AVE Trends in Intelligent Health Letters

Empowering Advanced Bone Cancer Detection Using ResNet-BEO Feature Extraction and AM-SE-CNN Classification

B. Rajalakshmi[1,*], B. Manjunatha[2], Thirumalraj Karthikeyan[3], S. Venkatasubramanian[4], Rahul Panakkal[5]

[1]Department of Computer Science and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India.
[2]Department of Mechanical Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India.
[3]Department of Artificial Intelligence, Quest Technologies, Trichy Research Labs, Tiruchirappalli, Tamil Nadu, India.
[4]Department of Computer Science and Business Systems, Saranathan College of Engineering, Panjapur, Trichy, Tamil Nadu, India.
[5]Department of Computer Science and Engineering, University of Illinois Urbana-Champaign, Champaign, Illinois, United States of America.
dr_rajalakshmi_imprint@yahoo.com[1], manjunatha.princi@gmail.com[2], thirumalraj.k@gmail.com[3], veeyes@saranathan.ac.in[4], rahulp8@illinois.edu[5]

*Corresponding author

Abstract: To enhance the outcomes of early diagnosis and treatment, research on the detection of bone cancer is necessary. The purpose of this research is to present a method for identifying bone cancer that uses sophisticated image preprocessing techniques, ResNet-based feature extraction, and hyperparameter tuning with the Black Eagle Optimiser (BEO). Utilising its deep architecture and residual learning methods, the ResNet model is utilised to extract discriminative features from medical imaging data. This is accomplished by capitalising on the model's capabilities. The retrieved features are then further refined and optimised using the BEO method, resulting in improved overall model performance and generalisation. In addition, it presents a novel Attention Mechanism-Squeeze-and-Excitation-Convolutional Neural Network (AM-SE-CNN) model for classification. This model blends attention mechanisms and channel-wise feature recalibration to enhance the model's discriminative power and interpretability. The proposed method offers improved performance in bone cancer identification when compared to cutting-edge approaches. This was accomplished through intensive experimentation on benchmark datasets. The proposed model achieved highly impressive performance metrics, including 99.62% accuracy, 99.47% precision, 99.35% recall, 99.21% specificity, and an F1 score of 99.17%, demonstrating its efficiency in identifying bone cancer.

Keywords: Bone Cancer; Residual Network; Black Eagle Optimiser; Squeeze and Excitation; Convolutional Neural Network; Diagnosis and Treatment; Image Preprocessing; Cancer Detection; Medical Imaging.

Cite as: B. Rajalakshmi, B. Manjunatha, T. Karthikeyan, S. Venkatasubramanian, and R. Panakkal, “Empowering Advanced Bone Cancer Detection Using ResNet-BEO Feature Extraction and AM-SE-CNN Classification,” *AVE Trends in Intelligent Health Letters*, vol. 2, no. 3, pp. 133–149, 2025.

Journal Homepage: https://avepubs.com/user/journals/details/ATIHL

Received on: 10/10/2024, **Revised on:** 27/01/2025, **Accepted on:** 09/04/2025, **Published on:** 07/09/2025

DOI: https://doi.org/10.64091/ATIHL.2025.000172

1. Introduction

The cells that comprise the body's bones can give rise to cancer, including bone cancer. It can be difficult to diagnose and treat [1]. Early detection is essential for improving patient outcomes and survival rates, given its varied forms and intricate

---

presentations [2]. The landscape of bone cancer detection has changed over time due to major advances in medical imaging, molecular diagnostics, and interdisciplinary collaboration. These developments have given patients and physicians new hope [3]. Understanding the nature of bone cancer is crucial before delving into its detection. Different types of bone cancer can appear, such as osteosarcoma, Ewing sarcoma, chondrosarcoma, and others, each with unique features and clinical behaviours. Although uncommon in comparison to other forms of cancer, bone cancer can be extremely aggressive and have disastrous outcomes if it is not identified and treated quickly [4]. The elusive nature of bone cancer, which frequently presents nonspecific symptoms in its early stages, is one of the main management challenges [5]. Patients may have fractures, oedema, or chronic bone pain that is easily linked to other musculoskeletal disorders. Because of this, it often takes longer to diagnose the condition, allowing it to advance to a point where there are few treatment options. Medical imaging methods are essential for identifying and diagnosing bone tumours [6]. Conventional radiography remains the first-line modality because it can detect suspicious lesions and provide detailed information about the skeletal system [7]. But the diagnosis of bone cancer has changed dramatically with the introduction of sophisticated Imaging procedures such as computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI).

MRI's unmatched soft-tissue resolution enables precise definition of tumour margins and determination of the local extent of the tumour [8]. By highlighting bone involvement and surrounding structures, CT scans help with surgical planning and provide comprehensive anatomical information. When PET imaging and CT are combined (PET/CT), it is possible to evaluate tumour metabolism functionally and identify distant metastases, which helps with prognosis and treatment selection [9]. By leveraging its ability to automatically identify complex patterns and features in medical imaging data, deep learning (DL) is highlighted as important for bone cancer detection [10]. DL algorithms can automatically derive hierarchical representations from unprocessed image inputs, thereby improving detection robustness and accuracy [11]. This contrasts with conventional machine learning methods, which rely on handcrafted features and require advanced domain expertise. Popular DL architectures, convolutional neural networks (CNNs), have demonstrated outstanding performance in tasks such as tumour segmentation, classification, and localisation across MRI, CT, and X-ray imaging modalities [12]. Furthermore, DL models are intrinsically more flexible and scalable than traditional machine learning techniques because they can generalise and adapt to a wide range of datasets. DL is a better option for bone cancer detection overall because of its innate ability to learn complex representations directly from data, without the need for hand-engineered features, providing increased accuracy and efficiency in clinical practice [13].

1.1. Motivation

Bone cancer detection using DL holds immense promise for revolutionising early diagnosis and treatment strategies. By leveraging advanced algorithms and medical imaging data, DL enables the extraction of intricate patterns and features indicative of bone malignancies with unprecedented accuracy and efficiency. Early detection is essential to improving patient satisfaction and survival rates, underscoring the urgency of developing robust, scalable DL solutions for timely intervention. Additionally, the non-invasive nature of DL-based approaches minimises patient discomfort and facilitates widespread adoption in clinical settings, offering new hope for patients battling this devastating disease.

1.2. Main Contributions

- **Feature Extraction with ResNet:** The utilisation of ResNet for feature extraction capitalises on its deep architecture and residual learning mechanisms, enabling the extraction of discriminative features from medical imaging data.
- **Optimisation with BEO:** The Black Eagle Optimiser (BEO) optimises the hyperparameters of the ResNet model, thereby enhancing performance and generalisation and improving detection accuracy.
- **Novel Classification Model:** The introduction of the AM-SE-CNN model for classification represents a novel contribution, leveraging attention mechanisms and channel-wise feature recalibration to improve classification accuracy.

2. Related Works

The study by Hu et al. [14] aimed to assess the ability of several MRI patterns, combined with a DL model, to detect cartilage damage associated with knee osteoarthritis (KOA). It used an enhanced wide residual network model for multiscale processing to compare image super-resolution algorithms with SSD, SRCNN, and EDSR. The outcomes of arthroscopic operations were compared with the MRI scans of 104 cartilage-damaged KOA patients: the recommended model displayed very little noise and artefacts, with excellent image quality. Lesions of grades I and II were mainly seen on the medial and lateral articular cartilage, whereas lesions of higher grade also extended to the patella and femoral trochlea. The 3D-DS-WE sequence showed the best diagnostic accuracy, correctly diagnosing over 95% of grade IV lesions. Consistency tests revealed good agreement between the T2 and 3D-DESS-WE are mapping sequences and arthroscopic results (Kappa values of 0.748 and 0.682, respectively). Breden et al. [15] used a Vision Transformer model to identify healthy and diseased regions in X-ray images. It used a pretrained model and substantial data augmentation to address the limited data. The terms median, standard deviation, and variance, and the continuous parameters by incidence and percentage ratio, characterised the discrete parameters. Sensitivity and specificity were calculated to evaluate the model's accuracy. Cross-validated accuracy, sensitivity, and specificity for the pathological category's two-entity classification for the test groups and the healthy control group were 89.1%, 82.2%, and 93.2%, respectively. To guarantee the accuracy of the predictions, Grad-CAMs were developed. By utilising cutting-edge DL techniques, the suggested model has demonstrated excellent results in identifying bone tumours on children's knee X-rays.

