% AUC	Knee Replacement
4	[23]	Nide Nasir et al.	97.8%	Covid-19 Classification
5	[43]	Aleksei et al.	66.71%	Knee Osteoarthritis Diagnosis

Table 2 Literature Review Table - Entries 6 to 25

Sr. No.	Ref No.	Author	Accuracy	Problem Addressed
6	[46]	Hua Wang et al.	90%	Ankle Fracture Detection
7	[20]	R. V. Manjunath et al.	98.61%	Liver Diseases Classification
8	[14]	Qingji Guan et al.	81.2% AUC	Chest X-ray Classification
9	[17]	Ahmed Khalid et al.	99.1%	Knee Osteoarthritis Grades
10	[38]	Dalila Say et al.	92%	Welding Defects
11	[10]	Mitra Daneshmand et al.	80.00%	Osteophytes Detection
12	[2]	Goram M. M. Alshmrani et al.	96.48%	Lung Diseases
13	[5]	Suganya Athisayamani et al.	98.85%	Brain Tumor Classification
14	[1]	Fahad Ahmed et al.	97.41%	Identification of Kidney Stones
15	[30]	Sun-Woo et al.	76%	Knee osteoarthritis
16	[24]	Soumya Ranjan Nayak et al.	98.33%	Chest X-ray Classification
17	[25]	Ming Ni et al.	93.00%	Classification of Calcaneofibular Ligament Injuries
18	[22]	Norio Nakata et al.	91.74%	Hepatic Masses Classification
19	[48]	Xin Xing et al.	88.00%	3D PET Brain Image Classification
20	[4]	Laith Alzubaidi et al.	99.2%	Shoulder Abnormalities Detection
21	[16]	Rohit Kumar Jain et al.	71.74%	Knee osteoarthritis
22	[47]	Yanling Xie et al.	86.50%	Fracture Classification
23	[18]	Suam Kim et al.	91.45% AUC	Ankle Radiographs Multiclass
24	[51]	Xiaoging Ying et al.	96.26%	Covid-19 Chest Image Classification
25	[32]	Fabi Prezja et al.	65.50%	Radiograph Knee Joint Osteoarthritis Classification

This comprehensive literature review has provided valuable insights into the current state of deep learning models for the detection and classification of knee osteoarthritis. By synthesizing and analyzing a wide range of research studies, we can gain insights into the strengths and limitations of various deep learning architectures in the context of classifying knee osteoarthritis.

3 Materials and methods

The strategy we will be using is a simple and efficient one. We will first pre-process the data set using different techniques to maximize the use the capacity of the dataset and avoid overfitting the CNNs model[7] while training them. Then, with the help of transfer learning and ensemble learning[52] train our models on the dataset and then test their performance based on different criteria. Fig.2 shows the overall flow of the work.

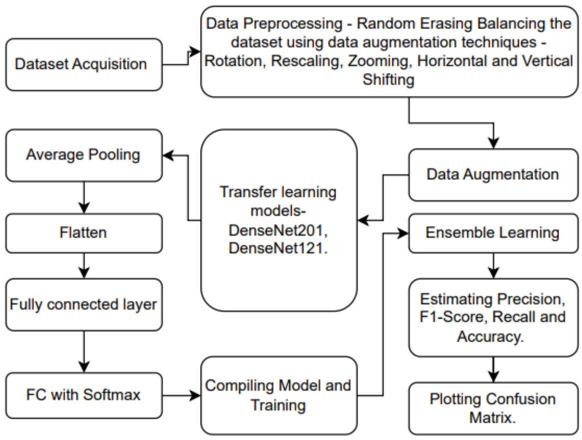

Fig. 1 Flowchart for 5 classes classification

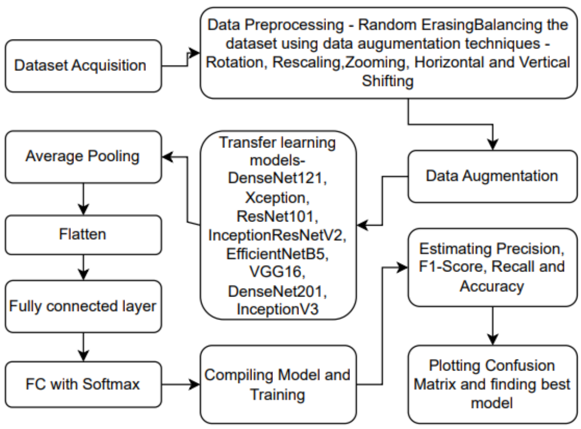

Fig. 2 Flowchart for 3 classes and Binary classification

3.1 Dataset Description

The dataset used here is the Knee Osteoarthritis Dataset with severity grading which consists of X-ray images of knee joints divided into 5 classes. The grading system used in dataset is Kellgren-Lawrence grading system which is a popular technique for determining the degree of osteoarthritis in the knee based on X-ray pictures. In the 1950s, Drs. John Kellgren and Jeffrey Lawrence created it. Based on the existence of specific radiographic characteristics, the method assigns five grades, from 0 to 4, to the severity of knee osteoarthritis. Grade 0 represents no indications of osteoarthritis. There are no osteophytes (bone spurs) or other anomalies, and the joint space is normal. Grade 1 represents doubtful condition that there may be osteophytic lipping and dubious joint narrowing, which could indicate very early osteoarthritis symptoms. Grade 2 represents minimal osteophytes are clearly present, indicating the onset of osteoarthritis. Additionally, joint space narrowing could occur. Grade 3 represents moderate condition that there are several osteophytes, noticeable narrowing of the joint space, and mild sclerosis (hardening) of the bone. This suggests a mild case of osteoarthritis. Grade 4 represents severe condition that there is severe sclerosis, large osteophytes are apparent, and there is noticeable joint space constriction.

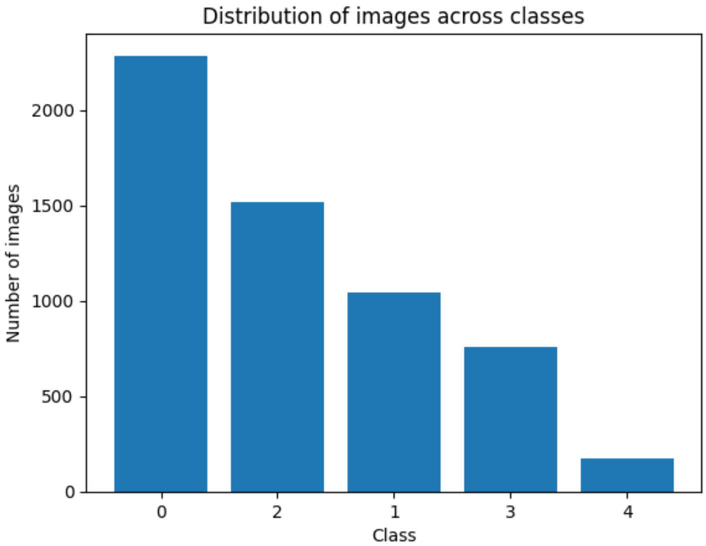

Fig. 3 Training Data Distribution

The dataset consists of four directories : Auto_test, Train, Test and Validation. Each of these directory consists of X-ray images belonging to five classes. Auto_test directory contains 604 images of Grade0, 275 images of Grade1, 403

