INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND MANAGEMENT (ISJEM) ISSN: 2583-6129
CONFERENCE PROCEEDINGS: VOLUME: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME116
AN INTERNATIONAL SCHOLARLY || MULTIDISCIPLINARY || OPEN ACCESS || INDEXING IN ALL MAJOR DATABASE & METADATA

An Ensemble Deep Learning Approach for Automated Knee Osteoarthritis Detection using X-Ray Images

Hema padmini Ganji
Dept of ECE-embedded systems,Sidhardha educational academy group of institutions, C.Gollapalli, Tirupati.
Affliated to JNTUA
Ananthapur,India.
ganjipadmini@gmail.com

Abstract— Knee osteoarthritis is a common type of joint problem where the cartilage in the knee slowly wears away over time. It often causes long-term pain, stiffness, and reduced movement, especially in older people. Finding KOA early and correctly assessing its severity is important for better treatment and to stop more joint damage from happening. Conventional diagnosis is mainly conducted via visual assessment of knee X-ray images utilizing the Kellgren-Lawrence (KL) grading system. Reading X-ray pictures by hand takes a lot of time and can vary depending on the doctor's skill and personal opinions. To address these limitations, this study proposes an automated deep learning-based framework for the detection and classification of knee osteoarthritis from radiographic images. The suggested method uses transfer learning with two strong convolutional neural network models, InceptionV3 and NASNetLarge, to get detailed features from knee X-ray images. Before starting the training of the model, certain image processing steps like resizing, normalizing, and augmenting the images are done. These steps help in making the data better quality and allow the model to perform well on different types of data. The system sorts knee images into various levels of seriousness according to the KL grading scale. Additionally, a group approach is used to bring together the predictions from both models, which helps make the classification more dependable and boosts the overall results.

Keywords— Transfer learning, Kellgren - Lawrence(KL) Grading, Convolutional Neural Networks(CNN)

I. INTRODUCTION

Knee osteoarthritis is a long-term condition that causes the breakdown of joint cartilage in the knee, affecting many people around the world. It occurs when the cartilage in the knee joint gradually deteriorates, leading to pain, stiffness, swelling, and decreased mobility.

It occurs when the cartilage in the knee joint gradually deteriorates, leading to pain, stiffness, swelling, and decreased mobility. This condition often affects older people and those who have certain risks, like being overweight, having had joint injuries, getting older, or having a family history of it. As the disease gets worse, it can greatly lower the quality of life by making it hard to do everyday physical tasks and leading to ongoing pain and discomfort. As the global population ages, the incidence of knee osteoarthritis is anticipated to increase in the future, positioning it as a significant public health issue.

Finding the disease early and correctly judging how bad it helps in taking care of knee osteoarthritis and stopping more harm to the joint. In real-world medical settings, X-ray imaging is commonly used to look at changes in the structure of the knee joint. Doctors often use the Kellgren-Lawrence grading system to figure out how bad osteoarthritis is. They look at things like how narrow the space in the joint is, whether there are bone spurs, and if there are any changes in the shape of the bones. The KL grading scale groups knee osteoarthritis into different levels, starting from normal to severe. However, manual interpretation of X-ray images can be time-consuming and may vary depending on the experience and judgment of radiologists, which can sometimes lead to inconsistent diagnoses. Recent developments in artificial intelligence, particularly deep learning, have created new opportunities for automated medical image analysis. Deep learning methods, especially convolutional neural networks (CNNs), have shown very good results in tasks like recognizing images, finding objects, and extracting features. These models are capable of automatically learning complex patterns from large datasets without requiring manual feature engineering.

INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND MANAGEMENT (ISJEM) ISSN: 2583-6129
CONFERENCE PROCEEDINGS: VOLUME: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME116
AN INTERNATIONAL SCHOLARLY || MULTIDISCIPLINARY || OPEN ACCESS || INDEXING IN ALL MAJOR DATABASE & METADATA

Because of this, deep learning has turned out to be a good method for creating computer-aided diagnosis tools that help doctors find diseases more precisely and quickly. In the area of medical imaging, deep learning models are commonly used to detect diseases from radiographic images. Research shows that CNNs can effectively detect and sort knee osteoarthritis using X-ray images by learning important details about the knee's structure. Transfer learning methods use models that were already trained on big sets of images. These methods help improve the performance of the systems by making it easier to extract useful features and speed up the training process. Popular models like InceptionV3 and NASNetLarge work well for many image recognition jobs because they have complex layers that help them understand detailed visual features.

