OSTEOARTHRITIS DISEASE DETECTION

Beebi Naseeba

School of Computer Science & Engineering

VIT-AP University,

Andhra Pradesh, India

beebi.naseeba@vitap.ac.in

Thanneeru Priyanka

School of Computer Science & Engineering

VIT-AP University,

Andhra Pradesh, India

priyankat3469@gmail.com

Gudiguntla Vyshnavi

School of Computer Science & Engineering

VIT-AP University,

Andhra Pradesh, India

gvyshnavi693@gmail.com

Maramareddy Meghana

School of Computer Science & Engineering

VIT-AP University,

Andhra Pradesh, India

maramreddymeghana03@gmail.com

Chidipothu Bhavya

School of Computer Science & Engineering

VIT-AP University,

Andhra Pradesh, India

chidipothubhavya@gmail.com

***Abstract*- Osteoarthritis (OA) is the most prevalent type of condition affecting the musculoskeletal system and its main cause of elderly people's decreased mobility. Evaluation of visual subjectivity still plagues OA. [1] An accurate prediction is a necessary first step to effectively identify and prevent severe osteoarthritis. Experts frequently use a manual examination of patient medical photos, typically gathered in hospitals, to diagnose osteoarthritis. For patients, determining whether osteoarthritis is present takes some time. The present studies also concentrate on using image-based deep learning systems to automatically identify osteoarthritis. Medical photographs of the patients are required, which necessitates hospital visits from the patients. In this study, a deep neural network was employed to identify the presence of osteoarthritis. Our study results on data from the publicly available Kaggle that the established method may provide results for the early identification of OA in the knee.[4]**

Index Terms— CNN Architectures, Early Stopping, L2 regularization, knee osteoarthritis detection, Feature representation learning, X-ray.

INTRODUCTION

Joint failure is a widespread, age-related, chronic, and slowly progressing joint condition called osteoarthritis. The joints that are most usually affected by osteoarthritis are those that are overused throughout life, such as the knees, neck, low back, and little joints at the ends of your fingers. The cartilage, a cushion made of silky rubber that covers the ends of the bones in the joint, is affected by this disorder. It has a few cells that secrete a complex matrix made of proteoglycans and hyaluronic acid.

Throughout life, worn-out matrix elements are constantly replaced even though cartilage has a very limited capacity for repair and adaptation. Menisci and other joint components are damaged, as well as the remodeling of 2 bones when cartilage breaks down in OA. OA is currently thought to be a disease brought on by the joints' dynamic response to various biomechanical and biochemical variables. When cartilage deterioration outpaces healing, it happens. The ligaments around the joint loosen, the joint capsule thickens, the joint fluid thins out, and the muscle deteriorates and weakens. Joint failure, a disorder similar to heart failure or kidney failure, is the result of these alterations. There is no discomfort in the cartilage. Bone and other tissues within and around the joint cause pain in OA. In OA, crystals and joint edoema are 2 major sources of discomfort.[18]

Around 10% of men and 13% of women over 60 have symptomatic knee OA, a disease that worsens with age and increases in frequency and prevalence. The number of persons who have OA that is the most severe in their age group is growing as the population ages. A primary contributor to morbidity and disability, OA also has high socioeconomic implications. The cost of arthritis in the

United States was projected to be $336 billion in 2004, accounting for 3% of the country's GDP, with osteoarthritis (OA) being the most prevalent type. About 4-6% of adults in India have osteoarthritis, which is one of the top 5 chronic disorders.
According to statistics, India annually reports more than 1 0 million instances of osteoarthritis. By 2025, the nation is anticipated to become the world's osteoarthritis capital with more than 60 million cases.

Source [21]

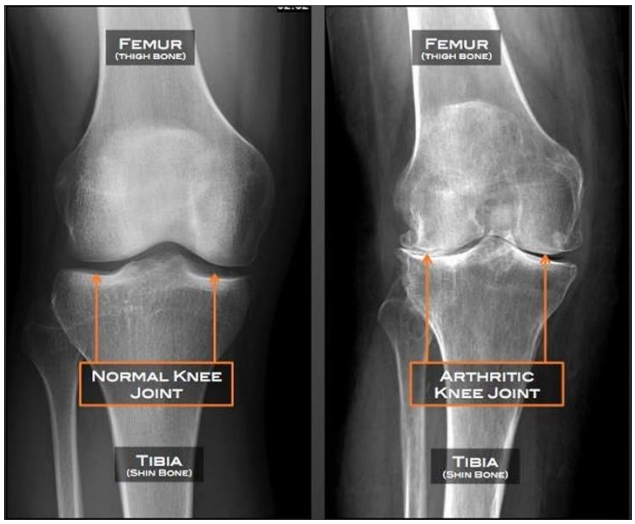


Fig 1 *Knee Joints of normal and diseased patient*

The most prevalent musculoskeletal disorder in the world is osteoarthritis (OA). According to the literature, hip and knee OA rank as the eleventh-highest worldwide disability factor1, placing a significant economic burden on society. According to reports, the average annual cost per patient for OA treatments is 19,000 euros. The current clinical failure to consistently detect the disease at an early stage, when it might still be possible to stop its course or at least lessen the severity of its future handicap, accounts for a portion of these expenditures. The only choices left to extend the patients' healthy years of life are an early diagnosis and behavioral interventions3 because there is no effective treatment for OA other than complete joint replacement surgery at the advanced stage.

