National College of Ireland

Recognition and Classification of Knee Osteoporosis and Osteoarthritis Severity using Deep Learning Techniques

MSc Research Project
Data Analytics

Tsai Shih Yang
Student ID: x21101825

School of Computing
National College of Ireland

Supervisor: Dr. Jorge Basilio

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Tsai Shih Yang
Student ID:	X21101825
Programme:	Data Analytics **Year:** 2022
Module:	Research Project
Supervisor:	Dr. Jorge Basilio
Submission Due Date:	19/09/2022
Project Title:	Recognition and Classification of Knee Osteoporosis and Osteoarthritis Severity using Deep Learning Techniques
Word Count:	7438 **Page Count** 24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Tsai Shih Yang
Date:	19/09/2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Contents

Contents

Recognition and Classification of Knee Osteoporosis and Osteoarthritis Severity using Deep Learning Techniques

Tsai Shih Yang
x21101825

Abstract

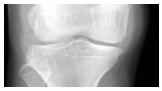
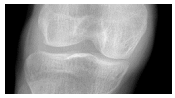
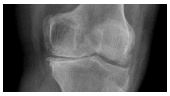
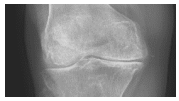
The Knee pain which is the common complaint and it affects different ages also is an irreversible disease cause problems and influence our life and future. There are two cases of knee disease in this study are the first case Knee Osteoarthritis which is the most common joint disorder based on Kellgren-Lawrence grading to distinguish from 0 to 4 levels of the severity and the second case is the Knee Osteoporosis which can progress without symptoms until a broken bone occurs. These two diseases the current diagnostic knee problems diagnostic systems are usually use X-ray, MRI, CT scan which require time and experienced physicians to identity knee diseases and diagnose X-Ray images from clinical data to prevent or early treatment also provide the appropriate medical diagnosis and treatment. This study aim to identify and classify knee diseases X-Ray images by using deep learning technique CNN, VGG16 and Late-Fusion model which methods this author had found great performance from researches of knee Osteoarthritis Severity X-ray images detection but in the Knee Osteoporosis case there is no research use deep learning method to do the X-ray image detective so this researcher assume the same method would bring the great performance assist the physicians identify the X-ray images and evaluate. This researcher had compared the results of models the Late-Fusion model bring the 77% accuracy for the OA dataset and the VGG16 model brings the 82% accuracy for OS data and propose the knee osteoporosis detection by using deep learning techniques.

1 Introduction

The first case of Knee Osteoarthritis is a degenerative joint disease which brings the bone rubbing on the bone and produce the painful bone spurs and the pain may from overuse of a joint, prolonged immobility and this causes thinning or destruction of the cartilage which covers the ends of bones and this knee disease would happened on any ages which ultimately affects activity ability and reducing quality of life. Normally the physicians diagnosis the knee osteoarthritis will check the joint for tenderness, swelling, redness and flexibility and the X-Ray imaging test is to get the pictures of affected knee joint that cartilage does not show up on images, but still can see the space between the bones in the joint has narrowed compare with healthy joint which means the cartilage loss. There are other diagnosing methods like the Magnetic Resonance Imaging (MRI), Blood tests which would not use these two methods in this study based on the limitation of the data collected and the deep learning techniques for image detection. The Knee Osteoarthritis severity dataset (*Digital Knee X-ray*, 2021) this author has got from Kaggle website the open source which provides public

datasets this dataset consists of 1650 digital X-ray images of knee joint and each radiographic knee X-ray images had labelled based on Kellgren and Lawrence grades by medical experts. In the Knee osteoarthritis severity, the Kellgren-Lawrence (KL) grading (IBJI, 2016) which is based on the radiographic features that can distinguish 0 to 4 different levels of Knee Osteoarthritis (OA) Severity shows on Table 1. Grade 0 (Normal): No signs of osteoarthritis and this is normal healthy knee without impairment or signs of joint damage. Grade 1 (Doubtful): The patients will not feel pain or discomfort but develop minor wear and tear and bone spur growth at the end of joint which is the early stage. Grade 2 (Mild): This stage will show more bone spur growth from the X-ray images although the space between bones distance looks normal but patients will begin experiencing symptoms of joint pain. Grade 3 (Moderate): The gap has obvious narrowing between bone joint and the bones develop spurs and it becomes rougher. Grade 4 (Severe): This level considered to be severe which bring greater pain and discomfort for the joint and the joint space has reduced causing the cartilage to wear off and bring the chronic inflammatory response.

Table 1: Digital Knee X-ray dataset (Osteoarthritis) sample images[1] (*Digital Knee X-ray*, 2021)

0 - Normal	1 - Doubtful	2 - Mild	3 - Moderate	4 - Severe
				

The second case Knee Osteoporosis which is a condition of the bones become less dense and more likely to fracture and this is the most common cause of knee disease in elderly people. This disease will cause loss of height, severe back pain also damage the ability to walk even bring the prolonged or permanent disability. The knee osteoporosis (NIHBone, 2018) is a silent disease which can progress undetected many years without any symptoms until fracture occurs. The physicians diagnosis of Knee Osteoporosis (radiologyinfo, 2021) may use X-ray, body CT or bone density scan to help diagnose the condition and assess risk for fracture. In this study only use the X-ray images for the image detection but there is a limitation to obtain the knee osteoporosis dataset as (Wani and Arora, 2021) they have mentioned very few datasets are available publically for research and the most of researches have created their own datasets without publically available for the researches but this author have found the dataset of Knee Osteoporosis from Kaggle (*Osteoporosis Knee X-ray Dataset | Kaggle*, 2022) and assume this dataset has labelled as normal knee and osteoporosis knee of total 372 digital X-ray images can see Table 2.

Table 2: Osteoporosis Knee X-ray Dataset sample images[2] (*Osteoporosis Knee X-ray Dataset | Kaggle*, 2022)

Normal	Osteoporosis

[1] Digital Knee X-ray dataset : https://www.kaggle.com/datasets/tommyngx/digital-knee-xray

[2] Osteoporosis Knee X-ray Dataset : https://www.kaggle.com/datasets/stevepython/osteoporosis-knee-xray-dataset


As these two knee diseases osteoporosis and osteoarthritis are two different medical conditions (NIHBone, 2018) and the condition develops differently with different symptoms so the author treats these as two cases in the study with the model and this author have choose the Convolutional Neural Networks (CNN) which is common use for image classification, VGG16 which called "Very Deep Convolutional Networks for Large-Scale Image Recognition" and Late-Fusion which is combining multiple classifiers approaches to enhance recognition accuracy. These methods have found from many researches get the good performance of image classification on Knee Osteoarthritis Severity and X-ray images which this author assume these models will bring the good accuracy to identify osteoporosis knee diseases these models result will evaluate through Precision, Recall and F1-score to get the better model to diagnosis osteoporosis which can assist physicians diagnosis to identify osteoporosis and other similar knee diseases based on X-ray images.

- Research Questions

Can this researcher use the similar deep learning technique methods for classifying in different feature knee diseases datasets based on X-ray images?

- Research Objectives

1. Investigate the Knee Osteoarthritis and the Knee Osteoporosis based on X- ray imaging dataset.
2. Implement and evaluate two deep learning CNN and VGG16 models based on each knee diseases X-ray imaging dataset
3. Merge two deep learning models and evaluate the performance

- Contribution

The contribution of this project is a novelty of Knee Osteoporosis x-ray image classification by using deep learning techniques method of CNN, VGG16, and Late-Fusion models which has known well performed for the Knee Osteoarthritis severity classification from the literature reviews and the result can show that these models assist the knee OA or OS diseases diagnosis get good performance in both datasets especial on OS dataset the VGG16 class weight with Epoch 100 can bring the 82% accuracy.

