LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 24, NO. 3 (2026)

Enhancing Local Public Health Services Through AI-Driven Knee Osteoarthritis Severity Assessment Using An Optimized Deep Vision Transformer

Sudha K [a*], Murugesan P [b] , Manjula devi R [c], Ramkumar M[d]

[a] Computer Science and Engineering (Cybersecurity), K. S. R. College of Engineering, Tiruchengode 637215, India
[b] Mechanical Engineering , K. S. R. College of Engineering, Tiruchengode 637215, India
[c] Computer Science and Engineering ,PSG Institute of Technology and Applied Research, Neelambur-641062, India
[d] Computer Science and Business Systems, Knowledge Institute of Technology, salem,India
Corresponding Author Email: srisudhan3@gmail.com

Abstract
Knee Osteoarthritis (Knee OA) is a prevalent condition that affects elderly individuals. Knee OA diagnosis includes knee X-ray images. Proposed Tetrachoric correlated AlexNet based Golden Jackal Optimized Deep Vision Transformer (TA-GJODVT) model is developed for efficient classification of knee osteoarthritis' severity levels. Numerous X-ray knee images gathered from the dataset in acquisition process. Then, the collected images are preprocessed using Adaptive Wiener Filter for eliminating the noise and enhancing the visibility without blurring. After that, segmentation process is performed by applying using Tetrachoric Correlated AlexNet model for segmenting joint space and extracting Region of Interest (RoI). Followed by this, Gradient Tuned Golden Jackal Optimized Deep Vision Transformer is employed to perform feature extraction, classification and fine-tuning. In addition, Gradient tuned Golden Jackal Optimization is applied to execute hyperparameter tuning of transformer model for improving the accuracy of severity grading with lesser error rate. Experimental evaluation is carried out on several factors. The quantitative analysis of proposed TA-GJODVT Model achieves superior performance in classifying knee OA classification with minimum time and error rate compared to existing techniques.

Keywords: classification, Knee OA, Transformer, Alex Net, Segmentation.

1.Introduction

Knee OA represents global health concern leading to pain and disability among aging population. You Only Look Once version 11 (YOLOv11) was designed in [1] to categorize KOA severity. But, hyperparameter was failed to be optimized. ResNet model was designed in [2] to classify knee OA. In [3], KOA grade assessment through improved Compact Convolutional Transformer (CCT) (KOA-CCTNet) was developed. Hybrid CNN and Visual Geometry Group 16 (CNN+VGG16) was introduced in [4]. Dense Neural Network (DNN) [5] and Genetic Algorithm (GA) enhanced deep ensemble method [6] were investigated. Attention enhanced DL and ML [7] and DL- [8] were presented. Densely Connected Fully Convolutional Network (DFCN) was introduced in [9]. Residual learning with CNN was presented in [10]. Ensemble of CNN, decision tree, K-nearest classifier model was designed in [11]. Fine-tuned CNN-based pre-trained models were introduced in [12]. Automated DL-based ordinal classification scheme was introduced in [13]. DL based X-ray techniques [14] and Transfer learning approach [15] were designed to enhance accuracy.

Knee Osteoarthritis is increased degenerative joint illness depicted through concerning the space between the tibia and femur bones. Early and precise diagnosis of KOA is significant to lessen its long-term effects and to make timely medical intervention. However, conventional diagnostic methods are huge time-consuming. Several deep learning models are complex patterns from knee X-ray images, facilitating accurate detection of KOA severity levels. But, traditional diagnostic methods often lack the precision and efficiency need for accurate KOA classification, especially at early stages. Also, extraction of most prominent features from x-ray images did not performed to increase overall classification performance of KOA. Additionally, the optimizing network training to obtain robust classification outcomes are not well concentrated in previously developed approaches. To overcome the issue, TA-GJODV is developed with higher PSNR, accuracy and lesser time and space. Innovation and contribution of proposed TA-GJODV is described as,

- To enhance classification accuracy of knee OA disease severity grading, the TA-GJODVT technique is developed by incorporating preprocessing, joint space segmentation, and classification processes.
- To eliminate the noise presented in the x-ray images and minimize the time consumed in severity classification, TA-GJODVT technique uses the Adaptive Wiener Filter for image preprocessing. Pixels are arranged by filtering window concept and local mean and variance are computed. Based on the maximum deviation between a pixel and its mean value, noisy pixels are identified. In this way, the knee image quality is enhanced.
- To precisely segment joint space images, Tetrachoric Correlated AlexNet model is employed to compute the association between pixels in the images for partitioning knee joint spaces.
- To classify the knee x-ray images into different severity classes, Gradient Tuned Golden Jackal Optimized Deep Vision Transformer is designed. The parameter used in the transformer model is fine-tuned using Golden Jackal Optimized Algorithm to minimize the classification error.
- Finally, an experimental evaluation is conducted to validate the performance of the TA-GJODVT technique using various metrics compared with other deep learning methods.

Structure of article is given as: Section 2 describes related words. Section 3 explains materials and methods. Section 4 and 5 outlines experiments and results and discussion. Section 6 summarizes conclusion.

2. Related works

Fine-tuning KOA diagnosis framework was designed in [16]. Automated decision-making approach [17] and CNN [18] were designed. Multi-Modal Learning Method (MMLM) was designed in [19]. Modified deep angular extreme learning ML was designed in [20]. Ensemble DL network [21] and sophisticated 12-layer CNN architecture [22] were presented. DL model was designed in [23]. Gaussian Aquila optimizer with dual CNNs was presented in [24]. Greylag Goose Optimization (bGGO) was designed in [25]. Transfer learning with CNNs was investigated in [26]. Automatic grading of knee OA [27] and Dense multi-scale CNN [28] were introduced. Optimal DNN [29] and Enhanced CenterNet were introduced in [30] for knee OA recognition. VGG16 with quantum Convolutional Neural Network (QCNN) was designed in [31] to classify knee osteoarthritis (OA) severity. But, the accuracy was lesser. Hierarchical classification method was examined in [32] to automatically estimate the severity of KOA. But, the error was not considered. Novel predictive classifier termed as Osteo-Fusion was introduced in [33] for deterring the presence or absence of KOA with higher accuracy. However, it failed to reduce the time.

3. Materials and methods

Effective classification of knee OA severity grades in knee X-ray images using TA-GJODVT is proposed. Architecture of TA-GJODVT is shown in Fig. 1.

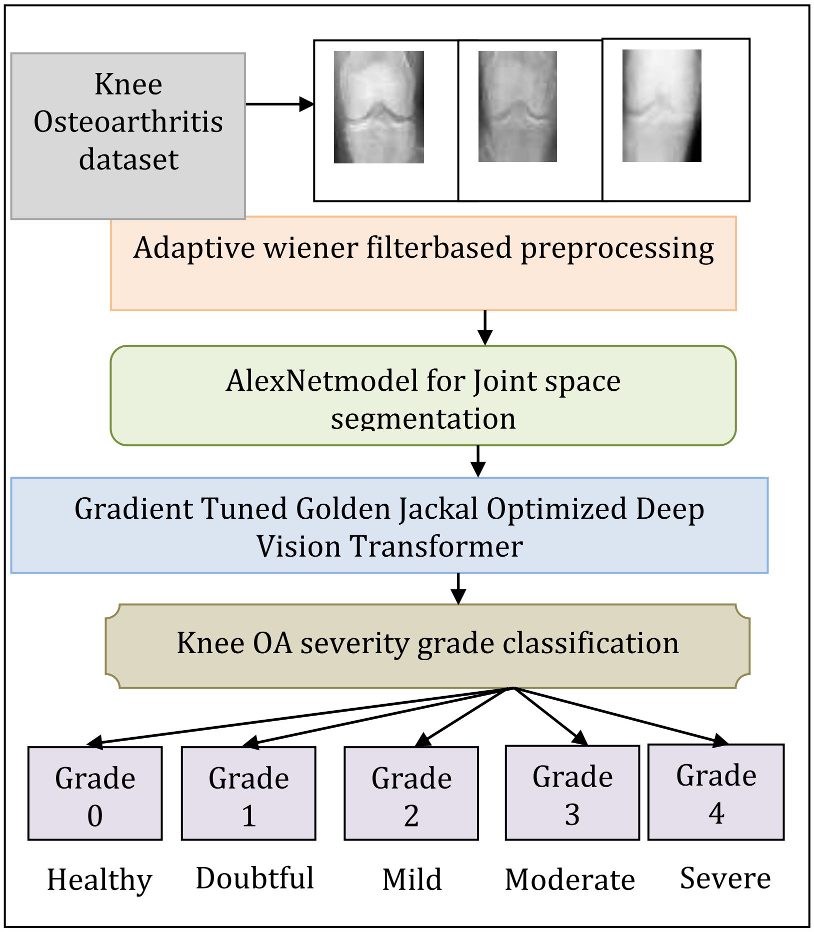


Fig. 1. Architecture of TA-GJODVT

3.1 Image Acquisition

Image acquisition on x-ray imaging dataset called Knee Osteoarthritis Dataset with Severity Grading performed. Dataset is gathered fromhttps://www.kaggle.com/datasets/shashwatwork/knee-osteoarthritis-dataset-with-severity. It comprises 9786 images. Knee OA KL grading descriptions are provided below.