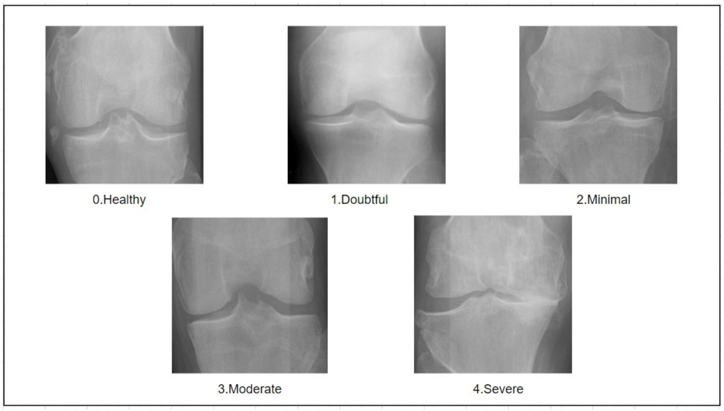

Fig. 4 Dataset Classes

images of Grade2, 200 images of Grade3, 44 images of Grade4. Train directory contains 2286 images of Grade0, 1046 images of Grade1, 1516 images of Grade2, 757 images of Grade3, 173 images of Grade4. Validation directory contains 328 images of Grade0, 153 images of Grade1, 212 images of Grade2, 106 images of Grade3, 27 images of Grade4. Test directory contains 639 images of Grade0, 296 images of Grade1, 447 images of Grade2, 223 images of Grade3, 51 images of Grade4.

3.2 Data Preprocessing

Data Preprocessing[26] is a big part when it comes to training a deep learning model. With the help of data augmentation, we can make use of data sets to their maximum extent.

3.2.1 Five Class Classification:

Random Erasing, a powerful data augmentation technique, was utilized to enhance the generalization of deep learning models in image classification tasks. This method involves randomly replacing a rectangular region in the input image with random pixel values, simulating occlusions or corruptions. By exposing the models to a diverse range of inputs during training, Random Erasing promotes robust feature learning and improves model performance on unseen data. Tuned hyperparameters optimized the application of Random Erasing, contributing significantly to the models' classification accuracy, especially in challenging tasks like the 5-class classification.

3.2.2 Three Class Classification:

We have performed three class classification by considering Grade 2, Grade 3, and Grade 4 classes. Random Erasing, a powerful data augmentation technique, was utilized to enhance the generalization of deep learning models in image classification tasks. This method involves randomly replacing a rectangular region in the input image with random pixel values, simulating occlusions or corruptions. By exposing the models to a diverse range of inputs during training, Random Erasing promotes robust feature learning and improves model performance on unseen data. Tuned hyperparameters optimized the application of Random Erasing, contributing significantly to the models' classification accuracy, especially in challenging tasks like the 5-class classification.

3.2.3 Binary Classification:

For binary classification, we have merged Grade0 and Grade1 into class 'Healthy' and Grade2, Grade3, Grade4 into class 'Unfit'. After classifying the whole dataset into two classes, we are trimming these two classes such that there are 500 images in each of them.Using the ImageDataGenerator class and its parameters—horizontal flip, rotation, width shift, height shift, and zoom range—the images for each class are balanced. Once every class that is taken into account has been balanced, Once every class that is taken into account has been balanced, Image Augmentation is applied on training and validation data in order to extract more insightful information from them.

Algorithm to Use Transfer Learning

Input: Dataset
Set hyperparameters like Optimizer, Learning Rate, Batch Size, epochs
Model → [Xception, ResNet101, VGG16, InceptionV3, EfficientNetB5, DenseNet201, DenseNet121, InceptionResNetV2]
Output: Predicted class of the image
Insert Required Libraries
Load dataset in notebook
Split Dataset into train, validation, and test
Explore dataset to visualize the sample images and find out the number of images in each class
Apply preprocessing techniques to enhance dataset and apply augmentation to avoid overfitting
Select a model given in the model list for transfer learning and use it as a fine-tune model and design a convolutional neural network architecture

While $i \neq$ epochs
  Compile the model on the train and validation dataset
End While
Test model on the test dataset and calculate performance parameters like confusion matrix, recall, precision, support, test accuracy, test loss, F1-Score

---

3.3 Transfer Learning

A machine learning technique called transfer learning[36] makes use of information from related activities to enhance a model's performance on a target task. When the source and target jobs are somewhat related but may differ in some specific areas, it is especially helpful. Transfer learning reduces the requirement for large amounts of labeled data and training time in the target domain by allowing knowledge from a prior task to be reused instead of starting from scratch when training a new model. Transfer learning in the context of deep learning entails leveraging pre-trained models that have been trained on sizable and varied datasets, such as Convolutional Neural Networks (CNNs) for visual tasks.

3.4 Ensemble Learning

Ensemble learning[52] is a machine learning technique that yields predictors that are more reliable and strong than any one model used alone through the use of several separate models together Ensemble learning's core principle is to take advantage of model variety in order to both minimize the shortcomings of individual models and maximize their combined strengths. Variations in the model design, learning techniques, or training data may be the source of this variance. We have implemented ensemble learning for five classes classification using two classifier models namely DenseNet121 and DenseNet201 using soft voting. Soft voting is an ensemble learning technique where the predictions of numerous independent models are integrated by averaging their predicted probabilities or confidence ratings for each class. It is sometimes referred to as weighted average or probability averaging. In contrast to hard voting, which just chooses the class label receiving the most votes, soft voting considers the degree of certainty or confidence in each model's forecast. Each base model gives each class for a specific instance a probability or confidence score in soft voting. The probabilities or scores indicate how confident the model is in its prediction. Next, a weighted average is used to combine the projected probabilities or scores from each of the basic models to arrive at the final forecast.

3.5 Deep Learning Architectures Implemented

The architectures that were implemented are as follows:

3.5.1 InceptionV3

A deep convolutional neural network (CNN) architecture called Inception V3[42][40] was created with image identification tasks in mind. It was created by Google researchers and is a noteworthy development in the field of computer vision. The architecture allows it to automatically learn hierarchical features from unprocessed visual input by utilizing many layers of convolutional and pooling procedures. The inventive usage of "Inception modules," which are made up of parallel convolutional layers with various filter sizes, is what sets Inception V3 apart. These modules improve the network's capacity to identify intricate patterns in images by enabling it to record a broad variety of characteristics at different spatial scales.

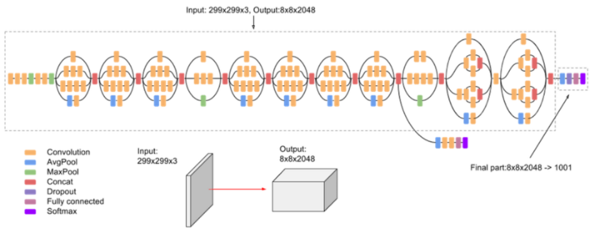

Fig. 5 InceptionV3 Model architecture[19]

3.5.2 InceptionResNetV2

Inception-ResNet-v2[11][39] is the fusion of two powerful architectures, Inception and ResNet and it capitalizes on each paradigm's advantages to provide even greater results. Inception-ResNet-v2, which was first presented by Szegedy et al. in 2016, retains the Inception architecture's multi-scale feature extraction capabilities while adding residual connections for better gradient flow and training convenience. The inception-residual blocks, which combine the residual connections of ResNet with the inception modules of the Inception architecture, are the main innovation of Inception-ResNet-v2. These blocks are made up of convolutional pathways that are parallel and have varying receptive fields.

