NTU Journal of Engineering and Technology (ICSDT2025) special issue
DOI: https://doi.org/10.56286/ntujet.icsdt2025

IRAQI Academic Scientific Journals

P-ISSN: 2788-9971 E-ISSN: 2788-998X
NTU Journal of Engineering and Technology
Available online at: https://journals.ntu.edu.iq/index.php/NTU-JET/index

Accurate Osteoporosis Diagnosis Model Proposing Genetic Algorithm Optimization for Convolutional Neural Network

Ahmed Khalid Abdullah[1, 2], Raid Rafi Omar Al-Nima[3], Khalid Ghanim Majeed[4]
[1]Technical Engineering College/Mosul, Northern Technical University (NTU), Mosul, Iraq,
[2]Directorate of Ninawa Health, Ministry of Health, Iraq,
[3]Technical Engineering College for Computer and Artificial Intelligence/Mosul, Northern Technical University, Mosul, Iraq
[4]Department of Medical Physiology, College of Medicine, Nineveh University, Mosul, Iraq
ahmedkh@ntu.edu.iq , raidrafi3@ntu.edu.iq , khalid.majeed@uoninevah.edu.iq

Article Informations

Received: 15-03- 2025,
Accepted: 28-04-2025,
Published online: 01-03-2026

Corresponding author:
Name: Ahmed Khalid Abdullah
Affiliation: Northern Technical University
Email: ahmedkh@ntu.edu.iq

Key Words:
Osteoporosis,
deep learning,
genetic algorithm,
hyperparameters optimization ,
healthcare.

A B S T R A C T

Osteoporosis is characterized by diminished bone mass and bone tissue loss, which results in weakened bones, decreased bone strength, and increased risk of fractures. This paper exploits a Medical Lumber Spine Images (MLSI) of Dual-Energy X-ray Absorptiometry (DEXA) clinic in Mosul / Iraq to be classified as either normal or having osteoporosis. It also presents the capability of optimizing the hyperparameters of a deep learning model, where the Genetic Algorithm (GA) is used for optimizing Convolutional Neural Network (CNN) hyperparameters. The proposed model essentially explores and optimizes 18 hyperparameters; it is named the Genetic Optimization for CNN (GOCNN). It combines the powerful of GA and CNN in order to provide best hyperparameters tuning, this would further decrease the manual tuning efforts. The real clinical dataset of MLSI is utilized and employed in this study. The proposed method can correctly diagnose 93% of osteoporosis cases from unseen data, with an Area Under Curve (AUC) of Receiver Operating Characteristic (ROC) equals to 0.98.

1. Introduction

Osteoporosis is characterized by decreased Bone Mineral Density (BMD) and microscopic deterioration of bone tissue [1]. The global prevalence of osteoporosis for women is 23.1%, while that for men is 11.7%, with wide variation in prevalence between countries. The most common method used to screen osteoporosis is the measurement of BMD by Dual-Energy X-ray Absorptiometry (DEXA). The World Health Organization (WHO) has delineated specific criteria for diagnosing osteoporosis based on BMD examinations, defining osteoporosis as a T-score exceeding 2.5 standard deviations below peak bone mass. These criteria pertain specifically to DEXA assessments at the lumbar spine, hip, or forearm [2]. DEXA is a rapid, non-invasive instrument that utilizes a minimal radiation dosage and has an accuracy precision error ranging from 1% to 2.5%. Precision is contingent upon the field of research, the patient's age, and the individual's physical and clinical states [3]. There are two types of DEXA: a central device and a peripheral device. Most of the devices used for DEXA are central devices, which are used to measure bone density in the hip and spine. They are usually located in hospitals and medical centers [4]. The central type of DEXA (see Fig. 1).

Additional equipment and modalities, such as Quantitative Computed Tomography (QCT) can also be employed to screen and diagnose osteoporosis. Magnetic Resonance Imaging (MRI) can be utilized to evaluate the trabecular architecture of peripheral bones such as the calcaneus, distal radius, and phalanx. [4]. Quantitative Ultrasound System (QUS) devices assess Broadband Ultrasonic Attenuation (BUA) and Speed of Sound (SOS), which are affected by density and structural characteristics. Both BUA and SOS are influenced by bone density and flexibility. QUS devices are capable of measuring ultrasonic velocity at various bone locations [5]. In this study, real clinical data of DEXA are used.

Deep learning, a form of artificial intelligence, has gained prominence in medical imaging due to its capacity to autonomously extract complex information from data. Convolutional Neural Networks (CNNs) are the most effective deep learning architectures for applications of images including disease detection, segmentation and classification. Through hierarchical feature extraction, CNNs may identify subtle patterns in medical images that may avoid human observation [6]. A CNN can differentiate between normal and osteoporotic bones by distinguishing and recognizing features associated with bone's density or structural integrity [7].

Fig. 1. Central DEXA modality for full body.

The goal of this paper is to optimize a robust CNN model for diagnosing osteoporosis by using real clinical and medical imaging DEXA data. The contributions here can be highlighted as proposing the Genetic Optimization for CNN (GOCNN) model by suggesting a CNN for classifying DEXA images into "normal" and "osteoporosis" categories, and applying Genetic Algorithm (GA) optimization to tune eighteen CNN hyperparameters in order to achieve the best performance.

Next sections will be organized as follows: the literature review is presented in Section 2, methodology of the proposed GOCNN model is explained in Section 3, results and discussions are given in Section 4, and the conclusion is provided in Section 4.

2. Related Literature Review

Recently, many studies have been conducted on the diagnosis of osteoporosis in different concepts. Such previous studies can be reviewed as follows:

In 2014, Tafraouti et al. focused on diagnosing osteoporosis using X-ray pictures of the calcaneus (heel bone) using fractal analysis. This work used fractal analysis, specifically the Fractional Brownian Motion (FBM) model, to analyze bone X-ray images to diagnose osteoporosis. The methodology encompassed several critical phases: To improve contrast and draw attention to important features, X-ray images are first quantized. Next, we project the images at various angles and divide them into sub-images. We handle the increments as Fractional Gaussian Noise (FGN) and model them as FBM processes. The Support Vector Machine (SVM) classifier then uses

a covariance matrix that was obtained from the FGN as input. With a 95% classification accuracy rate, the suggested approach successfully distinguished between those who were osteoporotic and those who were not. Additionally, the study examined how alternative SVM kernel functions, the number of sub images, and quantization levels affected classification accuracy, demonstrating the method's adaptability to a range of parameters [8].