2 Literature Survey

Knee Osteoporosis and Knee Osteoarthritis are the different diseases and the imaging classifying is the important to diagnosis process in the procedure. There are researches had used deep learning techniques to identify Knee Osteoarthritis Severity can see in the section 2.1.

2.1 Knee Osteoarthritis Severity

(Zhang et al., 2020) They have detected the knee joint from the radiographs by using Residual Neural Network (ResNet) first then combined with Convolutional Block Attention Module (CBAM) to classify and predict the KL-grade automatically and this ResNet-34 + CBAM model can approach the higher accuracy result because the pre-processing of knee joint localizer helps improving the performance for the model. In the other study (Marzorati *et al.*, 2020), the model they have mentioned is Deep convolutional neural networks (CNNs) with the U-Net architecture which can help decrease time and effort during CT image segmentation time and the model is trained by CNN model with the U-Net structure which get the high accuracy but other structure of CNN would still possibly to improve the accuracy. Their study (Abasolo and Rifai, 2021) has combined the machine learning methods and deep learning model by fusion method which better predicting Knee Osteoarthritis severity results. There is study which has provided a hybrid fully automatic segmentation method from (Memari and Moghbel, 2020), they use the random walkers with deep learning model to get the cartilage boundary and use the thickness map with Adaptive Boosting (AdaBoost) classifier the disease progression utilizing fractal image features based on MRI images of side of knee and the AdaBoost algorithm which they have mentioned is the simple and fast for medical imaging. (Dalia *et al.*, 2021) They have used the YOLOv5 Object detection algorithm to detective the knee joint and cropping out it to use VGG16 and Resnet to classify the KL grade and comparison of accuracy and this knee joint localized method can help VGG16 to improve and get the higher accuracy compare other model. There is other study (Wahyuningrum *et al.*, 2019) which has prepared the cropped images 400 x 100 for the classify knee OA severity KL grades the model they have used the new approach which is combines the pre-processing CNN with the Long Short-Term Memory (LSTM) as a CNN-LSTM method for classification the KL grade and compare with other models like VGG16, Resnet-10 and Densenet-161. They have found the VGG16 model performs the best by extracting the high level feature to enable let LSTM effectively discriminate the KL grades.
This author found the research (Bany Muhammad *et al.*, 2019) use radiometric images to score the severity of OA and they have used the R-CNN to crop the image and use new size of 180 x 180 image implement to their DL model for the knee OA severity detection also they have use hyper parameter tuning and Bayesian optimization to adjustment and ensemble three CNN models which can reach 80% for KI = 1. There is other study (Kumar and Saxena, 2019) use the Ahlbäck grading scale which also use for the knee OA grading but their study has measured the loss of articular cartilage is quantified by minimum joint space the loss of articular cartilage and implement into the different classifier methods and found the KNN performed the better accuracy compare other methods like random forest and SVM. This author also have found there is a research (Alexos *et al.*, 2020) use McMaster Universities Osteoarthritis Index (WOMAC) which is scores the pain from 0 to 100 and use the categorized data they and implement to the random forest which performance is 84.3% accuracy compare other algorithm performed well.

Table 3: Knee Osteoarthritis Severity

Author	Objects	Image	Method	Result
(Zhang et al., 2020)	Using trained CNN ResNet-18 to localize knee joint then ResNet-34 to classify KL-grade	Radiographs images Size: 2000 × 2000 KL-grade	Random cropping augmentation CNN ResNet-18 ResNet-34	The ResNet-34 with CBAM accuracy is 74.81%
(Marzorati *et al.*, 2020)	Using the deep CNNs model to detective the CT images and evaluate	CT scan images Size: 512x512	3D U-Net	3D U-Net shows better than 2D U-net CNN architectures
(Abasolo and Rifai, 2021)	Using Fusion model combine and evaluating machine learning models and deep learning models to improve the detection of Knee Osteoarthritis Severity	Cropped x-ray images 224x224 pixels KL grade severity	Random Forest Gradient Boosting Xtreme Gradient Boosting DenseNet201 InceptionResNetV2	Fusion model has provide higher accuracy than single model.
(Memari and Moghbel, 2020)	Providing automatic segmentation method to diagnosis knee Osteoarthritis	MRI images KL-grade	AdaBoost	The automated method can decrease segmentation processing time compared with the manual
(Dalia *et al.*, 2021)	Developing DL-based Clinical Decision Support System (CDSS) based on ROI segmentation	X-ray images KL-Scale	DL model YOLOv5 VGG ResNet	YOLOv5 get 93% recall and the **VGG16** got 69.8% accuracy
(Wahyuningrum *et al.*, 2019)	Proposing a new classify CNN-LSTM for knee OA severity from X-ray images	Radiographic cropped Images 400 x 100 KL grade	VGG-16 ResNet-50 DenseNet-121	**CNN** pre-trained architectures **VGG-16** model performs the best and the LSTM effectively discriminates

				between KL grades 0-4.
(Bany Muhammad *et al.*, 2019)	Automatic localization of the kneecap with CNN model for assessing and quantifying the OA severity	X-ray images resize 180 X 180 KL-Scale	Ensemble three CNN models (hyper parameter tuning and Bayesian Optimization) Fast R-CNN	**CNN** models performance can reach 80% if results for KI = 1
(Kumar and Saxena, 2019) Quantification of Cartilage loss for Automatic Detection and Classification of Osteoarthritis using Machine Learning approach	Using machine learning approach to predict knee OA severity from radiographs	MRI and X-ray images Ahlbäck grading	KNN SVM Random forest	KNN 97% SVM 84% Random forest 95%
(Alexos *et al.*, 2020) Prediction of pain in knee osteoarthritis patients using machine learning: Data from Osteoarthritis Initiative	Build a prognostic tool that will predict the KOA patients' pain progression	OAI database no images WOMAC grading	Decision Trees KNN SVM Random Forest XGBoost Naïve Bayes	Random Forest shows the best accuracy 82.5% in the 20 features

These researches have shown the knee osteoarthritis severity can use X-ray, radiometric, CT images to diagnosis with the OA knee severity grading which has different standard of the knee OA severity grading methods like Kellgren-Lawrence (KL), Ahlbäck, and WOMAC grading of images to detective the knee joint severity for different levels and these based on the physicians to diagnosis methods they use and the images scanned from equipment but this author can know the machine learning, deep leaning can help assist physicians diagnosis the identification of knee Osteoarthritis Severity by using CNN or transfer learning VGG get the great accuracy show on Table 3.

2.2 Knee Osteoporosis

They have used the system data-centric to collect the Knee X-ray Database for Osteoporosis Detection (Wani and Arora, 2021) that can make the research for finding the timely, cost effective, and accurate of the dataset see Table 4.