The purpose of this work by Kim-Wang et al. [16] was to modify the U-Net convolutional neural network architecture to automate the separation of the tibia and femur from knee magnetic resonance (MR) images with a twofold echo constant-state. Following, a musculoskeletal radiologist assessed the model. More than 4,000 magnetic resonance images from 34 people served as the training dataset, which were analysed and segmented by three experienced researchers. For the validation set, the dice coefficient was 0.985; for the test set, it was 0.984. It re-examined antecedent research on tibial cartilage stress and healing using the trained model. The analysis showed that across all subjects, the mean differences in cartilage strain between the actual bone models and the machine-learning bone models did not differ significantly (mean ± 95% confidence interval). This disparity is within the breadth of earlier studies on cartilage strain in the lab that used manual segmentation for measurement resolution. Zhan et al. [17] present a new intelligent auxiliary framework for segmenting medical images of bone lesions caused by cancerous tumours, incorporating a supervised edge-attention-guided selection module for border key points. The selection module for border key points is designed to monitor the model's acquisition of edge attention and maintain detailed edge feature information. For instance, it can precisely locate malignant tumours using segmentation networks by extracting tumour lesion feature maps from medical images. At the same time, mixed attention captures rich contextually dependent information from the feature map to better understand the boundary's ambiguity and uncertainty; edge attention learning instructs the segmentation network to focus on the fuzzy tumour and the region border. The model is validated through extensive research using actual medical data. The approach's superiority over the most recent segmentation techniques is demonstrated by its best results in the Dice Similarity Coefficient (0.967), accuracy (0.996), and precision (0.968).

In this study by Chen et al. [18], a modest internal dataset with one hundred cases of prostate and breast cancer, each with ten non-metastasis cases and ninety-nine bone metastases, was used for model training. Of the 100 patients, 50 had bone metastases, and the other 50 did not. All the images' labels were binary at first. Using the negative mining method, also known as thresholding, three classes were derived from the image labels to generate a non-metastasis mask. It employed a Double U-Net as the baseline model and changed the output activation function. SoftMax was used as an activation function to support multi-class segmentation. Some of the techniques used to enhance model performance were transfer learning to leverage feature similarities across datasets, adding negative samples, and background pre-processing to increase model accuracy and remove irrelevant information. Pixel-by-pixel performance analysis was conducted using 10-fold cross-validation. The best-performing model had 69.96% precision, 63.55% sensitivity, and 66.60% F1-score. This shows gains over the baseline model of 8.40%, 0.56%, and 4.33%, respectively, in terms of accuracy, sensitivity, and F1-score. In this paper by Li et al. [19], a Dependency on Space, the recommended network is the Multi-Task Transformer (SDMT) for landmark localisation and segmentation. It employed a common encoder to extract features, and then SDMT promoted the two tasks by utilising the spatial relationship among segmentation results and landmark location. Attention is classified into two categories based on whether it is focused on intra- or inter-task activities. This is referred to as a multi-head, task-hybrid attention mechanism. More specifically, SDMT uses spatial encoding to improve feature extraction. The two attention heads concentrate on correlation within a task and spatial dependence between two tasks, respectively. To ensure coordination, it produced a multi-task loss function with dynamic weights after training two assignments.

The suggested approach is validated using multi-task 3D knee MRI datasets. In the segmentation task, dice reached 83.91%, and in the renowned MRE localisation task, it was 2.12 mm, demonstrating its competitiveness and superiority over alternative cutting-edge single-task methods. Chung et al. [20] consists of 945 MRI pictures from patients who dropped their feet, whose CPN injury was verified by electrophysiologic testing (n = 42) and 1341 MRI pictures demonstrating non-traumatic knee pain (n = 107). An 8:1:1 ratio was used to split the data into training, validation, and test sets. To distinguish the CPN injury group from the other groups, an efficientNet-B5, ResNet152, and VGG19 convolution neural network-based algorithm was used. The receiver operating characteristic curve was used to gauge algorithm performance (AUC). AUC = 0.946 for CPN MR image classification showed that EfficientNet-B5 performed better than the other algorithms, with ResNet152 and VGG19 trailing closely behind. When comparing the three algorithms using additional performance metrics such as F1 score, recall, accuracy, and precision, EfficientNet-B5 performed better. The nerve region was the focus of the EfficientNet-B5 algorithm, as shown in a saliency map, to identify CPN injury. The AI-driven approach by Pan et al. [21] outperformed orthopaedic physicians in diagnosing proximal femoral bone cancers on plain hip radiographs. Employing sophisticated DL models, such as DenseNet, yielded exceptional precision, sensitivity, and accuracy. It increased productivity by automatically generating regions of interest (ROIs) and reducing the need for human labelling. DenseNet performed better than earlier models when compared to human physicians, showing notable improvements of more than 5% in accuracy, sensitivity, and precision. Tumour locations revealed by Grad-CAM (gradient-weighted class activation mapping) provided additional evidence of the framework's efficacy.

The objective of the study by Georgeanu et al. [22] was to use DL algorithms to determine whether a malignant bone tumour could be detected on magnetic resonance imaging (MRI). Supplies and Procedures: The cohort included nine men and 14 women, ages 15 to 80. Weighted T1 and T2 MRI scans are classified using two pretrained ResNet50 image classifiers. A tumour's malignancy is predicted using a clinical model. T1 and T2 classifier outputs, along with the patient's clinical data, are fed into the model. The model is a feed-forward neural network. Findings: The T1 classifier achieved 93.67% accuracy during training, while the T2 classifier achieved 86.67%. Both classifiers achieved 95.00% accuracy in validation. The clinical model's accuracy was 80.84% during the training phase and 80.56% during the validation phase. The clinical model's receiver operating characteristic (ROC) curve shows how well the algorithm can separate classes. Conclusions: The suggested approach removes the need for manual MRI image segmentation by utilising pretrained DL classifiers. These formulas can expedite the time required to diagnose and treat tumours by predicting whether they are malignant. Sharma et al. [23] first determine which edge detection algorithm is most appropriate, then prepare two feature sets: one with HOGs and another without. Along with support vector machines (SVMs) and two additional machine learning models, Random Forests are used to assess the efficacy of these feature sets. On these models, the HOG feature set performs notably better. Furthermore, the Random Forest score of 0.77 is lower than that of the hog feature set used to train the SVM model, which achieved 0.92.

2.1. Research Gaps

Despite the notable advancements in DL techniques for medical image analysis, several research gaps persist across the referenced studies. Firstly, there is a need for further investigation into the generalizability and robustness of proposed models across diverse patient populations and imaging modalities. Additionally, the interpretability of DL models remains a challenge, necessitating efforts to enhance model transparency and explainability for clinical adoption. Furthermore, standardised evaluation metrics and benchmark datasets are lacking, hindering direct comparisons between different methodologies. Addressing these gaps would facilitate the development of more reliable and clinically applicable DL solutions for bone cancer detection and diagnosis.

3. Proposed Methodology

Figure 1 shows the proposed workflow for bone cancer classification.

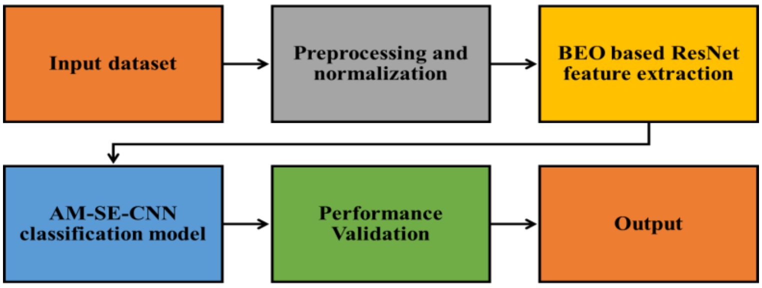

Figure 1: Block diagram

3.1. Dataset Description

Bone cancer is used to implement the suggested system.

Figure 2: Healthy bones

Dataset of X-ray images obtained from the Indian Institute of Engineering Science and Technology, Shibpur (IIEST) repository. The input dataset contains 100 images: 50 show healthy bones, and the remaining 50 show malignant bones [24]. The base size and uniformity of the input images are established by resizing them to 255 × 255 pixels. Examples of malignant and healthy bone images are shown in Figures 2 and 3.

Figure 3: Malignant bones

3.2. Data Preprocessing and Normalisation

Image preprocessing is the process of preparing images for use in model inference and training. Additional preprocessing steps include colour adjustments, resizing, and orientation. Preprocessing seeks to enhance image data by minimising inadvertent distortions and emphasising aspects of the image that are crucial for further processing. The photographs in the dataset used for this study have a resolution of 255 x 255 pixels. Images were resized to 255 × 255 pixels to speed up the calculations. In image processing, normalisation is a method for adjusting the luminance range within a pixel. Image normalisation generally involves converting a source image to pixel levels that the senses can tolerate or find more appetising. All that's present in the images is a composite of discrete pixel values ranging from 0 to 255. Large values require time-consuming and impractical operations, which call for more powerful computers. Nevertheless, the normalisation procedure lessens this burden by dividing the images by 255 [25].