In 2016, Kavitha et al. suggested an automated screening system which uses a hybrid Genetic Swarm Fuzzy (GSF) classifier for digital Dental Panoramic Radiographs (DPR) to classify osteoporosis or low BMD of females. DPR and DEXA for BMD were used to evaluate 141 participants of females, aged from 45 to 92. The approach here examined the geometrical characteristics of the mandibular cortical and trabecular bones. A combination of evolutionary algorithms and particle swarm optimization was used to optimize the classifier's Membership Functions (MFs) and Rule Sets (RSs). The overall accuracy is evidenced by Area Under the Curve (AUC) value of 0.986 [9].

In 2018, Lee et al. developed Computer-Assisted Diagnosis (CAD) systems based on Deep Convolutional Neural Networks (DCNNs) to identify osteoporosis using panoramic radiographs then contrast the results with expert oral and maxillofacial radiologists' diagnoses. The employed dataset was for panoramic radiographs, with osteoporosis and normal. Two experienced radiologists reviewed the radiographs and diagnosed osteoporosis based on the appearance of the mandibular inferior cortex. Then, three DCNN-based CAD systems were investigated. These are the single-column DCNN, single-column DCNN with data augmentation and multi-columns DCNN. Diagnostic performances of the three DCNN systems were assessed. The AUC values for these systems were recorded as 0.9763, 0.9991 and 0.9987, respectively. Such values demonstrate the high performances for all the three systems in detecting osteoporosis [10].

In 2019, Bhattacharya et al. proposed an automatic algorithm that can classify radiographic images into osteoporotic and normal. Machine learning classifiers like SVM and K-Nearest Neighbors (KNNs) were used. The applied method involved pre-processing of input images to enhance contrast and intensity. Also, it included the feature extraction of Gray Level Co-occurrence Matrix (GLCM) and Principal Component Analysis (PCA), which helped in identifying and quantifying relevant characteristics of the images. The study's results indicated that the suggested algorithm for osteoporosis detection using machine learning attained an exceptional average accuracy of 95 % utilizing SVM, surpassing many variants of KNN, which obtain a maximum accuracy of 83.77% [11].

In 2020, Yamamoto et al. applied hip radiographs to a deep learning model in order to determine whether adding patient clinical variables (age, sex, body mass index and history of hip fractures) may increase the diagnosis accuracy compared to using the image data alone. The dataset was collected from DEXA. Five CNN models were used to detect osteoporosis from radiological images. These include: Residual Network-18 layers (ResNet-18), Residual Network-34 layers (ResNet-34), Inception Version 1 (InceptionV1), Google Network (GoogleNet), Efficient Network (version b3), and EfficientNet (version b4). The EfficientNet (version b3) exhibited the highest performance with an accuracy of 88.5% [12].

In 2022, Mao et al. proposed a CNN model that could analyze regular X-ray images of the spine to detect osteoporosis, osteopenia and normal subjects. It was investigated that if adding information about the patient's age, gender and Body Mass Index (BMI) could improve the model's accuracy. The dataset was combined between X-ray images and T-scores, obtained from DEXA and used as a reference standard. Image preprocessing involved delineating Regions of Interest (ROIs) on lumbar vertebrae using precise rectangular frames marked by experienced radiologists. The ROIs were cropped and resized. Grayscale normalization, Gaussian filtering, histogram equalization, and pixel normalization were applied to enhance image consistency across varying scanning parameters. The CNN model was trained using the training data to classify patients into three groups: normal, osteopenia and osteoporosis. The results monitored that the CNN model achieved high diagnostic performance, particularly in identifying osteoporosis, with AUC values ranging from 0.909 to 0.937 for testing [13]. In 2023, Kang et al. developed a deep learning model which was capable of accurately predicting BMD, where osteoporosis is commonly diagnosed from CT radiographs. A dataset of lumber spine bone CT images was gathered and BMD values were measured by DEXA. A deep learning model was then constructed and trained on the dataset, utilizing a sophisticated architecture designed to extract relevant features from the CT images. The system's testing accuracy was 86%, and the correlation coefficient between its estimated BMD and those determined by DEXA was 0.9 [14].

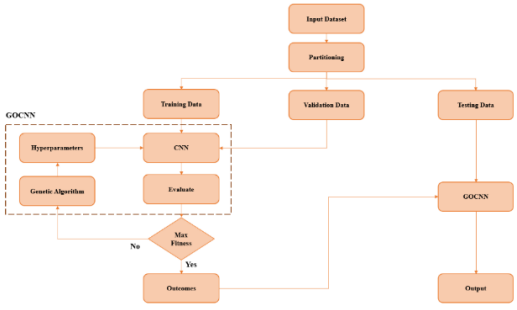

Fig. 2. Full structure of the suggested method.

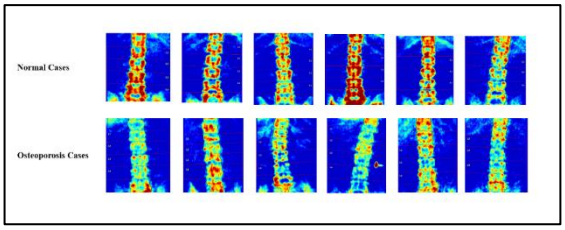

Fig. 3. Different samples of lumber spine images from the MLSI dataset.

In the same year, Mane et al. investigated the diagnosis of osteoporosis from X-ray images of the human spine using deep learning methods. To improve diagnostic accuracy, the authors use random forest classifiers and transfer learning to implement several CNN models, such as VGG16, ResNet50, and InceptionV3. The dataset was a grayscale photo classified as normal and osteoporosis. the study shows notable gains in classification ability, using the VGG16 model in conjunction with random forest to achieve an accuracy of 95% and an AUC of 0.95 [15].