Clinically, OA can be diagnosed early, but this needs the costly magnetic resonance imaging (MRI) that is now only available in specialized facilities or private practice. Additionally, this modality misses changes in bone architecture, which could be a sign of early OA progression.

LITERATURE STUDY

In this study, they have developed a system that is based on auto-encoders called the discriminative regularised auto-encoder (DRAE). In the cost function, a discriminative penalty term was added to increase the separability between healthy and OA situations. The obtained results support previous findings from research indicating the medial compartment of the knee contains the most discriminative regions. This study demonstrates that classical auto-encoders can still be enhanced for better extraction of discriminative patterns related to OA.[1]

In order to identify the most optimal feature set, we performed a principal component analysis (PCA) on the feature space. The most crucial information is compressed onto a subspace with lower dimensionality than the original space by PCA, which projects data onto a new space in which successive dimensions contain decreasing amounts of the variance of the original dataspace. They used 10-fold cross-validation to evaluate the feature space with 5–100% of the anticipated subspace to determine how many principal components were required to achieve full performance. They explored the use of four machine learning methods to learn the mapping function between the CDI feature space and OA severity denoted by KL, JSM, and JSN grades. The four machine learning methods are artificial neural network (ANN), support vector machine (SVM), random forest, and naive Bayes [2]

First, using the deep-stacked transformation technique, they thoroughly preprocess MRIs and radiographs to remove any potential artifacts and noise that would have hidden pictures for domain generalization. The region of interest (ROIs) are then extracted using U-Net architecture with ResNet as the backbone. They trained DenseNet and VGG architectures on the retrieved ROIs to categorize the cohorts. They then used gradient-guided class activation maps (Grad-CAM++) and layer-wise relevance propagation (LRP) to highlight regions that distinguish between different classes, and we then explain the predictions in terms that are understandable to humans.[3]

Deep learning approaches based on various types of data have been employed in multiple research to identify a variety of issues [6,7,8,9,10]. These include research projects using a variety of techniques and datasets to detect or predict OA using machine learning or deep learning. ANN version for osteoarthritic bone fracture chance classification and a decision-tree model for the prediction of antibody-mediated kidney transplant rejection have been proposed by shaikhina et al.[11]. Despite their great accuracy, they only utilized limited datasets (35 bone specimens and 80 kidney transplants). With 25 accessible samples, Kovanova et al[12] .'s NN study of trabecular bone in OA examined 25 samples. Convolution Neural Network (CNN) was used by Antony et al. [13] to asses the severity of knee OA.

They turned parameters on photos of knee OA and trained a dataset using ImageNet. With the aid of 2-D and 3-D X-ray data, Kobashi et al. [14] examined a post-operative knee function prediction model (52 OA patients). A machine learning algorithm's capacity to categorize magnetic resonance imaging (MRI) of 65 OA patients was assessed by Ashinsky et al. [15]. In order to forecast the incidence of knee OA over the next 30 months utilizing a variety of data, including clinical characteristics, questionnaires, biochemical markers, and pictures, Lazzarini et al. [16] suggested an analytical pipeline based on machine learning. 420 hip X-ray images were used by Xue et al. [17] to test

the diagnostic efficacy of CNN in identifying hip OA. Tiulpin et al[18] .'s analysis of X-ray images from 3000 individuals included CNN to identify knee OA.

Using NN, SVM, and random forest among other machine learning methods, Du et al. [19] investigated knee MRI characteristics for OA prediction with PCA. Using 986 pictures, Hirvasniemi used machine learning to predict the prevalence of radiographic hip OA. Using knee X-rays, Brahim et al. [20] used machine getting to know (multivariate linear regression) to perceive early oa inside the knee. Most earlier research centered on deep learning or machine learning with image data from sources like MRI or X-rays. To the best of our knowledge, there isn't much research that combines machine learning or deep learning with statistical data for prediction.

Our method outperforms comparable state-of-the-art methods in extensive testing based on DL models, producing up to 90% classification accuracy.