3.3. Feature Extraction

3.3.1. Residual Neural Networks

To address the problem of, ResNets are constructed with a modified CNN architecture, "vanishing or exploding gradients" in extraordinarily deep convents, where a very small gradient causes the inability to update weights. The training error begins to increase rather than decrease due to this problem. This problem was resolved by including residual blocks in the design. In addition to establishing a correct link between layers, residual blocks also introduce a skip connection, or shortcut, that enables us to transfer activation from a shallow layer to a much deeper layer. Sequential stacking of these residual blocks results in a ResNet. ResNet contains several residual blocks. A ResNet block is illustrated in the equation below:

$$y = f(x) + x \tag{1}$$

Where x is the residual block's input, and f(x) is the result of applying different operations. Consequently, the residual block's output is f(x) + x, illustrating how information flows from the top to the bottom layer. Within a residual block, the convolutional layers are followed by the batch normalisation and ReLU layers. To prevent premature activation function saturation, preprocessed and typically normalised data for DL follow a normal distribution. However, because the distribution changes continuously following activation, the internal covariate shift issue persists in the intermediate layers. To normalise each layer's output, which will serve as an input to a subsequent layer, the batch normalisation layer follows the activation layer. The input is normalised using the previously computed batch statistics, and the mean and variance are determined [26]. The architecture selected for this paper consists of one fully connected layer and seventeen convolutional layers, collectively referred to as ResNet-18. The result of 16 convolutional layers is eight leftover blocks. In the hierarchy, each convolutional layer is followed by an activation and a batch normalisation layer. There are two pooling layers after the first and last residual blocks, which link to the dense layer.

3.3.2. Hyperparameter Tuning Using BEO

The predatory, migratory, courtship, and breeding behaviours of black eagles will be discussed in the following, and these behaviours will be utilised to build a mathematical model of the BEO algorithm for hyperparameter tuning of the ResNet model.

3.3.3. Black Eagle

The black eagle is a medium-sized raptor distinguished by its black-brown feathers. Black eagles typically live in forests that are close to wetlands and grasslands. They feed on birds and small mammals and are found in river valleys, grasslands, moorlands, deserts, hills, mountains, and forests [27]. Black eagles typically stalk their prey by standing on the shoreline or in a tree, and then they quickly descend to capture the prey by hovering near it. The black eagle occasionally steals other birds' hunting kills. Most black eagles migrate once a year, and the precise date and route of their journey can change based on the location and climate. Black eagles typically return from colder regions to their warmer birthplaces in the autumn and winter to get ready to breed. Typically, black eagles start courting after the winter has passed. Black eagles use flight displays, body displays, food presentations, and aerial displays during courtship to demonstrate to potential mates that they are healthy and capable of procreating. These are elegant and distinctive gestures and behaviours. Following the conclusion of winter, black eagles typically commence their breeding season in the spring. They prefer to build their nests in more protected areas, such as the tops of tall trees or the edges of sheer cliffs. The male black eagle provides food and protection, while the female hatches the young [28].

3.3.4. Mathematical Model and Algorithm

It will go into more detail about the black eagle's biological behaviour and use that information to build a mathematical model for the BEO algorithm in the sections that follow. The black eagle's primary means of survival throughout the year is hunting. The hunt starts with the search for prey. Black eagles typically stand atop objects like rocks, trees, or cliffs to have a better view of their prey when searching for it. This makes it easier for them to identify conspicuous prey. This is referred to as stalking. It first established the quantity of black eagles at n, the variable's dimension that needs to be optimised to d, and modelled the location of the black eagles using a matrix:

$$\mathrm{X} = \begin{bmatrix} \mathrm{x}_{11} & \mathrm{x}_{12} & \cdots & \mathrm{x}_{1n} \\ \mathrm{x}_{21} & \mathrm{x}_{22} & \cdots & \mathrm{x}_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{x}_{d1} & \mathrm{x}_{d2} & \cdots & \mathrm{x}_{dn} \end{bmatrix} \tag{2}$$

Next, a single black eagle's fitness value is represented by $\mathrm{f_x}$. The level of fitness within a black eagle population, $\mathrm{F_x}$, can be illustrated by the subsequent vectors:

$$\mathrm{F_X} = [\mathrm{f}([\mathrm{x}_{11}, \mathrm{x}_{21}, \cdots \mathrm{x}_{d1}]^{\mathrm{T}}) \quad \mathrm{f}([\mathrm{x}_{12}, \mathrm{x}_{22}, \cdots \mathrm{x}_{d2}]^{\mathrm{T}}) \quad \cdots \quad \mathrm{f}([\mathrm{x}_{1n}, \mathrm{x}_{2n}, \cdots \mathrm{x}_{dn}]^{\mathrm{T}})] \tag{3}$$

Thus, it proposed a mathematical method, called mirror search, to simulate the stalking behaviour of black eagles, reflecting their characteristic search for prey at high locations within their territories and their preference for conspicuous prey. The search space is split in half in that dimension where the dimensional element with the largest absolute value of X represents those noticeable prey tracks, in accordance with the mirror search rule. Next, a thorough search of the upper half of the area was conducted, simulating how black eagles hunt while perched on high ground. After that, the upper half of the space is fully explored by mapping the search path through it. Stalking (global exploration phase 1):

$$\alpha = \mathrm{e}^{-\frac{\left|\mathrm{x}^{t} - \mathrm{x}^{t}_{\mathrm{best}}\right|}{\mathrm{D}}} \tag{4}$$

$$\mathrm{X}^{t}_{r} = \mathrm{e}^{-0.2\left(\frac{t}{T}\right)}\left(\mathrm{lb}^{*} + \mathrm{r}_1(\mathrm{lb}^{*} + \mathrm{ub}^{*})\right) \tag{5}$$

$$\mathrm{X}^{t+1}_{1} = \mathrm{X}^{t}_{r} - \alpha \cdot \mathrm{r}_1 \cdot (\mathrm{t}_1 \cdot \mathrm{X}^{t}_{\mathrm{best}} - \mathrm{X}^{t}_{k}) \tag{6}$$

$$\mathrm{X}^{t+1}_{2} = \mathrm{ub} + \mathrm{lb} - \mathrm{X}^{t+1}_{1} \tag{7}$$

$$\mathrm{X}^{t+1} = \text{ Minimise fitness } \textstyle\sum_{1}^{n} (\mathrm{X}^{t+1}_{1}, \mathrm{X}^{t+1}_{2}, \mathrm{X}^{t}) \tag{8}$$

where $\mathrm{X_k}$ is where a black eagle appears out of nowhere, $\mathrm{X_{best}}$ is the prey's location and the world's best position, and D is the distance at which $\mathrm{X_{best}}$ from the perimeter of the search. In the dimension where the absolute value of the search space is split in half, $\mathrm{X_{best}}$ is the largest, and its upper and lower boundaries are lb^* and ub^*, respectively. $\mathrm{X_1}, \mathrm{X_2}$ are the updated group locations in the corresponding upper and lower search spaces. $\mathrm{r_1}$ is an arbitrary value that falls between 0 and 1, α is the step factor, and $\mathrm{t_1}$ is a random number from 0 to 1 formed by the Tent mapping. It is the current iteration number. Eq. (8) shows the revised standing, determined by the n individuals with the lowest fitness. The black eagle will hover in midair, swooping rapidly as it approaches, and quickly adjusting its stance to ensure an effective prey grab once it spots prey. Black eagles occasionally steal prey from other birds. When black eagles spot other birds catching food, they move quickly to get close to the target bird and scare it into giving up the food. It's interesting to note that a black eagle is frequently attacked or warned when it flies into another black eagle's hunting area. A high-pitched chirp, a show of strong claws, or a lifting of the chest and wings to demonstrate intimidation and the ability to attack in defence of its territory are typically used as warning signs.