In 2024, Chen et al. enhanced and validated a CNN model called X1 Artificial Intelligence Osteoporosis (X1AI-Osteo), which utilized a configurable feature layer and image preprocessing techniques to predict the T-score in the proximal hip region and detect osteoporosis using standard hip radiographs. The dataset comprised unilateral hip pictures and was partitioned into training, validation, and testing sets. The X1AI-Osteo model demonstrated an AUC of 0.96% and it outperformed other models like ResNet50 and InceptionV3. Furthermore, it showed strong consistency with DEXA for T-score, when considering factors as age, body mass index and sex [16].

In the same year, Sarmadi et al. evaluated the effectiveness of Vision Transformers (ViTs) in the context of medical image analysis, specifically for diagnosing osteoporosis. The dataset here involved knee X-ray images. It separated into three categories: normal, osteopenia and osteoporotic. Two models were considered in this study: a CNN, which is the conventional method for image classification and processes images pixel by pixel, and a ViT, the deep learning architecture that processes images as a series of patches. In the identification of osteoporosis by X-ray radiographs, it was found that ViT surpassed CNN, yielding superior results relative to CNN [17].

Based on the literature, it appears that there were less efforts to utilize DEXA dataset imaging for osteoporosis diagnosis. Furthermore, there is deficiency in employing the genetic algorithm for optimizing CNN. Such are fulfilled in this study.

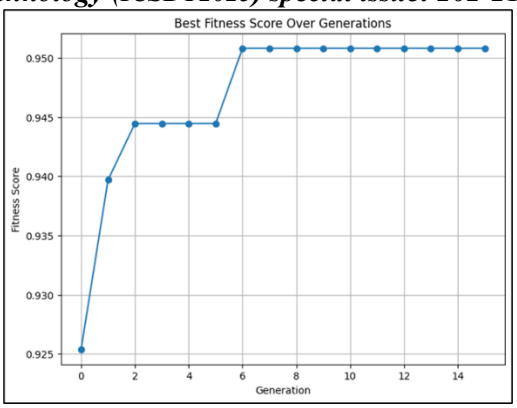

Fig. 4. Improvements of fitness score curve over generations.

3. Methodology

3.1. Suggested method

We propose the GOCNN model, which is suggested using it to classify lumber spine images into normal and osteoporotic categories. It consists of a CNN with fixed structure and initial hyperparameters for numbers of filters, convolutional kernel sizes, convolutional kernel strides, pooling block sizes, pooling block strides, numbers of dense neurons and dropout rates. These parameters are mainly for providing well extracting features and classifications, aiming to achieve the maximum accuracy. The full structure of the suggested method (see Fig. 2). First of all, the dataset of images serves as the input. Then, it is divided into training, validation and testing with 60%, 20% and 20%, respectively. Consequently, a CNN is established and its hyper parameters are initialized. Then, the GA is exploited to change the hyperparameters in order to find the maximum fitness function. So, the best set of hyperparameters are selected. As mentioned, this is called the GOCNN, where it is then tested for obtaining the evaluated accuracy.

3.2. CNN bases

The input layer of a CNN accepts images. Each image is of format Joint Photographic Group (JPG) and of type colored (or Red, Green and Blue (RGB)). In other words, the inputs are colored images, each one has three channels of the RGB.

The convolutional layer includes the essential physical process in the CNN. Basically, Two-Dimensional (2D) convolution operations are executed for each input channel. This layer contains a collection of image channels referred to as feature maps (or filters). The convolutional layer is crucial for extracting various features from an image [18] [19]. It is executed by convolving a kernel with an image channel. The convolution function is expressed by the following equation:

Table 1. The considered GOCNN hyperparameters with their configuration of initial values in GA.

Index	Hyperparameters Type	Initial Values	Type of choosing
1	Number of filters in layer 1	32 to 256	Random Integer
2	Kernel size in layer 1	3, 4 or 5	Random choice
3	Kernel stride in layer 1	1or 2	Random choice
4	Block size in layer 2	2, 3 or 4	Random choice
5	Block stride in layer 2	1or 2	Random choice
6	Number of filters in layer 3	32 to 256	Random Integer
7	Kernel size in layer 3	3, 4 or 5	Random choice
8	Kernel stride in layer 3	1or 2	Random choice
9	Block size in layer 4	2, 3 or 4	Random choice
10	Block stride in layer 4	1or 2	Random choice
11	Number of filters in layer 5	32 to 256	Random Integer
12	Kernel size in layer 5	3, 4 or 5	Random choice
13	Kernel stride in layer 5	1or 2	Random choice
14	Block size in layer 6	2, 3 or 4	Random choice
15	Block stride in layer 6	1or 2	Random choice
16	Number of dense neurons in layer 7	64 to 512	Random Integer
17	Number of dense neurons in layer 8	64 to 512	Random Integer
18	Dropout rate in layers 7 and 8	0.1 to 0.5	Random float

$$Z(x,y) = \sum_{m=1}^{M}\sum_{n=1}^{N} R(m,n) \times W(x-m, y-n) \quad (1)$$

Where Z(x,y) represents a computed convolution, M and N represent the width and height of an 2D input channel, respectively, R(m,n) represents the input channel, and W(x-m ,y-n) represents a weight value [20] [21].

The ReLU is utilized after a convolutional layer and it is a prominent activation function in CNN. The ReLU activation function retains positive values and eliminates negative values from any preceding channel (feature map). Consequently, it facilitates non-linear processing [22]. The ReLU activation function can be expressed by the following equation:

$$ReLU(x,y) = max(0,(x,y)) \quad (2)$$

Where ReLU(x,y) represents a computed ReLU, and max represents the maximum operation [23].

The pooling layer serves a vital role in reducing the dimensions of channels. It generally considers the maximum or average value within a block in the preceding channel. The benefit is minimizing spatial dimensions of feature maps while keeping the most salient qualities. The maximum pooling computation is performed according to the following equation:

$$P(i,j) = max\ (\mathbf{V}) \quad (3)$$

where P(i,j) represents a computed pooling and V is a small matrix (block) in a previous channel [18].