Table 4: Knee Osteoporosis

Author	Objects	Images	Method	Result
(Wani and Arora, 2021)	Provide the detection of osteoporosis database which can use for the research of finding the cost-effective, timely, and accurate osteoporotic detection system data-centric.	X-ray	QUS system	The dataset provides early osteoporosis diagnosis with help of knee x-ray images to make the computer-aided diagnostic system

In the knee osteoporosis study section, (Wani and Arora, 2021) their study which is related the dataset collect for the research of finding the osteoporotic detection system data-centric this author has searched related of Knee osteoporosis studies by using deep learning techniques but cannot find the other studies so this author has assumed the Knee Osteoporosis detection can use the same method of Knee OA severity x-ray images detection that performed well in the Knee OA severity x-ray images detection studies can see Table 3. which shows great accuracy of in previous literature reviews section to classify images to get some ideas of image classification.

2.2.1 Image classification methodology of Convolutional Neural Network

There is an image classification study (Latha et al., 2021). They used CNN to recognize and classify the various fruits such as Apple, Blueberry, Cherry, Grape, and other fruits images. The CNN model accuracy obtained is 97.4%. In their study, the CNN model performed well in classifying the images. There is an image detection of Fake Rice (Mahmud Ridoy et al., 2021). They used the ANN and CNN to identify the real and fake rice images. The CNN model showed 98% accuracy, ANN showed 60% accuracy. This author can assume the CNN model perfromed well compare with the ANN model. The study (Dandavate and Patodkar, 2020) used CNN to classify four different fruits by using a Deep learning methods into three categories. The CNN model accuracy is 97.74% in 8 epochs. From the (Eryigit and Tugrul, 2021) study they have compare MobileNet and VGG19 accuracy of classify the grass seed that the accuracy of each model depends on the size of the images. The VGG model and MobileNet accuracy-related on the images size which this author has to consider. (Sallam et al., 2021) This research has used a lot pre-trained models such as AlexNet, VGG11, VGG16, VGG19, GoogleNet (Inception V1), ResNET-18, ResNET50, ResNET-101 and ResNet-152 to detective glaucoma and compared accuracy and the ResNET-152 model achieves the highest accuracy 86.9% which provide stronger detail on image detection. Their study Author(Jayakumar, 2020) used CNN, VGG16, VGG19, Random Forest, XGBoost, SVM. Gradient Boost and 5-Layer CNN these models in detecting and classifying maize plant disease. The CNN based transfer learning technique VGG19 can identify leaf disease.

Table 5: CNN images

Author	Objects	Method	Result
(Latha *et al.*, 2021)	Identify the fake rice in the market	ANN CNN	**CNN** 98% accuracy ANN 60% accuracy
(Mahmud Ridoy *et al.*, 2021)	Proposing fruit classification with CNN model	CNN	**CNN** 97.4% accuracy
(Dandavate and Patodkar, 2020)	Based Fruit Classification Model	CNN(8 epochs)	**CNN** model 97.74% accuracy in 8 epochs
(Eryigit and Tugrul, 2021)	Compare VGG and MobileNet model	CNN MobileNet VGG19	VGG19 256 x 256 accuracy 94.84% MobileNet 64x64 accuracy 91.88%
(Sallam *et al.*, 2021)	Proposing glaucoma detective method using pre-trained model to find the best model	Alex Net VGG-11 VGG-16 VGG-19 Googlenet-V1 ResNet-18 ResNet-50 ResNet-101 Resnet-152	ResNet-152 achieved 86.9% accuracy the higher performance
(Jayakumar, 2020)	Evaluating XGBoost, Gradient Boost, CNN(VGG16, VGG19)	VGG19 VGG16 Random Forest XGBoost SVM GradientBoost 5-Layer CNN	VGG19 shows the best accuracy but take a lot of time for training.

There is a lot of the image classification methods such as the Machine learning method, CNN model, Transfer learning models can bring good result with epoch adjustment, different image size, pooling method will influence accuracy can see Table 5.

2.2.2 X-Ray Image classification methodology of Convolutional Neural Network

The X-Ray images are the target images use for this researcher and will focus on the X-ray images classification methods by using CNN models in this section. For the x-ray image detection, there is a study (Rao et al., 2020) used Backpropagation Neural Network(BPNN) and CNN with filter which can remove the noise from images and use K means to get the region of interest to extract the feature get the accuracy of CNN is 91% and BPNN is 89% and there is a similar method used for the study (Education *et al.*, 2020) they have use the CNN with mean filtering M3 to get the central pixel which can prepare the FCM to get the ROI prepared for extraction segmentation and use the statical method to get the feature the models have CNN, BPNN, RBFNN, KNN and SVM to get the better model. The study (Wei, Chen and Zhang, 2020) of Pneumonia X-ray images detection by comparing with CNN models like LeNet5, VGG16 and Residual Network50 with the heat map Class Activation

Map (CAM) which can get the feature and they have found the number of the layers increases then LeNet5 accuracy below 75%, VGG16 accuracy over 80%, Resnet 50 accuracy approached near 85%. The research (Wu *et al.*, 2020) they have used the Muti-Feature-Enhanced R-CNN to detective the different domain classification of each body part X-ray image which is a feature enhancement model. There is a study (Nafiiyah and Setyati, 2021) use chest X-ray images enhancement CLAHE to contrast the medical images and their CNN model with 8 different architectural models that the highest accuracy is 82.53% which has 35 layers.

Table 6: X-Ray image CNN

Author	Objects	Method	Result
(Rao *et al.*, 2020)	Bone fracture detection use CNN and BPNN	CNN BPNN	**CNN** 91% accuracy BPNN 89% accuracy
(Education *et al.*, 2020)	CNN to classify X-Ray image	CNN BPNN KNN SVM RBFNN	**CNN** 94% accuracy BPNN 87% accuracy KNN 83% accuracy SVM 82% accuracy RBFNN 86% accuracy
(Wei, Chen and Zhang, 2020)	Compared the CNN model's by using the chest x-ray images.	LeNet5 VGG16 Resnet 50	The **VGG16** get the highest accuracy which is over 80%
(Wu *et al.*, 2020)	Proposes a novel Multi-domain Fracture Detection Network(MFDN) to classify each body part of X-ray images	Faster R-CNN Faster R-CNN + Multiple R-CNN Faster R-CNN + FEM-1 Faster R-CNN + FEM-1 +FEM-2 MFDN	The MFDN model can achieves accuracy 98%.

(Nafiiyah and Setyati, 2021)	Proving chest X-ray image enhancement with CLAHE to contrast the medical images to diagnosing Pneumonia	CNN 15 layers CNN 19 layers CNN 27 layers CNN 23 layers CNN 35 layers CNN 31 layers	The **CNN** with 35 layer show accuracy is 82.53%

Summary in this project of literature reviews there are numerous studies have used different deep learning techniques for classifying knee osteoarthritis severity can see Table 3. but in knee osteoporosis case only found one research (Wani and Arora, 2021) which is provides the good quality of Knee osteoporosis database method can see Table 4. As this author cannot find the other research instead and assumed the knee osteoarthritis severity classification method can get the great result in Knee osteoporosis classification case so this author has searched the deep learning technique CNN method for image classification can see Table 5. The CNN and VGG16 can bring the good performance but this author has to consider these images are X-ray images so this author has searched CNN method with X-ray image can see Table 6. found the deep learning technique CNN and the transfer learning VGG16 brought higher accuracy and effect compare other models. The novelty in this research is the knee osteoporosis has not been used deep learning techniques for classifying the osteoporosis.

3 Research Methodology

The KDD methodology which is based on Knowledge Discovery in Databases (KDD) in Figure 1. This author has decided this method which is more focused on discovering patterns and finding information within the given data would help us to improve the method for image identification.