A concatenation operation and a residual connection come next. Rich hi-

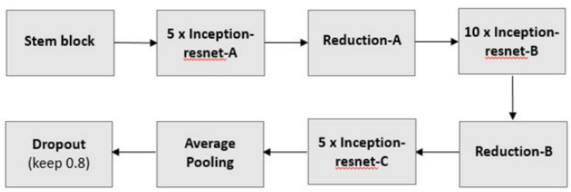

Fig. 6 A basic architecture of InceptionResNetV2[12]

erarchical properties may be captured by the model because to its design, which also guarantees effective training and scalability to deeper architectures. Inception-ResNet-v2 employs other methods in addition to its architectural design, including aggressive data augmentation, dropout regularization, and batch normalization. All things considered, Inception-ResNet-v2 is a monument to the ongoing innovation in deep learning architectures, providing cutting edge results on a range of computer vision applications while preserving computational efficiency and scalability.

3.5.3 DenseNet121

DenseNet121[28], first presented by Huang et al. in 2017, expands on the core ideas of deep learning architectures by presenting new connectivity patterns that facilitate more effective gradient flow and feature reuse. DenseNet's primary innovation is found in its densely connected blocks, which create direct connections between every layer inside a block. Because of this dense connection, it is possible to create deeper networks with fewer parameters and better gradient flow during training by promoting feature reuse across layers. DenseNet121's densely connected blocks let each layer to receive inputs from all preceding levels inside the same block, in contrast to standard CNN architectures where each layer receives inputs only from the preceding layer. This structure of dense connectivity promotes feature reuse and improves feature propagation, which improves information flow across the network. The transition layers that separate the numerous dense blocks that make up the DenseNet121 architecture regulate the development of feature maps and ease the transition between various spatial resolutions. Every dense block consists of several convolutional layers with rectified linear unit (ReLU) activation functions and batch normalization. DenseNet121 offers a powerful blend of dense connectivity, parameter efficiency, and computational efficacy, marking a substantial leap in deep learning architectures. Due to its creative design and outstanding performance, DenseNet121 is now considered a useful tool for many different computer vision jobs and applications.

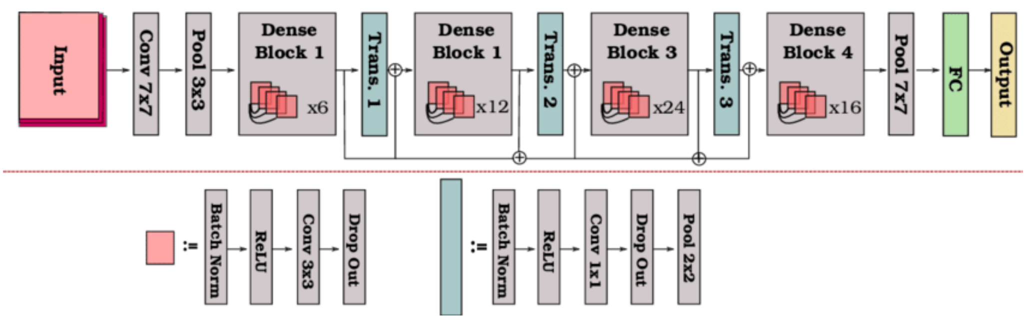

Fig. 7 DenseNet121 architecture [49]

3.5.4 DenseNet201

As a strong convolutional neural network, the DenseNet201[28] deep learning model—a variation of the DenseNet architecture—is renowned for its dense connection patterns and outstanding performance in image classification tasks. DenseNet201, which has 201 layers overall, is distinguished by its distinctive architecture, which encourages feature reuse and gradient flow across the network. DenseNet201 promotes feature propagation and facilitates effective information transmission by establishing feed-forward connections between every layer and every other layer. This enhances learning capabilities and model accuracy.
In addition to improving the model's representative capability, this dense connection topology also helps to mitigate the vanishing gradient issue that deep neural networks sometimes face. Because of this feature, DenseNet201 can train deep networks efficiently even when they have a lot of layers.

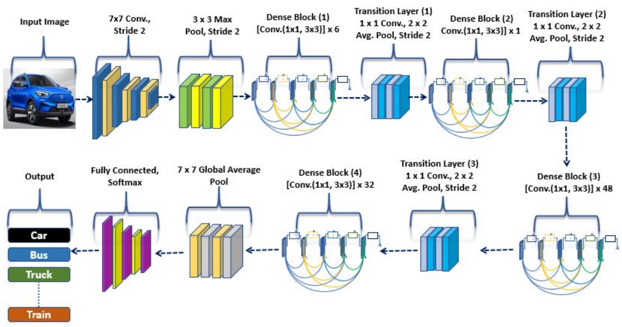

Fig. 8 Densenet201 Model Architecture[37]

3.5.5 EfficientNetB5

EfficientNetB5[6] is a significant advancement in deep learning architectures, namely in the field of convolutional neural networks (CNNs). It belongs to the EfficientNet family and is distinguished by a unique compound scaling technique that strikes the ideal balance between model breadth, depth, and resolution in order to achieve the highest levels of accuracy and efficiency. Building on this foundation, EfficientNetB5 has a hierarchical structure made up of efficient building blocks such as squeeze-and-excitation modules and inverted residual blocks. EfficientNetB5 is able to achieve a desirable balance between model capacity and computational efficiency because of these blocks' meticulous design, which maximizes feature extraction while minimizing computational cost. Compound scaling, which methodically expands the network's depth, width, and resolution in accordance with a predetermined scaling factor, is one of EfficientNetB5's primary advances. By ensuring that the model's capacity scales in accordance with the available computational resources, this method improves performance without imposing an undue amount of computational cost. In comparison to other modern architectures, EfficientNetB5's architectural design, effective building blocks, and compound scaling strategy allow it to achieve state-of-the-art performance on a variety of computer vision tasks with a far smaller number of parameters and computational resources. Transfer learning features and pre-trained models further increase EfficientNetB5's usefulness in the real world by enabling academics and practitioners to tailor the system to certain tasks with a minimal amount of data and processing power.

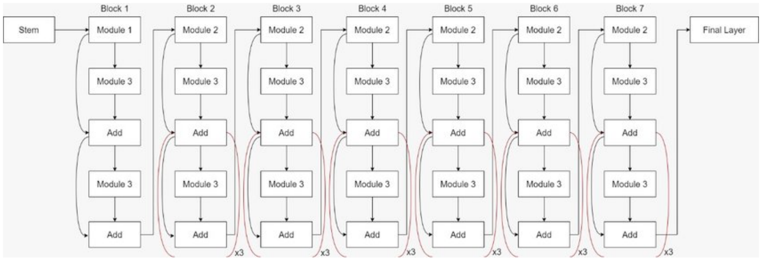

Fig. 9 EfficientNetB5 Model Architecture [27]

3.5.6 Xception

Convolutional neural network (CNN) architecture Xception[31][8] is a trailblazer in deep learning techniques, especially for image identification and computer vision applications. François Chollet, the brains behind the Keras deep learning framework, created Xception, which is notable for its creative feature extraction method and efficient use of parameters. The main building element of Xception is depthwise separable convolutions, in contrast to ordinary CNNs, which usually use standard convolutions. By breaking down the convolutional process into two separate stages—depthwise convolutions and pointwise convolutions—this architectural innovation achieves its goal. Without sacrificing performance, Xception dramatically lowers the number of parameters needed and computational complexity by separating spatial and channel-wise correlations. Xception's depthwise separable convolutions allow it to perform with unmatched effectiveness and efficiency across a range of image recognition applications. Its superiority over earlier systems has been continuously shown empirically, with impressive accuracy and scalability across test datasets. Furthermore, because of its efficiency, Xception is a good fit for implementation in situations with limited resources, such as embedded and mobile devices.