Inspired by these developments, this study introduces a deep learning approach to automatically detect and classify knee osteoarthritis using knee X-ray images. The suggested method uses transfer learning with two powerful convolutional neural network models, InceptionV3 and NASNetLarge, to get useful information from X-ray images. Image preprocessing techniques such as resizing, normalization, and augmentation are used to enhance data quality and improve model generalization. Additionally, a group approach is used to merge the results from both models to enhance the accuracy and trustworthiness of the classification.

The main goal of this project is to create a smart and dependable system that helps doctors spot knee osteoarthritis and figure out how serious it is by looking at X-ray pictures. By using deep learning along with medical image analysis, the suggested system is designed to help doctors get quicker and more reliable diagnostic assistance. Such automated systems have the potential to enhance early detection, decrease diagnostic workload, and aid in making better clinical decisions for the management of knee osteoarthritis.

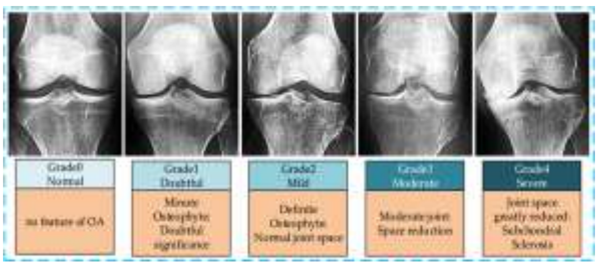

Fig. 1. Kellgren-Lawrence Grading System.

II. LITERATURE REVIEW

Automated detection of knee osteoarthritis (KOA) using medical imaging has gained considerable attention in recent years due to the rapid development of artificial intelligence and deep learning techniques. Doctors usually check for knee osteoarthritis by looking at X-ray pictures with their eyes and experience, using a system called Kellgren-Lawrence to rate the condition. However, when people evaluate medical images by hand, their opinions can vary, and this can cause differences in results, especially when looking at a lot of images. As a result, various diagnostic methods have been developed to enhance the accuracy and efficiency of osteoarthritis detection. Early studies concentrated on classical machine learning techniques for the classification of knee osteoarthritis. These methods usually required people to manually extract features like texture details, shape descriptions, and statistical data from X-ray images. After taking out the important features, classifiers like Support Vector Machines (SVM), Random Forests, and Artificial Neural Networks were used to determine the levels of osteoarthritis severity. Even though these methods had some good results, their effectiveness mostly relied on the quality of the features that were created by people, which made it hard for them to work well with different sets of data. With the advancement of deep learning, convolutional neural networks (CNNs) have become widely used for medical image analysis due to their capability to automatically learn discriminative features from raw images. CNN-based models have been effectively used to find changes in the structure of knee joints, such as narrowing of the joint space and the development of bone spurs, which are important signs of osteoarthritis. These models greatly minimize the need for manual feature engineering and enhance the reliability of automated diagnosis systems. Many scientists have studied deep learning methods to automatically classify knee osteoarthritis. Research shows that CNN models work well for looking at knee X-rays and can accurately predict how serious a disease is. Deep neural networks trained on large datasets are capable of capturing complex visual patterns and identifying subtle differences between different stages of osteoarthritis. Deep neural networks trained on large datasets are capable of capturing complex visual patterns and identifying subtle differences between different stages of osteoarthritis.

Deep learning methods have been added to computer-aided diagnosis tools to help doctors spot issues in X-ray images. Transfer learning has helped deep learning models work better for tasks in medical imaging. In this approach, pre-trained neural network models, which have been developed using large-scale image datasets, are adapted for specific medical applications. Some well-known network designs like VGG, ResNet, Inception, and DenseNet have worked well for identifying and categorizing osteoarthritis. These models are able to extract high-level image features and significantly cut down the training time needed for model development.