First, it uses the rotation method of points in high-dimensional space to simulate the hovering behaviour of the black eagle. In this method, 12 dimensions of the image of a 30-dimensional point are displayed after rotating around a fixed point by $\pi/12$ 120 times. This method is inspired mathematically by the rotation of two-dimensional points. Next, the phenomenon of the black eagle flexibly adjusting the capture angle and way based on its own and its prey's current positions is simulated using the position-based adaptive capture formula. In the end, it reduces the snatching behaviour to the point-moving curved path of the black eagle during its curved flight from one location to another. With the point movement of curve jumping, it mimics the black eagle's snatching behaviour. The Poisson distribution was utilised to model particle movement during predation and snatching, mimicking the black eagle's warning system. The blue point, which is extremely near the boundary in dimension Y, represents the current local optimal position in three dimensions. In dimension Y, no further position-updating rules are applied when such a scenario occurs, and the Poisson warning mechanism is activated. Rather, the particle motion is guided by the Poisson warning mechanism. Hovering (global exploration phase 2):

$$a = r_2 \cdot 2\pi \tag{9}$$

$$m = \begin{bmatrix} \cos a & -\sin a & 0 & \cdots & 0 \\ \sin a & \cos a & -\sin a & \ddots & \vdots \\ 0 & \sin a & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \cos a & -\sin a \\ 0 & \cdots & 0 & \sin a & \cos a \end{bmatrix} \tag{10}$$

$$X^{t+1} = X_{best}^{t} + m(X^{t} - X_{best}^{t}) \tag{11}$$

where r_2 is a random number from 0 to 1, a is the hovering angle, m is a d-dimensional hovering matrix. Catching (local exploration phase 1):

$$d_j = \sqrt{\left(X_j - \frac{\sum_{j=1}^{n} x_j}{n}\right)^2} \tag{12}$$

$$D_1 = \frac{\sum_{j=1}^{n} d_j}{n} \tag{13}$$

$$D_2 = 0.8 + \frac{1.95(t-0.1T)}{9T} \tag{14}$$

$$X^{t+1} = 2X^{t} - X_{best}^{t} + s_0 \cdot D_1 + D_2 \cdot (2X_{best}^{t} - 2X^{t} - s_0 \cdot D_1) \tag{15}$$

Where d is the separation between a person and the group's centre, Position adjustment factors D1 and D2 are represented, and s_0 is a column vector of dimension d with normally distributed elements. Snatching (local exploration phase 2):

$$X^{t+1} = X_{best}^{t} + e^{\frac{r3}{\pi}}(X_{best}^{t} - X^{t}) \tag{16}$$

where r_3 is a dimension D column vector whose values for each dimension element follow a normal distribution. Poisson warning (an anti-escape mechanism):

$$d_s^* = \frac{d_s - d_{s\,min}}{d_{s\,max} - d_{s\,min}} \tag{17}$$

$$\lambda = -10 d_s^* + 15 \tag{18}$$

$$P(j) = \frac{\lambda^j}{j!} e^{-\lambda} \ (j = 1,2 \cdots n) \tag{19}$$

$$\lambda_n = \frac{\lambda - 5}{10} \tag{20}$$

$$\beta = \sin\left(\lambda_n \cdot \frac{\pi}{2}\right) \cdot 0.063 + 0.89 \tag{21}$$

$$X_P^{t+1}(i,j) = X_{best}^{t}(i,j) + (X_d^t(i,j) - X_{best}^{t}(i,j)) \cdot (0.9 + P(j)) \tag{22}$$

$$X^{t+1}(i) = X_P^{t+1}(i) + \beta(X_{best}^{t}(i) - X_P^{t+1}(i)) \tag{23}$$

When the distance of X_{best} If the centre of the search space in dimension i exceeds the 0.4 (ub-lb) threshold, the Poisson warning is triggered. Where d_s is the distance between X and the centre of the search space in dimension i $(i \in [1, d])$, d_{smin}, d_{smax} are the minimum and maximum d_s, d_s^* is the normalised $d_s. \lambda$ is the parameter of the Poisson distribution, P is a weighting factor $(j \in [1, n])$, and β is an adaptive adjustment factor. X_d is the position matrix after X has been moved, placing the items in the new order of proximity to the X_{best} . X_P The Poisson distribution that directs the position matrix. The Poisson warning's benefit is that when X_{best} Suppose the particle is near the boundary in each dimension. In that case, the particle swarm in that dimension adheres to the Poisson distribution law, thereby obstructing particle egress from the search space. Black eagles from the north migrate to the warmer south in the autumn because there is more food there for them to survive. The black eagle's environmental adaptation is represented by the fitness value, which it also uses to create the migration function $z(f)$. The black eagle migration law was fitted well by the migration function: the greater the distance of migration, the less acclimated the individual is to the present surroundings. Migration (local exploration phase 3):

$$z = \frac{1}{2e^{\frac{-f_{best}}{f(j)+\varepsilon}}} \tag{24}$$

$$X^{t+1} = X_{best}^{t} + z \cdot s_1 \cdot (X^t - t_2 \cdot X_{best}^{t}) \tag{25}$$

where f_{best} is the greatest fitness value available right now, f(j) is the worth of fitness of the j-th individual, s_1 is a column vector of dimension d, with elements between -1 and 1, and t_2 is a random number from 0.4 to 1 formed by the Tent mixing mapping. In most places, black eagles will start courting during the winter. To attract the other sex, they congregate at designated breeding sites and engage in a variety of courtship behaviours. Spreading their wings, giving and receiving gifts, phoning, dancing, and crouching while clutching each other's paws are a few courtship rituals. It replicated the interaction behaviour of female black eagles during courtship using similar fluctuations in the sine-cosine function and alternation between odd and even numbers. Courtship (local exploitation phase 1):

$$k = \frac{1}{0.9\left(1+e^{\frac{-14t}{T}-7}\right)} + 0.5 \tag{26}$$

$$\begin{cases} X_j^{t+1} = X_j^t + r_4 \cdot k \cdot \left(X_{best}^{t} - X_j^t\right) + r_5 \cdot \cos\left(\frac{\pi j}{n}\right) \cdot \left(X_{best}^{t} - X_j^t\right) & \text{if} \quad j\%2 = 1 \\ X_j^{t+1} = X_j^t + r_6 \cdot k \cdot \left(X_{best}^{t} - X_j^t\right) + r_7 \cdot \sin\left(\frac{\pi j}{n}\right) \cdot \left(X_{best}^{t} - X_j^t\right) & \text{if} \quad j\%2 = 0 \end{cases} \tag{27}$$

where r_4, r_6 are random numbers from 0 to 1, and r_5, r_7 are dimension d column vectors whose elements are regularly distributed. After being drawn to one another, the two black eagles will undergo a series of tests to ensure they are compatible before building a nest and raising their young. The male black eagle guards the area surrounding the nest for the duration of the chicks' incubation, except at times when he ventures out to hunt. It employs an adaptive normal distribution to guide particle movement, a technique known as adaptive normal flight, to replicate the phenomenon that male black eagles are more active near the nest and less active farther from it. Hatching (local exploitation phase 2):

$$\sigma = \frac{1}{3\left(1+e^{\frac{-14t}{T}-10}\right)} + 0.9 \tag{28}$$

$$\mu = 0.2\sin(r_8 \cdot 2\pi) \tag{29}$$

$$Q(j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(j-\mu)^2}{2\sigma^2}} \tag{30}$$

$$X^{t+1} = X_{best}^{t} + (X_d^t - X_{best}^{t}) \cdot (0.95 + Q) \tag{31}$$

Where σ is the black eagle's activity frequency coefficient, μ is the activity track's centre's deviation coefficient from the nest, and Q is the weight factor. And X_d is the black eagle's individual group position matrix X rearranged in accordance with the order of increasing proximity to the nest X_{best} . Notably, it observes that particle swarm diversity decreases under the common boundary restriction rule, which sets particles beyond the search space's boundary to the boundary. Consequently, it put forth the bouncing boundary rule, which states that particles that cross the boundary are redirected into it at the same distance and angle. This is how it is expressed mathematically:

$$X_j = \begin{cases} ub - X_j\%ub \text{ if } X_j > ub \\ lb - X_j\%ub \text{ if } X_j < lb \end{cases} \tag{32}$$

Thus far, it has developed a corresponding mathematical model that accounts for the behaviours of the black eagle during courtship, hatching, stalking, hovering, catching, snatching, warning, and circling. Based on this model, the BEO optimisation algorithm has been developed. The current best solution is indicated by the blue point in the schematic diagram of particle motion at each stage. Algorithm 1 presents the pseudo-code for the BEO algorithm proposed in this paper, based on the previous discussion.