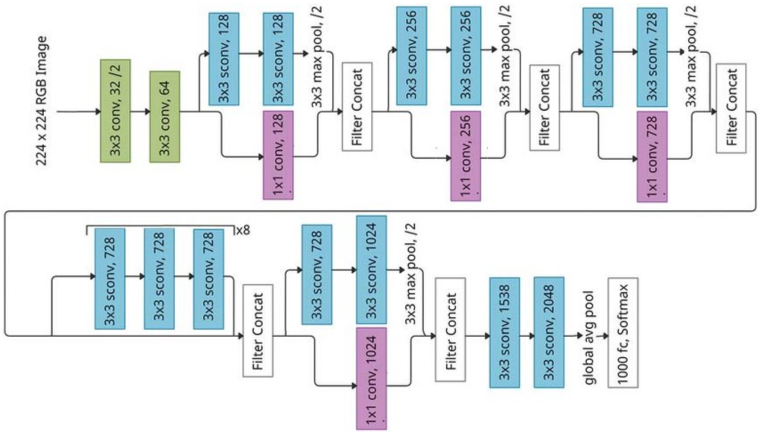

Fig. 10 Xception Model Architecture [34]

3.5.7 ResNet101

ResNet101[41], an acronym for Residual Network with 101 layers, is a significant development in deep learning, especially when it comes to convolutional neural networks (CNNs) used for computer vision and image recognition applications. ResNet101[53], which was unveiled by Kaiming He et al., is an advancement on the original ResNet architecture, which introduced skip connections or residual connections to overcome the difficulties associated with training extremely deep neural networks. The idea for ResNet101 was inspired by the finding that deeper neural networks do not always mean better performance. In actuality, the vanishing gradient problem frequently affects deep networks, making it challenging to train them efficiently. ResNet101 introduces residual connections to solve this problem by enabling direct layer-to-layer information transfer without degradation. The vanishing gradient issue is lessened and the training of incredibly deep networks is made easier by these links. ResNet101 is built as a sequence of residual blocks, each of which has several convolutional layers. The incorporation of skip connections, which omit one or more block layers, is the primary innovation. As a result, the network is able to learn residual mappings, capturing fine-grained information and making deeper network training easier.

3.5.8 VGG16

Convolutional neural network (CNN) architecture VGG16[21][33] is a seminal model in deep learning, especially for image recognition applications. VGG16, which was created by the University of Oxford's Visual Geometry Group (VGG), is a notable breakthrough in CNN design because of its efficiency and simplicity. VGG16 has a 16-layer architecture made up of 3 fully linked layers and 13 convolutional layers. Little 3x3 filters with a stride of 1 and padding

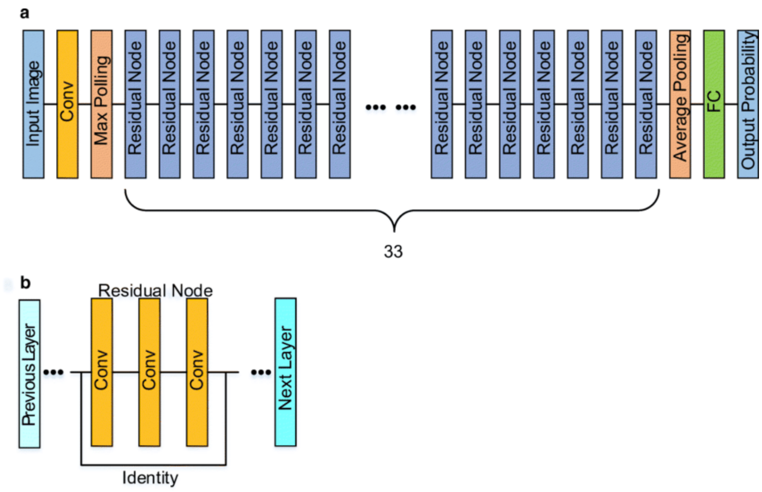

Fig. 11 ResNet101 Model Architecture [13]

are used in the convolutional layers to preserve the input's spatial dimensions. The design is made simpler and training is made easier by the homogeneity of the filter size and spatial resolution. The focus that VGG16 places on depth allows the model to learn ever more complicated features and representations, which is one of its main contributions. The VGG16 model achieves better results in image recognition tasks by learning hierarchical features of increasing abstraction through the stacking of numerous convolutional layers. In a number of benchmarks, including as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), where it performed exceptionally well in picture classification, VGG16 has shown impressive achievement. Its solid performance, ease of use, and efficacy have led to its widespread use in academic and industrial settings.

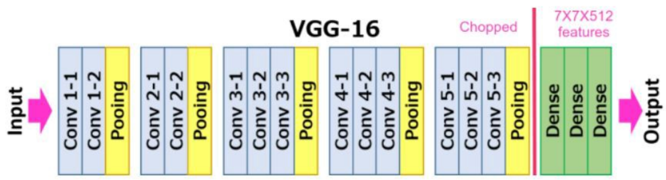

Fig. 12 VGG16 Model Architecture [15]

4 Experiments and Results

This section is dedicated to presenting the results obtained by training and testing the current dataset on using deep-learning architectures.

4.1 Hardware and Software Setup

Google Colab GPU and Kaggle P100 GPU used for training along with TensorFlow, Keras, and Scikit-learn libraries. Python

4.2 Data Division Scheme

Here the dataset is divided into different categories like training dataset, validation dataset and testing dataset. Table 3 describes the division of images in different classes.

Table 3 Number of sample images for the experiments

Class Name	Total Images in a class	No. of Images in Training Set	No. of Images in Validation Set	No. of Images in Test Set
Normal	3253	2286	328	639
Doubtful	1495	1046	158	296
Minimal	2175	1516	212	447
Moderate	392	106	63	223
Severe	251	173	27	51

4.3 Evaluation Criteria

The results are evaluated on various important metrics like accuracy and confusion matrix. The definition of all metrics are given here:

Accuracy : It is a fraction of the total correct predictions.[3]

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

Results were recorded and compared against other models. We have used the Confusion Matrix for the evaluation of the models tested.

A confusion matrix is an NxN matrix where N represents the number of classes being predicted.

5 Results and Discussion

5.1 5-Class Classification

In the 5-class classification task, the deep learning models exhibited varying levels of performance. DenseNet201 (Figure 14) achieved an accuracy of 69.63% after 69 epochs of training with a learning rate of 0.0001 and a batch size of 32. This result (Figure 13) showcases the model's ability to effectively capture and distinguish between the nuances of the five distinct classes.