Table 1. Knee Osteoarthritis Dataset with Severity Grading

Grades	**Label**	**Description**
Grade 0	Healthy	Healthy Knee image
Grade 1	Doubtful	Joint constriction is doubtful, and osteophytic lipping may be present.
Grade 2	Minimal	Osteophytes are definitely present, and feasible joint space narrowing
Grade 3	Moderate	Numerous osteophytes, obvious narrowing of the joint area, and mild sclerosis

Grade 4	Severe	Significant joint narrowing, large osteophytes, and serious sclerosis

3.2 Preprocessing

Proposed TA-GJODVT uses adaptive wiener filtering technique to perform preprocessing. Wiener filtering is to filter irrelevant information in input images. Weiner filtering is employed for eradicating noise to enhance the image quality. Adaptive Wiener filtering is chosen for image preprocessing over other denoising techniques to balance noise suppression. It calculates local mean and variance to adjust the filter intensity within a pixel neighborhood while preserving edges in high-variance, detailed regions. It is designed to minimize the MSE between the estimated image and the original image. It also provides higher PSNR. Knee X-ray images '$I_i = I_1, I_2, .., I_n$' from Knee Osteoarthritis Dataset with Severity Grading is used as input. Each image includes number of pixels denoted by $p_{1,} p_2, p_3, \ldots p_k$.

$p_{1,}$	p_2	p_3
p_4	p_5	p_6
p_7	p_8	p_9

Fig. 2. Filtering window

Fig. 2 demisters 3x3 filtering window, where image pixels $p_{1,} p_2, p_3, \ldots p_k$ placed into rows and columns. Upon arranging pixels in filtering window, local adaptive filter gets improved quality image by minimizing mean square error in **Eq. (1).**

$$DI_s = \frac{\sigma^2 (x,y)}{\sigma^2 (x,y) + \sigma^2{}_n(x,y)} [n(x,y) - m(x,y)] + m(x,y) \qquad (1)$$

Where 'DI_s' is denoised images, '$\sigma^2(x,y)$' is local variance of pixels, '$\sigma^2{}_n(x,y)$' is variance of noise, '$n(x,y)$' is observed noisy pixels, '$m(x,y)$' is mean of pixels, 'DI_s' is estimated true pixel value of the image after decreasing the noise. Filter adapts based on local mean and variance to maintain image information while decreasing noise.

// **Algorithm 1:** Adaptive wiener filter based image preprocessing
Input: Dataset 'DS', Number of knee x-ray images $I_1, I_2, I_3, \ldots, I_n$ **Output**: Enhance image quality
Begin   1. **Collect** number of natural images $I_1, I_2, I_3, \ldots, I_n$ from dataset   2. **For each** image I_i 3. Locate pixels $p_{1,} p_2, p_3, \ldots p_k$ in filtering window   4. **For each** p_i 5. Compute local mean and deviation   6. **End for**   7. **End for**   8. Get filtered output using (1)   9. **Return** (quality enhanced image) **End**

3.3 ROI joint space extraction and segmentation

ROI extraction and segmentation are concerned in image processing process. Image segmentation is to separate preprocessed image into its essential parts. After segmentation, binary image is formed where all background pixels include gray level value and all object pixels include white level value. This helps to get image into preferred region of knee area and thereby regions are removed from image. Segmentation algorithm is performed by Tetrachoric correlated AlexNet model for semantically segment ROI knee joint spaces. AlexNet is adapted for image segmentation to segment the Regions of Interest (ROIs) in knee joints by transforming it into a Fully Convolutional Network (FCN). In our work, the segmentation algorithm is performed by applying Tetrachoric correlated AlexNet model in proposed TA-GJODVT model for semantically segment ROI knee joint spaces. Tetrachoric Correlation coefficient is a statistical technique used in AlexNetto estimate the association between the pixels in ROI images viaCramér's phi test. Knee x-ray images are segmented into several similar ROI regions with lesser time.

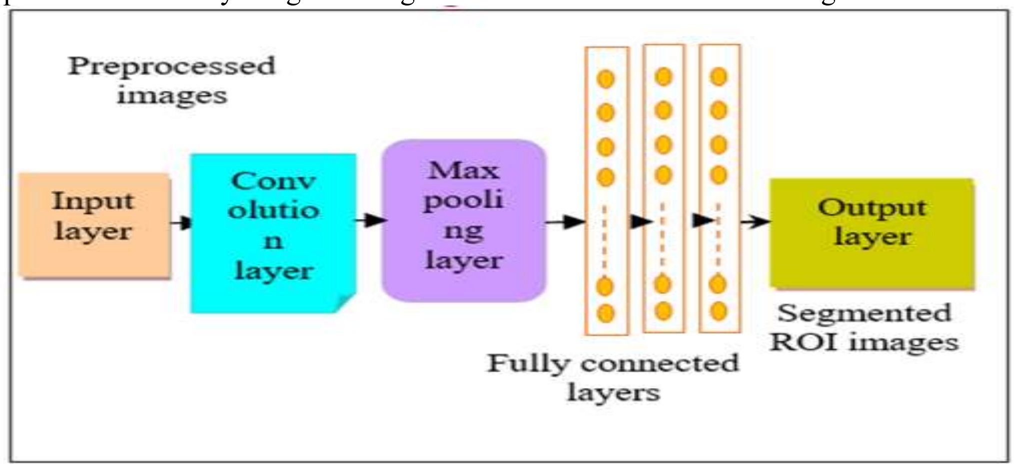


Fig. 3. Schematic structure of Tetrachoric Correlated AlexNet

Fig. 3 illustrates schematic structure of TetrachoricCorrelated AlexNet. AlexNet comprises different types of layers such as input layer, one or more hidden layers, and output layer. Hidden layer includes several sublayers such as convolutional layers, max-pooling layers and fully connected layers. Input layer takes preprocessed images as input and it is forwarded to the convolutional layers. Five convolutional layers are to extract ROI portions from preprocessed images. ROI is determined through summing up all pixels within region and it is given by Eq. (2),

$$Q = \sum_i \sum_j p_{ij} \quad (2)$$

Where, 'Q' refers area, '$\sum_i \sum_j P_{ij}$' refers summing up all pixels in that region. RoI is identified. Max pooling layers are employed to decrease the spatial dimensions (width and height) of extracted ROI. In Alexnet, Max pooling diminish spatial dimensions of extracted ROI at same time maintain most significant features. Pooling operation for resultant ROI image is given Eq. (3) as,

$$max_{pool} = \max\{ROI_{ij} : i \in [k, k+1], j \in [l, l+1]\} \quad (3)$$

From (3), 'max_{pool}' represents the output of the max pooling operation at position (k, l) in output ROI, ROI_{ij} denotes extracted ROI in the input preprocessed image, $\max$ refers to the operation of choosing maximum value from a particular set of values of an input ROI area. By pooling layers, spatial dimension reduced extracted ROI are subjected to fully connected layers for segmenting knee joints spaces. In that layer, Tetrachoric Correlated Cramér's phi test is used to analyze correlation between pixels in ROI images. Tetrachoric Correlation coefficient is a statistical technique used to compute association between the pixels in ROI images using Eq. (4) by means of Cramér's phi test.

$$C_{coe} = \left[\frac{\sum_{i=1}^{k} \sum_{j=1}^{l} |p_i - p_j|^2}{(s-1)+(o-1)}\right] \quad (4)$$

Where 'C_{coe}' denotes Correlation coefficient result, 'p_i' refers pixel, 'p_j refers a neighboring pixel, 'e' and 'o' are sample sizes. Result of correlation coefficient provides value from 0 to 1 where '0' indicates no association between the pixels and '1' is absolute association between pixels in Eq. (5).

$$C_{coe} = \begin{cases} 1\,; & segmentROIarea \\ 0; & nosegmenation \end{cases} \quad (5)$$

From (5), if correlation between pixels is higher, segmentation of ROI joint space is performed. Otherwise, process is repeated. Knee x-ray images segmented into several similar ROI regions with minimum time.

3.4 Gradient Tuned Golden Jackal Optimized Deep Vision Transformer (GJODVT)

Joint space segmented area is subjected to the Gradient tuned golden jackal optimized deep vision transformer to perform feature extraction and classification for accurate knee OA severity grading. Vision transformer is worked based on the deep learning model to handle huge volume of images. The deep vision transformer comprises patch embedding, positional embedding, transformer encoder, and classification stages. A vision transformer (ViT) is a transformer developed for computer vision. Vision transformer is worked based on the deep learning model to handle huge volume of images. The deep vision transformer comprises patch embedding, positional embedding, transformer encoder, and classification stages. Knee joint space segmented ROI area is provided as input to the deep vision transformer. Segmented ROI image is divided into number of uniform patches in the patch embedding stage. Then, the positional embedding is performed for spatial arrangement. These patches are forwarded to the multiple layers of the Transformer encoder to extract the appropriate features for knee OA severity classification. Followed by this, Multilayer Perceptron (MLP) is applied to classify the features of knee x-ray images. Lastly, Gradient Tuned Golden Jackal Optimization is applied to tune the hyperparameters of transformer model for improving the accuracy of knee OA severity classification with lesser error.