More recently, ensemble learning techniques have been introduced to improve the robustness and accuracy of osteoarthritis detection systems. Ensemble models use predictions from several different deep learning models to get better results when applying them to new data. Using different models together can help cut down on mistakes and make medical diagnoses more dependable.

Despite the significant progress made in this area, challenges still exist in developing highly accurate and efficient automated systems for knee osteoarthritis detection. Different image quality, not enough data available, and how diseases develop differently can all impact how well a model works. So, we need better systems that use powerful deep learning methods and combine multiple techniques to make classification more accurate.

In response to these challenges, this work presents a deep learning framework that employs transfer learning with two strong convolutional neural network architectures, InceptionV3 and NASNetLarge, for the automated detection and classification of knee osteoarthritis from X-ray images. By using the features from both models together in an ensemble method, the system is designed to enhance the accuracy of diagnosis and offer dependable support for making medical decisions.

III. METHODS AND METHODOLOGY

The proposed framework employs deep learning techniques to automatically detect and classify knee osteoarthritis (KOA) from radiographic images. The model looks at X-ray pictures of the knee and guesses how bad osteoarthritis is by using the Kellgren-Lawrence grading system. This system groups the disease into five different levels, starting from normal and going up to the most severe. In this work, a transfer learning-based convolutional neural network framework is created using two advanced architectures: InceptionV3 and NASNetLarge. These models are able to extract high-level image features from radiographs and detect structural abnormalities linked to osteoarthritis. Another approach is used to combine the predictions from both networks, which helps make the classification more accurate and reliable.

A. Input Data

The study used a dataset with 1650 knee X-ray images that show various levels of osteoarthritis severity. The dataset is categorized based on the KL grading scale, with the distribution comprising 514 images of Grade 0, 477 images of Grade 1, 232 images of Grade 2, 221 images of Grade 3, and 206 images of Grade 4.

To make sure the model is trained and tested properly, the dataset is split into three different parts:

Training set (70%) - used for learning the model's parameters.

Validation set (20%) - used for hyperparameter tuning and performance monitoring.

Test set (10%) - this is used to check how well the final model works with data it has never seen before.

This way of splitting the data helps stop the model from learning too much from the examples and makes it better at working with new X-ray pictures.

B. Image Pre-Processing

Before using the images in the deep learning model, some steps are taken to prepare the data, making it more consistent and helping the model work better.

Image Resizing

All the X-ray images are changed to a size of 224 by 224 pixels, which is the size that the deep learning models need as input.

Normalization

Pixel intensity values are normalized to ensure consistent input ranges and enhance the stability of the training process.

Data Augmentation

To make the dataset more varied and help the model work better with different kinds of data, techniques like rotating images, flipping them left to right, and zooming in or out are used. These methods create different image angles and help prevent the model from learning too much from specific examples.

c. Feature Extraction using Transfer Learning

Deep learning models generally need extensive datasets to achieve optimal performance. However, medical imaging datasets are usually not very large. To solve this problem, the suggested approach uses transfer learning. Transfer learning uses a neural network that was already trained on a big set of images and then modifies it to work for a particular job.

INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND MANAGEMENT (ISJEM) ISSN: 2583-6129
CONFERENCE PROCEEDINGS: VOLUME: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME116
AN INTERNATIONAL SCHOLARLY || MULTIDISCIPLINARY || OPEN ACCESS || INDEXING IN ALL MAJOR DATABASE & METADATA

In this research, two well-known deep convolutional neural networks, InceptionV3 and NASNetLarge, are utilized as feature extractors.

In this research, two well-known deep convolutional neural networks, InceptionV3 and NASNetLarge, are utilized as feature extractors. These models include multiple convolutional layers that can learn hierarchical representations of images. The first layers of the model detect basic features like edges and textures, whereas the deeper layers understand more complicated patterns that relate to the structure of the knee joint.