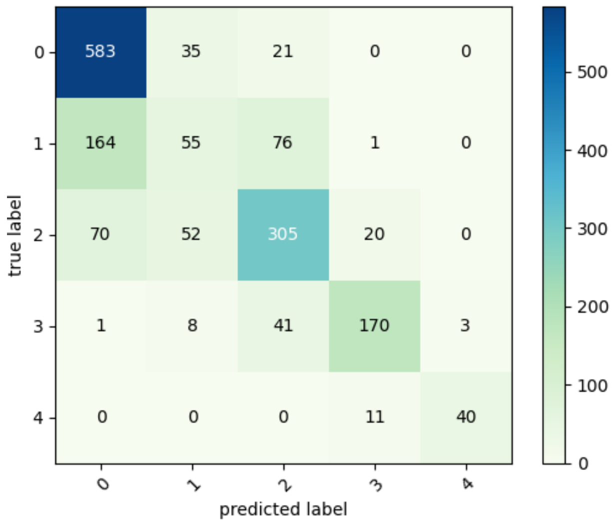

Fig. 13 Confusion Matrix for Densenet201 (5 Classes)

Closely following DenseNet201, DenseNet121 achieved an accuracy of 68.78% under similar training conditions. This suggests that the DenseNet architecture, with its efficient feature extraction and aggregation capabilities, is well-suited for handling multi-class classification problems.

The InceptionV3 model, on the other hand, demonstrated a slightly lower accuracy of 67.57% for the 5-class task. While still a respectable performance, this outcome indicates that the Inception-based architecture may not have been as adept at capturing the intricate relationships between the five classes as the DenseNet models.

A soft voting strategy of Densenet201 and DenseNet121 was also performed. It resulted in an accuracy (Figure 15) of 70.2%, surpassing the solo performing DenseNet201 model. This demonstrates the promise of leveraging model ensembling techniques to enhance classification performance.

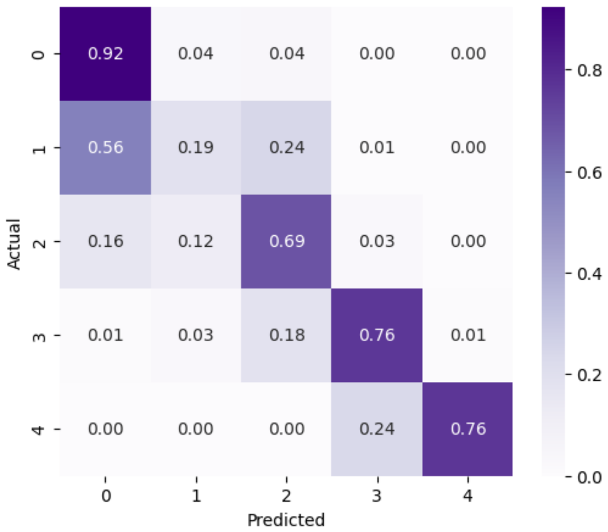

Fig. 15 Confusion Matrix for Stacked Model

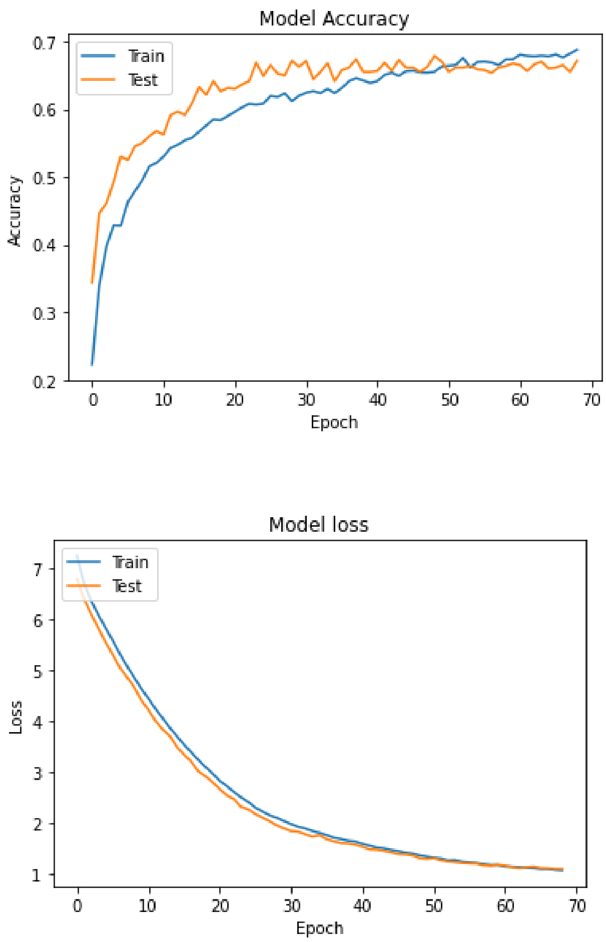

Fig. 14 Training and Validation Graph for Densenet201 (5 Classes)

It also suggests that there may be complementary information or feature representations that can be exploited to improve the overall classification accuracy. This approach opens up avenues for further research into ensemble methods and model combination strategies to unlock the full potential of deep learning architectures.

5.2 3-Class Classification

When the models were adapted for the 3-class classification task, a remarkable improvement in performance was observed. DenseNet201 achieved an exceptional accuracy of 89.5% after 48 epochs, with a learning rate of 0.0001 and a batch size of 32.

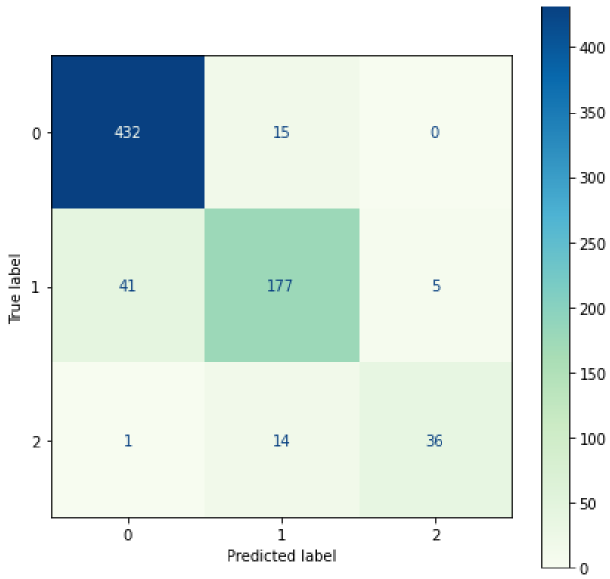

Fig. 16 Confusion Matrix for Densenet201 (3 Classes)

This significant increase in accuracy highlights the models' ability to effectively handle a reduced number of classes, leveraging their robust feature extraction and classification capabilities to achieve near-human-level performance. The consistent performance of the DenseNet models across the 3-class and 5-class tasks underscores their versatility and suitability for a wide range of multi-class classification problems.

5.3 2-Class Classification

In the 2-class classification task, a diverse set of models were evaluated, including DenseNet121, Xception, ResNet101, InceptionNetV3, InceptionResNetV2, and EfficientNetB5. Among these, EfficientNetB5 (Figure 19) demonstrated the highest accuracy of 83.94% after 40 epochs, utilizing an adaptive learning rate of 0.01 and a batch size of 32.