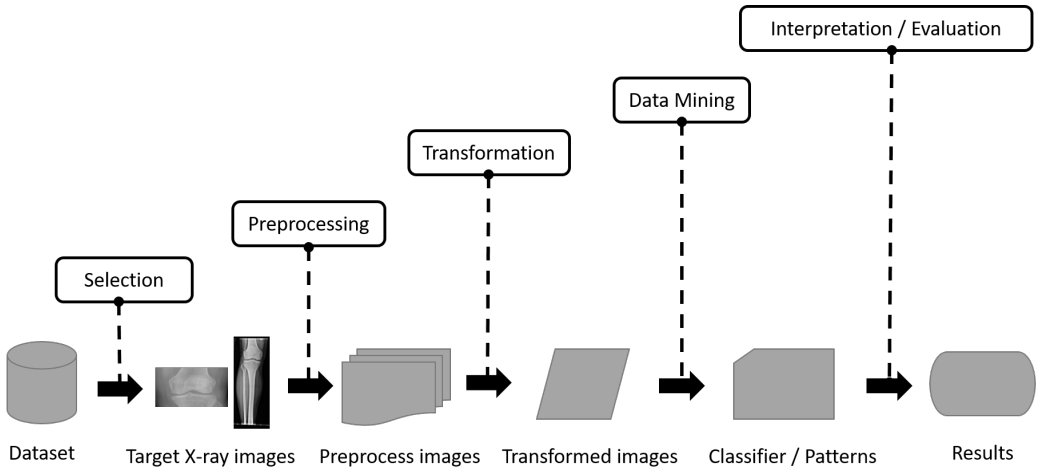


Figure 1. Steps of the KDD Process

3.1 Data Collection

This research is using the X-Ray images which are not easy to collect with the lack of data also the data images have been taken from medical equipments without any medical permission and the relative. The Knee Osteoarthritis Severity (*Digital Knee X-ray*, 2021) which has taken by the profession medical institution and all the X-ray images have labelled based on Kellgren-Lawrence (KL) grading as 0 to 4 different stage of severity by experts and the Osteoporosis Knee X-ray Dataset (*Osteoporosis Knee X-ray Dataset | Kaggle*, 2022) which have been labeled as the normal knee and the osteoporosis knee as well and these two datasets this author has found in Kaggle website and the dataset are suitable for this study which has required for this task to identify knee diseases X-ray images by deep learning models.

3.1.1 Ethics concern

All the process of data have to consider the Ethics concern which is very important and related with the personal information and the privacy even the law. The data this author has been collected from Kaggle this website provided the X-ray images of leg or knee joint which have not provided any patients information from these images and this is not possible can track with personal information from these knee diseases X-ray images which means these datasets are safety to use and the well preserve confidentiality patients data.

3.2 Data Pre-processing and transformation

All the images have to pre-process and transformed for the model. There are studies show that localizing images of knee joints can improve the accuracy of the Osteoarthritis detection result (Dalia et al., 2021) has used the Region of Interest (ROI) segmentation to crop out the knee joint and (Wahyuningrum et al., 2019) they have cropped the image into 400 x 100 pixels at the knee joint. The data augmentation technique which can increase the dataset (Latha et al., 2021) resized images then shifted and flipped randomly by using ImageDataGenerator to augment their dataset and (Dalia et al., 2021) have split into the ratio of 7:2:1 for training, validation, and testing on each dataset.

3.3 Data Mining

The deep learning techniques have many models can use for image classification and based on the literature reviews Neural Convolutional Network (CNN) model and the transfer learning VGG16 model bring the great performance can see Table 3 and Table 6 and the Fusion model (Abasolo and Rifai, 2021) can provide higher accuracy than single model see Table 3.

3.3.1 Convolutional Neural Network (CNN) model

Convolutional Neural Network (CNN) model used for classifying normal images or X-ray image has been very effective such as (Latha et al., 2021) have used 2 convolutional layers with Average and Max pooling layers to extract features the accuracy is 98% to detective fake rice images. The traditional CNN model can see Figure 2 has a single connection to each layer it also can adjust layers with different numbers of nodes, pooling layers (Nafiiyah and Setyati, 2021) CNN model for identifying Pneumonia in lung X-Ray Image accuracy is 82.53% with image size 224 x 224 and 8 convolution layers 32,32,64,64,128,128,256, 256 and 7 times max-pooling layer.

- Convolution layer which is use for extract features from the image by the filter and this receive the input X-ray images into the pixel value and extract low level characteristics of image like curve, edge and generate to a feature map.
- The pooling layer is used to reduce the image dimensions and the max pooling can select the maximum value in a region from the filter size.
- The fully connected layer (Dense) is the part of final layer of CNN this can capable of recognizing features to obtain the result of the previous pooling layer.
- Dropout layer can drop certain set of neurons randomly to reduce the model get overfitting.
- Softmax helps input X-ray images classify into different classes based on the feature.

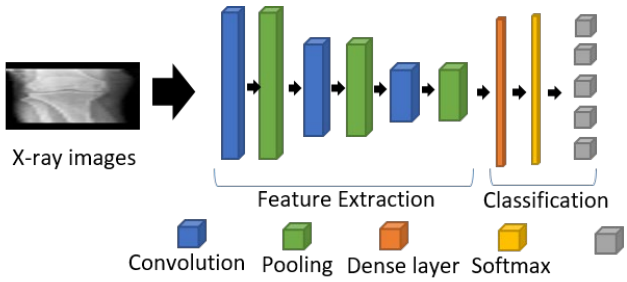


Figure 2: Convolutional Neural Network (CNN) architecture

3.3.2 Transfer learning (Pre-trained) model

Transfer learning model (Wahyuningrum *et al.*, 2019) they used VGG-16, ResNet-50 and DenseNet-161 to get the feature extraction, and (Dalia *et al.*, 2021) have compared VGG16 and Resnet-152 and the result of VGG16 show the better accuracy 69.8% for the OA severity classification. The VGG-16 which has13 convolutional layers with 5 max pooling layers and the final layer 4096 neurons and Softmax layer is used for the classification can see Figure 3.

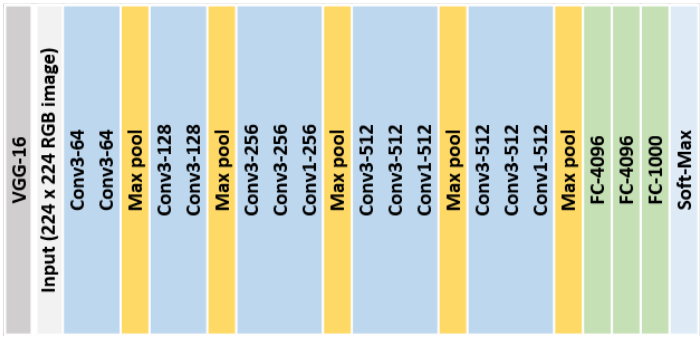


Figure 3: VGG-16 architecture

3.3.3 Late Fusion model

The Late Fusion method that can ensemble multiple models to improve the performance (Abasolo and Rifai, 2021) Fusion model provide higher accuracy compare single model. Another (Bany Muhammad et al., 2019) study has used Ensemble three CNN models to get the KL result and the accuracy can reach 80%. The late fusion is to fusion different models and the rule to fusion is to use weight, majority voting, averaging, or a meta classifier and the strategy can see in Figure 4.