Algorithm 1: The Framework of the BEO

```
Input: T: the maximum iterations.
N: the number of black eagles.
Output: Optimal position $x_{best}$ and its fitness value $f_{best}$.
while $t < T$ do
Initialize a population of black eagles.
for $j = 1$ to N do
Using Eq. (5), (6),(7),(8) update the Individual position;
end for
for $j = 1$ to N do
for $i = 1$ to d do
if $X_{best}^{t}(i) < 0.8ub(i)$ or $X_{best}^{t}(i) > 0.8lb(i)$ then
Using Eqs. (11), (15), and (16) update the Individual position;
else
Using Eqs. (23) update the Individual position;
end if
end for
end for
While $t > 0.15T$ do
for $j = 1$ to N do
Using Eq. (25) update the Individual position;
end for
end while
While $t > 0.4T$ do
for $j = 1$ to N do
Using Eq. (27),(31) update the Individual position;
end for
end while
Get the current new position.
If the new position is better than before, update it.
$t = t + 1$.
end while
return $x_{best}$ and $f_{best}$.
```

3.4. Classification Using AM-SE-CNN

3.4.1. Minimalist Attention-Based Network Architecture

Using ResNet topology, A CNN model based on lightweight attention was developed. The residual blocks 3 and 4 were where the attention modules were embedded (after the 16th layer) in this 20-layer structure. A list of the base model's specific parameters is presented in Table 1 [28]. To reduce overall network parameters and make the base network lighter and more portable, each layer's kernel filter was 4 times more compact than in the standard ResNet architecture. To further reduce network complexity, only 20 convolutional layers could be added. Conv1 layer featured sixteen kernel filters with large patch sizes after batch normalisation, a maximum pooling layer, and a rectifier linear unit (ReLU) activation layer (Max. pooling) (7×7). This resulted in feature maps that were only half the size of the input image. Blocks 1 through 4 are combined into a convolutional block consisting of three convolutional layers (conv.), a batch normalisation (BN) layer, and an activation (ReLU) layer. In contrast, the identity block comes after the convolutional block in residual blocks 2 and 3. Except for the shortcut path, the identity block's structure resembled that of the convolutional block. Before a dense output layer, a global average pooling (also known as global average pooling) layer transformed 2D feature maps into 1D. All attention modules were placed at the same location within the base network. The required maximum pooling and the spatial dimensions of both feature maps, input and output, were altered by the addition of zero-padding layers.

Table 1: Detailed ResNet20 base network architecture specifications

Block	**Sub-Block**	**Layer**	**Kernel Size, Stride and Number**	**Output Shape**
Input image		Input		$256 \times 256 \times 3$
	Convolutional block	Convolutional	$7 \times 7,2,16$	$128 \times 128 \times 16$
			$1 \times 1,2,16$	
			$3 \times 3,1,16$	
Residual block1			$1 \times 1,1,64$	$63 \times 63 \times 64$
	Shortcut		$1 \times 1,2,64$	
			$1 \times 1,2,32$	
	Convolutional block		$3 \times 3,1,32$	
			$1 \times 1,1,128$	
Residual block2	Shortcut		$1 \times 1,2,128$	$32 \times 32 \times 128$
			$1 \times 1,1,32$	
	Identity block		$3 \times 3,1,32$	
			$1 \times 1,1,128$	
			$1 \times 1,2,64$	
	Convolutional block		$3 \times 3,1,64$	$16 \times 16 \times 256$
			$1 \times 1,1,256$	
Residual block3	Shortcut		$1 \times 1,2,256$	
			$1 \times 1,1,64$	
	Identity block		$3 \times 3,1,64$	
			$1 \times 1,1,256$	
			$1 \times 1,2,256$	
Residual block4	Convolutional block		$3 \times 3,1,256$	
			$1 \times 1,1,1024$	$8 \times 8 \times 1024$
	Shortcut		$1 \times 1,2,1024$	
Global average pooling		Global average pooling		1024
Dense		Output		11

3.4.2. Convolutional Block Attention Module (CBAM)

Channel attention and spatial attention are the two attention modules used by CBAM; the spatial attention module is last. The module for channel attention produced two feature maps by utilising the maximum and average pooling layers from the intermediate layer. The shared multilayer perceptron (MLP) was then fed both feature maps, and the sigmoid function was used to combine the earlier feature maps, normalising the data. The spatial attention module was used to locate important features in the image by applying the multiplied features from the convolutional layer, along with the channel attention module. Equations (33) and (34) provide the modules for the final spatial and channel attention feature maps:

$$\mathrm{CA(x)} = \sigma(\mathrm{MLP(AvgPool(x))} + \mathrm{MLP(MaxPool(x))}) \qquad (33)$$

$$\mathrm{SA(x)} = \sigma\left(\mathrm{f}^{7\times7}\left(\left[\mathrm{F}^{\mathrm{s}}_{\mathrm{avg}}\,;\mathrm{F}^{\mathrm{s}}_{\mathrm{max}}\,\right]\right)\right) \qquad (34)$$

Wherein $\mathrm{CA(x)}$ depicts the feature maps of channel attention, $\mathrm{SA(x)}$ is the feature maps of spatial attention, σ indicates the feature maps' sigmoid function, $\mathrm{f}^{7\times7}$ the 7x7 convolutional operation; a multilayer perceptron is called an MLP, $\mathrm{AvgPool(x)}$

is the average input pooling x, $\mathrm{MaxPool(x)}$ is the greatest possible input pooling x, $\mathrm{F}^{\mathrm{s}}_{\mathrm{max}}$ does the maximum pooling technique produce the feature maps, and $\mathrm{F}^{\mathrm{s}}_{\mathrm{avg}}$ is the maximum pooling operation's feature maps.

3.4.3. Attention Module for Squeeze-and-Excitation (SE)

The input feature map's dimensions were reduced to $1 \times 1 \times \mathrm{C}$, and an excitation block was built using sigmoid activation layers, a rectifier linear unit (ReLU), two fully connected (FC) layers, and a global pooling operation. The input feature maps were forwarded to the subsequent layer, where SE feature maps were generated by pairwise element-wise multiplication [29]. Equation (35) provides a mathematical expression for the SE module's computational operation. Ultimately, the SE features were multiplied by a factor to incorporate them into the main network's feature maps:

$$\mathrm{SE(x)} = \mathrm{F_{ex}(x, W)} = \sigma\big(\mathrm{W_2}\delta(\mathrm{W_1 x})\big) \tag{35}$$

Where $\mathrm{SE(x)}$ symbolises the feature maps of squeeze and excitation, $\mathrm{F_{ex}}$ is global pooled features or squeezed features, x are the feature maps that are input, W is the SE networks' weight, σ is the sigmoid function, δ is the ReLU operation, and $\mathrm{W_1}$ and $\mathrm{W_2}$ are, respectively, the initial and secondary dense layer weights [30].

3.4.4. Self-Attention (SA) Module

The network's architecture incorporates a self-attention module. Three convolutional layers with ReLU activations were arranged in parallel to extract unique features from the input images. To create an attention map, a SoftMax layer was applied to the element-wise product of the outputs of the two convolutional layers. The third convolutional branch's feature maps were transposed, and the attention maps were multiplied to create self-attention feature maps. Ultimately, Equation (36) shows that the input and output feature maps were generated by combining scaled attention maps:

$$\mathrm{Out(x)} = \mu\mathrm{SA(x)} + \mathrm{In(x)} = \mu(\mathrm{S(x)} \cdot \mathrm{T(x)}) + \mathrm{In(x)} = \mu(\mathrm{Soft}(\mathrm{P(x)} \cdot \mathrm{Q(x)}) \cdot \mathrm{R(x)^T}) + \mathrm{In(x)} \tag{36}$$

Wherein $\mathrm{out(x)}$ is output feature maps following the module for self-attention (SA), $\mathrm{SA(x)}$ is the feature maps following the module on self-attention, $\mathrm{In(x)}$ is the input feature maps and is the softmax operation, μ is a scaling factor, $\mathrm{P(x)}, \mathrm{Q(x)}$, and $\mathrm{R(x)}$ the feature maps produced by the SA module's three convolutional paths running parallel to each other, $\mathrm{S(x)}$ following softmax operation, the feature maps and $\mathrm{T(x)}$ features maps that have been transposed from the $\mathrm{P(x)} \cdot \mathrm{Q(x)}$.