The superior performance of EfficientNetB5 (Table 18) in the binary classification context can be attributed to its efficient architecture and effective feature extraction capabilities. This model's ability to achieve high accuracy with relatively fewer training epochs and a smaller batch size highlights its suitability for applications where computational resources may be limited.

Additionally, the Xception, ResNet101, InceptionNetV3, and InceptionResNetV2 models also exhibited respectable accuracies of 78.01%, 74.76%,

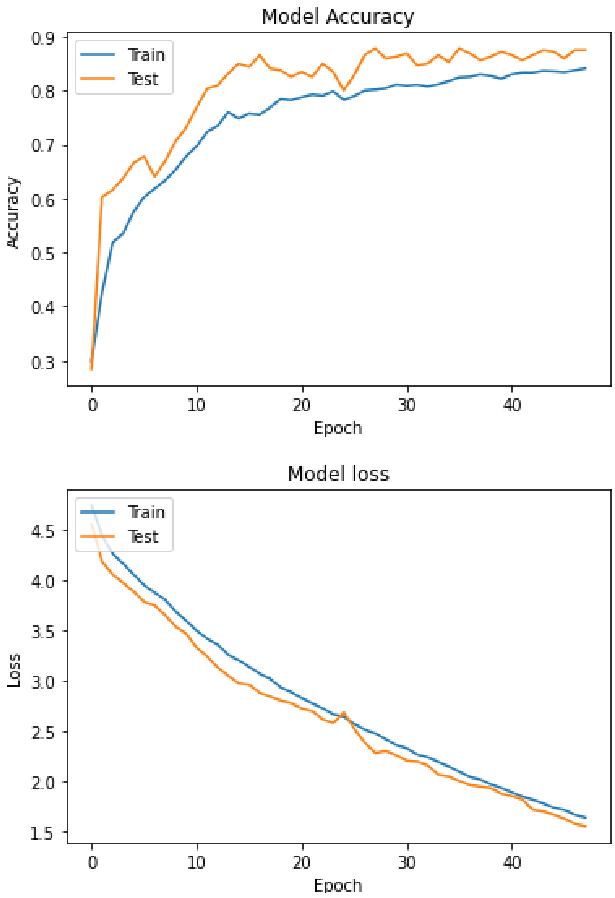

Fig. 17 Training and Validation Graph for Densenet201 (3 Classes)

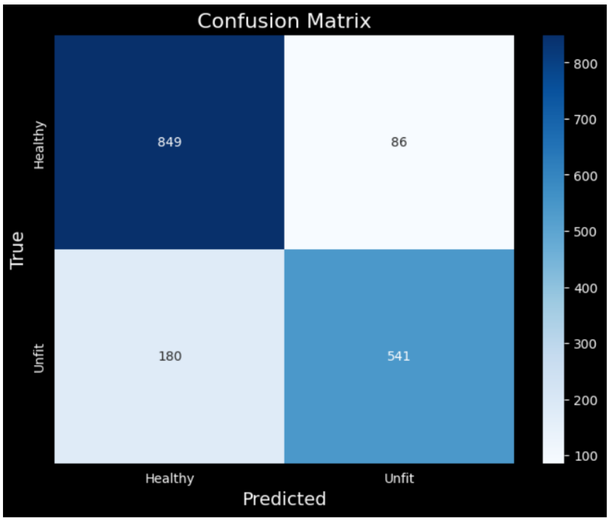

Fig. 18 Confusion Matrix for EfficientNetB5 (2 Classes)

73.73%, and 81.40%, respectively, showcasing the diversity of deep learning architectures that can be effectively employed for binary classification tasks.

The comparison between the mentioned models and the proposed models can be seen in the Table 4.

5.4 Discussion

The results demonstrate the performance of various deep learning models on 2-class, 3-class, and 5-class classification tasks. For the 2-class classification, the EfficientNet model achieved an accuracy of 83.9%, the InceptionNet model achieved 78.3%, and the NasNet model had a lower accuracy of 58.3%. In

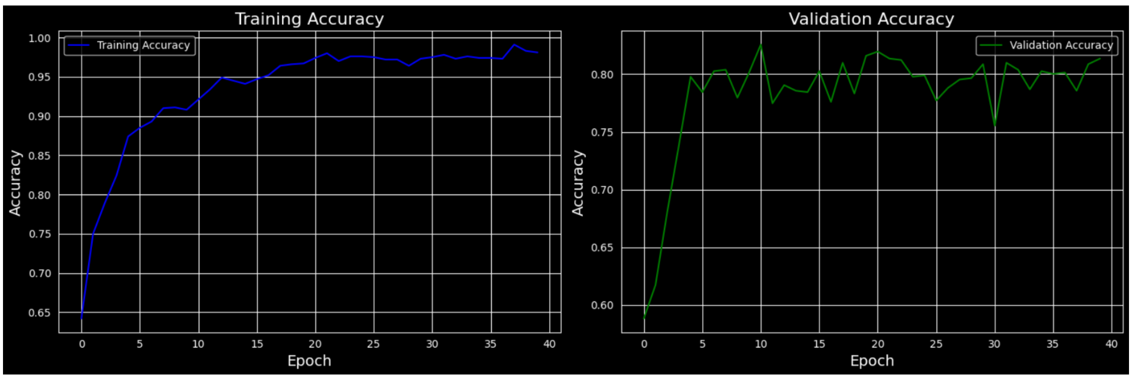

Fig. 19 Training and Validation Graph for EfficientNetB5 (2 Classes)

Table 4 Results Comparision Table

Model	Classes	Learning Rate	Epochs	Batch Size	Accuracy
2-Class Models					
DenseNet121	2	0.01 (Adaptive)	40	20	79.4%
Xception	2	0.01 (Adaptive)	40	20	78.01%
ResNet101	2	0.01 (Adaptive)	40	20	74.76%
InceptionNetV3	2	0.01 (Adaptive)	40	20	73.73%
InceptionResNetV2	2	0.01 (Adaptive)	40	32	81.40%
EfficientNetB5	2	0.01 (Adaptive)	40	32	**83.94%**
3-Class Models					
DenseNet201	3	0.0001	48	32	**89.5%**
DenseNet121	3	0.0001	48	32	86%
InceptionResNetV2	3	0.01 (Adaptive)	40	20	84.74%
VGG16	3	0.01 (Adaptive)	40	20	83.35%
5-Class Models					
DenseNet201	5	0.0001	69	32	69.63%
DenseNet121	5	0.0001	98	32	68.78%
InceptionV3	5	0.0001	95	32	67.57%
Soft Voting Model					
DenseNet201 + DenseNet121	5	-	-	-	**70.2%**

the 3-class classification task, the InceptionNet model performed very well, achieving an accuracy of 92.09%. For the more challenging 5-class classification, the InceptionNet model attained an accuracy of 74.8%. These findings suggest that the complexity of the classification task can significantly impact the performance of deep learning models, and that careful model selection and hyperparameter tuning may be necessary to achieve optimal results across different problem domains. The results also highlight the trade-offs between model complexity and classification accuracy, with simpler models like EfficientNet performing well on the 2-class task, while more complex architectures like InceptionNet excelling on the 3-class and 5-class problems.