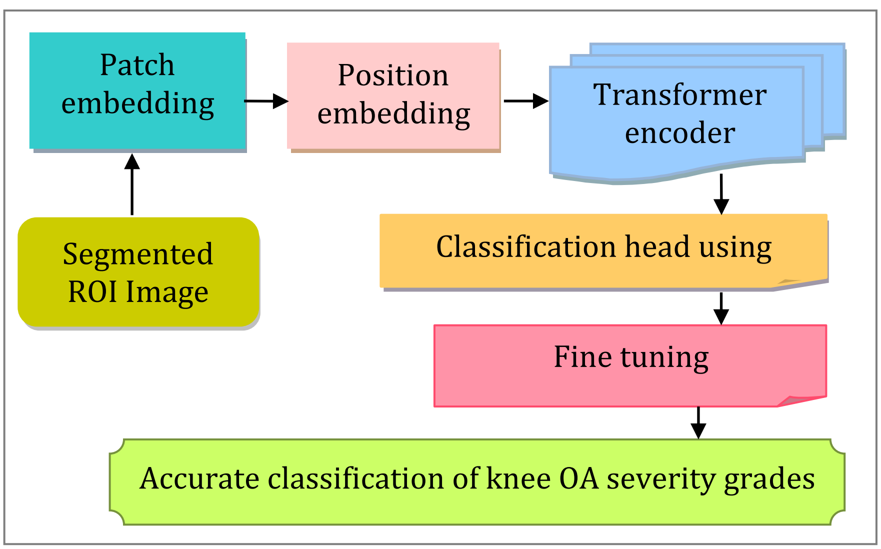


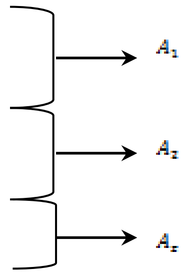
Fig. 4. Structure of gradient tuned golden jackal optimized deep vision transformer model

Fig. 4 demonstrates structure of gradient tuned golden jackal optimized deep vision transformer model to classify knee OA severity levels. In deep vision transformer, patch embedding converts the segmented images into series of smaller and fixed-size patches. These patches flattened, embedded to lower-dimensional feature space and making it suitable for processing. In patch embedding process, ROI image partitioned into number of patches with uniform sizes such as height, weight and color by Eq. (6).

$$SI \in A^{\square * w * c} \quad (6)$$

Where Q^{h*w*c} is image patches with equal height 'h', width 'w' and color (RGB) 'c'. Each patch is referred consistently with similar dimension across the embedding process. Each patch matrix is build to one-dimensional vector. To perform further process in patch embedding, transformation recorded patches into vectorized form using Eq. (7 and 8).

$$A = \begin{bmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{j1} \\ p_{12} \\ \vdots \\ p_{j2} \\ \vdots \\ p_{jk} \end{bmatrix} \quad (7)$$



$$A = [A_1, A_2, A_3, \dots A_z] \quad (8)$$

From (8), 'A' is number of patches of images. Position embedding is performed to encode positional information of input patches. Absolute positional embedding is used to organize spatial correlation of each image patches. It includes sine and cosine functions to encode position of each patch. Sine function is employed for even position and cosine function for odd position. Each position is represented in Eq. (9 and 10) through combination of sine and cosine waves.

$$O \sin\left(\frac{po_z}{10^4}\frac{g}{\left(\frac{d}{4}\right)}\right) \quad (9)$$

$$E = \cos\left(\frac{po_z}{10^4}\frac{g}{\left(\frac{d}{4}\right)}\right) \quad (10)$$

Where, 'O' is position embedding for odd patches 'A', E is position embedding for even patches, 'g' is length of patch, 'z' is total number of patches, d indicates the dimensionality.

Results of position embedding are given to transformer encoder layers. Transformer encoder layers include multiple components such as Self-Attention Mechanism and Position-Wise Feed-Forward Networks. Self-attention mechanism enables model to focus on diverse region of input patches and extracting relevant features while discarding other regions. LSTM is applied to self-attention mechanisms to extract highly relevant or pertinent features from input images. The self-attention mechanism uses Hoover index similarity to estimate the similarity between the pixels within the image patches to prioritize important features.Depended on the similarity, the vital features such as edges, textures, and shapes are extracted. LSTM network includes various components such as forget gate 'Fg_t', input gate 'Ip_t', candidate state 'Cd_t' and output gate 'Op_t'. Input image patches are provided in Eq. (11) to LSTM network at current time step't'.

$$Ip_t = \varphi[(\delta_i . A_t) + b_i] \quad (11)$$

Where 'Ip_t' is input gate, 'φ' is softmax activation function, 'δ_i' is weights 'A_t' current input image patches and 'b_i' is bias. Results of 'δ_i' is varies between 0 and 1. Feature extraction is performed in forget gate to keep relevant information. Hoover index similarity calculates similarity between pixels in Eq. (12) within image patches.

$$H_r = \frac{1}{2}\frac{\sum_i |p_i - p_j|}{\sum_i p_i} \quad (12)$$

From (12), 'H_r' is Hoover index similarity coefficient between pixels p_i andp_j. Pixels with high similarity are retained while others are discarded. Based on similar pixels in patches, significant features such as edges, textures, and shapes are extracted. Texture feature 'TX' is extracted for identifying regularity, smoothness and coarseness of pixel intensities in an image by Eq. (13).

$$T_{fr} = \frac{\sum_i \sum_j (p_i - \mu_i)(p_j - \mu_j)}{\sigma_i \sigma_j} \quad (13)$$

From (13), μ_i and $\mu_j$are mean of pixelsp_i, p_j, σ_i and σ_j denotes deviation of pixels. Distance is computed to determine object boundary from the origin in Eq. (4).

$$Distance = \sqrt{(u_2 - u_1)^2 - (v_2 - v_1)^2} \quad (14)$$

Where, point '(u_1,v_1)' is origin. '(u_2,v_2)' is object boundary. Shape of object is identified. Edge smoothing concept using Difference of Gaussian (DoG) is to identify regions (like joints). Image patches are partitioned into similar pixels based on spatial coordinates '(u,v)' at particular resolution. Edge detection is performed in Eq. (15).

$$IA_{edge} = G_{\sigma 1}(u,v) - G_{\sigma 2}(u,v) \quad (15)$$

From (15), 'IA_{edge}' is edge detection output, 'σ' is deviation, 'u' is distance from origin in horizontal axis, 'v' refers to distance from origin in perpendicular axis. Image quality is improved. Output gate provides extracted feature vector by Eq. (16).

$$Op_t = \varphi[(\delta_o.Ef_t) + b_o] \quad (16)$$

Where, Op_t is output Gate, φ is sigmoid activation function, $\beta_o$is weights, Ef_t extracted feature vector, and b_o is bias at output gate. Results extracted features are given to feed-forward layer and then to fully connected layers to classify various types of knee OA severity grades. Fully connected layers obtain information from the transformer encoder's outputs and learn higher-level representations to carry out knee OA severity classification. Fully connected (FC) layers in a deep learning consist of an input layer, one or more hidden layers, and an output layer. These layers are used to interpret flattened feature maps to make final classification predictions for knee osteoarthritis (OA) severity classification. Kernelized Multilayer Perceptron (MLP) networks are implemented for more accurate Knee OA severity classification. The input layer is used to extract features from the image patches as input. Input is broadcasted to the hidden layer to classify the images into different severity grades. In the hidden layer, similarity between testing and training features are investigated during the Gaussian kernel function.Kernel output offers value ranges from 0 to 1. The higher similarity between the features denotes that the x-ray knee image is categorized into certain severity grade.