The fully connected layer is also executed. This layer consists of a number of neurons, each of which is interconnected with all neurons in the preceding layer. The quantity of neurons in any CNN may correspond to the number of its requisite classes (C). The fundamental equation of the fully connected layer can be represented as follows:

$$FC(c) = \sum_{l=1}^{p}(FW(c,l) \times P(l)) + B(c) \quad (4)$$

Where FC(c) represents a calculated value in the fully connected layer, c represents a neuron's number in the current layer, P represents the number of neurons in the previous layer, FW(c,l) represents a connection weight value between the previous and current layers and B(c) represents a bias value [22] [24].

A softmax layer trained on input images will produce distinct probabilities for each class, with the total summing to one. Consequently, the softmax activation function in the output layer of a deep neural network serves to represent a categorical distribution across class labels. Consequently, the probabilities for each input element associated with a label are derived.

$$S(c) = \frac{exp(FC(c))}{\sum_{lo=1}^{C} exp(FC(lo))} \quad (5)$$

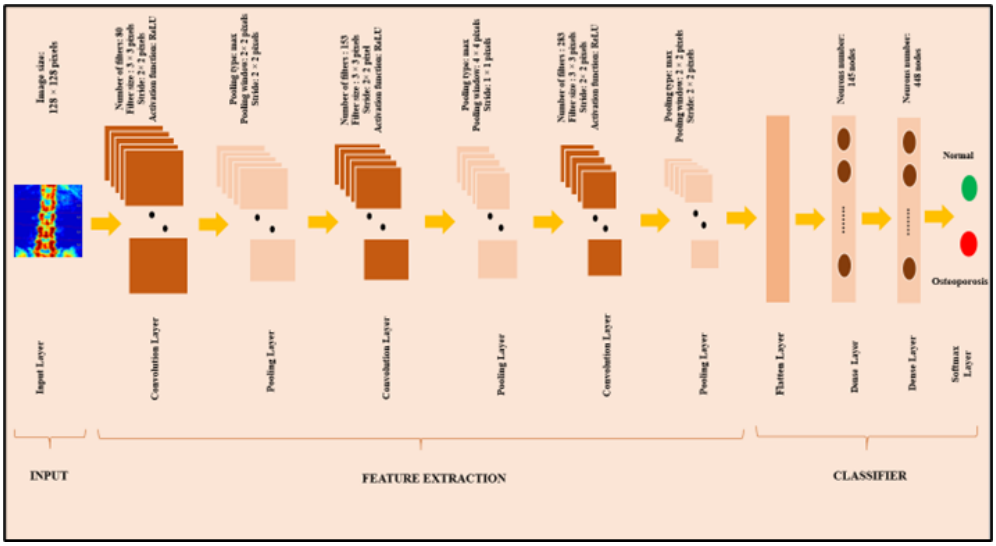

Fig. 5. Full architecture of the proposed GOCNN model.

Where S(c) represents a calculated softmax value and exp is the exponential [23] [25].

3.3. GA fundamentals

GA is a technique that utilizes for addressing both constrained and unconstrained optimization challenges. GA emulates the actions of human genes (selection, crossover and mutation of individual chromosomes) to solve a certain fitness function. A fitness function evaluates the closeness of a particular solution to the optimal solution of the target problem. GA employ selection to choose individuals from the existing population for reproduction, striving to preserve diversity while prioritizing those with superior fitness score. The crossover and mutation operators ensure the enhancement of original solutions to achieve a global optimum [26].

The initial stage of a GA involves the establishment of a population, comprising a collection of individuals, commonly known as chromosomes. Each individual has a potential solution to the issue and it can be encoded by a particular representation, such as binary, real-valued or permutation-based encoding. The population size significantly affects the algorithm's effectiveness, as a larger population increases diversity while requiring more processing resources. Generally, individuals are initialized randomly to guarantee an extensive searching space [27].

The fitness function is an essential element of GA as it is used to assess the quality of individuals within each population. The fitness values direct the selection process, guaranteeing that individuals with superior fitness are more probable to contribute to the subsequent generation [28].

Selection is the process of choosing individuals from the existing population for reproduction. The objective is to choose individuals with superior fitness scores while preserving variety within the population.

Commonly employed selection procedures include roulette wheel selection, tournament selection and rank-based selection. In roulette wheel selection, individuals are selected probabilistically according to their fitness proportion. Tournament selection identifies the optimal solution from a randomly selected subset, fostering robust candidates while preserving genetic variety. Rank selection allocates selection probabilities according to ranked fitness instead of absolute values, so averting early convergence caused by the dominance of a few number of high-fitness individuals [29].
Crossover is the GA process that generates offspring by amalgamating genetic material from two parent individuals. This procedure improves genetic variety and expedites the attainment of optimal solutions. Many crossover procedures are available, including single-point crossover, which involves selecting one cut point in chromosomes; two-point crossover, which incorporates two cut points; and scattered crossover, where genes are randomly exchanged between parents. GA generally sustain elevated crossover rates of 70% to 90% to efficiently merge parental solutions, preserve diversity, and avert premature convergence during solution space exploration [30].

Mutation is a fundamental mechanism in GA that introduces random modifications to the genetic material of individuals. It preserves genetic diversity and can avert the algorithm's stagnation. Mutation can be implemented using one of several methods, including bit-flipping in binary encoding, Gaussian perturbation in real-valued encoding and swap mutation in permutation-based problems.

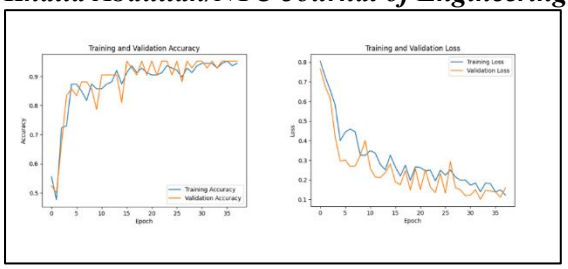

Fig. 6. Curves of training and validation. Training and validation curves of accuracies are shown at the left, whilst, training and validation curves of losses are given at the right.

The mutation rate is often maintained at a low level (about 1% to 5%) to prevent excessive randomness while facilitating adequate exploration of the solution space [29].