PROPOSED WORK

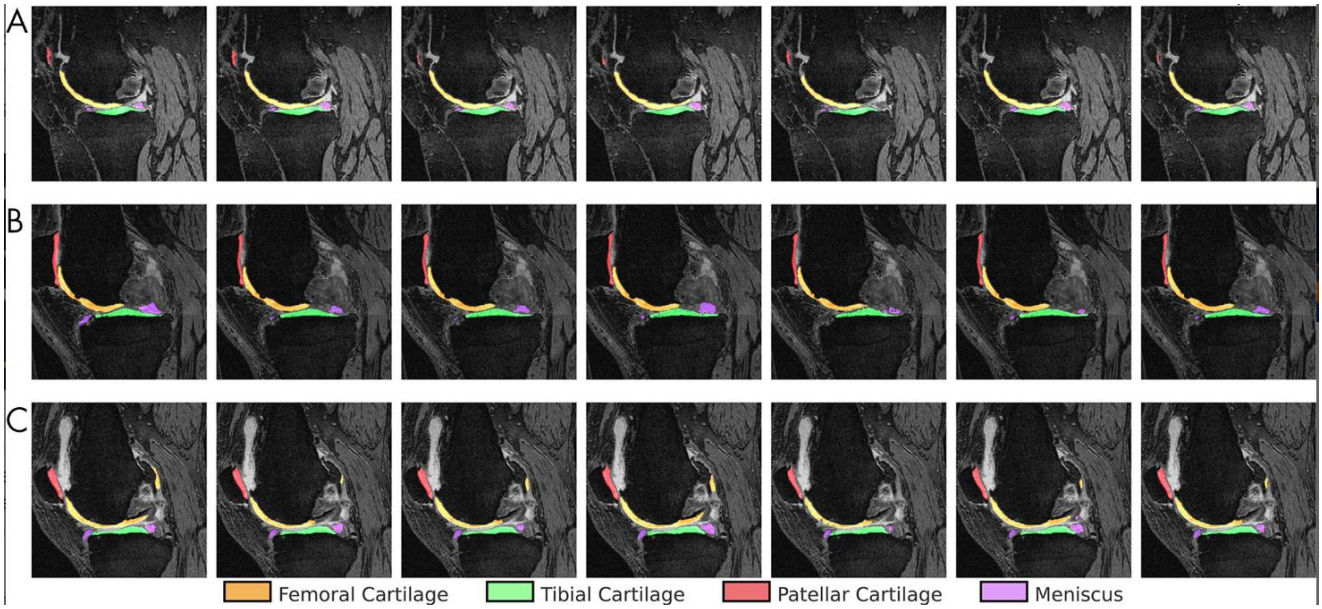


Fig 2 *Osteoarthritis Dataset*

DATA AUGMENTATION

Data augmentation is the method of creating extra data points from contemporary facts with the purpose of artificially booming the number of records of data. Expanding the data set may require minor tweaks to the data or the use of machine learning models to generate new data points in the latent space of the original data. Many photos that are similar to one another can be produced via data augmentation. Increasing the dataset size in this way helps to decrease overfitting. The cause of this is that as more data is added, the model is unable to overfit all samples and is instead compelled to generalize comparable cutting-edge methods.

Our model essentially applies rotation, horizontal flip, height shift, width shift, and shear to the dataset where noise is added. However, prediction is precise.

NORMALIZING

A set of data is transformed to be on a similar scale through normalization. Depending on the data itself, the purpose of machine learning models is typically to recenter and rescale our data so that it is between 0 and 1 or -1 and 1. Normalization is a technique used in data processing to scale the values of numeric columns in a dataset to a common scale when the features in the data have diverse ranges.

The pixel values for RGB images range from 0 to 255. Incorrect values may appear when an algorithm is used. The values of the pixels were translated to the 0–1 range, to obtain clear values.

Several benefits of normalization include:

- Improving training by reducing internal covariate shift

- Preventing or reducing bias in the network by scaling each feature to a similar range.
- Accelerating the optimization process by limiting the range of the weights and preventing them from exploding in all directions.
- Reducing network overfitting by promoting regularisation.

EARLY STOPPING

We want to train our model for a longer period of time by training it for a greater number of epochs. An epoch in deep learning is the number of times our model has been trained using the full dataset. The number of epochs must be carefully chosen to produce an ideal model. A model will be underfitting if the number of epochs is very low, and it will be overfitted if the number of epochs is very large.

Early halting can be used to select the ideal number of epochs. We picked 100(high number) epochs, and early stopping will halt training as soon as the data begins to fit the model too well. Due to the size of our dataset, we used early stopping to shorten training time and improve accuracy.

The objective of such tasks is to approximate a regression function, $f\rho$ provided by for a given input space, X, output space, Y, and samples taken from an unknown probability measure, rho, on $Z=X*Y$.

Where ρ (p/x) is the conditional distribution at x induced by ρ.

$$f_\rho(x) = \int_Y y \, d\rho(y \mid x),\ x \in X,$$

L2 REGULARIZATION

Due to the size of the dataset, there is a possibility of overfitting, which renders the model ineffective. Therefore, we employed L2 regularisation to prevent the issue. In L2, you add the data to the model equation to be the sum of the vector's squared times the regularisation parameter's () value, which can be any large number over the amount of the data (m), where (n) is a number of features.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)}) + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

ACCURACY

A metric of the model's performance across all classes is accuracy. When all classes are equally important, it is helpful. The number of accurate forecasts divided by the total number of predictions is used to compute it. The evaluation process takes into account measures for MRI segmentation like accuracy. This measure is calculated using the confusion matrix's True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

The classification metric used to assess performance is accuracy.