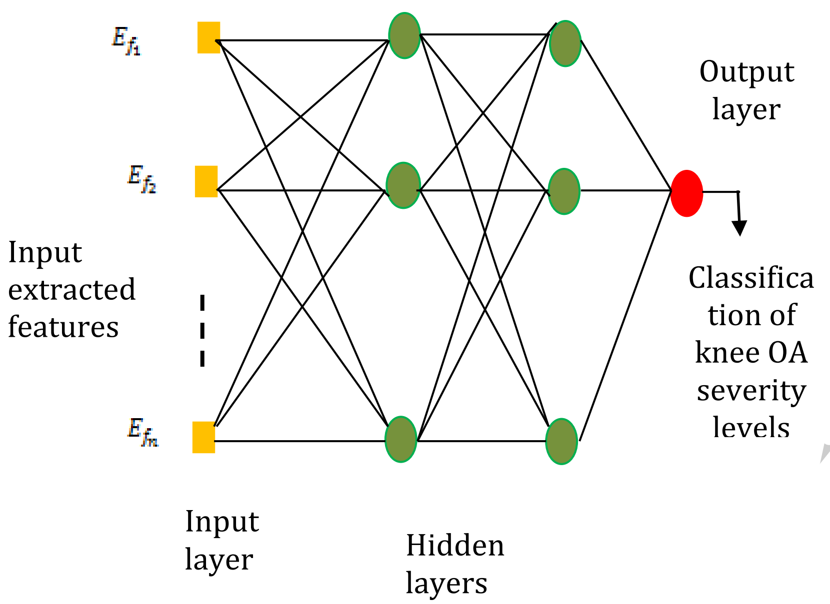


Fig. 5. Structure of MLP

Fig. 5 shows schematic structure of MLP. Input layer gets extracted features from image patches as input. Features in input layer associated in Eq. (17) with weight $w$and bias β.

$$\alpha_l = \sum_{l=1}^{k}(Ef_n * w_{ih}) + \beta \quad (17)$$

From (17), 'α_l' is activity of neuron at input layer of MLP network, $w_{ih}$is weights between input and hidden layer, input image patches 'Ef', 'β' is bias that stored value is '1. Input is transmitted to hidden layer for classifying images into different severity grades. In that layer, similarity between testing and training features are examined through Gaussian kernel. Kernel function identifies class of training feature value whose similarity is closest to testing feature value for knee OA severity classification in Eq. (18).

$$G_{kernel} = exp\left(-\frac{|Ef_t - Ef_r|^2}{2\sigma^2}\right) \quad (18)$$

From (18) 'G_{kernel}' is Gaussian kernel function, Ef_r is training feature (extracted feature), Ef_t is testing feature, σ is deviation. Kernel output provides ranges from 0 to 1. Higher similarity between features represents that x-ray knee image is classified into certain severity grade. After classification, Error rate 'ε' is computed in Eq. (19).

$$\varepsilon = (Y^{act} - Y^{pre})^2 \quad (19)$$

From (19), 'Y^{act}' and 'Y^{pre}' is actual and predicted knee OA severity grading Weights between layers are gets updated to minimize error rate in Eq. (20) using stochastic gradient function.

$$w_{(t+1)} = w_t - \gamma\left[\frac{\partial\varepsilon}{\partial w_t}\right] \quad (20)$$

Where, $w_{(t+1)}$ is updated weight, w is current weight, γ is learning rate, $\left[\frac{\partial\varepsilon}{\partial w_t}\right]$ is first-order derivative function regarding error 'ε' and weight 'w'. Upon updating weights, optimal weight value is determined using Golden Jackal Optimization to get accurate severity grading results. Traditional gradient-based optimizations methods namely SGD and Adam are easily get stuck in local optima and not balancing the exploration and exploitation and failed to provide global capabilities. Contrary to existing SGD and Adam, Golden Jackal Optimization is employed to faster convergence and higher accuracy in finding the optimal weights for specific tasks such as medical image analysis and severity grading. It offers robust global search capabilities. Golden Jackal Optimization algorithm is utilized in deep learning for severity grading of Knee OA. It is a swarm-based approach that mimics the hunting activity of the golden jackal in its nature. It updates the network weights through treating each solution as a jackal and searching optimal weight that reduces the error rate in the classification process. It has metaheuristic optimizer than conventional gradient-based methods such as Adam and SGD in certain complex scenarios. In optimization, initial candidate solution of jackals (multiple weights 'w_i') arbitrarily generated in search region by Eq. (21 and 22).

$$w_i = w_1, w_2, w_3, \dots, w_i \quad (21)$$

Fitness '$\mathrm{Fit}(w_i)$' is calculated.

$$\mathrm{Fit}(w_i) = \arg\min \varepsilon \quad (22)$$

From (22), $arg\ min$ denotes argument minimal function. Based on fitness, current two best jackals (male and female jackal) selected (lowest and second lowest error value). Exploration and exploitation performed to update weight value to get global best solution. In exploration step, jackals identify and track their prey. Jackals wait and hunt for new prey (weights) using Eq. (23 and 24).

$$G_1(t) = G_M(t) - E|G_M(t) - rl * prey\ (t)| \quad (23)$$

$$G_2(t) = G_{FM}(t) - E|G_{FM}(t) - rl * prey\ (t)| \quad (24)$$

Where $prey\ (t)$ refers current weight, $G_M(t)$ and $G_{FM}(t)$ refers male and female jackal position in search space, rl refers arbitrary random vector, E refers escape energy of prey. In exploitation step, evading capability of prey is diminished when it is harassed by jackals. Jackal pairs encircle prey detected in prior phase. Upon encircling, jackals close to their prey and refine weights in Eq. (25 and 26).

$$G_1(t) = G_M(t) - E|rl\ .\ G_M(t) - prey\ (t)| \quad (25)$$

$$G_2(t) = G_{FM}(t) - E|rl\ .\ G_{FM}(t) - prey\ (t)| \quad (26)$$

Where, 'rl' is random vectors in exploitation step. This ignores local optimum. New update position of golden jackal computed in Eq. (27).

$$G(t+1) = \frac{G_1(t) + G_2(t)}{2} \quad (27)$$

From (27), '$G(t+1)$' is newly updated position. This process is repetitive until attaining maximum iteration. Optimal weight is chosen to decrease error in output layer. Accurate classification of knee OA severity classification is achieved.

//Algorithm 2: Tetrachoric correlated AlexNet based Golden Jackal Optimized Deep Vision Transformer
Input:Segmented ROI joint space images **Output:** Accurate knee OA severity grade classification
Begin **1:** Collect segmented ROI image as input **2**: **For each**image segmented ROI **3:** Execute patch embedding using (6) **4**: Perform position embedding using (9) & (10) **5**: **End for** **6**: Transform position embedding results into transformer layers **7: For each** patch A_z **8:** Compute hoover index using (12) **9**: Extract the features from patches using (13), (14) & (15) **11**: **End for** **12:** Get extracted features as input in MLP input layer **13:** Calculate Gaussian kernel function using (18) **14: If** $G_{kernel} = 1$ **then** **15**: classify the images into particular severity grade **16**: **Else** **17:** Images are not classified **14. End if** **15.** Obtain final classification results **16: For each classificationresult** **17:** Determine the error rate ε **18:** Update the weight **19:** Initialize the population ofweights **20: For** each w_i **21:** Calculate the fitness **22:** Update the position of jackal **23:** Determine optimal weight **24:** Obtain final classification results **End**

4. Experimental setup

Proposed TA-GJODVT is conducted in Python 3.10.3 with keras and tensorflow libraries by Knee Osteoarthritis Dataset with Severity Grading dataset. The dataset is divided into two sets such as training and testing. Training set contains 5778 images and testing set contains 2047 images. Table 2, 3, 4 and 5 describes the proposed TA-GJODVT hyperparameter.

Table 2 Hyperparameters for Tetrachoric correlated AlexNet model

Parameter	**Value**
Input image size	224 × 224 × 3
Convolutional layers Filters	256

Convolutional layers Kernel Size	5 × 5
Max Pooling layer Size	3 × 3
Fully Connected layer	4096 neurons
Dropout rate	0.5

Table 3 Hyperparameters for deep vision transformer

Parameter	Value
Patch Dimension	16 × 16
Embedding Size	768
Encoder Blocks in Transformer Layers	12
Multi-Head Attention Heads	12
Multi-Head Dimension	64
MLP Hidden Layer Dimension	3072
Dropout Rate	0.1
Classification Head Output Classes	5 (KL Grades 0–4)
Activation Function	Softmax

Table 4 Hyperparameters for Golden Jackal Optimization

Parameter	Value
Population Size	50
Maximum Iterations	50–100
Optimized Variables	Learning rate, batch size, dropout rate, number of transformer layers, attention heads
Fitness Function	Validation accuracy / F1-score
Exploration–Exploitation Strategy	Adaptive coefficient
Termination Condition	Max iterations or stagnation (10 iterations)

Table 5 Hyperparameters for training process

Parameter	Value
Optimizer	Adam / AdamW
Initial Learning Rate	0.0004
Batch Size	16–32 (GJO-optimized)
Loss Function	Categorical cross-entropy
Epochs	50
Learning Rate Scheduler	Cosine annealing
Early Stopping	Patience = 10 epochs
Framework	Python (TensorFlow / PyTorch)
Hardware	CPU

5. Results and discussion

The Results of TA-GJODVT and [1], [2], [3], [4] and [5] are compared using different metrics. Accuracy 'ACC' is proportion of accurately classified knee x-ray images in Eq. (28).

$$ACC = \left[\frac{Tru_{pos}+Tru_{neg}}{Tru_{pos}+Tru_{neg}+Fal_{pos}+Fal_{neg}}\right] * 100 \quad (28)$$

From (28), 'Tru_{pos}' is true positive, 'Fal_{pos}' is false positive, 'Tru_{neg}' is true negative, 'Fal_{neg}' is false negative.