A GA progresses through many generations until a termination criterion is reached. Typical termination criteria encompass attaining a certain number of generations, obtaining a solution with acceptable fitness or noting a lack of substantial gain in fitness across multiple iterations [27]

3.4. Proposed Model

As mentioned, the GOCNN model is proposed. Here, the GA is utilized to optimize the hyperparameters of a CNN for osteoporosis classification by iteratively refining the model configurations via selection, crossover and mutation. The fitness function in this context is designed to evaluate the model performance during optimization, incorporating both training and validation accuracies to helps balance model's generalization. It is formulated as a weighted sum of training accuracy and validation accuracy. This weighting scheme prioritizes the model's ability to learn effectively from the training data while still considering its generalization capability on validation data. The involvement of validation accuracy helps mitigate the risk of overfitting. The fitness function of the GOCNN can be represented by the following equation:

$$F = (\alpha \times Ta) + (\beta \times Va) \quad (6)$$

Where F represents the fitness function, α and β are adjustable factors of weights, Ta represents the training accuracy, and Va represents the validation accuracy.

The training accuracy component measures how well the model learns from the training data, while validation accuracy assesses the model's generalization to untrained data during training. By evaluating 18 hyperparameters, GA explores the best CNN optimizing hyperparameters includes the number of filters, convolutional kernel size, convolutional kernel stride, pooling block size, pooling block stride, numbers of dense neurons and dropout rates as listed in table 1. Hyperparameters optimization improves convergence speed, decreases training time and reduces computational costs. Such proposed optimization is crucial in deep learning; it facilitates improved prediction performance and results in a more efficient GOCNN model for osteoporosis diagnoses.

4. Results and Discussion

4.1. Employed dataset

First of all, our own dataset has been acquired and named as the Medical Lumber Spine Images (MLSI). It is for osteoporosis recognition and it is of DEXA. It consists of 211 lumber spine images. Any image in the dataset has been cropped from the DEXA screen for the dimension of 453 × 519 × 3 pixels. Each image has the format of Joint Photographic Experts Group (JPEG). The images have been collected and confirmed by Dr. Khalid Ghanim Majeed in his clinic in Mosul-Iraq. They are classified into osteoporotic and normal accordingly to their BMD measurements. Various samples of images from this dataset (see Fig. 3).

4.2. Standarizations

All images in the dataset are divided into three groups: 60% for training, 20% for verification and 20% for testing. Each image has been resized into a uniform dimension of 128×128×3 pixels. We have also performed augmentations to data, this enhances the network's performance and prevents the overfitting [31]. Rescaling, shearing by a factor of 0.2, zooming by a factor of 0.2 and horizontal flipping are employed.

The experiments here are implemented on a laptop computer which has the following specifications: type Lenovo, Intel Core i5 processor with 2.20 GHz speed and 12 GB computer memory. Google Colab pro pulse utilizes Python (version 3.9.16) and a Central Processing Unit (CPU) with a maximum of 50 GB of Random Access Memory (RAM).

4.3. GA optimization

As previously mentioned, the proposed GOCNN model uses GA to optimize its hyperparameters. Improvements of fitness score curve over generations (see Fig. 4).

This figure demonstrates the progress of the best fitness score over 15 generations during the GA optimization. Initially, the fitness score is somewhat low at the start, about 0.925, but it rises rapidly over the first few generations, implying that the method finds better answers fast. The decline in advancements between the 2nd and 6th generations indicates that the searching process is honing concepts rather than yielding significant breakthroughs.

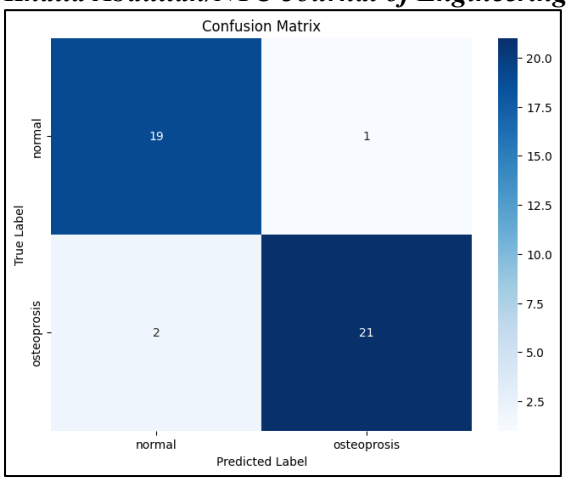

Fig. 7. Confusion matrix of testing the GOCNN.

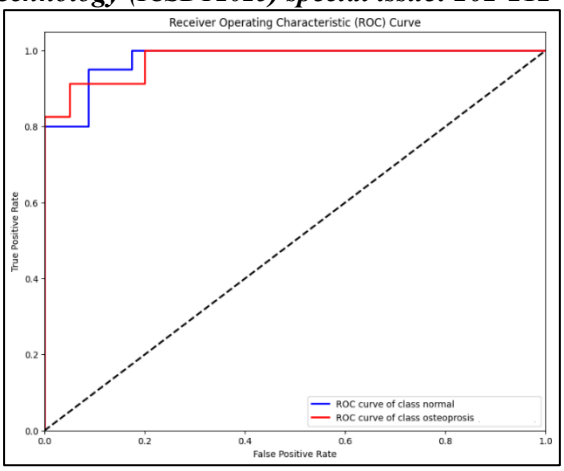

Fig. 8. Demonstrate ROC curve and AUC for normal and osteoporosis classifications.

This stage involves meticulous adjustments, as the algorithm investigates minor alterations to enhance current solutions. The curve reaches a peak around the 7th generation, where the fitness score reaches approximately 0.95. Then, it stabilizes, indicating convergence to an optimal solution. This illustrates GA's efficacy in enhancing the model's performance.