The last parts of the already trained networks are changed to fit the five-class classification task related to the severity of knee osteoarthritis. By utilizing knowledge from pre-trained networks, the model can learn meaningful features more efficiently and achieve improved performance even with limited training data.

D. Convolutional Layers

The convolutional layer is the main part of a convolutional neural network. It uses a group of filters that can be learned to look for patterns in the image and pick out important details. Each filter moves over the image and calculates the dot product between the filter's weights and the part of the image it's covering, creating a feature map that shows key structures.

In knee osteoarthritis detection, convolutional layers help identify radiographic characteristics such as:

- joint space narrowing
- bone texture variations
- osteophyte formation
- cartilage degradation patterns

By stacking several convolutional layers, the neural network slowly learns more and more complex ways to represent the structures of the knee joint.

E. Activation Layer (Leaky ReLU)

Activation functions add non-linear elements to neural networks, allowing them to understand and learn more complicated patterns in data. In this study, the Leaky Rectified Linear Unit (Leaky ReLU) activation function is employed.

Unlike regular ReLU, which gives zero when the input is negative, Leaky ReLU lets a tiny amount of gradient pass through for negative values. This characteristic prevents the "vanishing gradient" problem, where neurons stop updating during training due to zero gradients.

Using Leaky ReLU offers several benefits:

- improved gradient flow during backpropagation
- enhanced learning stability in deep networks
- better representation of subtle image features

In the case of analysing knee osteoarthritis, this activation function helps the network notice small changes in the images that could show early signs of joint wear and damage.

F. Pooling Layer

Pooling layers help shrink the size of feature maps but keep the important details. This process reduces computational complexity and enhances the model's efficiency.

The most common way to pool in CNNs is called max pooling, and it picks the largest value from a certain area of the feature map. When you do this action, the network pays more attention to the most important features that the convolutional filters have found.

Pooling helps the model recognize key features even if those features are in slightly different spots in the image. This makes the model's performance stable regardless of where the features appear.

G. Fully Connected Layer

After using convolutional and pooling layers to extract features, the feature maps are turned into a single line of numbers. This vector is then passed through fully connected layers, which handle high-level reasoning and classification.

Each neuron in the fully connected layer gets input from every neuron in the layer before it, which lets the network mix the features it has found and make predictions. The last layer of the model creates a probability distribution for the five KL grades. The class with the highest probability is selected as the predicted severity level of knee osteoarthritis.

H. Ensemble Learning Strategy

This study uses an ensemble learning method to boost the accuracy and reliability of classification. Instead of using just one neural network model, the predictions from both InceptionV3 and NASNetLarge are put together to create the final result.The group model uses a soft voting method, where the predicted chances from both networks are combined to decide the final result. This approach uses the best parts of both designs to make it less likely that the predictions will be wrong.

Ensemble learning provides several advantages:

- improved prediction stability

- reduced model variance
- better generalization performance

I. Model Optimization

The training of the proposed model uses the Adam optimizer, which changes the network's weights by using learning rates that adjust automatically. Adam is commonly used in deep learning because it helps models reach good results quicker and keeps the training process steady and reliable.

To enhance model generalization, regularization techniques such as dropout layers are incorporated. During training, dropout randomly turns off some neurons, which stops the network from relying too much on any single feature.

Hyperparameters like learning rate, batch size, and training epochs are closely adjusted using the validation data to get the best possible results in classification.

IV. RESULTS

This part shows how well the new deep learning system works for automatically sorting out knee osteoarthritis from X-ray pictures. The proposed system combines InceptionV3 and NASNetLarge architectures using an ensemble learning approach to enhance classification accuracy and reliability. The models' performance was checked using measures like accuracy, precision, recall, Fl-score, and confusion matrix analysis.

A. Training and Validation Accuracy

The accuracy curves for training and validation were looked at to check how well the networks were learning during the training process.During the implementation of the InceptionV3 model, the training accuracy gradually improved with an increase in the number of epochs, suggesting that the model effectively learned key features from the knee X-ray images. The validation accuracy kept getting better and finally reached around 95 to 96 percent, showing that the model works well with new, unseen data. The small difference between the training and validation accuracy shows that the model isn't overfitting much.