Precision'Pre' is referred as ratio of relative amount of true positives in Eq. (29).

$$Pre = \frac{Tru_{pos}}{Tru_{pos}+Fal_{pos}} * 100 \quad (29)$$

Recall 'Rec' correctly predict all images in Eq. (30).

$$Rec = \frac{Tru_{pos}}{Tru_{pos}+Fal_{neg}} * 100 \quad (30)$$

Specificity'Spe' accurately differentiate between healthy and severity images in Eq. (31).

$$Spe = \frac{Tru_{neg}}{Tru_{neg}+Fal_{pos}} * 100 \quad (31)$$

F1-score ($\mathrm{F1-S}$) is estimated as average of precision and recall in Eq. (32).

$$\mathrm{F1-S} = \left[2 * \frac{Pre * Rec}{Pre + Rec}\right] * 100 \quad (32)$$

Error rate 'Err'is ratio of incorrect predictions done by Eq. (33),

$$Err = \left(\frac{Fal_{pos}+Fal_{neg}}{Tru_{pos}+Tru_{neg}+Fal_{pos}+Fal_{neg}}\right) * 100 \quad (33)$$

Mathew Correlation Co-efficient(MCC) is measured based on TP, TN, FP and FN in Eq. (34).

$$MCC = \frac{Tru_{pos}-(Fal_{pos}*Fal_{neg})}{\sqrt{(Tru_{pos}+Fal_{pos})(Tru_{pos}+Fal_{neg})(Tru_{neg}+Fal_{pos})(Tru_{neg}+Fal_{neg})}} \quad (34)$$

Classification Time 'CT' is measured as time used to classify knee OA severity in Eq. (35).

$CT = \sum_{i=1}^{n} EI_i * T[KOA]$ (35) Where, $T\,[KOA]$ is time for knee OA severity detection of single image I_i.

5.1 Analysis of Deep learning algorithms

Results of different DL models are validated using training set images to classify knee images into five diverse severity grades.

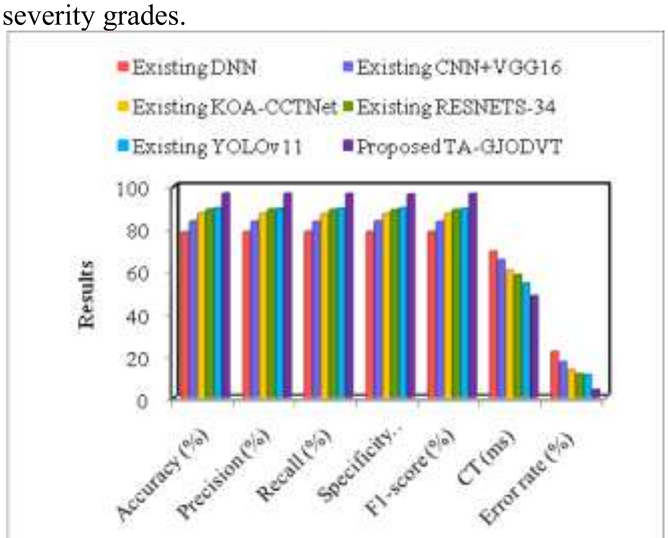


Fig. 6. Results of detecting severity grade-0 using Training images

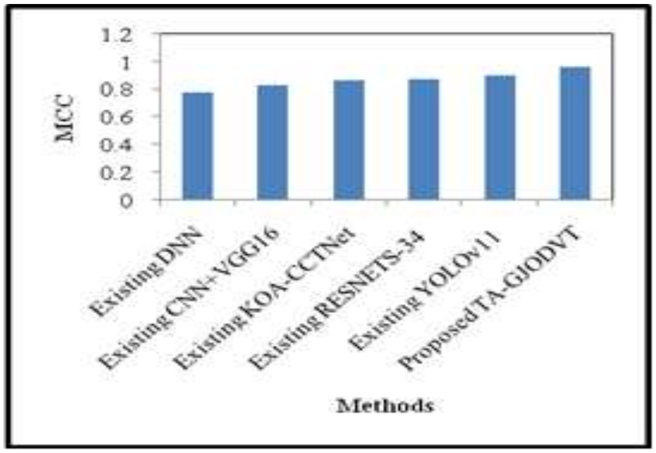


Fig. 7. MCC results of detecting severity grade-0 using Training images

Fig. 6 and 7 illustrate comparison of TA-GJODVT with various existing DL models. TA-GJODVT attains accuracy of 96.5%, precision of 96.4% and recall of 96.3% and specificity of 96.1%, F1-score of 96.3%, time used as 48ms, MCC is 0.96. The main reason behind this improvement is to use deep vision transformer in proposed TA-GJODVT model. With this, more pertinent features are extracted and the classified using MLP network. Moreover, the fine-tuning process using optimization helps to enhance the overall performance of grading.

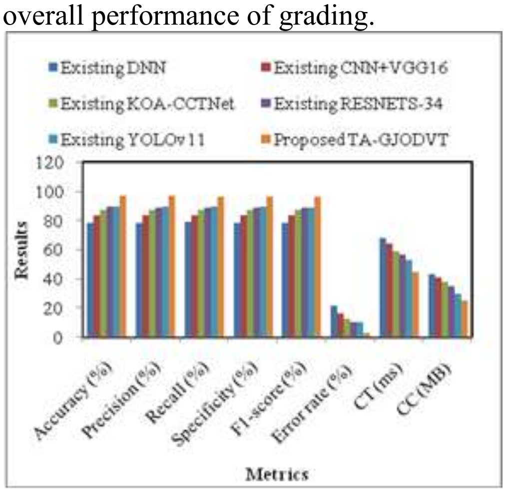


Fig. 8. Results of detecting severity grade-1 using training images

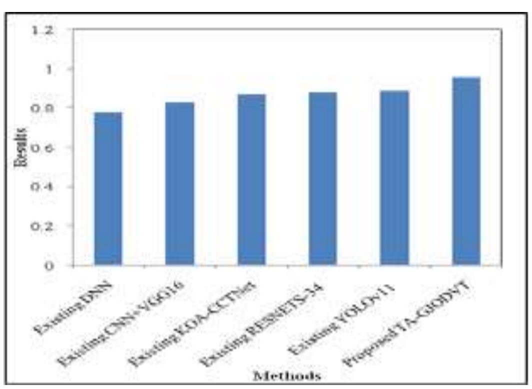


Fig. 9. MCC results of detecting severity grade-1 using Training images

Fig. 8 and 9 illustrates results of identifying severity grade-1. TA-GJODVT provides higher accuracy of 96.7%, 96.6% of precision, 96.5% recall, 96.3% of specificity, F1-score of 96.5%. MCC is 0.96. This demonstrates the superior effectiveness and efficiency of the proposed TA-GJODVT model for early knee OA severity grading (particularly grade 1). Proposed deep vision transformer uses LSTMs with self-attention layers that efficiently extract the robust features for knee OA detection. In addition, application of kernelized MLP provides better classification performance for categorizing the severity levels with higher accuracy and lesser error.

Fig. 10. Results of detecting severity grade-2 using Training images

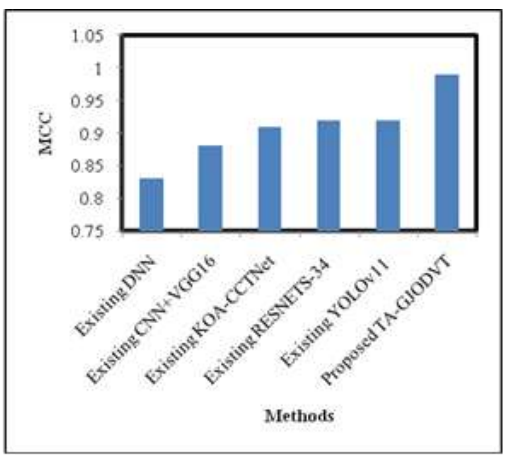


Fig. 11. MCC results of detecting severity grade-2 using Training images

Fig. 10 and 11 demonstrates detection of knee OA severity grade-2. TA-GJODVT achieves accuracy as 96.6%, precision as 96.4%, recall as 96.4%, and specificity as 99.2%. It records F1-score as 99.4% ,and the MCC is 0.97. On the contrary to existing models, TA-GJODVT model uses the Optimized Deep vision transformer to perform feature extraction and classification. Moreover, the model incorporates the stochastic gradient function for weight updates. The Golden Jackal Optimization is designed to detect the optimal weight by means of reducing classification errors and improves performance of knee OA severity classification.

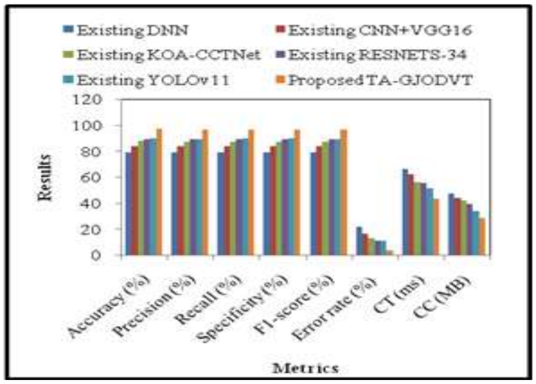


Fig. 12. Results of detecting severity grade-3 using Training images

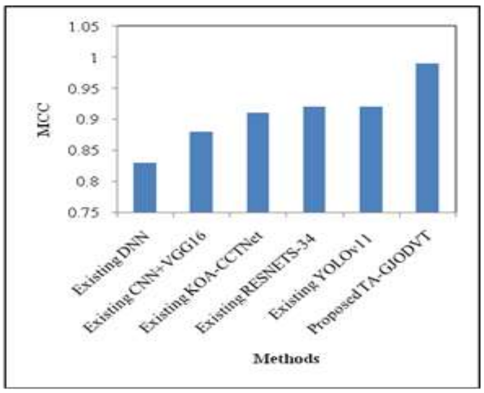


Fig. 13. MCC results of detecting severity grade-3 using Training images

Fig. 12 and 13 demonstrates identifying severity grade-3. TA-GJODVT achieves higher accuracy, precision and recall, Specificity, F1-score as 96.8%, 96.7% and 96.6%, 96.3%, 96.6% . MCC is 0.97 and time is 43ms measured.