4.4. GOCNN specifications

After many experiments, we proposed a CNN model to classify MLIS into normal and osteoporosis cases. The CNN model follows an architecture consisting of 11 layers that systematically extract and classify osteoporosis features. The layers include input layer, 1st convolution layer, 1st pooling layer, 2nd convolution layer, 2nd pooling layer, 3rd convolution layer, 3rd pooling layer, flatten layer, 1st dense layer, 2nd dense layer and softmax layer. For GA, specific setting is also considered. Population size of 5 is employed, GA introduces genetic variation by random mutation at a rate of 5% in every generation. With K=2, a "tournament" approach is used for selection whereby two individuals are randomly selected. The individual with the best fitness within each group is chosen as a parent to create the next generation. By choosing two random sites inside the chromosomes of two parents and swapping the segments between those spots, the "two points" crossover technique promotes the recombination of genetic material and produces two kids inheriting features from both parents. During the optimization process, these options control the exploration and use of overall searching space.

For the GOCNN, GA optimizes 18 hyperparameters for the CNN, which are distributed in the essential layers of convolutions, poolings and denses. Each convolution layer involves the parameters: number of filters, kernel size and kernel stride. Each pooling layer has the parameters: block size and block stride.

Any dense layer includes the parameters: number of neurons and dropout rate. Table 1 lists the 18 hyperparameters, which are encoded using a real-coded representation with their configuration of initial values in GA. Each hyperparameter is represented by its actual numerical value within a defined range. Real-coded optimization often preferred because it allow for more natural representation of the search space compared to binary encoding, which can suffer from limited precision.

Best hyperparameters are determined by exploiting the power of GA, where obtaining the maximum fitness value is considered. The final optimal GOCNN architecture is demonstrated (see Fig.5). It has the best achieving values of the considered hyperparameters. That is, for the 1st convolution layer, number of filters = 80 filters, kernel size = 3 × 3 pixels and kernel stride = 2 × 2 pixels. For the 1st pooling layer, block size = 2 × 2 pixels and block stride = 2 × 2 pixel. For the 2nd convolution layer, number of filters = 153 filters, kernel size = 3 × 3 pixels and kernel stride = 2 × 2 pixels. For the 2nd pooling layer, block size = 4 × 4 and block stride = 1 × 1 pixel. For the 3rd convolution layer, number of filters = 283 filters, kernel size = 3 × 3 pixels and kernel stride = 2 × 2 pixels. For the 3rd pooling layer, block size = 2 × 2 pixels and block stride = 2 × 2 pixels. For the 1st dense layer, number of neurons = 145 nodes and dropout = 0.438. For the 2nd dense layer, number of neurons = 448 nodes and dropout = 0.438.

4.5. GOCNN training and validation

The GOCNN has been trained using the optimal hyperparameters obtained with the following settings: Adaptive Moment Estimation (ADAM) optimizer, fixed learning rate value of 0.001, batch size value of 32, maximum epochs of 60, and early stopping monitor for validation loss of 5 epochs. Monitoring the validation loss can prevent overfitting and save time.

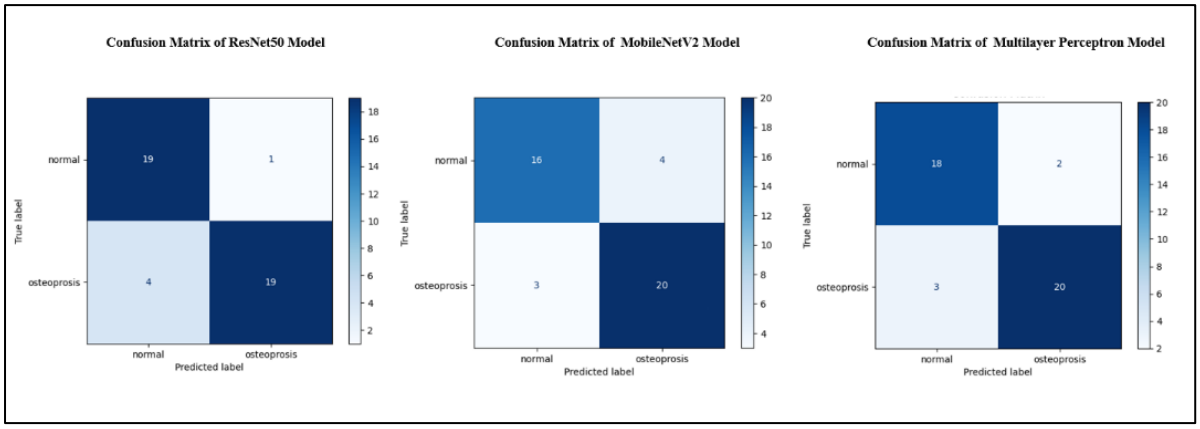

Fig. 9. Confusion matrices of ResNet50, MobileNetV2 and MLP models

The success of the training and validation processes (see Fig. 6). It shows the training and validation accuracies of the GOCNN across roughly 37 epochs. In the initial stages, both accuracies increase rapidly, indicating fast learning. Although some fluctuations occur, particularly between epochs 10 and 20, the validation accuracy closely follows the training accuracy throughout the process. At the later epochs, both converge above 0.9, suggesting minimal overfitting and strong generalization. The relatively stable validation accuracy towards the end highlights the model's ability to effectively finish its learning.

Also in the same figure, the training and validation losses are given along the epochs. Initially, the losses were high, reflecting the model's early-stage learning process. While training progresses, both training and validation losses decrease significantly, demonstrating successful convergence.

4.6. GOCNN Testing

Finally, unseen images from the MLSI dataset have been utilized in the testing phase of the GOCNN model. The confusion matrix of the model's performance in differentiating between normal and osteoporosis cases (see Fig. 7). There are two valuable measurements that can well describe attained results. These are the False Positive (FP) and False Negative (FN). In this work, our approach correctly identified 19 normal cases out of 21. Also, it successfully identified 21 osteoporosis patients out of 22. Therefore, the FP equals 1 and FN equals 2. This can be considered as remarkable performance. This leading to a recall of 95% for normal cases. The precision for osteoporosis detection has attained excellent percentage of 95%, meaning every predicted osteoporosis is correct. Overall accuracy has been benchmarked as 93%, meaning that our proposed GOCNN model can well classify between normal and osteoporosis cases for real clinical data.

Moreover, this high accuracy shows that the GOCNN learns significant patterns from the training data and can confidently recognize whether osteoporosis exists or not.