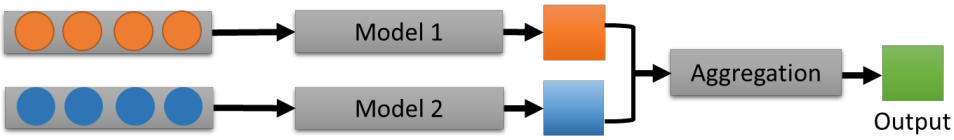


Figure 4: Late Fusion Strategy

3.4 Data Interpretation / Evaluation

The evaluation of model performance includes accuracy, precision-recall, and f-1 score. (Jayakumar, 2020) has compared their models of Accuracy, Precision, Recall and F1-Score and the confusion matrix. The confusion matrix which is used for evaluating model performance can see in Figure 5

True Positive (TP): True positive value be predicted to positive value.
False Positive (FP): True positive value be predicted to negative value.
True Negative (TN): True negative value be predicted to negative value.
False Negative (FN): True negative value be predicted to positive value.

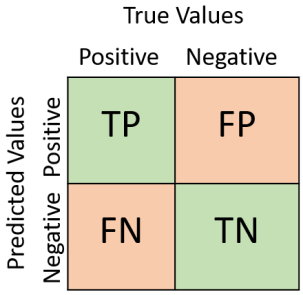


Figure 5: Confusion Matrix

From the confusion matrix this researcher can get the accuracy, precision-recall, and f-1 score which can help us understand the model performance and effect.

Accuracy is an important metric which is the fraction of predictions the model got right
Accuracy Score = (TP + TN)/ (TP + FN + TN + FP)
Precision is a proportion of positive results that are true positives
Precision Score = TP / (TP + FP)
The recall is a proportion of actual positives that were identified correctly
Recall Score = TP / (TP +FN)
F1-Score is a harmonic mean of precision and recall score
F1 Score = 2* ((Precision Score * Recall Score) / (Precision Score + Recall Score))

4 Design Specification

In the project, this author has designed the architecture to classify the Knee Osteoporosis and Knee Osteoarthritis Severity in the different dataset by using the same deep learning technique method to improve and evaluate the models can see in Figure 6

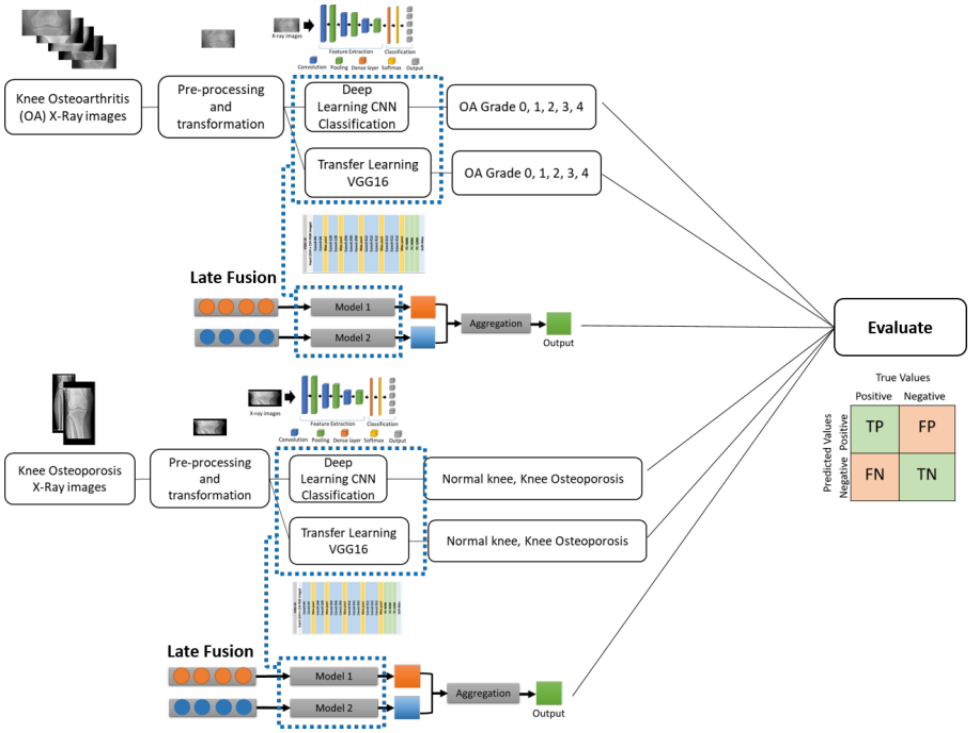


Figure 6 : Design architecture

5 Implementation

In this section, this researcher has 2 different cases of knee disease Knee Osteoarthritis Severity and Knee Osteoporosis two different datasets by proposing the same deep learning technique CNN, VGG16 and Late Fusion method to identify the knee X-ray images by using deep learning method which explain and follow the design flow and evaluate the result of models.

The software of platform this author has decided use Google Colaboratory (Colab) which allows anybody to write and execute arbitrary python code through the browser and this platform has pre-built with lots of python library and no charge for GPU usage which is convenient can run code for 24 hours and saved in google drive.

5.1 Data Collection – Knee Osteoarthritis severity and Knee Osteoporosis

The data collection of both knee diseases dataset alreay collected from Kaggle which is the open source website and as this study request the Knee Osteoarthritis severity dataset (Digital Knee X-ray, 2021) and the Knee Osteoporosis dataset (Osteoporosis Knee X-ray Dataset | Kaggle, 2022) the imaging data this author decides to use the Kaggle API to download the datasets into Google Colaboratory which needs the personal Kaggle account to download the API credentials and the file name will be kaggle.json this can help to connect the dataset from Kaggle to Google Colab.

5.2 Knee Osteoarthritis severity

5.2.1 Knee Osteoarthritis severity Dataset

The first dataset is Knee Osteoarthritis severity X-ray images this author has got from the kaggle website which have already been cropped to size 300 x 162 pixels as the region of interest(ROI) the knee joint view and this author has decided to resize images to 224 x 224 that is common use for the transfer learning model as VGG16 model and the image size has to be considered for the training model and this author will use same size for the CNN model as well. In this OA severity dataset this author has extracted the image and explored the data and found the total images are 1650 with 5 classes which this author has split into (Dalia et al., 2021) the ratio of 7:2:1 for train, valid and test data can see Figure. 7. and the check few images Figure. 8.

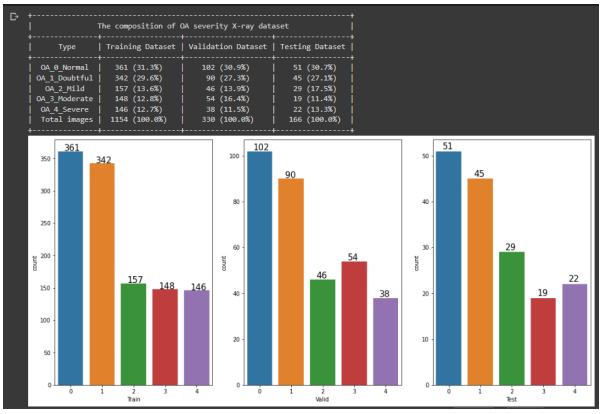


Figure 7: Knee Osteoarthritis Severity X-Ray image data distribution

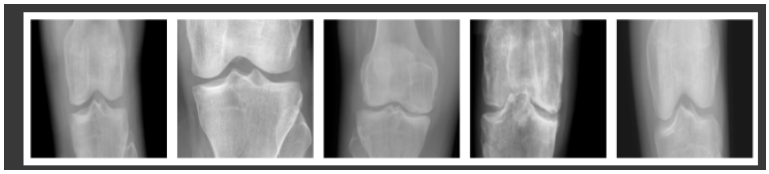


Figure 8 : Knee Osteoarthritis Severity X-Ray images sample

5.2.2 Data Pre-process and Transformation

This author has found OA training data has imbalance issue which has considered to stack the lack of classes data or use the class weight method to balance the dataset and this author found the class weight see method shows better after few times run in the same structure and all the images have been resize to 224 x 224 and split to train, valid and test data can see Figure 7. Prepared for the training model.