3.4.5. Dual Attention (DA) Module

For scene segmentation, the authors proposed a dual-attention approach utilising two attention networks: the channel attention (CA) and position attention (PA) networks. Apart from the layers of activation and the application of various alternative attention map generation techniques, comparable to the self-attention module, is the position attention network. The channel attention network is another feature of the DA module, which performs an addition, a SoftMax, and two multiplications. Equation (37) shows the entire set of calculations performed in the dual attention module, and the mathematical operations carried out in the PA and CA networks are provided by Equations (38) and (39):

$$\mathrm{DA(x)} = \mathrm{PA(x)} + \mathrm{CA(x)} \tag{37}$$

$$\mathrm{PA(x)} = \mathrm{U(x)} + \mathrm{In(x)} = \mathrm{R(x)}.\mathrm{S(x)} + \mathrm{In(x)} = \mathrm{Soft}\big(\mathrm{P(x)^T}.\mathrm{Q(x)}\big) + \mathrm{In(x)} \tag{38}$$

$$\mathrm{CA(x)} = \mathrm{N(x)} + \mathrm{In(x)} = \mathrm{Soft}\big(\mathrm{M(x)}\big).\mathrm{In(x)} + \mathrm{In(x)} = \mathrm{Soft}\big(\mathrm{In(x)^T}.\mathrm{In(x)}\big).\mathrm{In(x)} + \mathrm{In(x)} \tag{39}$$

Wherein $\mathrm{DA(x)}$ is the feature maps of the dual attentions, $\mathrm{PA(x)}$: feature maps of position attentions, CA (x) : feature maps of channel attentions, $\mathrm{In(x)}$: input feature maps, $\mathrm{P(x)}, \mathrm{Q(x)}$, and $\mathrm{R(x)}$: Three parallel convolutional operations' feature maps, $\mathrm{T(x)}$: transposed $\mathrm{P(x)} \cdot \mathrm{Q(x)}, \mathrm{S(x)}$: following the softmax operation, feature maps of $\mathrm{T(x)}, \mathrm{U(x)}$: multiplied feature maps of $\mathrm{S(x)}$ and $\mathrm{R(x)}, \mathrm{CA(x)}$: feature maps of channel attention, $\mathrm{N(x)}$: multiplied feature of $\mathrm{softmax(M(x))}$ and input feature maps, $\mathrm{M(x)}$: input feature maps that have been reshaped, transposed, and multiplied, and . and + are the feature maps' multiplication and element-wise product.

4. Results and Discussion

4.1. Experimental Setup

Every model was tested, verified, and trained using the same hardware and software setup. The system used was an Intel 10th-generation Core i9-10900K workstation with 64 GB DDR4 RAM, an NVIDIA RTX 2070 GPU (8 GB dedicated memory), and Windows 10 Pro, Redmond, WA, USA. This study used Python 3.8.3 (Wilmington, DE, USA); CUDA Toolkit 11.0 (Santa Clara, CA, USA); cuDNN v8.2.0 (Santa Clara, CA, USA); and Keras with TensorFlow 2.4.1 (Santa Clara, CA, USA) at the backend.

4.2. Performance Metrics

A measure of the proposed work's success is the level of acceptance it receives. It states that:

$$\text{Accuracy (ACC)} = \frac{\text{No.of correctly expressions}}{\text{Total no.of images}} \times 100 \tag{40}$$

$$\text{precision (PR)} = \frac{\text{TP}}{\text{TP+FP}} \times 100 \tag{41}$$

$$\text{F1} - \text{score (F1)} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \times 100 \tag{42}$$

$$\text{Recall (RC)} = \frac{\text{TP}}{\text{TP+FN}} \times 100 \tag{43}$$

$$\text{Specificity (SP)} = \frac{\text{TN}}{\text{TN+FP}} \times 100 \tag{44}$$

4.3. Feature Extraction Validation

Table 2 shows the validation results for ResNet feature extraction of the suggested model compared with other models.

Table 2: Feature extraction validation of the ResNet model

Models	ACC	PR	RC	SPEC	F1
AlexNet	91.61	92.13	91.15	91.33	91.23
VGG	92.52	93.24	92.47	92.55	92.34
UNet	93.33	94.57	93.33	93.62	93.26
GoogleNet	94.64	96.68	94.65	94.56	94.61
Proposed ResNet Model	95.62	95.47	95.35	95.22	95.17

Table 2 and Figure 4 illustrate the results of feature-extraction validation across different models, including AlexNet, VGG, UNet, GoogleNet, and the proposed ResNet model, for bone cancer detection.

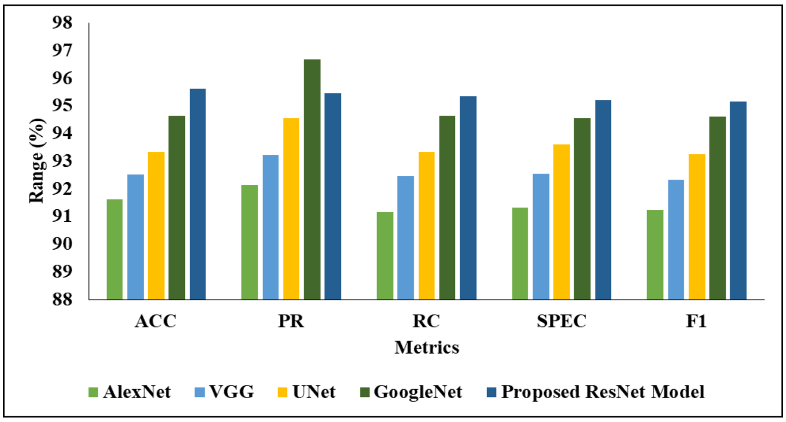

Figure 4: ResNet analysis using performance metrics

Among these models, the proposed ResNet achieves the highest accuracy (ACC) of 95.62%, demonstrating its effectiveness in extracting discriminative features from medical imaging data. It achieves precision (PR) and recall (RC) scores of 95.47% and 95.35%, respectively, indicating its ability to classify positive instances accurately while capturing true positives. The ResNet model's specificity (SPEC) is 95.22%, indicating its ability to accurately identify negative instances. Comparatively, AlexNet achieves 91.61%, VGG 92.52%, UNet 93.33%, and GoogleNet 94.64%. These results underscore the superior performance of the proposed ResNet model in feature-extraction validation for bone cancer detection, thereby demonstrating its potential to enhance diagnostic accuracy in clinical settings. Table 3 and Figure 5 present an analysis of various optimisation models applied to the proposed ResNet model for bone cancer detection. These optimisation models include ResNet-PSO, ResNet-CSO, ResNet-HBO, ResNet-LOA, and the proposed ResNet-BEO model. The ResNet-BEO model achieves the highest accuracy (ACC) of 99.62%, indicating its superior ability to classify instances. Table 3 presents the validation results for various optimisation models using the proposed ResNet feature extraction model.

Table 3: Various optimisation model analyses with the proposed ResNet model

Models	ACC	PR	RC	SPEC	F1
ResNet-PSO	92.62	92.62	92.42	92.36	92.58
ResNet-CSO	94.53	94.26	94.53	94.29	94.49
ResNet-HBO	96.27	95.93	95.43	96.45	96.53
ResNet-LOA	98.33	97.98	97.55	98.63	98.84
Proposed ResNet-BEO Model	99.62	99.47	99.35	99.21	99.17

It achieves precision (PR) and recall (RC) scores of 99.47% and 99.35%, respectively, demonstrating its ability to identify positive instances while effectively capturing true positives.

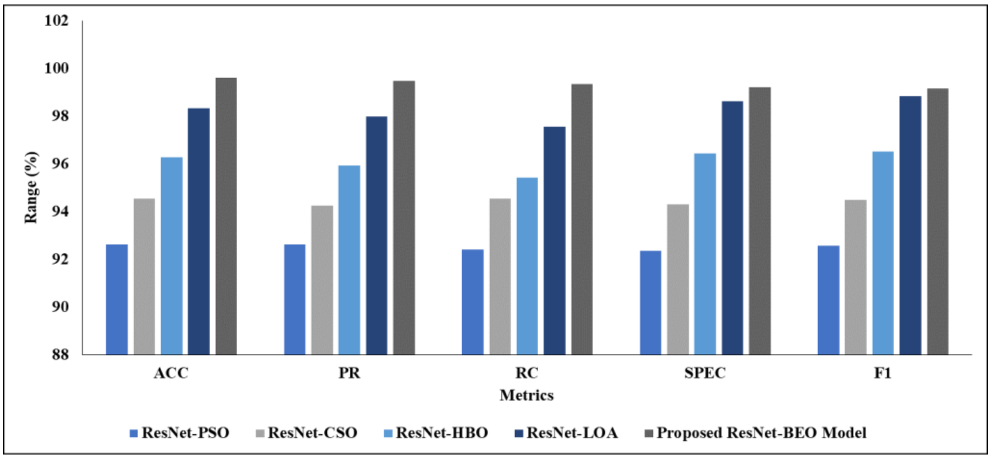

Figure 5: ResNet-BEO validation

The specificity (SPEC) of the ResNet-BEO model is 99.21%, highlighting its accuracy in identifying negative instances. In comparison, ResNet-PSO achieves 92.62%, ResNet-CSO 94.53%, ResNet-HBO 96.27%, and ResNet-LOA 98.33%. These results underscore the efficacy of the proposed ResNet-BEO model in optimising ResNet performance for bone cancer detection, demonstrating its potential to enhance diagnostic accuracy and clinical utility.