Also, the Receiver Operating Characteristic (ROC) curve, (see Fig. 8), indicates the GOCNN's exceptional classification efficacy in differentiating between normal and osteoporosis cases. It is resulted with the high AUC value of 0.98, which means that the model can be so valuable in diagnosing osteoporosis. Finally, due to its exceptional performances of the GOCNN, it can be concluded that it is suitable for real-world implementation.

Based on the above performance, the GOCNN model exhibits promising accuracy in classifying osteoporosis, potentially leading to substantial advancements in early detection and intervention approaches. Using this model in real-life settings like radiology centers and hospitals, specialized mobile apps can be invented. This app may significantly help doctors in making faster and more accurate diagnoses. Furthermore, it can allow tracking patients' treatment progress, which means reducing manual and slow evaluations. Still, the implementation needs more testing in the form of large clinical trials, regulatory approvals and hardware developments to make sure it can work well and reliable for a wide range of patients.

4.7. Comparisons with other models

After knowing the performance of the proposed GOCNN model, comparisons are made with other models accordingly to performance. The three models tested are respectively: ResNet50, MobileNetV2 and a Multilayer Perceptron (MLP). The result of confusion matrices of ResNet50, MobileNetV2 and MLP models (see Fig. 9).

It shows that the confusion matrices offer an in-depth analysis of the classification performance of each model. The ResNet50 model provides an accuracy of 88.4%. The MobileNetV2 model achieves an accuracy of 83.7%. The

Multilayer Perceptron (MLP) model obtains an accuracy of 88.4%. Such results reveal that our proposed model attains superior performance.

5. Conclusion

In this study, we utilized a MLSI dataset of DEXA clinic in Mosul / Iraq to be classified as either normal or having osteoporosis. The model of GOCNN was proposed. It shows the capability of optimizing 18 hyperparameters of a CNN by GA in order to achieve a maximum possible accuracy. The results also underline how choosing suitable CNN hyperparameters efficiently affects its performances, especially for medical diagnosis.

A total of 211 real clinical images in MLSI have been exploited, and partitioned into training, validation and testing sets. The various hyperparameters that have been explored include numbers of filters, convolutional kernel sizes, convolutional kernel strides, pooling block sizes, pooling block strides, numbers of dense neurons and dropout rates. High testing accuracy of 93% has been achieved for the best configuration of the proposed GOCNN model. Furthermore, very high AUC value of 0.98 has been attained for distinguishing any class of normal and osteoporosis. The precision has reached an excellent percentage of 95%. The recall has obtained a reasonable result of 91.3%.

In future, more real clinical data can be collected in order to further validate the model's performance. Moreover, investigating and comparing between various metaheuristic optimization methods such as Red Fox Optimization (RFO) and Whale Optimization Algorithm (WOA) seem good suggestions for future work directions.

References

[1] K. J. Chun, "Bone densitometry," Seminars in Nuclear Medicine, vol. 41, no. 3, pp. 220–228, 2011, doi: 10.1053/j.semnuclmed.2010.12.002.

[2] J. Damilakis, K. Perisinakis, and N. Gourtsoyiannis, "Imaging ultrasonometry of the calcaneus: Optimum T-score thresholds for the identification of osteoporotic subjects," Calcified Tissue International, vol. 68, no. 4, pp. 219–224, 2001, doi: 10.1007/s002230020014.

[3] John Damilakis, Thomas G. Maris, and Apostolos H. Karantanas, "An update on the assessment of osteoporosis using radiologic techniques," European Radiology, vol. 17, no. 6, pp. 1591–1602, 2007, doi: 10.1007/s00330-006-0511-z.

[4] R. M. Lorente Ramos, J. Azpeitia Armán, N. Arévalo Galeano, A. Muñoz Hernández, J. M. García Gómez, and J. Gredilla Molinero, "Dual energy X-ray absorptimetry: Fundamentals, methodology, and clinical applications," Radiologia, vol. 54, no. 5, pp. 410–423, 2012, doi: 10.1016/j.rx.2011.09.023.

[5] C. C. Gluer, C. Y. Wu, M. Jergas, S. A. Goldstein, and H. K. Genant, "Three quantitative ultrasound parameters reflect bone structure," Calcified Tissue International, vol. 55, no. 1. pp. 46–52, 1994, doi: 10.1007/BF00310168.

[6] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015, doi: 10.1038/nature14539.

[7] E. Ahishakiye, M. B. Van Gijzen, J. Tumwiine, R. Wario, and J. Obungoloch, "A survey on deep learning in medical image reconstruction," Intelligent Medicine, vol. 1, no. 3. pp. 118–127, 2021, doi: 10.1016/j.imed.2021.03.003.

[8] A. Tafraouti, M. El Hassouni, H. Toumi, E. Lespessailles, and R. Jennane, "Osteoporosis diagnosis using fractal analysis and support vector machine," in Proceedings - 10th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2014, 2014, pp. 73–77, doi: 10.1109/SITIS.2014.49.

[9] Muthu Subash Kavitha, Pugalendhi Ganesh Kumar, Soon-Yong Park, Kyung-Hoe Huh, Min- Suk Heo, Takio Kurita, Akira Asano, Seo-Yong An, and Sung-Il Chien, "Automatic detection of osteoporosis based on hybrid genetic swarm fuzzy classifier approaches," Dentomaxillofacial Radiology, 2016, [Online]. Available: British Institute of Radiology.

[10] J. S. Lee, S. Adhikari, L. Liu, H. G. Jeong, H. Kim, and S. J. Yoon, "Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: A preliminary study," Dentomaxillofacial Radiology, vol. 48, no. 1. 2018, doi: 10.1259/dmfr.20170344.

[11] S. Bhattacharya, D. Nair, A. Bhan, and A. Goyal, "Computer Based Automatic Detection and Classification of Osteoporosis in Bone Radiographs," in 2019 6th International Conference on Signal Processing and Integrated Networks, SPIN 2019, 2019, pp. 1047–1052, doi: 10.1109/SPIN.2019.8711616.