5.2.3 CNN model

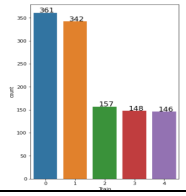
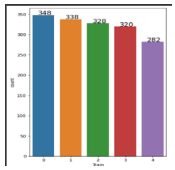
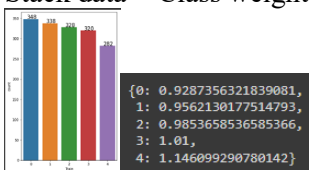
The data has been prepared to the CNN model and this author has coompared three different CNN structure for the OA data to see Table . the structure 1 (Bany Muhammad et al., 2019) their network architecture of the base model-2 bring the better performance compare with other two structures by implementing the OA datasets with same 20 epochs and this author assume this structure 1 is suited in this project can see Table. 7.

Table 7: CNN structures comparison

CNN structures	Structure 1 32,64,96,128,256, 5 Maxpooling 2 dropout(0.1) activation relu and dense 5 with 'softmax' (Bany Muhammad *et al.*, 2019) Network architecture of the base model-2.	Structure 2 32,32,64,64,128,128,256 with 7 maxpooling activation relu and dense 5 with 'softmax' (Nafiiyah and Setyati, 2021) CNN 35 Layers Architecture	Structure 3 32,64,128, 2 Maxpooling and dense 256,512 relu and dense 5 with 'softmax' (This author created)
Accruacy	**49%**	40%	41%

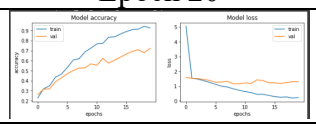
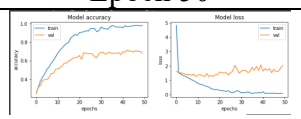
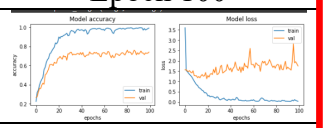
After this author has decided the CNN structures and the data imbalance issues this author has tried put original data, stack data, class weight and both to test which way can bring the better result can see Table 8. as this author found using stack with class weight can get 74% accuracy.

Table 8: CNN Data imbalanced method comparison

Structure 1 With epoch 20	Imbalance data	Stack data	Class weight	Stack data + Class weight
			{0: 0.6393351800554017, 1: 0.6748538011695906, 2: 1.4700636942675158, 3: 1.5594594594594595, 4: 1.580821917808219 2}	
Accuracy	49%	65%	50%	**74%**

Through the CNN Structure 1 with stack data and class weight which bring the 74% accuracy for the OA severity data and this author has adjusted the Epochs and compared the epochs 20, 50 and 100 can to see the performance of model see Table 9.

Table 9: CNN Epochs comparison

	Epoch 20	Epoch 50	Epoch 100
CNN Structure 1 Stack data + Class weight			
Accuracy	**74%**	69%	**74%**

5.2.4 VGG-16 model

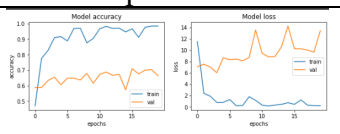
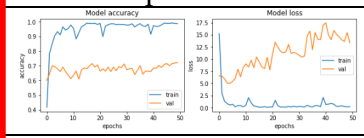
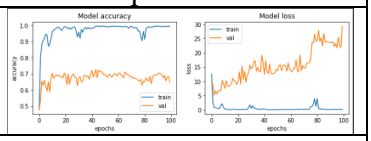
The VGG16 model this author imports the basic model and add extra flatten layer and implement with the data which has resized as 224 x 224 and this author also has compared the data imbalance method and this author found the OA data use the class weight for the VGG16 model bring better performance and the accuracy show 52% which has showed the different effect on CNN model can see Table 10.

Table 10: VGG-16 Data imbalanced method comparison

VGG-16 Epochs 20	Stack data	Class weight	Stack data + Class weight
Accuracy	41%	**52%**	48%

The class weight effect well in VGG16 model and this author also has compared different epochs of the model as 20, 50 and 100 by using VGG16 and class weight can see Table 11.

Table 11: VGG-16 Epochs comparison

	Epoch 20	Epoch 50	Epoch 100
VGG-16			
Accuracy	52%	**61%**	56%

5.2.5 Late-Fusion model

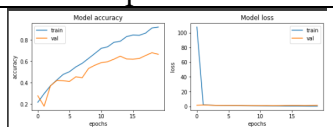
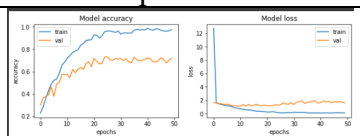
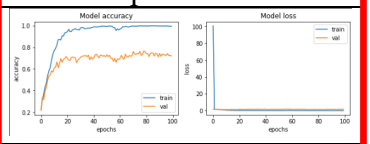
The Late-Fusion (Abasolo and Rifai, 2021) idea which is ensemble two model features and improve the model performance and in this project this author has used the first model is CNN with structure 1 which has compared can see Table 7 use concatenate method with second model VGG16 and the data this author has input two same size 224 x 224 of training data with two validation data for feeding the Fusion model and compared with different data imbalance method for the merged model can see Table 12.

Table 12: Late Fusion model Data imbalanced method comparison

Late-Fusion CNN + VGG16 Epochs 20	Stack data	Class weight	Stack data + Class weight
Accuracy	72%	46%	**72%**

The Stack data + Class weight for the epochs test because this researcher consider the ensemble model which with CNN structure inside as this author had test the CNN single model the Stack data + Class weight can bring the better performance can see Table 12 . and the next step have compared different epochs 20, 50 and 100 and found the Epoch 100 bring the 77% accuracy see Table 13.

Table 13: Late Fusion model Epochs comparison

	Epoch 20	Epoch 50	Epoch 100
Late-Fusion CNN + VGG16			
Accuracy	72%	71%	**77%**

5.3 Knee Osteoporosis

5.3.1 Knee Osteoporosis Dataset

The second dataset is Knee Osteoporosis X-ray images also from the kaggle website and this author has found this dataset has total 372 images which has included normal knee 186 images and osteoporosis 186 images without cropped as this author has known knee osteoporosis disease is lack of bone density so it could happened in any part of the knee bones. Following the design architecture the same way to implement the data same as the OA data resize to 224 x 224 for the model but this author has extracted the image data to ratio of 7:2:1 for train, valid and test data can see Figure 9. and the check few images after this author has resized Figure. 10.