4.4. Classification Validation

Table 4 compares the training and validation results of the proposed AM-SE-CNN model with those of other existing models.

Table 4: Training analysis of the proposed AM-SE-CNN model

Models	Accuracy	Precision	Recall	F1-Score	Specificity
DBN	85.54	87.32	86.67	86.41	86.26
ANN	87.72	91.23	87.66	88.55	87.53
RNN	91.53	91.58	91.58	91.54	95.34
DNN	93.34	93.57	93.29	93.43	93.11
Proposed AM-SE-CNN model	95.25	95.46	96.38	95.12	96.22

Table 4 and Figure 6 present the training analysis results of various models, including Deep Belief Network (DBN), Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Deep Neural Network (DNN), and the proposed Attention

Mechanism-Squeeze and Excitation-CNN (AM-SE-CNN) model, for bone cancer detection. The AM-SE-CNN model outperforms all other models with an accuracy of 95.25%. It achieves 95.46% precision and 96.38% recall, indicating its superior ability to classify positive instances while minimising false positives correctly. The F1-score of the proposed model is 95.12%, reflecting its balanced performance between precision and recall. Additionally, the AM-SE-CNN model achieves 96.22% specificity, demonstrating its ability to accurately identify negative instances. Comparatively, the DBN model achieves an accuracy of 85.54%, while the ANN, RNN, and DNN models attain accuracies of 87.72%, 91.53%, and 93.34% respectively. These results underscore the effectiveness of the proposed AM-SE-CNN model in enhancing bone cancer detection performance, thus highlighting its potential for clinical applications.

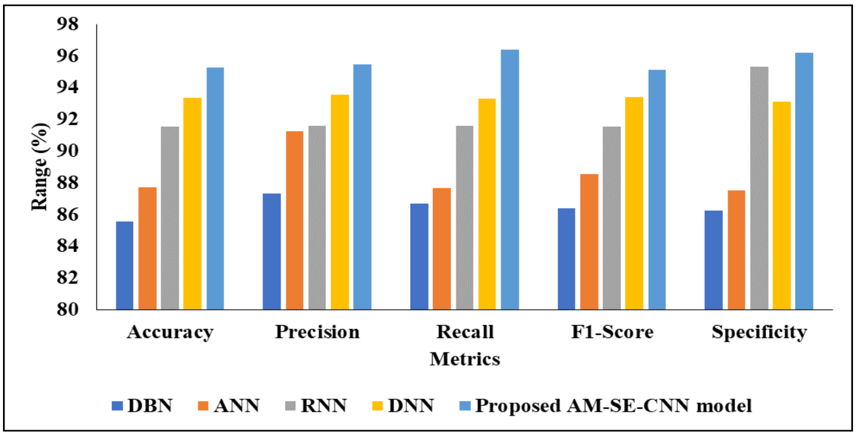

Figure 6: Training validation of the AM-SE-CNN model

Table 5 shows the validation results for the proposed AM-SE-CNN model compared with other existing models.

Table 5: Testing classification analysis

Models	ACC (%)	PR (%)	RC (%)	F1 (%)	SP (%)
DBN	95.63	95.38	95.28	95.28	95.14
ANN	96.74	96.89	96.78	96.76	96.66
RNN	97.47	97.47	97.38	97.38	97.42
DNN	98.22	98.97	98.88	98.82	98.75
Proposed AM-SE-CNN model	99.62	99.47	99.35	99.21	99.17

Table 5 and Figure 7 present the testing classification analysis results of different models, including Deep Belief Network (DBN), Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Deep Neural Network (DNN), and the proposed Attention Mechanism-Squeeze and Excitation-CNN (AM-SE-CNN) model, for bone cancer detection.

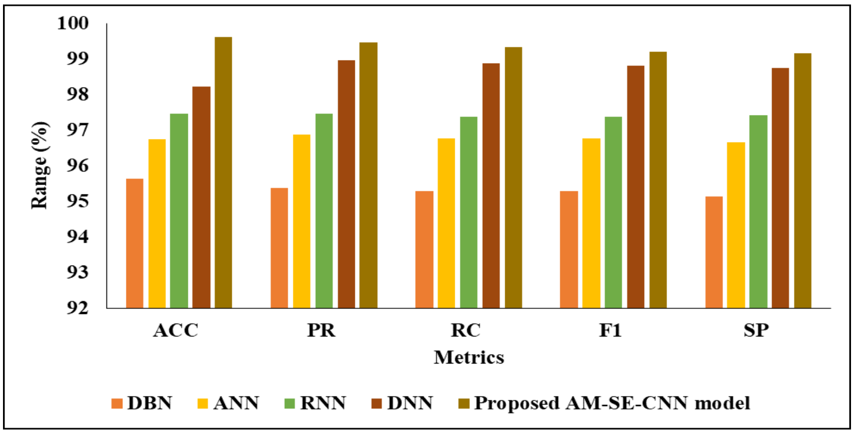

Figure 7: Testing validation of the AM-SE-CNN model

The AM-SE-CNN model achieves an accuracy (ACC) of 99.62%, demonstrating its ability to correctly classify instances. It achieves precision (PR) and recall (RC) scores of 99.47% and 99.35% respectively, reflecting its high precision in positive predictions and ability to capture true positive instances. The F1-score of the proposed model is 99.21%, indicating a balanced trade-off between precision and recall. Additionally, the AM-SE-CNN model's specificity (SP) is 99.17%, indicating its ability to identify negative instances accurately. In comparison, the DBN, ANN, RNN, and DNN models achieve accuracies of 95.63%, 96.74%, 97.47%, and 98.22% respectively. These results underscore the effectiveness of the proposed AM-SE-CNN model in improving bone cancer detection performance on test data, demonstrating its potential for reliable clinical deployment.

5. Conclusion

This research proposes a unique and complete methodology for bone cancer diagnosis by merging advanced techniques in image preprocessing, deep feature extraction, hyperparameter optimisation, and intelligent classification. The preprocessing stage plays a vital role in improving the quality of medical images by suppressing noise and enhancing structural clarity, thereby providing reliable inputs for subsequent analysis. This procedure greatly contributes to maintaining diagnostically useful information while avoiding artefacts that may limit accurate detection. For feature extraction, the proposed system leverages the ResNet architecture, widely known for its strong deep learning capabilities and effectiveness at learning hierarchical representations from complex image data. To further boost model performance, the ResNet network's hyperparameters are tuned using the Black Eagle Optimiser (BEO). This nature-inspired metaheuristic efficiently explores the search space to maximise detection accuracy. This optimisation technique improves convergence, reduces overfitting, and enhances generalisation. The classification stage employs an AM-SE-CNN model that integrates attention mechanisms and squeeze-and-excitation-based feature recalibration. This architecture enables the model to focus on the most informative parts of the input images while adaptively stressing crucial feature channels, hence enhancing discriminative power and classification robustness. The synergy between improved deep features and attention-driven categorisation leads to superior diagnostic performance. An experimental evaluation conducted on standard benchmark datasets indicates the usefulness of the proposed approach. The model achieves remarkable performance metrics, including 99.62% accuracy, 99.47% precision, 99.35% recall, 99.21% specificity, and 99.17% F1 score, exceeding the state of the art in bone cancer detection. These results illustrate the robustness, reproducibility, and therapeutic significance of the proposed paradigm. Future studies will focus on rigorous validation using larger, more diverse datasets to test generalisability and resilience better. Such efforts are important for fully understanding the technique's clinical usefulness and for advancing its translation into real-world diagnostic tools for early and accurate bone cancer detection.

Acknowledgement: The authors gratefully acknowledge New Horizon College of Engineering, Quest Technologies Trichy Research Labs, Saranathan College of Engineering, and the University of Illinois Urbana-Champaign for their collaborative assistance. Their dedicated support played a significant role in shaping the outcomes of this study.

Data Availability Statement: The datasets supporting this study, Bone Cancer Detection Using ResNet–BEO Feature Extraction and AM-SE-CNN Classification, are publicly available.

Funding Statement: The authors confirm that this research was conducted without any external financial support or funding assistance.

Conflicts of Interest Statement: The authors declare that they have no competing interests related to this work. All sources of information, citations, and references have been properly acknowledged.

Ethics and Consent Statement: Ethical procedures were strictly followed in this study. Approval was obtained from the relevant organisation, and informed consent was collected from all participating individuals before data acquisition.