[12] Norio Yamamoto, Shintaro Sukegawa, Akira Kitamura, Ryosuke Goto, Tomoyuki Noda, Keisuke Nakano, Kiyofumi Takabatake, Hotaka Kawai, Hitoshi Nagatsuka, Keisuke Kawasaki, Yoshihiko Furuki, and Toshifumi Ozaki, "Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates," Biomolecules, vol. 10, no. 11, pp. 1–13, 2020, doi: 10.3390/biom10111534.

[13] Liting Mao, Ziqiang Xia, Liang Pan, Jun Chen, Xian Liu, Zhiqiang Li, Zhaoxian Yan, Gengbin Lin, Huisen Wen, and Bo Liu, "Deep learning for screening primary osteopenia and osteoporosis using spine radiographs and patient clinical covariates in a Chinese population," Frontiers in Endocrinology, vol. 13, 2022, doi: 10.3389/fendo.2022.971877.

[14] J. W. Kang, C. Park, D. E. Lee, J. H. Yoo, and M. W. Kim, "Prediction of bone mineral density in CT using deep learning with explainability," Frontiers in Physiology, vol. 13. 2023, doi: 10.3389/fphys.2022.1061911.

[15] P. R. Mane, J. Vemulapalli, N. S. Reddy, N. Anudeep, and G. Prabhu, "Osteoporosis

Detection Using Deep Learning on X-Ray images of Human Spine," in Journal of Physics: Conference Series, 2023, vol. 2571, no. 1, doi: 10.1088/1742-6596/2571/1/012017.

[16] Y. P. Chen, W. P. Chan, H. W. Zhang, and Z. R. Tsai, "Automated osteoporosis classification and T-score prediction using hip radiographs via deep learning algorithm," Therapeutic Advances in Musculoskeletal Disease, vol. 16, 2024, doi: 10.1177/1759720X241237872.

[17] Ali Sarmadi, Zahra Sadat Razavi, Andre J. van Wijnen, and Madjid Soltani, "Comparative analysis of vision transformers and convolutional neural networks in osteoporosis detection from X-ray images," Scientific Reports, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-69119-7.

[18] Sarah O.Ali, Raid R. Al-Nima, and Emad A. Mohammed, "Individual Recognition with Deep Earprint Learning," in International Conference on Communication and Information Technology, 2021, pp. 304–309, doi: 10.1109/ICICT52195.2021.9568410.

[19] Adnan A. Ismael, Abdulnaser A. Ahmed, Raid Rafi Omar Al-Nima, and Mohammed Khaire Hussain, "Prediction of the discharge coefficient of steeply," NTU Journal of Engineering and Technology, vol. 3, no. 4, pp. 41–44, 2023, doi: 10.56286/ntujet.v2i4.778 Prediction.

[20] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi, "Understanding of a convolutional neural network," Proceedings of 2017 International Conference on Engineering and Technology,. pp. 1–6, 2017, doi: 10.1109/ICEngTechnol.2017.8308186.

[21] Saba Qasim Hasan, "Shallow Model and Deep Learning Model for Features," NTU Journal of Engineering and Technology (2023), vol. 2, no. 3, pp. 1–8, 2023, doi: 10.56286/ntujet.v2i3.449.

[22] J. van de Wolfshaar, M. F. Karaaba, and M. A. Wiering, "Deep Convolutional Neural Networks and Support Vector Machines for Gender Recognition," in IEEE Symposium Series on Computational Intelligence (SSCI), 2015, pp. 188–195, doi: 10.1109/SSCI.2015.37.

[23] Sagheer Abbas, Yousef Alhwaiti, Areej Fatima, Muhammad A. Khan, Muhammad Adnan Khan, Taher M. Ghazal, Asma Kanwal, Munir Ahmad, and Nouh Sabri Elmitwally, "Convolutional Neural Network Based Intelligent Handwritten Document Recognition," Computers, Materials and Continua, vol. 70, no. 3, pp. 4563–4581, 2022, doi: 10.32604/cmc.2022.021102.

[24] Duaa Mowafak Hameed, and Raid Rafi Omar Al-Nima , "High-PerformanceCharacter Recognition System Utilizing Deep Convolutional Neural Networks," NTU Journal of Engineering and Technology (2024), vol. 3, no. 4, pp. 42–51, 2024, doi: 10.56286/ntujet.v3i4.1086.

[25] Raid Rafi Omar Al-Nima, Saba Qasim Hasan, and Sahar Esmaiel Mahmood, "Utilizing Fingerphotos with Deep Learning Techniques to Recognize Individuals," NTU Journal of Engineering and Technology, vol. 2, no. 1, pp. 39–45, 2023, doi: 10.56286/ntujet.v2i1.318.

[26] Raid Rafi Omar Al-Nima, Tingting Han, Saadoon Awad Mohammed Al-Sumaidaee, Taolue Chen and Wai Lok Woo, "Robustness and performance of Deep Reinforcement Learning," Applied Soft Computing Journal, vol. 105, 2021, doi: https://doi.org/10.1016/j.asoc.2021.107295.

[27] John H. Holland, Adaptation in Natural and Artificial Systems. MIT Press, 1992.

[28] Sue Ellen Haupt, "Introduction to Genetic Algorithms," in Artificial Intelligence Methods in the Environmental Sciences, Springer, 2016, pp. 103–125.

[29] D. E. Goldberg, "Genetic algorithms in search, optimization, and machine learning," Choice Reviews Online, vol. 27, no. 02, pp. 27-0936-27–0936, 1989, doi: 10.5860/choice.27-0936.

[30] Namrata Karlupia, Palak Mahajan, Pawanesh Abrol, and Parveen K. Lehana, "A Genetic Algorithm based Optimized Convolutional Neural Network for Face Recognition," International Journal of Applied and Computational Mathematics, vol. 33, no. 1, pp. 21–31, 2023, doi: 10.34768/amcs-2023-0002.

[31] Raid Rafi Omar Al-Nima, Lubab H. Albak, and Arwa H. Al-Hamdany, "Translating Sumerian Symbols into French.pdf," NTU Journal of Engineering and Technology, 2024, doi: 10.56286/ntujet.v3i1.660.