ADAM OPTIMIZER

The Adam optimizer combines the following two methods of gradient descent: By taking into account the "exponentially weighted average" of the gradients, this approach is used to speed up the gradient descent algorithm. The technique converges quicker towards the minima when averages are used. Adam can be viewed as an RMSprop and stochastic gradient descent with a momentum combination. It scales the learning rate using squared gradients, similar to RMSprop and leverages momentum by using the gradient's moving average rather than the gradient itself, similar to SGD with momentum.

GLOBAL AVERAGE POOLING

A pooling method called global average pooling is intended to take the place of fully connected layers in conventional CNNs. The goal is to produce one feature map in the final mlp (multi-layer perceptron) Conv layer for each category that corresponds to the classification problem. We average each feature map, and the resulting vector is sent straight into the softmax layer, rather than constructing fully connected layers on top of the feature maps. By requiring correspondences between feature maps and categories, global average pooling is more native to the convolution structure than completely connected layers, which is one advantage.

As a result, the feature maps are simply categories of confidence maps. Another benefit of global average pooling is that overfitting is prevented at this layer because there are no parameters to optimize. Additionally, because global average pooling sums up the spatial data, it is more resistant to input spatial translations.

RESIDUAL NETWORK

This architecture introduces the idea of residual blocks to solve the vanishing/explosive gradient problem. We use a technique known as connection skipping on this network. Skip connections connect layer activations to subsequent layers, bypassing some intermediate layers. This will generate a block of residuals. These remaining blocks are stacked to create resnets.

The strategy behind this network is to let the network fit the residual mapping rather than have layers learn the underlying mapping. Therefore, let the network fit**, F(x):= H(x) - x**, which results in **H(x):= F(x) + x**, be used instead of, say, the original mapping H(x).

BATCH NORMALIZATION

The contributions to a layer for each mini-batch are normalized when using the batch normalization technique to train very deep neural networks. As a result, the learning process is stabilized, and the quantity of training epochs needed to train deep neural networks is dramatically reduced. Creating a deep neural network with several layers is necessary for deep learning since these networks might be sensitive to the learning algorithm's architecture and underlying initial random weights.

One possible cause of this issue is that when the weights are refreshed after each mini-batch, the distribution of the inputs to layers below in the network may shift. This can force the learning system to pursue a moving target continuously.

Internal covariate shift, a technical term for this modification in the input distribution to the network's layers, has been mentioned. The challenge is that the model is updated layer-by-layer in the opposite order from input to output, with an error estimate that presumes that the weights of the layer preceding the current layer remain constant. Batch normalization offers a comprehensive method for parameterizing practically any deep neural network. Reparameterization greatly reduces the difficulty of revising plans at several levels.

ReLU

The rectified linear activation function, or ReLU for short, will produce output zero if the input is negative and the input directly if it is positive. Because a model that utilizes it is simpler to train and frequently performs better, it has evolved into the standard activation function for many different kinds of neural networks.

The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time

$$RELU(x) = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x >= 0 \end{cases}$$

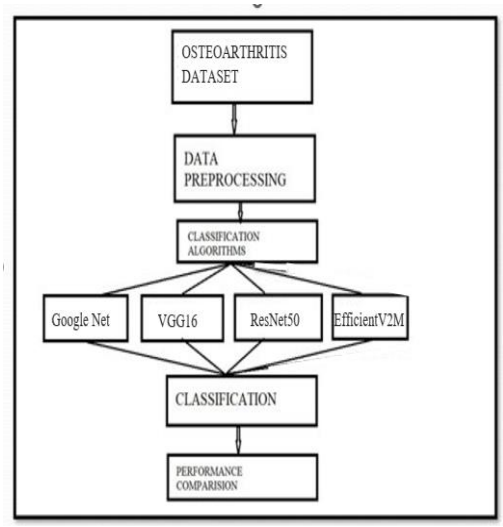
SIGMOID

A Sigmoid function is a mathematical function that has a characteristic S-shaped curve. There are a number of common sigmoid functions, such as the logistic function, the hyperbolic tangent, and the arctangent

In machine learning, the term sigmoid function is normally used to refer specifically to the logistic function, also called the logistic sigmoid function. One application of a sigmoid function is to transform a real value into one that can be understood as a probability because all sigmoid functions have the property that they map the entire number line into a tiny range, such as between 0 and 1, or -1 and 1.