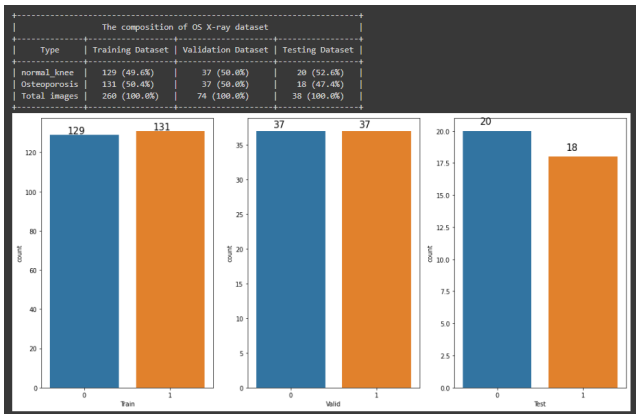


Figure 9: Knee Osteoporosis X-Ray image data distribution

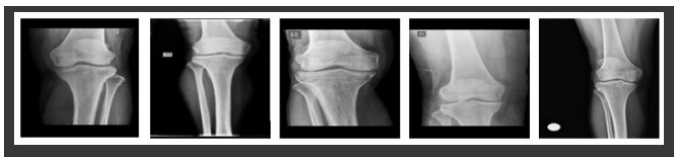


Figure 10: Knee Osteoporosis X-Ray images sample

5.3.2 Data Pre-process and Transformation

In the OS training data only has little imbalance issue which can test later with the class weight method to balance the and all the images have been resize to 224 x 224 with pretty table shows all the images classes and split to train, valid and test data prepare for the models can see Figure 9. The OS data class weight can see Figure 11.

```
{0: 1.0655737704918034, 1: 0.9420289855072463}
```

Figure 11: OS data class weight

5.3.3 CNN model

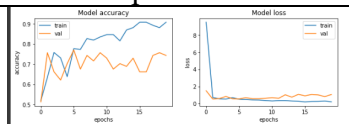
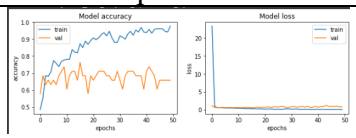
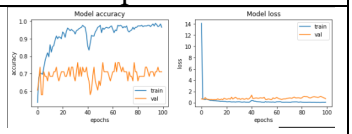
The OS data for CNN model also have compared the different structures and this author has found the Structure 1 and 2 bring the same and good performance as 71% accuracy but considered to use the class weight to check the effect and got the 76% accuracy for the CNN structure 1. See table 14.

Table 14: CNN structures comparison

	Structure 1	Structure 2	Structure 3	Structure 1 + class weight
Accuracy	71%	71%	63%	**76%**

The CNN structure 1 + class weight got the 76% accuracy from the epochs 20 compared with 50 and 100 for the OS data. See table 15.

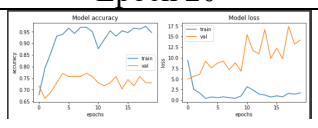
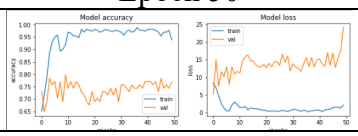
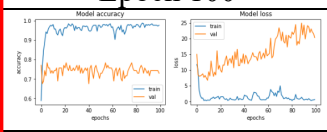
Table 15: CNN model Epochs comparison

	Epoch 20	Epoch 50	Epoch 100
CNN Structure 1 Stack data + Class weight			
Accuracy	**76%**	63%	71%

5.3.4 VGG-16 model

The VGG16 model for the OS data also have used the class weight and test different epochs which got the 82% accuracy from the epoch 100 . See Table 16.

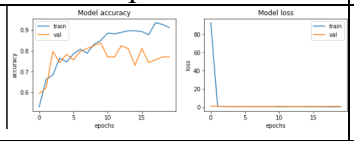
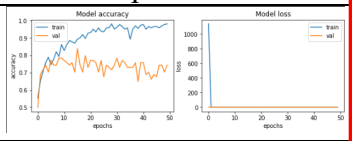
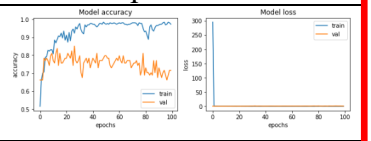
Table 16: VGG-16 Epochs comparison

	Epoch 20	Epoch 50	Epoch 100
VGG-16 class weight Epochs 20			
Accuracy	71%	71%	**82%**

5.3.5 Late-Fusion model

Our OS data in the Late-Fusion model the accuracy 71% with Epoch 100 is better compared with Epochs 20 and 50. See Table 17.

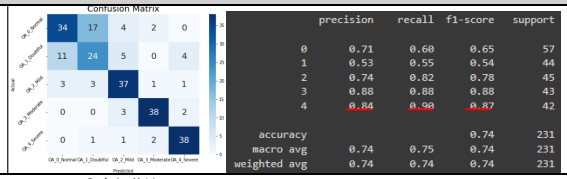
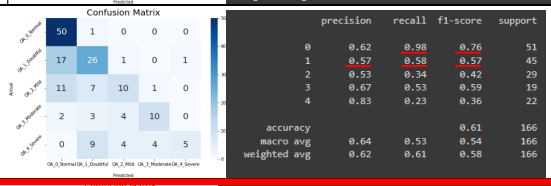
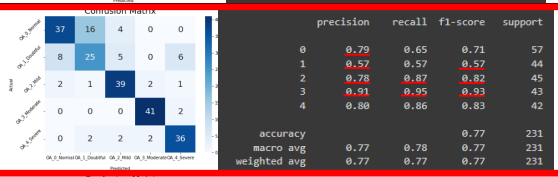
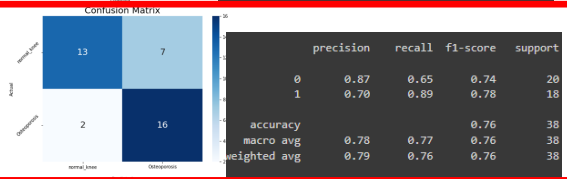
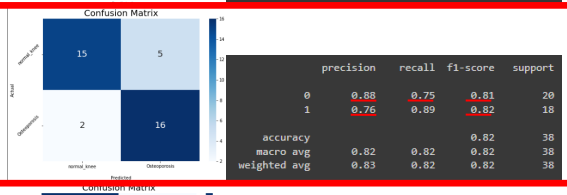
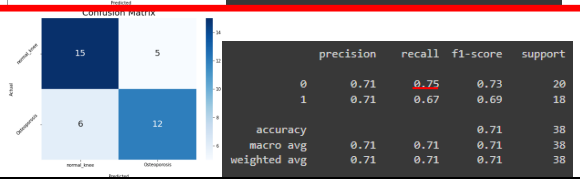
Table 17: Late Fusion model Epochs comparison

	Epoch 20	Epoch 50	Epoch 100
Late-Fusion CNN + VGG16			
Accuracy	63%	68%	**71%**

6 Evaluation

In the evaluation section, the confusion matrix can help to get the X-ray images classified result from the model to evaluate the model performance with these two different cases of dataset. Table 18.

Table 18: CNN, VGG16 and Fusion models comparison

Dataset	Model	Confusion matrix with information	Accuracy
Case 1 OA severity dataset	CNN structure 1 Stack data + Class weight Epoch 100		74%
	VGG16 Class weight Epoch 50		61%
	Late-Fusion model Stack data + Class weight Epoch 100		77%
Case 2 OS dataset	CNN structure 1 Class weight Epoch 20		76%
	VGG16 Class weight Epoch 100		82%
	Late-Fusion model Class weight Epoch 100		71%

The correct classify is very important in this research, the F1-score which is a harmonic mean of precision and recall score in Table 18 this researcher compared each models with marked the higher scores of each classes in precision, recall, and f1-score. In the OA data in f1-score section, the fusion model can get the classes 1 to 3 classify correctly and in the class 2 VGG16 get the same 57% also in the class 0 the VGG16 model get 76% and the CNN model get the 87% in class 4. The OS data in f1-score the VGG16 model classify correctly for the both classes the class 0 has 81% and the class 1 is 82% better than other models.