The NASNetLarge model also showed good learning results during training. The training accuracy steadily increased and reached nearly 96-97%, while the validation accuracy reached approximately 97-98%. The NASNetLarge model was able to extract features better because it has a deeper structure and a more efficient design from neural architecture search. The fact that the training and validation accuracy curves come together shows that the model successfully learned features that help tell the difference in how severe osteoarthritis is.

The gradual progression of both curves indicates that the applied preprocessing methods, such as data augmentation and normalization, enhanced model stability and led to better generalization performance.

B. Confusion Matrix Analysis

The confusion matrix gives a clear look at how well the ensemble model performed in classifying the five Kellgren-Lawrence (KL) grades.

The diagonal elements in the confusion matrix show how many samples were classified correctly, and the elements not on the diagonal show how many samples were classified incorrectly.

The results show that the model got most of the images right in each category.

Normal (Grade 0): Only a small percentage of the 492 photos were incorrectly identified as adjacent classes.

Doubtful (Grade 1): There was little misunderstanding with Grade 0 and Grade 3, and 476 photos were accurately identified.

Mild (Grade 2): 230 photos were accurately categorized, indicating excellent performance in identifying early signs of osteoarthritis.

Moderate (Grade 3): There were relatively few incorrect classifications and 220 photos were correctly predicted.

Severe (Grade 4): The model performed exceptionally well in recognizing cases of advanced osteoarthritis, as all 206 photos were accurately identified.

The majority of misclassifications happened across adjacent severity ratings, which makes sense given the minute variations between successive phases of osteoarthritis. Despite these difficulties, the model showed a high degree of class discrimination.