References

1. D. Shrivastava, S. Sanyal, A. K. Maji, and D. Kandar, “Bone cancer detection using machine learning techniques,” in Smart Healthcare for Disease Diagnosis and Prevention, *Academic Press*, Massachusetts, United States of America, 2020.
2. M. Avula, N. P. Lakkakula, and M. P. Raja, “Bone cancer detection from MRI scan imagery using mean pixel intensity,” in *2014 8th Asia Modelling Symposium,* Taipei, Taiwan, 2014.
3. B. Jabber, M. Shankar, P. V. Rao, A. Krishna, and C. Z. Basha, “SVM model based computerized bone cancer detection,” in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA),* Tamil Nadu, India, 2020.
4. M. M. Ranjitha, N. L. Taranath, C. N. Arpitha, and C. K. Subbaraya, “Bone cancer detection using K-means

segmentation and KNN classification," in *2019 1st International Conference on Advances in Information Technology (ICAIT). IEEE,* Karnataka, India, 2019.

5. D. Kumar, N. Jain, A. Khurana, S. Mittal, S. C. Satapathy, R. Senkerik, and J. D. Hemanth, "Automatic detection of white blood cancer from bone marrow microscopic images using convolutional neural networks," *IEEE Access*, vol. 8, no. 7, pp. 142521–142531, 2020.
6. R. Almajalid, J. Shan, M. Zhang, G. Stonis, and M. Zhang, "Knee bone segmentation on three-dimensional MRI," in *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE*, Florida, United States of America, 2019.
7. F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative," *Medical Image Analysis*, vol. 52, no. 3, pp. 109–118, 2019.
8. F. Kordon, P. Fischer, M. Privalov, B. Swartman, M. Schnetzke, J. Franke, R. Lasowski, A. Maier, and H. Kunze, "Multi-task localization and segmentation for X-ray guided planning in knee surgery," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference,* Shenzhen, China, 2019.
9. S. J. MacDessi, W. Griffiths-Jones, I. A. Harris, J. Bellemans, and D. B. Chen, "Coronal plane alignment of the knee (CPAK) classification: A new system for describing knee phenotypes," *The Bone and Joint Journal,* vol. 103, no. 2, pp. 329–337, 2021.
10. R. S. Hegadi, D. I. Navale, T. D. Pawar, and D. D. Ruikar, "Multi feature-based classification of osteoarthritis in knee joint X-ray images," in *Medical Imaging: Artificial Intelligence, Image Recognition, and Machine Learning Techniques, CRC Press,* Florida, United States of America, 2019.
11. N. T. Do, S. T. Jung, H. J. Yang, and S. H. Kim, "Multi-level Seg-UNet model with global and patch-based X-ray images for knee bone tumor detection," *Diagnostics*, vol. 11, no. 4, pp. 1–22, 2021.
12. H. S. Gan, M. H. Ramlee, A. A. Wahab, Y. S. Lee, and A. Shimizu, "From classical to deep learning: Review on cartilage and bone segmentation techniques in knee osteoarthritis research," *Artificial Intelligence Review*, vol. 54, no. 4, pp. 2445–2494, 2021.
13. S. M. Ahmed and R. J. Mstafa, "A comprehensive survey on bone segmentation techniques in knee osteoarthritis research: From conventional methods to deep learning," *Diagnostics*, vol. 12, no. 3, pp. 1–26, 2022.
14. Y. Hu, J. Tang, S. Zhao, and Y. Li, "Deep learning-based multimodal 3T MRI for the diagnosis of knee osteoarthritis," *Computational and Mathematical Methods in Medicine*, vol. 2022, no. 5, pp. 1–13, 2022.
15. S. Breden, F. Hinterwimmer, S. Consalvo, J. Neumann, C. Knebel, R. von Eisenhart-Rothe, R. H. Burgkart, and U. Lenze, "Deep learning-based detection of bone tumors around the knee in X-rays of children," *Journal of Clinical Medicine*, vol. 12, no. 18, pp. 1–9, 2023.
16. S. Y. Kim-Wang, P. X. Bradley, H. C. Cutcliffe, A. T. Collins, B. S. Crook, C. S. Paranjape, C. E. Spritzer, and L. E. DeFrate, "Auto-segmentation of the tibia and femur from knee MR images via deep learning and its application to cartilage strain and recovery," *Journal of Biomechanics*, vol. 149, no. 3, p. 111473, 2023.
17. X. Zhan, J. Liu, H. Long, J. Zhu, H. Tang, F. Gou, and J. Wu, "An intelligent auxiliary framework for bone malignant tumor lesion segmentation in medical image analysis," *Diagnostics*, vol. 13, no. 2, pp. 1–22, 2023.
18. Y. Y. Chen, P. N. Yu, Y. C. Lai, T. C. Hsieh, and D. C. Cheng, "Bone metastases lesion segmentation on breast cancer bone scan images with negative sample training," *Diagnostics*, vol. 13, no. 19, pp. 1–19, 2023.
19. X. Li, S. Lv, M. Li, J. Zhang, Y. Jiang, Y. Qin, H. Luo, and S. Yin, "SDMT: Spatial dependence multi-task transformer network for 3D knee MRI segmentation and landmark localization," *IEEE Transactions on Medical Imaging*, vol. 42, no. 8, pp. 2274–2285, 2023.
20. K. M. Chung, H. Yu, J. -H. Kim, J. J. Lee, J. -H. Sohn, S. -H. Lee, J. H. Sung, S. -W. Han, J. S. Yang, and C. Kim, "Deep learning-based knee MRI classification for common peroneal nerve palsy with foot drop," *Biomedicines*, vol. 11, no. 12, pp. 1–9, 2023.
21. C. Pan, L. Lian, J. Chen, and R. Huang, "FemurTumorNet: Bone tumor classification in the proximal femur using DenseNet model based on radiographs," *Journal of Bone Oncology*, vol. 42, no. 10, pp. 1–7, 2023.
22. V. A. Georgeanu, M. Mămuleanu, S. Ghiea, and D. Selișteanu, "Malignant bone tumors diagnosis using magnetic resonance imaging based on deep learning algorithms," *Medicina*, vol. 58, no. 5, pp. 1–16, 2022.
23. A. Sharma, D. P. Yadav, H. Garg, M. Kumar, B. Sharma, and D. Koundal, "Bone cancer detection using feature extraction-based machine learning model," *Computational and Mathematical Methods in Medicine*, vol. 2021, no. 2, pp. 1–13, 2021.
24. G. Suganeshwari, R. Balakumar, K. Karuppanan, S. B. Prathiba, S. Anbalagan, and G. Raja, "DTBV: A deep transfer-based bone cancer diagnosis system using VGG16 feature extraction," *Diagnostics*, vol. 13, no. 4, pp. 1–12, 2023.
25. M. A. A. Walid, S. Mollick, P. C. Shill, M. K. Baowaly, M. R. Islam, M. M. Ahamad, M. A. Othman, and M. A. Samad, "Adapted deep ensemble learning-based voting classifier for osteosarcoma cancer classification," *Diagnostics,* vol. 13, no. 19, pp. 1–25, 2023.
26. M. Ahmed, K. Kamal, T. A. H. Ratlamwala, G. Hussain, M. Alqahtani, M. Alkahtani, M. Alatefi, and A. Alzabidi, "Tool health monitoring of a milling process using acoustic emissions and a ResNet deep learning model," *Sensors*,

vol. 23, no. 6, pp. 1–18, 2023.

27. H. Zhang, H. San, J. Chen, H. Sun, L. Ding, and X. Wu, “Black eagle optimizer: A metaheuristic optimization method for solving engineering optimization problems,” *Cluster Computing*, vol. 27, no. 9, pp. 12361–12393, 2024.
28. A. Bhujel, N. E. Kim, E. Arulmozhi, J. K. Basak, and H. T. Kim, “A lightweight attention-based convolutional neural networks for tomato leaf disease classification,” *Agriculture*, vol. 12, no. 2, pp. 1–18, 2022.
29. T. M. Aruna, P. Kumar, N. N. Srinidhi, G. N. Divyaraj, K. Asha, A. Thirumalraj, and E. Naresh, “Effective utilisation of geospatial data for peer-to-peer communication among autonomous vehicles using optimized machine learning algorithm,” *Research Square*, 2023. Available: https://www.researchsquare.com/article/rs-3698525/v1 [Accessed by 12/08/2024].
30. A. Thirumalraj and T. Rajesh, “An improved ARO model for task offloading in vehicular cloud computing in VANET,” *Research Square*, 2023. Available: https://assets-eu.researchsquare.com/files/rs-3291507/v1_covered_2b8 de597-e0d0-42be-b5d7-f6de74e7356f.pdf [Accessed by 22/08/2024].