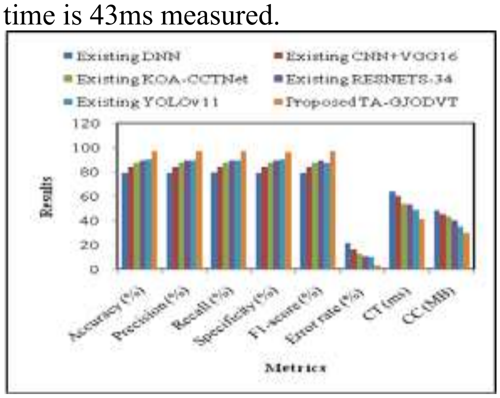


Fig. 14. Results of detecting severity grade-4 using Training images

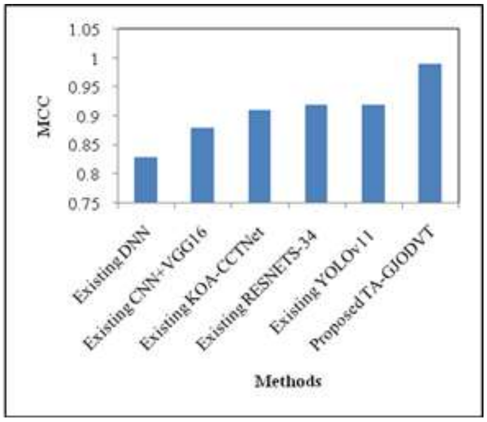


Fig. 15. Results of detecting severity grade-4 using Training images

Fig. 14 and 15 demonstrates result of classifying severity grade-4 of knee osteoarthritis. TA-GJODVT achieved higher accuracy, precision, recall, specificity, F1-score, error rate, CT, and MCC and CC as 96.9%, 96.8%, 96.7%, 96.5%, 96.7%, 3.1%, 41ms. Results of DL analyzed using testing images for detecting severity grades.

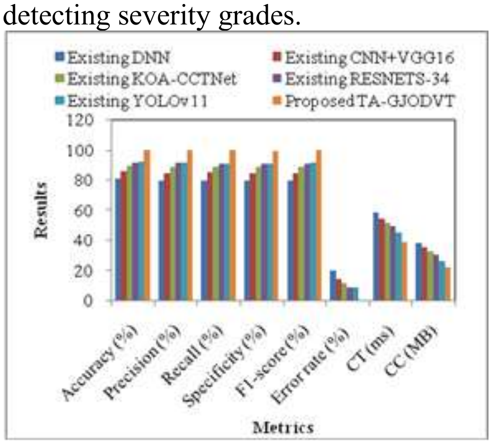


Fig. 16. Results of detecting severity grade-0 using testing images

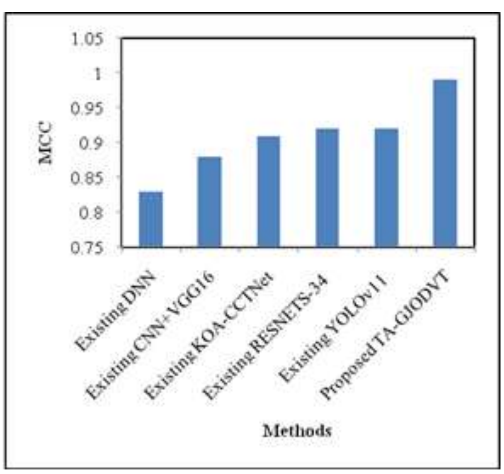


Fig. 17. Results of detecting severity grade-0 using testing images

Fig. 16 and 17 demonstrates knee-severity grade '0'. TA-GJODVT achieved higher accuracy, precision, recall, specificity, F1-score, error rate, CT and MCC and CC as 99.7%, 99.5%, 99.4%, 99.1%, 99.4%, 0.3%, 39ms.

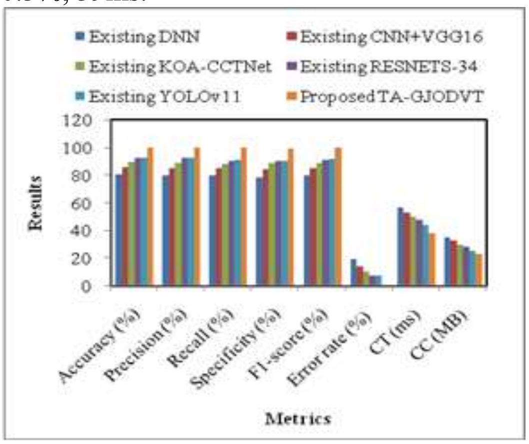


Fig. 18. Results of detecting severity grade-1 using testing images

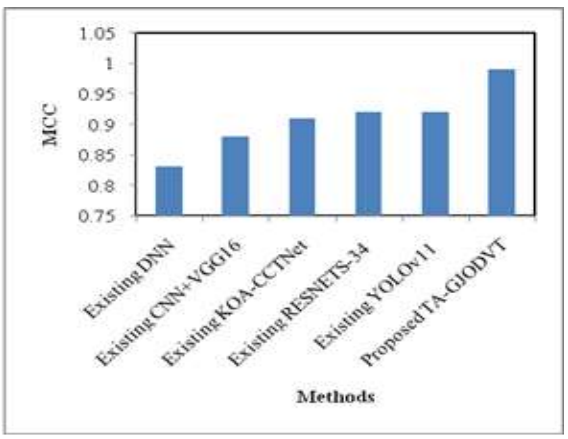


Fig. 19. Results of detecting severity grade-1 using testing images

Fig. 18 and 19 shows result of detecting severity grade-1. TA-GJODVT attained higher accuracy, precision, recall, specificity, F1-score, MCC and CC with lesser error and time as 99.8%, 99.6%, 99.5%, 99.3%, 99.5%, 0.99, 0.2%, and 38 ms.

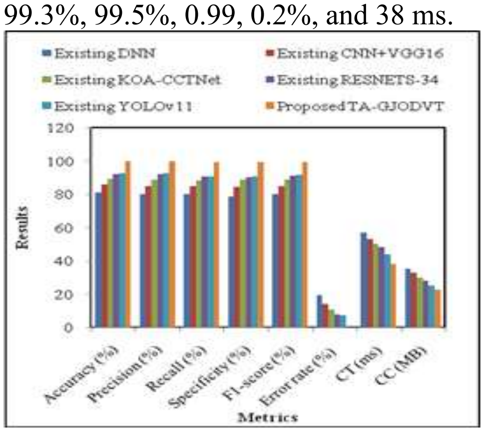


Fig. 20. Results of detecting severity grade-2 using testing images

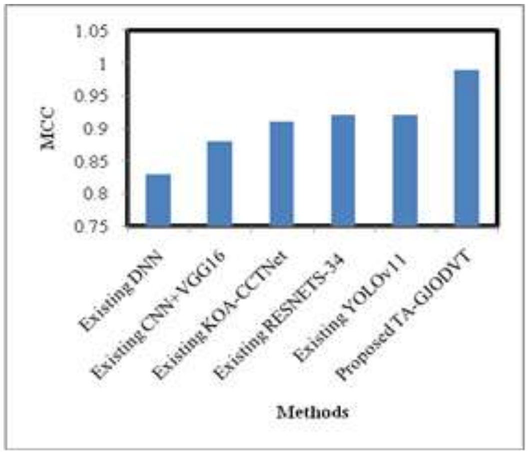


Fig . 21. MCC results of detecting severity grade-2 using training images

Fig. 20 and 21 provides severity grade-2 of knee osteoarthritis using testing images. TA-GJODVT higher accuracy, precision, recall, specificity and F1-score as 99.7%, 99.5%, 99.4%, 99.1% and 99.3% and lowest error and time and CC as 0.3% and 40ms.