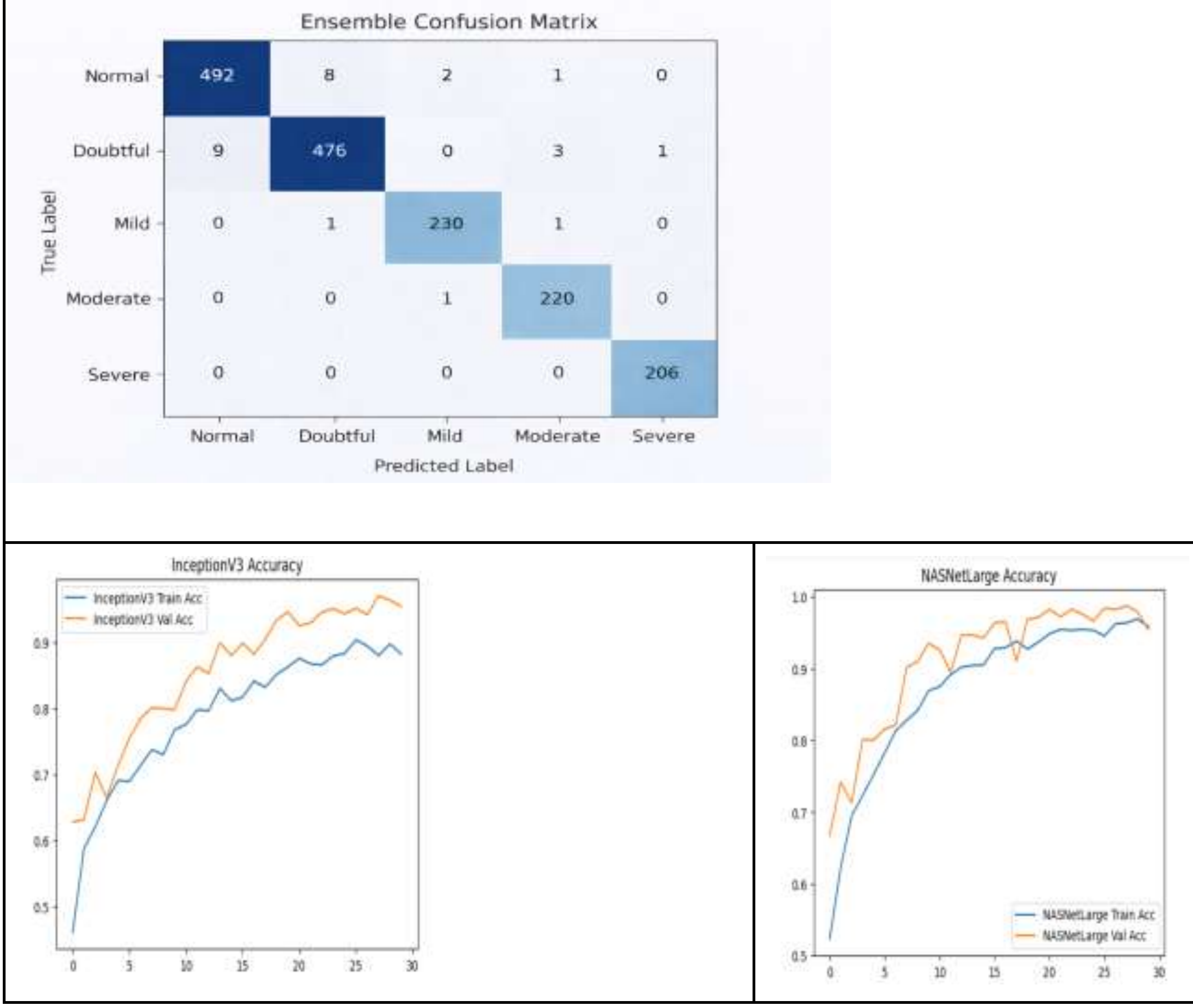

Fig2. Confusion Matrix of InceptionV3 and NASNetLarge and Training Loss and Accuracy of Inception v3 and NASNetLarge

C. Performance in Classification

The classification report provides more evidence of the suggested model's efficacy. Precision, recall, and F1-score are the evaluation measures for every class.
The outcomes demonstrate that the model continuously performed well in every class:

- Grade 0 (Normal): Recall = 0.98, Precision = 0.98
- Grade 1 (Doubtful): Recall = 0.97, Precision = 0.98
- Grade 2 (Mild): Recall = 0.99, Precision = 0.99
- Grade 3 (Moderate): Recall = 1.00, Precision = 0.98
- Grade 4 (Severe): Recall = 1.00, Precision = 1.00

For the majority of classes, the F1-score ranges from 0.98 to 1.00, indicating a performance that strikes a balance between recall and precision. The model's overall ensemble accuracy of 98.36% shows how well predictions from several deep learning architectures may be combined.
Furthermore, the weighted-average and macro-average scores are both roughly 0.98–0.99, suggesting that the model functions consistently at various levels of osteoarthritis severity.

D. Discussion of Model Performance

The high classification accuracy achieved by the proposed framework can be attributed to several key factors.
First, using transfer learning let the model use knowledge from big image datasets, which helped it get useful features even when the medical dataset wasn't very large.
Second, using the Leaky ReLU activation function helped the neural network learn better by stopping the issue of neurons not working properly and keeping the gradient flow during the backpropagation process.
Third, data augmentation techniques increased the diversity of the training dataset and helped the model generalize better to unseen radiographic images.

In the end, using an ensemble learning method greatly improved the classification results by bringing together the best parts of both InceptionV3 and NASNetLarge models. Ensemble models usually do better than single models because they lower the chance of wrong predictions and make the results more reliable.

E. Summary of Results

The experimental evaluation shows that the proposed deep learning framework is highly effective for automated knee osteoarthritis detection. The model got an overall classification accuracy of 98.36%, and it had very good precision, recall, and Fl-score results for all KL grades.
The results show that deep learning models are able to effectively recognize radiographic signs linked to the severity of osteoarthritis and can offer dependable predictions for use in clinical settings.