Formula: $$f(x) = \frac{1}{1 + e^{-(x)}}$$

System Design



Various CNN predefined architectures are used here, including VGG16, RESNET50, EfficientNetV2L, and GOOGLE NET. Here, we evaluate the dataset's precision and effectiveness across all built-in architectures. Furthermore, depending on our training and testing datasets, these techniques produce varying degrees of accuracy.

Model	Size (MB)	Top-5 Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
VGG16	528	90.1%	138.4M	16	69.5	4.2
ResNet50	98	92.1%	25.6M	107	58.2	4.4
GoogleNet	92	93.7%	23.9M	189	42.2	6.9
EfficientNet V2M	220	97.4%	54.4M	-	-	-

Table 1: Models Used

1. GoogleNet

Understanding CNN Architecture and the GoogleNet Model In 2014, Google researchers introduced Google Net (also known as Inception V1) in the research article "Going Deeper with Convolutions" with the assistance of several universities. This architecture employs methods like global average pooling and 1–1 convolutions in the middle of the architecture. What GoogleNet offers:

In comparison to earlier state-of-the-art architectures like AlexNet and ZF-Net, Google's architecture is quite distinctive. It employs a variety of techniques, including global average pooling and 1-1 convolution, to build deeper architecture.

This architecture includes vanishing and exploding gradients built-in.

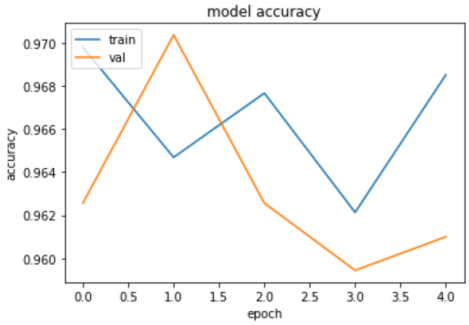
Vanishing Gradient

The gradients frequently decrease and finally approach zero as the backpropagation algorithm moves from the output layer to the input layer downward (or backward), essentially maintaining the weights of the lower or beginning layers. Because of this, the gradient decline never reaches the ideal state. The vanishing gradients dilemma describes this situation.

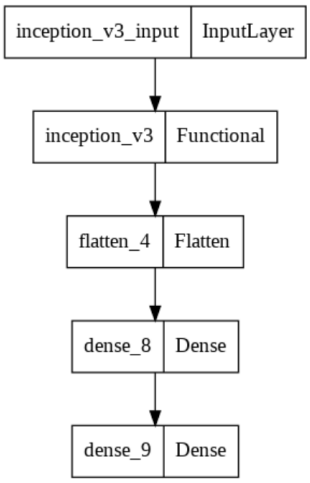
Exploding Gradient

On the other hand, in some instances, the gradients continue to grow as the backpropagation method advances. The gradient descent then diverges as a result of the very huge weight updates. The exploding gradients problem describes this situation.

Graph of Accuracy Vs Epoch (Train and Validation data)



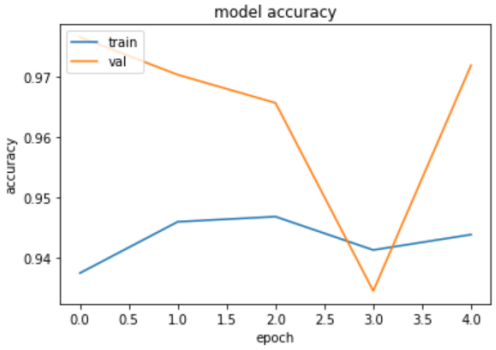
System Architecture



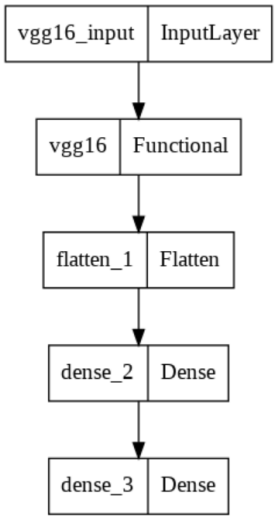
2. VGG16

The Visual Geometry Group is known as VGG. A convolutional neural network with 16 layers is called VGG-16. It is utilized for extremely large-scale image identification, and in the 2014 ILSVRC challenge, this model took first and second place in the aforementioned category. The pretrained network can categorize photos into 1000 different object categories, including several animals, a keyboard, a mouse, and a pencilAs a result, the network now has comprehensive feature representations for a range of photos.