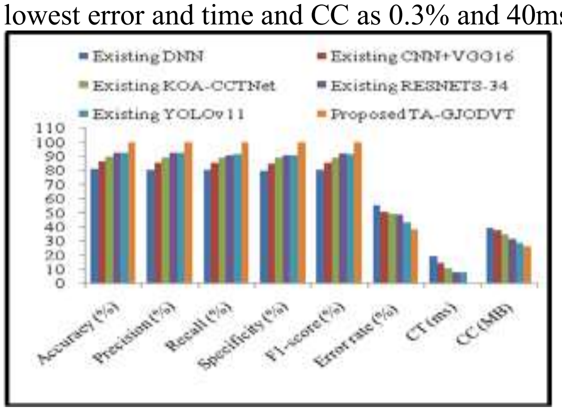


Fig. 22. Results of detecting severity grade-3 using testing images

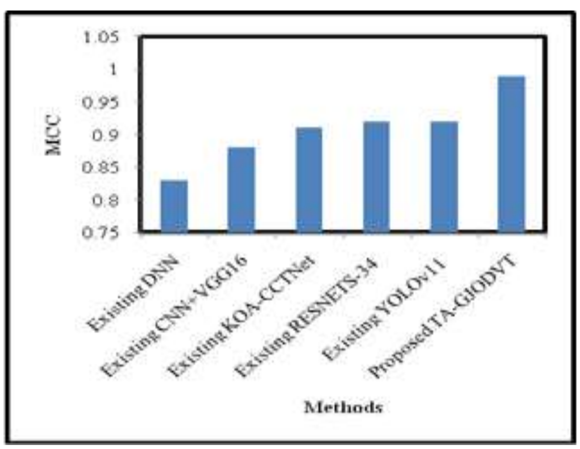


Fig. 23. MCC results of detecting severity grade-3 using training images

Fig. 22 and 23 shows result of detecting severity grade-3 of knee osteoarthritis. TA-GJODVT higher accuracy, precision, recall, specificity, MCC, F1-score with lesser error rate, and time and CC as 99.8%, 99.7%, 99.6%, 99.4%, 0.99, 99.6%, 0.2%, and 39ms.

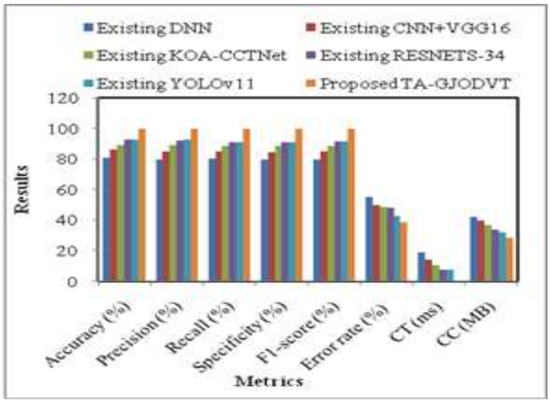


Fig. 24. Results of detecting severity grade-4 using testing images

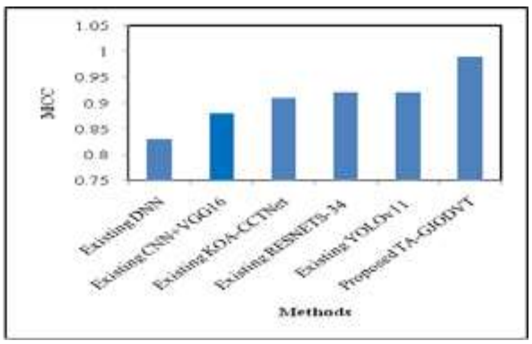


Fig. 25. MCC results of detecting severity grade-4 using testing images

Fig. 24 and 25 shows results of detecting severity grade-4.TA-GJODVT achieves higher accuracy, precision, recall, specificity, F1-score, CT and MCC as 99.9%, 99.8%, 99.7%, 99.6%, 99.7%, 38ms, 0.1 % and 0.99 and .

5.2 Ablation study

The ablation study is carried out in Tetrachoric correlated AlexNet based Golden Jackal Optimized Deep Vision Transformer to find out the contribution of preprocessing segmentation, feature extraction, and classification and fine-tuning implemented into AlexNet, deep vision transformer to enhance overall performance.

Table 6 Ablation Study for proposed TA-GJODVT to detect severity grades

Methods/ Metrics	Accuracy (%)				
	Grade-0	Grade-1	Grade-2	Grade-3	Grade-4
Existing DNN	79.2	80.1	80.3	80.5	80.7
Existing CNN+VGG16	84.3	85.1	85.3	85.5	85.7
Existing KOA-CCTNet	88.4	88.6	88.6	88.9	89.1
Existing RESNETS-34	90	90.6	90.9	91.1	91.8
Existing YOLOv11	90.3	91.2	91.4	91.6	92.1
Proposed Tetrachoric correlation+ AlexNet+Golden Jackal Optimization+ Deep Vision	98.5	99.1	99.2	99.4	99.7

Transformer					

Table 6 shows the performance of three methods Tetrachoric correlation+ AlexNet+Golden Jackal Optimization+ Deep Vision Transformer for Knee Osteoarthritis Dataset.The proposed TA-GJODVT method attained higher accuracy of 98.5%, 99.1%, 99.2%, 99.4% and 99.7% for severity grades 0, 1, 2, 3, and 4.

5.3Anova test

ANOVA (Analysis of Variance) is a statistical test employed to assess whether there is a significant difference between two or more groups within a dataset. In the context of knee images classified in medical, ANOVA is applied to assess whether the mean values of a dependent variable (accuracy) differ significantly across different groups defined by a categorical variable (severity grade class). In our work, theone-sample hypothesis test is used to conduct the experiments for proposed TA-GJODVT method.

Table 7 ANOVA test results for accuracy with grade-wise

Severity Grade	**Accuracy (%)**
Grade 0	96.5
Grade 1	96.7
Grade 2	96.6
Grade 3	96.8
Grade 4	96.9

Table 8 ANOVA test results for accuracy with grade-wise

Metric	**Value**
Mean Accuracy (%)	96.70
Standard deviation	0.063
95% Confidence interval lower	96.58
95% Confidence interval upper	96.82
p-value	0.034

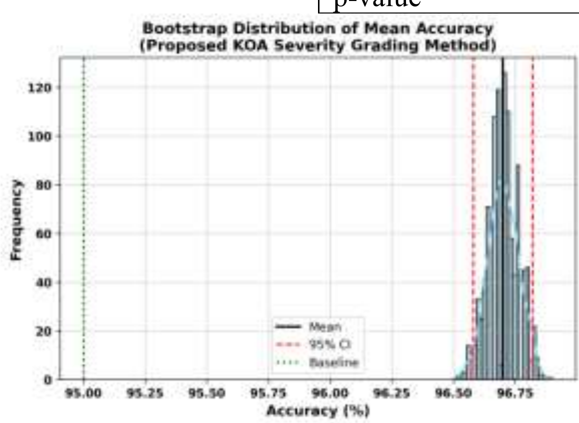


Fig. 26. ANOVA test results graph for accuracy with grade-wise

Table 7, 8 and Fig. 26 present the results of the ANOVA statistical test. The x-axis denoted as accuracy for proposed TA-GJODVT method and y-axis represented as frequency. From the results, the p-value is obtained as 0.034. If the p-value is lower than the significance level (0.05), it indicates that there is a statistically significant difference between the five classes.

6. CONCLUSION

Proposed TA-GJODVT performs automated classification of knee OA severity. TA-GJODVT carry outs image preprocessing, extraction and segmentation and classification. To achieve accurate classification results, fine-tuning performed using gradient tuned golden jackal optimization algorithm. Accurate detection of knee OA severity grades is achieved with higher accuracy and lesser time. The limitation of proposed method is considered only one dataset. In future work, the proposed method is extended to apply novel deep learning and optimization method for classification of knee OA severity. In addition, the two or more datasets will be considered to enhance the deep learning performance.