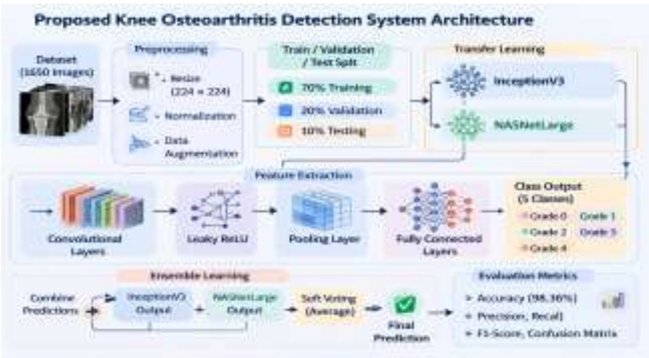

Fig3. Proposed Knee Osteoarthritis Detection System Architecture.

V. CONCLUSION

This study introduced a system that automatically detects and sorts out knee osteoarthritis (KOA) by using deep learning methods on knee X-ray images. The proposed approach combines transfer learning with an ensemble strategy to enhance classification performance across various osteoarthritis severity levels. Two strong convolutional neural network models, InceptionV3 and NASNetLarge, were used to extract features and perform classification. These models were combined using a soft voting ensemble method to use the best parts of both designs and make the predictions more trustworthy.
The experimental results show that the proposed ensemble model works well in recognizing different levels of osteoarthritis, as defined by the Kellgren-Lawrence system, which goes from Grade 0, which means normal, up to Grade 4, which is the most severe. The confusion matrix and performance metrics show that the model attains high classification accuracy and enhanced robustness compared to individual deep learning models. Data preprocessing methods like normalization, resizing, and augmentation helped the model generalize better and avoid overfitting during training.
The suggested framework has good promise as a tool to help doctors detect and evaluate the seriousness of knee osteoarthritis early on. By using automation to analyse X-ray images, the system can help make diagnoses more consistent and allow doctors to make quicker decisions in patient care. Future efforts might include growing the dataset, using more advanced deep learning models, and adding attention mechanisms or explainable Al methods to improve how well the system can be understood and used in real medical situations.

REFERENCES

[1] V. Vijaya Kishore and others: Predicting Knee Osteoarthritis Using Deep Learning and the Kellgren-Lawrence Grading System. Proceedings of Engineering Sciences, Vol. 05, No. 3, pp. 475-484, 2023.

[2] Kalpana, V., and Hemanth Kumar G.: Assessing How Well Deep Learning Models Work for Predicting Knee Osteoarthritis Using the Kellgren-Lawrence Grading System, ePrime Advances in Electrical, Engineering, Electronics, and Energy, 2023.

[3] Tariq, Z. Suhail and Z. Nawaz: Detecting and classifying knee osteoarthritis using X-rays, published in IEEE Access, volume 11, pages 48292 to 48303, in 2023.

[4] J. Kaur & C.C., Alvaro & M. Kermani & R. A comprehensive survey on the implementations, attacks, and countermeasures of the current NIST standards. lightweight cryptography standard (2023).

[5] S.C. Sangapu, K.S.N. Prasad, R.J. Kannan and others studied the effect of class imbalance in the VeReMi dataset on detecting misbehavior in autonomous vehicles, as published in Soft Computing in 2023, reference.

[6] C.C., Alvaro & J. Kaur & M.K., Mehran & R. Algorithmic security is insufficient: a comprehensive survey on implementation attacks threatening post-quantum security (2023).

[7] A. Tiwari, M. Poduval, V. Bagaria: Assessing artificial intelligence models for knee osteoarthritis

with deep learning algorithms applied to orthopedic X-rays, World J. Orthop. 13 (6) (2022) 603-614.

[8] The Knee OsteoArthritis Prediction (KNOAP2020) challenge: An image analysis challenge to predict incident symptomatic radiographic knee osteoarthritis from MRI and X-ray images. S10-S11, 2022.

[9] Y. Himeur, M. Elnour, F. Fadli and others wrote a survey titled "Al-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives" in the journal Artif. Intell. Rev. 56 (2023) 4929-5021.

[10] M. Yang, C. Colak, K.K. Chundru and others: Automated knee cartilage segmentation for diverse clinical MRI scans using generative adversarial networks with transfer learning, Quantitative Imaging in Medicine and Surgery, volume 12, issue 5, pages 2620-2633, 2022.