6 Conclusion and future Scope

In this study, we adopted an effective method combining prior knowledge and transfer learning to train deep learning models for knee osteoarthritis classification. Our analysis reveals important facts about various divisions of labor. More specifically, the EfficientNet model achieved 83.94% accuracy in two-class classification, while the DenseNet201 model achieved 89.5% accuracy in three classes. Additionally, the Ensemble model(DenseNet201 and DenseNet121) achieved an accuracy of 70.2% on the challenging five-class task. These findings show the effectiveness of the de9ep learning method in correctly classifying the level of knee osteoarthritis. Simple models such as EfficientNet perform well in binary classification, while more complex models such as DenseNet201 and Ensemble show the best accuracy in multi-class. Training analysis and validation of trends, as well as analysis of confusion matrices, provide insight into the model and operational nuances. This highlights the importance of considering model selection and hyperparameter optimization to achieve best results in different tasks. Check the possibility of diagnosis. Using the latest technology and rigorous testing, we are able to produce accurate and reliable automatic diagnostic systems in the medical field. Going forward, continued research aimed at improving design methods and discovering new techniques will hold great promise for improving clinical imaging in musculoskeletal therapy.

The future scope of early detection of knee osteoarthritis using transfer learning holds significant potential. Firstly, refining transfer learning techniques could optimize model performance for this specific application. Additionally, integrating multimodal data could enhance the accuracy and reliability of predictions. Personalized risk assessment models tailored to individual patient characteristics offer promise for targeted interventions. Integration into clinical decision support systems could aid healthcare professionals in making informed decisions. Validation studies are crucial for assessing performance and generalizability across diverse clinical settings. Successful validation could pave the way for clinical translation and widespread implementation. Overall, these advancements have the potential to revolutionize knee osteoarthritis management and improve patient outcomes.

With many new algorithms and models being developed every day, deep learning and visual details seem to have a bright future. Almost every new year brings a new breakthrough in the application of deep learning in our daily lives.

7 Declarations

Conficts of interest The authors declare that they have no confict of interest.
Funding No funding was received to assist with the preparation of this manuscript.
Data Availability The paper uses the publicly available dataset for Knee Osteoarthritis Severity Dataset. The dataset is openly available at https://www.kaggle.com/datasets/tommyngx/kneeoa

References

1. Ahmed, F., Abbas, S., Athar, A., Shahzad, T., Khan, W.A., Alharbi, M., Khan, M.A., Ahmed, A.: Identification of kidney stones in kub x-ray images using vgg16 empowered with explainable artificial intelligence. Scientific Reports **14**(1), 6173 (2024)
2. Alshmrani, G.M.M., Ni, Q., Jiang, R., Pervaiz, H., Elshennawy, N.M.: A deep learning architecture for multi-class lung diseases classification using chest x-ray (cxr) images. Alexandria Engineering Journal **64**, 923–935 (2023)
3. Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., Christy, W., Cooke, T., Greenwald, R., Hochberg, M., et al.: Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology **29**(8), 1039–1049 (1986)
4. Alzubaidi, L., Salhi, A., A. Fadhel, M., Bai, J., Hollman, F., Italia, K., Pareyon, R., Albahri, A., Ouyang, C., Santamaría, J., et al.: Trustworthy deep learning framework for the detection of abnormalities in x-ray shoulder images. Plos one **19**(3), e0299545 (2024)

5. Athisayamani, S., Antonyswamy, R.S., Sarveshwaran, V., Almeshari, M., Alzamil, Y., Ravi, V.: Feature extraction using a residual deep convolutional neural network (resnet-152) and optimized feature dimension reduction for mri brain tumor classification. Diagnostics **13**(4), 668 (2023)
6. Bhawarkar, Y., Bhure, K., Chaudhary, V., Alte, B.: Diabetic retinopathy detection from fundus images using multi-tasking model with efficientnet b5. In: ITM web of conferences, vol. 44, p. 03027. EDP Sciences (2022)
7. Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K.G., Murphy, K.: Deep learning for chest x-ray analysis: A survey. Medical Image Analysis **72**, 102125 (2021)
8. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251–1258 (2017)
9. Chouhan, V., Singh, S.K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R., De Albuquerque, V.H.C.: A novel transfer learning based approach for pneumonia detection in chest x-ray images. Applied Sciences **10**(2), 559 (2020)
10. Daneshmand, M., Panfilov, E., Bayramoglu, N., Korhonen, R.K., Saarakkala, S.: Deep learning based detection of osteophytes in radiographs and magnetic resonance imagings of the knee using 2d and 3d morphology. Journal of Orthopaedic Research® (2024)
11. Dash, S., Sethy, P.K., Behera, S.K.: Cervical transformation zone segmentation and classification based on improved inception-resnet-v2 using colposcopy images. Cancer Informatics **22**, 11769351231161477 (2023)
12. DEMİR, A., YILMAZ, F.: Inception-resnet-v2 with leakyrelu and averagepooling for more reliable and accurate classification of chest x-ray images. In: 2020 Medical Technologies Congress (TIPTEKNO), pp. 1–4. IEEE (2020)
13. Dhibar, S.: Resnet101 and dae for enhance quality and classification accuracy in skin cancer imaging. arXiv preprint arXiv:2403.14248 (2024)
14. Guan, Q., Chen, Q., Huang, Y.: An improved heteroscedastic modeling method for chest x-ray image classification with noisy labels. Algorithms **16**(5), 239 (2023)
15. Izdihar, N., Rahayu, S.B., Venkatesan, K.: Comparison analysis of cxr images in detecting pneumonia using vgg16 and resnet50 convolution neural network model. JOIV: International Journal on Informatics Visualization **8**(1), 326–332 (2024)
16. Jain, R.K., Sharma, P.K., Gaj, S., Sur, A., Ghosh, P.: Knee osteoarthritis severity prediction using an attentive multi-scale deep convolutional neural network. Multimedia Tools and Applications **83**(3), 6925–6942 (2024)
17. Khalid, A., Senan, E.M., Al-Wagih, K., Ali Al-Azzam, M.M., Alkhraisha, Z.M.: Hybrid techniques of x-ray analysis to predict knee osteoarthritis grades based on fusion features of cnn and handcrafted. Diagnostics **13**(9), 1609 (2023)
18. Kim, S., Rebmann, P., Tran, P.H., Kellner, E., Reisert, M., Steybe, D., Bayer, J., Bamberg, F., Kotter, E., Russe, M.: Multiclass datasets expand neural network utility: an example on ankle radiographs. International journal of computer assisted radiology and surgery **18**(5), 819–826 (2023)
19. Lin, C., Li, L., Luo, W., Wang, K.C., Guo, J.: Transfer learning based traffic sign recognition using inception-v3 model. Periodica Polytechnica Transportation Engineering **47**(3), 242–250 (2019)
20. Manjunath, R., Ghanshala, A., Kwadiki, K.: Deep learning algorithm performance evaluation in detection and classification of liver disease using ct images. Multimedia Tools and Applications **83**(1), 2773–2790 (2024)
21. Mateen, M., Wen, J., Nasrullah, Song, S., Huang, Z.: Fundus image classification using vgg-19 architecture with pca and svd. Symmetry **11**(1), 1 (2018)
22. Nakata, N., Siina, T.: Ensemble learning of multiple models using deep learning for multiclass classification of ultrasound images of hepatic masses. Bioengineering **10**(1), 69 (2023)
23. Nasir, N., Kansal, A., Barneih, F., Al-Shaltone, O., Bonny, T., Al-Shabi, M., Al Shammaa, A.: Multi-modal image classification of covid-19 cases using computed tomography and x-rays scans. Intelligent Systems with Applications **17**, 200160 (2023)
24. Nayak, S.R., Nayak, J., Sinha, U., Arora, V., Ghosh, U., Satapathy, S.C.: An automated lightweight deep neural network for diagnosis of covid-19 from chest x-ray images. Arabian journal for science and engineering **48**(8), 11085–11102 (2023)