REFERENCES

[1] L. Chen, Z. Su, Novel attention-enhanced multi-task deep learning for knee osteoarthritis (KOA) grading and localization in X-ray imaging of basketball players, Journal of Radiation Research and Applied Sciences. 18 (2025) 1–12. https://doi.org/10.1016/j.jrras.2025.101442
[2] S. Olsson, E. Akbarian, A. Lind, A.S. Razavian, M. Gordon, Automating classification of osteoarthritis according to Kellgren-Lawrence in the knee using deep learning in an unfiltered adult population, BMC Musculoskeletal Disorders. 22 (2021) 1–8. https://doi.org/10.1186/s12891-021-04722-7
[3] M. Jahan, Z. Md. Hasan, I.J. Samia, K. Fatema, A.H. Md. Rony, M.S. Arefin, A. Moustafa, KOA-CCTNet: An enhanced knee osteoarthritis grade assessment framework using modified compact convolutional transformer model, IEEE Access. 12 (2024) 107719–107741. https://doi.org/10.1109/ACCESS.2024.3435572
[4] S. Rehman, V. Gruhn, A sequential VGG16+CNN-based automated approach with adaptive input for efficient detection of knee osteoarthritis stages, IEEE Access. (2024) 62407–62415. https://doi.org/10.1109/ACCESS.2024.3395062
[5] S. Moustakidis, N.I. Papandrianos, E. Christodolou, E. Papageorgiou, D. Tsaopoulos, Dense neural networks in knee osteoarthritis classification: A study on accuracy and fairness, Neural Computing and Applications. 35 (2023) 21–33. https://doi.org/10.1007/s00521-020-05459-5
[6] T.B. Nguyen-Tat, T.P. Nguyen-Duong, Optimizing knee osteoarthritis severity diagnostics: A GA-enhanced deep ensemble approach in medical imaging, Alexandria Engineering Journal. 16 (2025) 1–15. https://doi.org/10.1016/j.asej.2025.103524
[7] X. Wang, T. Wang, Z. Su, Attention-enhanced deep learning and machine learning framework for knee osteoarthritis severity detection in football players using X-ray images, Journal of Radiation Research and Applied Sciences. 18 (2025) 1–12. https://doi.org/10.1016/j.jrras.2025.101428
[8] D. Preethi, V. Govindaraj, S. Dhanasekar, K.M. Sagayam, S.I. Ansarullah, F. Amin, I.D. Díez, C.O. García, A.P.E. Barrera, F.S. Alshammari, Deep learning-assisted 3D model for the detection and classification of knee arthritis, Image and Vision Computing. 160 (2025) 1–10. https://doi.org/10.1016/j.imavis.2025.105574
[9] A.R. Patil, S.S. Salunkhe, Classification and risk estimation of osteoarthritis using deep learning methods, Measurement: Sensors. 35 (2024) 1–8. https://doi.org/10.1016/j.measen.2024.101279
[10] S.M. Naguib, M.A. Kassem, H.M. Hamza, M.M. Fouda, M.K. Saleh, K.M. Hosny, Automated system for classifying uni-bicompartmental knee osteoarthritis by using redefined residual learning with convolutional neural network, Heliyon. 10 (2024) 1–10. https://doi.org/10.1016/j.heliyon.2024.e31017
[11] M.S. Islam, M.T.A. Rony, CDK: A novel high-performance transfer feature technique for early detection of osteoarthritis, Journal of Pathology Informatics. 15 (2024) 1–15. https://doi.org/10.1016/j.jpi.2024.100382
[12] R. Ahmed, A.S. Imran, Knee osteoarthritis analysis using deep learning and XAI on X-rays, IEEE Access. 12 (2024) 68870–68879. https://doi.org/10.1109/ACCESS.2024.3400987
[13] T. Tariq, Z. Suhail, Z. Nawaz, Knee osteoarthritis detection and classification using X-rays, IEEE Access. 11 (2023) 48292–48303. https://doi.org/10.1109/ACCESS.2023.3276810

[14] H. Zhao, L. Ou, Z. Zhang, L. Zhang, K. Liu, J. Kuang, The value of deep learning-based X-ray techniques in detecting and classifying K-L grades of knee osteoarthritis: A systematic review and meta-analysis, European Radiology. 35 (2025) 327–340. https://doi.org/10.1007/s00330-024-10928-9
[15] S. Kinger, Deep learning for automatic knee osteoarthritis severity grading and classification, Indian Journal of Orthopaedics. 58 (2025) 1458–1473. https://doi.org/10.1007/s43465-024-01259-4
[16] S.A. El-Ghany, M. Elmogy, A. El-Aziz, A fully automatic fine-tuned deep learning model for knee osteoarthritis detection and progression analysis, Egyptian Informatics Journal. 24 (2023) 229–240. https://doi.org/10.1016/j.eij.2023.03.005
[17] K. Fatema, A.H. Md Rony, S. Azam, S.H. Md Mukta, A. Karim, Z. Md Hasan, M. Jonkman, Development of an automated optimal distance feature-based decision system for diagnosing knee osteoarthritis using segmented X-ray images, Heliyon. 9 (2023) 1–27. https://doi.org/10.1016/j.heliyon.2023.e21703
[18] J. Wang, T.A.G. Hall, O. Musbahi, G.G. Jones, R.J. Arkel, Predicting hip-knee-ankle and femorotibial angles from knee radiographs with deep learning, The Knee. 42 (2023) 281–288. https://doi.org/10.1016/j.knee.2023.03.010
[19] L. Liu, J. Chang, P. Zhang, Q. Ma, H. Zhang, T. Sun, H. Qiao, A joint multi-modal learning method for early-stage knee osteoarthritis disease classification, Heliyon. 9 (2023) 1–14. https://doi.org/10.1016/j.heliyon.2023.e15461
[20] S.Y. Malathi, G.R. Bharamagoudar, S.K. Shiragudikar, Diagnosing and grading knee osteoarthritis from X-ray images using deep neural angular extreme learning machine, Proceedings of the Indian National Science Academy. 91 (2025) 95–108. https://doi.org/10.1007/s43538-024-00366-y
[21] S.W. Pi, B.D. Lee, M.S. Lee, H.J. Lee, Ensemble deep learning networks for automated osteoarthritis grading in knee X-ray images, Scientific Reports. 13 (2023) 1–17. https://doi.org/10.1038/s41598-023-50210-4
[22] S. Pandey, “Adaptive digital twin framework for real-time asset monitoring and predictive maintenance,” International Journal of Digital Twin Systems and Computing, vol. 1, no. 1, pp. 15–22, 2025.
[23] D.W. Lee, D.S. Song, H.S. Han, D.H. Ro, Accurate automated classification of radiographic knee osteoarthritis severity using a novel method of deep learning: Plug-in modules, Knee Surgery & Related Research. 36 (2024) 1–11. https://doi.org/10.1186/s43019-024-00228-3
[24] B. Subha, V. Jeyakumar, S.N. Deepa, Gaussian Aquila optimizer-based dual convolutional neural networks for identification and grading of osteoarthritis using knee joint images, Scientific Reports. 14 (2024) 1–23. https://doi.org/10.1038/s41598-024-57002-4
[25] A.G. Diab, E.S.M. El-Kenawy, N.F.F. Areed, H.M. Amer, M. El-Seddek, A metaheuristic optimization-based approach for accurate prediction and classification of knee osteoarthritis, Scientific Reports. 15 (2025) 1–28. https://doi.org/10.1038/s41598-025-99460-4
[26] A.M. Sarhan, M. Gobara, S. Yasser, Z. Elsayed, G. Sherif, N. Moataz, Y. Yasir, E. Moustafa, S. Ibrahim, H.A. Ali, Knee osteoporosis diagnosis based on deep learning, International Journal of Computational Intelligence Systems. 17 (2024) 1–33. https://doi.org/10.1007/s44196-024-00615-4
[27] W. Li, J. Liu, Z. Xiao, D. Zhu, J. Liao, W. Yu, J. Feng, B. Qian, Y. Fang, S. Li, Automatic grading of knee osteoarthritis with a plain radiograph radiomics model combining anteroposterior and lateral images, Insights into Imaging. 15 (2024) 1–12. https://doi.org/10.1186/s13244-024-01719-3
[28] D. Zhang, Y. Dong, Q. Xu, Y. Qian, M. Ye, Q. Yuan, J. Luo, Enhancing knee osteoarthritis diagnosis with DMS: A novel dense multi-scale convolutional neural network approach, Journal of Orthopaedic Surgery and Research. 19 (2024) 1–10. https://doi.org/10.1186/s13018-024-05352-0
[29] A. Haseeb, M.A. Khan, F. Shehzad, M. Alhaisoni, J.A. Khan, T. Kim, J.H. Cha, Knee osteoarthritis classification using X-ray images based on optimal deep neural network, Computer Systems Science and Engineering. 47 (2023) 1–19. https://doi.org/10.32604/csse.2023.040529
[30] S. Aladhadh, R. Mahum, Knee osteoarthritis detection using an improved CenterNet with pixel-wise voting scheme, IEEE Access. 11 (2023) 22283–22296. https://doi.org/10.1109/ACCESS.2023.3247502

[31] D. Tikariha, A. Moomin, D. Jeyamani, P. Rukmani, Osteoarthritis classification using hybrid quantum convolutional neural network, IEEE Access. 13 (2025) 145060–145070. https://doi.org/10.1109/ACCESS.2025.3599679
[32] J. Pan, Y. Wu, Z. Tang, K. Sun, M. Li, J. Sun, J. Liu, J. Tian, B. Shen, Automatic knee osteoarthritis severity grading based on X-ray images using a hierarchical classification method, Arthritis Research & Therapy. 26 (2024) 1–14. https://doi.org/10.1186/s13075-024-03416-4
[33] N. Sharma, R. Sapra, S. Gulia, P. Dhaliwal, Osteo-fusion: A multimodal decision-chaining approach for automated knee osteoarthritis detection and severity classification, Intelligence-Based Medicine. 12 (2025) 1–9. https://doi.org/10.1016/j.ibmed.2025.100268

Nomenclature

Abbreviations	Definition
Knee OA	Knee Osteoarthritis
KL	Kellgren-Lawrence
DL	Deep Learning
TA-GJODVT	Tetrachoric correlated AlexNet based Golden Jackal Optimized Deep Vision Transformer
RoI	Region of Interest
YOLOv11	You Only Look Once version 11
KOA-CCTNet	KOA grade assessment through the improved Compact Convolutional Transformer
CNN+VGG16	Hybrid CNN and Visual Geometry Group 16
DNN	Dense Neural Network
GA	Genetic Algorithm
DFCN	Densely Connected Fully Convolutional Network
CDK	K-nearest classifier model
MMLM	Multi-Modal Learning Method
BGGO	Greylag Goose Optimization
QCNN	VGG16 with quantum Convolutional Neural Network