Graph of Accuracy Vs Epoch (Train and Validation data)



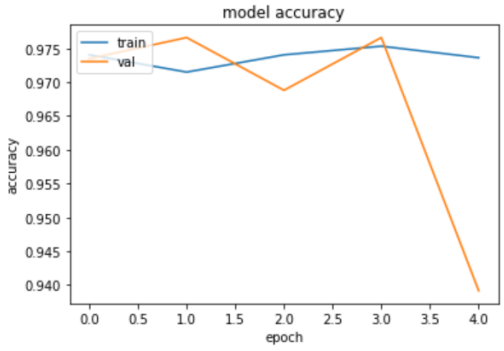
System Architecture



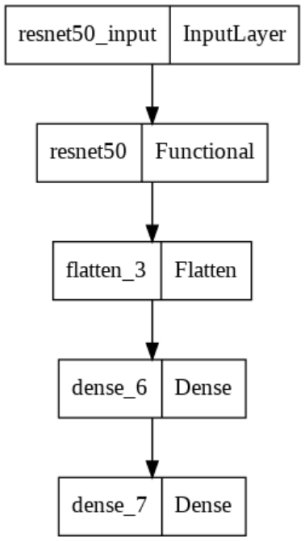
3. ResNet50

Deep residual networks are convolutional neural networks (CNNs) with more than 50 layers, like the well-known ResNet-50 model. A Residual Neural Network (ResNet) is a type of Artificial Neural Network (ANN) that builds a network by piling residual blocks on top of one another. There are around 23 million trainable parameters in the ResNet-50. A ResNet model version called ResNet50 contains 48 Convolution layers, 1 MaxPool layer, and 1 Average Pool layer. There are 3.8 x 109 floating point operations available.

Graph of Accuracy Vs Epoch (Train and Validation data)



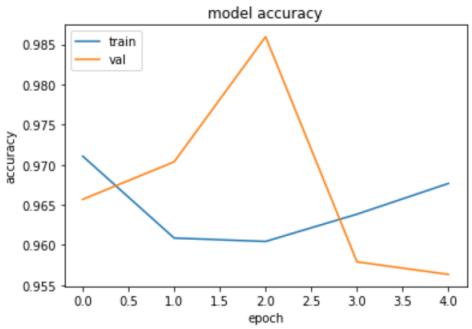
System Architecture



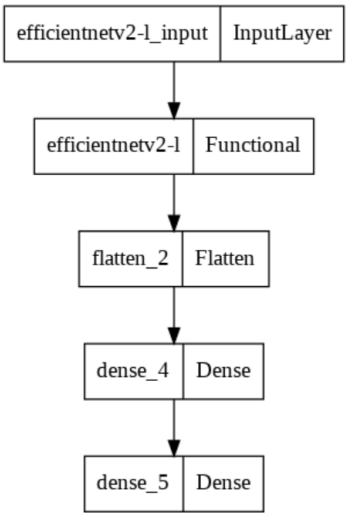
4. EfficientNetV2M

A family of image classification models called EfficientNetV2 outperforms earlier works in terms of parameter efficiency and training speed. Our EfficientNetV2M models are scaled up for quicker training and inference speed and are based on EfficientNetV1 and use neural architecture search (NAS) to simultaneously maximize model size and training speed. Mobile inverted bottle net convolution is employed (MBConv).

Graph of Accuracy Vs Epoch (Train and Validation data)



System Architecture



RESULTS

I. DATASET:

A dataset partition comprising ground truth articular (femoral, tibial, and patellar) cartilage and meniscus segmentations was created using three-dimensional knee MRI scans from 88 retrospective individuals at baseline and one year afterward. On a holdout test set, the Dice score, average symmetric surface distance, volumetric overlap error, and coefficient of variation were used to compare challenge submissions and a majority-vote ensemble to ground truth segmentations. Pairwise Dice coefficient correlations were used to gauge similarities in automated segmentations. The thickness of the articular cartilage was calculated longitudinally and using scans. The Pearson correlation coefficient was used to calculate the relationship between thickness error and segmentation measures. In order to calculate two empirical upper bounds for ensemble performance, model output combinations that combined true positives and true negatives were used. [4]

II. DATA PREPROCESSING

Segmentations from the unreleased ground truth test set were used to evaluate networks. The only available evaluation measures for the challenge were the average Dice score (from 0 to 1) for each tissue individually. The volumetric overlap error (VOE) (range, 0-1), coefficient of variation (CV) (range, 0–∞), and average symmetric surface distance (ASSD) (range, 0–∞) in millimeters were also computed as additional pixel-wise segmentation metrics. With the exception of Dice, a lower number denoted greater accuracy.

The pairwise Dice correlations among test set predictions from all networks were computed to determine how comparable segmentation results were across all networks. Equation (1) gave the following definition for the dice correlation (ρDice) between network A and network b segmentation ($f_A(x)$ and $f_B(x)$):

$$Dice\left(f_A(x), f_B(x)\right) = \frac{2 \cdot f_A(x) \cdot f_B(x)}{f_A(x) + f_B(x)} \quad (1).$$

III. CLASSIFICATION PERFORMANCE USING DIFFERENT ARCHITECTURE MODELS

We compared four deep-learning algorithms: including VGG16, ResNet50, EfficientNetV2L, and Google Net These algorithms were performed in the binary classification case.