7 Conclusion and Future Work

The aim of this research was to use similar model method for the X-ray images of OA severity and OS knee disease detection to understand the model performance for detecting different feature. This research proposes the Knee Osteoporosis x-ray image classification by using deep learning techniques method of CNN, VGG16, and Late-Fusion models which is the novelty in the research. Results demonstrate that for accuracy the Late-fusion model detective OA severity show good performance but for detective OS the VGG16 model is better than CNN and Late-fusion model. The limitation of the study is the X-ray images insufficient not truly classify all knee severity and osteoporosis in each classes even this researcher get the good performance but still possible misdiagnosis by the model so if use it in the real-world application still have to improve. This research can potentially enhance

assist the physicians identify the knee X-ray images and detective normal knee and disease knee by using deep learning models. This work can be improved by getting more X-ray images and test with the clinical detection result to adjustment the model.

References

Abasolo, K. H. and Rifai, H. (2021) 'Detection of Knee Osteoarthritis Severity using a Fusion of Machine and Deep Learning models'.

Alexos, A. *et al.* (2020) 'Prediction of pain in knee osteoarthritis patients using machine learning: Data from Osteoarthritis Initiative', *11th International Conference on Information, Intelligence, Systems and Applications, IISA 2020*. doi: 10.1109/IISA50023.2020.9284379.

Bany Muhammad, M. *et al.* (2019) 'Deep ensemble network for quantification and severity assessment of knee osteoarthritis', *Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019*, pp. 951–957. doi: 10.1109/ICMLA.2019.00163.

Dalia, Y. *et al.* (2021) 'DeepOA: Clinical Decision Support System for Early Detection and Severity Grading of Knee Osteoarthritis', *2021 5th International Conference on Computer, Communication, and Signal Processing, ICCCSP 2021*, pp. 250–255. doi: 10.1109/ICCCSP52374.2021.9465522.

Dandavate, R. and Patodkar, V. (2020) 'CNN and data augmentation based fruit classification model', *Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 2020*, pp. 784–787. doi: 10.1109/I-SMAC49090.2020.9243440.
Digital Knee X-ray (2021). Available at: https://www.kaggle.com/datasets/tommyngx/digital-

knee-xray (Accessed: 15 March 2022).
Education, K. L. *et al.* (2020) 'X-ray images with Fuzzy C means and', pp. 919–923.

Eryigit, R. and Tugrul, B. (2021) 'Comparison of VGG and MobileNet Models for Grass Seed Dataset', pp. 255–259. doi: 10.1109/icicos53627.2021.9651865.

IBJI (2016) *Arthritis in Knee: 4 Stages of Osteoarthritis - IBJI*, *Illinois Bone and Joint Institute*. Available at: https://www.ibji.com/blog/orthopedic-care/arthritis-in-knee-4-stages-of-osteoarthritis/.

Jayakumar, A. (2020) 'Detection and Classification of Leaf Diseases in Maize Plant using Machine Learning'. Available at: http://trap.ncirl.ie/id/eprint/4278.

Kumar, A. and Saxena, P. (2019) 'Quantification of Cartilage loss for Automatic Detection and Classification of Osteoarthritis using Machine Learning approach', *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*, pp. 6–11. doi: 10.1109/ICCCNT45670.2019.8944538.

Latha, R. S. *et al.* (2021) 'Automatic Fruit Detection System using Multilayer Deep Convolution Neural Network', *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*. doi: 10.1109/ICCCI50826.2021.9402513.

Mahmud Ridoy, M. A. *et al.* (2021) 'Identifying Fake Rice Using Computer Vision in Perspective of Bangladesh', pp. 1–6. doi: 10.1109/icecct52121.2021.9616715.

Marzorati, D. *et al.* (2020) 'Deep 3D convolutional networks to segment bones affected by severe osteoarthritis in CT scans for PSI-based knee surgical planning', *IEEE Access*, 8, pp. 196394–196407. doi: 10.1109/ACCESS.2020.3034418.

Memari, N. and Moghbel, M. (2020) 'Computer-Aided Diagnosis (CAD) of Knee Osteoarthritis based on Magnetic Resonance Imaging for Quantitative Pathogenesis Analysis and Visualization', *ISCAIE 2020 - IEEE 10th Symposium on Computer Applications and Industrial Electronics*, pp. 192–197. doi: 10.1109/ISCAIE47305.2020.9108837.

Nafiiyah, N. and Setyati, E. (2021) 'Lung X-Ray Image Enhancement to Identify Pneumonia with CNN', *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIConCIT 2021*, pp. 421–426. doi: 10.1109/EIConCIT50028.2021.9431856.

NIHBone (2018) 'Osteoporosis and Arthritis: Two Common but Different Conditions', *NIHBone*, free TTY(April), pp. 202–223. Available at: https://www.niams.nih.gov/health_info/bone/Osteoporosis/Conditions_Behaviors/osteoporosis_arthritis.asp.

Osteoporosis Knee X-ray Dataset | Kaggle (2022). Available at: https://www.kaggle.com/stevepython/osteoporosis-knee-xray-dataset (Accessed: 15 March 2022).

radiologyinfo (2021) *Osteoporosis | Bone Density Loss - Diagnosis, Evaluation and Treatment*. Available at: https://www.radiologyinfo.org/en/info/osteoporosis.

Rao, L. J. *et al.* (2020) 'Advanced computerized classification of X-ray images using CNN', *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, (Icssit), pp. 1247–1251. doi: 10.1109/ICSSIT48917.2020.9214136.

Sallam, A. *et al.* (2021) 'Early Detection of Glaucoma using Transfer Learning from Pre-trained CNN Models', *2021 International Conference of Technology, Science and Administration, ICTSA 2021*. doi: 10.1109/ICTSA52017.2021.9406522.

Wahyuningrum, R. T. *et al.* (2019) 'A New Approach to Classify Knee Osteoarthritis Severity from Radiographic Images based on CNN-LSTM Method', *2019 IEEE 10th International Conference on Awareness Science and Technology, iCAST 2019 - Proceedings*. doi: 10.1109/ICAwST.2019.8923284.

Wani, I. M. and Arora, S. (2021) 'A Knee X-ray Database for Osteoporosis Detection', pp. 1–5. doi: 10.1109/icrito51393.2021.9596444.

Wei, X., Chen, Y. and Zhang, Z. (2020) 'Comparative Experiment of Convolutional Neural Network (CNN) Models Based on Pneumonia X-ray Images Detection', *Proceedings - 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence, MLBDBI 2020*, pp. 449–454. doi: 10.1109/MLBDBI51377.2020.00095.

Wu, S. *et al.* (2020) ‘AN END-TO-END NETWORK FOR DETECTING MULTI-DOMAIN FRACTURES ON X-RAY IMAGES’, pp. 8–12.

Zhang, B. *et al.* (2020) ‘Attention-based CNN for KL Grade Classification: Data from the Osteoarthritis Initiative’, *Proceedings - International Symposium on Biomedical Imaging*, 2020-April, pp. 731–735. doi: 10.1109/ISBI45749.2020.9098456.