25. Ni, M., Zhao, Y., Wen, X., Lang, N., Wang, Q., Chen, W., Zeng, X., Yuan, H.: Deep learning-assisted classification of calcaneofibular ligament injuries in the ankle joint. Quantitative Imaging in Medicine and Surgery **13**(1), 80 (2023)
26. Nomoto, K., Hanada, M., Hotta, K., Matsuyama, Y.: Distribution of coronal plane alignment of the knee classification does not change as knee osteoarthritis progresses: a longitudinal study from the toei study. Knee Surgery, Sports Traumatology, Arthroscopy **31**(12), 5507–5513 (2023)
27. Özdemır, Z., Keleş, H.Y.: Covid-19 detection in chest x-ray images with deep learning. In: 2021 29th Signal Processing and Communications Applications Conference (SIU), pp. 1–4. IEEE (2021)
28. Pappula, S., Nadendla, T., Lomadugu, N.B., Nalla, S.R.: Detection and classification of pneumonia using deep learning by the dense net-121 model. In: 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 1671–1675. IEEE (2023)
29. Paul, A., Tang, Y.X., Shen, T.C., Summers, R.M.: Discriminative ensemble learning for few-shot chest x-ray diagnosis. Medical image analysis **68**, 101911 (2021)
30. Pi, S.W., Lee, B.D., Lee, M.S., Lee, H.J.: Ensemble deep-learning networks for automated osteoarthritis grading in knee x-ray images. Scientific Reports **13**(1), 22887 (2023)
31. Polat, Ö.: Detection of covid-19 from chest ct images using xception architecture: a deep transfer learning based approach. Sakarya University Journal of Science **25**(3), 800–810 (2021)
32. Prezja, F., Annala, L., Kiiskinen, S., Ojala, T.: Exploring the efficacy of base data augmentation methods in deep learning-based radiograph classification of knee joint osteoarthritis. Algorithms **17**(1), 8 (2023)
33. Qassim, H., Verma, A., Feinzimer, D.: Compressed residual-vgg16 cnn model for big data places image recognition. In: 2018 IEEE 8th annual computing and communication workshop and conference (CCWC), pp. 169–175. IEEE (2018)
34. Rahimzadeh, M., Attar, A.: A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2. Informatics in medicine unlocked **19**, 100360 (2020)
35. Rajamohan, H.R., Wang, T., Leung, K., Chang, G., Cho, K., Kijowski, R., Deniz, C.M.: Prediction of total knee replacement using deep learning analysis of knee mri. Scientific reports **13**(1), 6922 (2023)
36. Salehi, A.W., Khan, S., Gupta, G., Alabduallah, B.I., Almjally, A., Alsolai, H., Siddiqui, T., Mellit, A.: A study of cnn and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability **15**(7), 5930 (2023)
37. Sanghvi, H.A., Patel, R.H., Agarwal, A., Gupta, S., Sawhney, V., Pandya, A.S.: A deep learning approach for classification of covid and pneumonia using densenet-201. International Journal of Imaging Systems and Technology **33**(1), 18–38 (2023)
38. Say, D., Zidi, S., Qaisar, S.M., Krichen, M.: Automated categorization of multiclass welding defects using the x-ray image augmentation and convolutional neural network. Sensors **23**(14), 6422 (2023)
39. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI conference on artificial intelligence, vol. 31 (2017)
40. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826 (2016)
41. Targ, S., Almeida, D., Lyman, K.: Resnet in resnet: Generalizing residual architectures. arXiv preprint arXiv:1603.08029 (2016)
42. Thangaraj, R., Pandiyan, P., Ramakrishnan, J., Nallakumar, R., Eswaran, S.: A deep convolution neural network for automated covid-19 disease detection using chest x-ray images. Healthcare Analytics **4**, 100278 (2023)
43. Tiulpin, A., Thevenot, J., Rahtu, E., Lehenkari, P., Saarakkala, S.: Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. Scientific reports **8**(1), 1727 (2018)

44. Toyooka, S., Osaki, Y., Masuda, H., Arai, N., Miyamoto, W., Ando, S., Kawano, H., Nakagawa, T.: Distribution of coronal plane alignment of the knee classification in patients with knee osteoarthritis in japan. The Journal of Knee Surgery **36**(07), 738–743 (2023)
45. Wang, C.T., Huang, B., Thogiti, N., Zhu, W.X., Chang, C.H., Pao, J.L., Lai, F.: Successful real-world application of an osteoarthritis classification deep-learning model using 9210 knees—an orthopedic surgeon's view. Journal of Orthopaedic Research® **41**(4), 737–746 (2023)
46. Wang, H., Ying, J., Liu, J., Yu, T., Huang, D.: Harnessing resnet50 and senet for enhanced ankle fracture identification. BMC Musculoskeletal Disorders **25**(1), 250 (2024)
47. Xie, Y., Li, X., Chen, F., Wen, R., Jing, Y., Liu, C., Wang, J.: Artificial intelligence diagnostic model for multi-site fracture x-ray images of extremities based on deep convolutional neural networks. Quantitative Imaging in Medicine and Surgery **14**(2), 1930 (2024)
48. Xing, X., Rafique, M.U., Liang, G., Blanton, H., Zhang, Y., Wang, C., Jacobs, N., Lin, A.L.: Efficient training on alzheimer's disease diagnosis with learnable weighted pooling for 3d pet brain image classification. Electronics **12**(2), 467 (2023)
49. Yan, F., Huang, X., Yao, Y., Lu, M., Li, M.: Combining lstm and densenet for automatic annotation and classification of chest x-ray images. IEEE Access **7**, 74181–74189 (2019)
50. Yeoh, P.S.Q., Lai, K.W., Goh, S.L., Hasikin, K., Wu, X., Li, P.: Transfer learning-assisted 3d deep learning models for knee osteoarthritis detection: Data from the osteoarthritis initiative. Frontiers in Bioengineering and Biotechnology **11**, 1164655 (2023)
51. Ying, X., Liu, H., Huang, R.: Covid-19 chest x-ray image classification in the presence of noisy labels. Displays **77**, 102370 (2023)
52. Zeng, T., Wu, L., Peduto, D., Glade, T., Hayakawa, Y.S., Yin, K.: Ensemble learning framework for landslide susceptibility mapping: Different basic classifier and ensemble strategy. Geoscience Frontiers **14**(6), 101645 (2023)
53. Zhang, Q.: A novel resnet101 model based on dense dilated convolution for image classification. SN Applied Sciences **4**, 1–13 (2022)