Source [5]

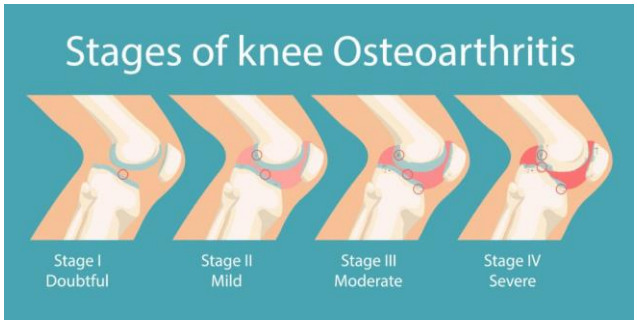


Fig 2

"Have you ever had osteoarthritis (OA) diagnosed by a doctor?" this question was utilized to identify patients with OA. Gender, age, region, marital status, education, and household income can be included as independent factors, along with other demographic and personal traits. Following data training, the stages of osteoarthritis (OA) of the knee are predicted. The estimated OA probability for the photos are as follows: probabilities below 0.3 are regarded as questionable, probabilities between 0.3 and 0.5 as mild, probabilities between 0.5 and 0.8 as moderate, and probabilities over 0.8 as severe.

The objective of this study is to choose the model that performs the best in terms of classification. Table 2 presents that the VGG16 model, which has a maximum voting accuracy of 90.4%, produced the best results. We also obtained accuracy values of 83.8% for Google Net, 81.9% for ResNet50, and 88.5% for EfficientV2L. These findings demonstrate that, in comparison to other algorithms, the VGG16 model generally produced greater performance.

Architecture	Accuracy	Loss
Google Net	0.8380952477	0.46964722871
VGG16	0.9047619104	0.27701595425
ResNet50	0.8190476298	1.04587388038
EfficientV2L	0.8857142925	0.21474921703

Table 2

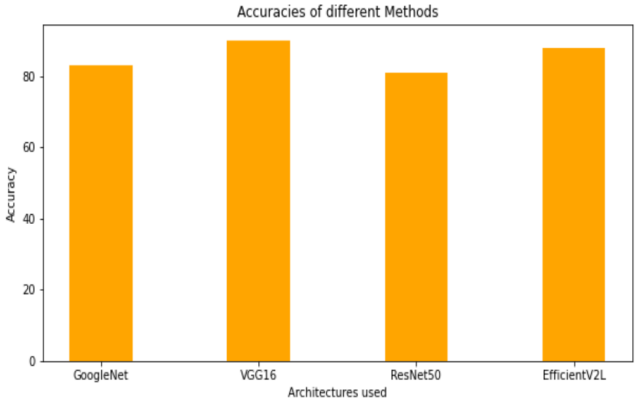
Conclusion

In this study, we provided a novel representation learning model for X-ray image-based knee OA identification. The multicentric OAI database experiment results show that our technique is resistant to artifacts and data-collecting conditions. The obtained results support previous findings from other studies indicating the medial compartment of the knee contained the most discriminative regions.

This study demonstrates that accuracy can still be increased using CNN architectures. Therefore, adding layers promises to improve performance. Each weight is updated so that it moves along the gradient of weight direction, which is the direction along which the loss diminishes. This causes the gradient to gradually decrease, which results in relatively minor modifications to the original layers. As a result, VGG's training period is greatly extended.

Another sort of convolutional neural network architecture (CNN) with 50 layers is known as deep residual networks (ResNets), including the well-known ResNet-50 model. ResNets are less complicated than VGGNets because they contain fewer filters. ResNet, also known as Residual Network, prevents the vanishing gradient issue from happening. Whereas GoogLeNet is a convolutional neural network that is 22 layers deep and has inception modules and EfficientNetV2L is a type of convolutional neural efficiency than previous models.

Graph



We got an accuracy of 83% for Google Net, 90% for VGG16, 81% for ResNet50, and 88% for EfficientV2L. These previously mentioned parameters help us achieve the maximum accuracy for the VGG16 Architecture in this work. Overall, these findings demonstrate that accuracy levels are broadly similar, making it more difficult to select the optimum strategy for the categorization problem.

References:

[1] Y. Nasser, R. Jennane, A. Chetouani, E. Lespessailles and M. E. Hassouni, "Discriminative Regularized Auto-Encoder for Early Detection of Knee OsteoArthritis: Data from the Osteoarthritis Initiative," in IEEE Transactions on Medical Imaging, vol. 39, no. 9, pp. 2976-2984, Sept. 2020, doi: 10.1109/TMI.2020.2985861.

[2] Progression on MRI Using Machine Learning Methods," in *IEEE Transactions on NanoBioscience*, vol. 17, no. 3, pp. 228-236, July 2018, doi: 10.1109/TNB.2018.2840082.

[3] M. R. Karim et al., "DeepKneeExplainer: Explainable Knee Osteoarthritis Diagnosis From Radiographs and Magnetic Resonance Imaging," in IEEE Access, vol. 9, pp. 39757-39780, 2021, doi: 10.1109/ACCESS.2021.3062493.

[4] The International Workshop on Osteoarthritis Imaging Knee MRI Segmentation Challenge: A Multi-Institute Evaluation and Analysis Framework on a Standardized Dataset Arjun D. Desai, Francesco Caliva, Claudia Iriondo, Aliasghar Mortazi, Sachin Jambawalikar, Ulas Bagci, Mathias Perslev, Christian Igel, Erik B. Dam, Sibaji Gaj, Mingrui Yang, Xiaojuan Li, Cem M. Deniz, Vladimir Juras, Ravinder Regatte, Garry E. Gold, Brian A. Hargreaves, Valentina Pedoia, Akshay S. Chaudhari, on behalf of the IWOAI Segmentation Challenge Writing Group, Naji Khosravan, Drew Torigian, Jutta Ellermann, Mehmet Akcakaya, Radhika Tibrewala, Io Flament, Matthew O'Brien, Sharmila Majumdar, Kunio Nakamura, and Akshay Pai Radiology: Artificial Intelligence 2021 3:3

[5] https://healthlibrary.askapollo.com/what-is-osteoarthritis-of-the-knee-stages-symptoms-causes-diagnosis-and-treatment/

[6] Ma, X.; Dai, Z.; He, Z.; Ma, J.; Wang, Y.; Wang, Y. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. Sensors 2017, 17, 818.

[7] Izadpanahkakhk, M.; Razavi, S.M.; Taghipour-Gorjikolaie, M.; Zahiri, S.H.; Uncini, A.; Taghipour-Gorjikolaie, M.; Zahiri, S.; Uncini, A. Deep region of interest and feature extraction models for palmprint verification using convolutional neural networks transfer learning. Appl. Sci. 2018, 8, 1210.

[8] Xu, K.; Feng, D.; Mi, H. Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. Molecules 2017, 22, 2054.

[9] Steven Eyobu, O.; Han, D.S. Feature representation and data augmentation for human activity classification based on wearable imu sensor data using a deep lstm neural network. Sensors 2018, 18, 2892.

[10] Jing, L.; Wang, T.; Zhao, M.; Wang, P. An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox. Sensors 2017, 17, 414.

[11] Shaikhina, T.; Lowe, D.; Daga, S.; Briggs, D.; Higgins, R.; Khovanova, N. Machine learning for predictive modelling based on small data in biomedical engineering. IFAC–PapersOnLine 2015, 48, 469–474.

[12] Khovanova, N.A.; Shaikhina, T.; Mallick, K.K. Neural networks for analysis of trabecular bone in osteoarthritis. Bioinspired Biomim. Nanobiomater. 2015, 4, 90–100.

[13] Antony, J.; McGuinness, K.; O'Connor, N.E.; Moran, K. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 1195–1200.

[14] Kobashi, S.; Hossain, B.; Nii, M.; Kambara, S.; Morooka, T.; Okuno, M.; Yoshiya, S. Prediction of post-operative implanted knee function using machine learning in clinical big data. In Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), Jeju, Korea, 10–13 July 2016; pp. 195–200.

[15] Ashinsky, B.G.; Bouhrara, M.; Coletta, C.E.; Lehallier, B.; Urish, K.L.; Lin, P.C.; Goldberg, I.G.; Spencer, R.G. Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative. J. Orthop. Res. 2017, 35, 2243–2250.

[16] Lazzarini, N.; Runhaar, J.; Bay-Jensen, A.C.; Thudium, C.S.; Bierma-Zeinstra, S.M.A.; Henrotin, Y.; Bacardit, J. A machine learning approach for the identification of new biomarkers for knee osteoarthritis development in overweight and obese women. Osteoarthr. Cartil. 2017, 25, 2014–2021.

[17] Xue, Y.; Zhang, R.; Deng, Y.; Chen, K.; Jiang, T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. PLoS ONE 2017, 12, e0178992.

[18] Tiulpin, A.; Thevenot, J.; Rahtu, E.; Lehenkari, P.; Saarakkala, S. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. Sci. Rep. 2018, 8, 1727.

[19] Du, Y.; Almajalid, R.; Shan, J.; Zhang, M. A novel method to predict knee osteoarthritis progression on MRI using machine learning methods. IEEE Trans. Nanobiosci. 2018, 17, 228–236.

[20] Brahim, A.; Jennane, R.; Riad, R.; Janvier, T.; Khedher, L.; Toumi, H.; Lespessailles, E. A decision support tool for early detection of knee osteoarthritis using x-ray imaging and machine learning: Data from the osteoarthritis initiative. Comput. Med. Imaging Graph. 2019, 73, 11–18.

[21] https://healthjade.com/osteoarthritis/

[22] https://artificialintelligencestechnology.com/wp-content/uploads/2021/05/sigmoid-function.png