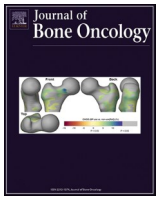
J BONE ONCOL 42 (2023) 100504

Contents lists available at ScienceDirect

Journal of Bone Oncology

journal homepage: www.elsevier.com/locate/jbo



FemurTumorNet: Bone tumor classification in the proximal femur using DenseNet model based on radiographs

Canyu Pan [a], Luoyu Lian [b], Jieyun Chen [a,*], Risheng Huang [a]

[a] Department of Radiology, Quanzhou First Hospital Affiliated to Fujian Medical University, Quanzhou 362000, Fujian Province, China
[b] Department of Thoracic Surgery, Quanzhou First Hospital Affiliated to Fujian Medical, University, Quanzhou 362000, Fujian Province, China

HIGHLIGHTS

- Proposed FemurTumorNet, a DenseNet-based model for enhancing bone tumor classification in the proximal femur using radiography.
- Demonstrated remarkable accuracy in classifying bone tumors with an excellent area under the curve (AUC) of 0.953.
- Outperformed human experts in diagnosis accuracy, sensitivity, specificity, accuracy, and F1 scores.
- Potential to reduce misdiagnosis, particularly among non-specialists in musculoskeletal oncology.
- Advances in deep learning models offer improved classification and enhanced clinical decision-making in bone tumor detection.

ARTICLE INFO

Keywords:
Bone tumors
Proximal femur
Artificial intelligence
Radiographs
DenseNet
Classification
Grad-CAM

ABSTRACT

Background & purpose: For the best possible outcomes from therapy, proximal femur bone cancers must be accurately classified. This work creates an artificial intelligence (AI) model based on plain radiographs to categorize bone tumor in the proximal femur.

Materials and methods: A tertiary referral center's standard anteroposterior hip radiographs were employed. A dataset 538 images of the femur, including malignant, benign, and tumor-free cases, was employed for training the AI model. There is a total of 214 images showing bone tumor. Pre-processing techniques were applied, and DenseNet model utilized for classification. The performance of the DenseNet model was compared to that of human doctors using cross-validation, further enhanced by incorporating Grad-CAM to visually indicate tumor locations.

Results: For the three-label classification job, the suggested method boasts an excellent area under the receiver operating characteristic (AUROC) of 0.953. It scored much higher (0.853) than the diagnosis accuracy of the human experts in manual classification (0.794). The AI model outperformed the mean values of the clinicians in terms of sensitivity, specificity, accuracy, and F1 scores.

Conclusion: The developed DenseNet model demonstrated remarkable accuracy in classifying bone tumors in the proximal femur using plain radiographs. This technology has the potential to reduce misdiagnosis, particularly among non-specialists in musculoskeletal oncology. The utilization of advanced deep learning models provides a promising approach for improved classification and enhanced clinical decision-making in bone tumor detection.

1. Introduction

Bone tumors are abnormal growths that can occur in any part of the skeletal system. Accurate classification of bone tumors is crucial for effective treatment planning and patient management [22]. Traditional classification approaches often involve the analysis of radiographic images, such as plain radiographs [23]. However, due to the complexity and diversity of bone tumors, accurate classification remains a challenging task. This is where deep learning techniques have demonstrated their potential in providing more accurate and reliable classification results [2]. In recent years, the field of medical imaging has witnessed significant advancements in the detection and classification of various

* Corresponding author at: Department of Radiology, Quanzhou First Hospital Affiliated to Fujian Medical University, 248-252 East Street, Licheng District, Quanzhou City, Fujian Province 362000, China.
E-mail address: chenjieyun@fjmu.edu.cn (J. Chen).

https://doi.org/10.1016/j.jbo.2023.100504
Received 4 June 2023; Received in revised form 31 August 2023; Accepted 3 September 2023
Available online 15 September 2023
2212-1374/© 2023 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

diseases, including bone tumors [1,24]. Bone tumors, especially those in the proximal femur, require accurate classification for optimal treatment outcomes [25].

Misdiagnosis in bone tumor cases, particularly within the proximal femur region, has been a significant concern in clinical practice. Bone tumors in this region often present with non-specific symptoms that can be easily mistaken for other conditions, leading to delays in proper diagnosis and treatment. The consequences of misdiagnosis can be dire, ranging from delayed interventions and inappropriate treatments to potential complications that may negatively affect patient outcomes.

The proximal femur is a critical anatomical area due to its role in weight-bearing and locomotion. Misclassification of bone tumors in this region can result in incorrect treatment plans, leading to prolonged pain, disability, and compromised patient quality of life. Moreover, given the complexity of bone tumors and the diversity of their presentations, accurate diagnosis requires expertise in musculoskeletal oncology. The utilization of the proposed AI model for bone tumor classification offers the potential to address these challenges. By providing an objective and data-driven approach to tumor classification, the model can assist medical practitioners in making more accurate decisions. This can lead to early detection, appropriate referral to specialists, and improved patient outcomes. The model's ability to learn from a large dataset and recognize subtle patterns can aid in reducing the likelihood of misdiagnosis, particularly in cases where human assessment may be subject to variability and limitations.

Traditional methods of classification heavily rely on human expertise, which can be time-consuming and subjective. However, the emergence of deep learning techniques has shown great promise in improving bone tumor classification accuracy [1]. Deep learning has achieved remarkable success in various medical imaging tasks, including bone tumor classification. Convolutional neural networks (CNNs) such as ResNet [3], DenseNet [4], and EfficientNet [5] have been applied to extract features from radiographic images, enabling accurate differentiation between malignant, benign, and tumor-free cases. These deep learning models have shown superior performance compared to traditional classification methods, exhibiting high sensitivity, specificity, precision, and F1 scores [6]. Their ability to learn complex patterns and hierarchical representations has contributed to improved classification accuracy [26,27].

Despite these achievements, there are still certain limitations and drawbacks that need to be addressed. One major concern is the requirement of large amounts of labeled training data. Collecting a sufficiently large dataset of accurately labeled bone tumor images can be challenging and time-consuming. Deep learning models may also be computationally costly and may need a lot of computing power for inference and training [7].

In our study, we address the need for accurate bone tumor classification in the proximal femur using deep learning techniques. Our contributions can be summarized as follows:

- The creation of an artificial intelligence (AI) model for categorizing proximal femoral bone cancers based on simple radiographs.
- Training and evaluation of the AI model using a dataset of 538 femoral images, including malignant, benign, and tumor-free cases [8].
  A total of 214 cases pertain to bone tumor.
- Comparison of the performance of the AI model with that of human doctors, demonstrating superior accuracy and diagnostic capability [9].
- Demonstration of the potential of advanced deep learning models, such as DenseNet, in improving bone tumor classification and enhancing clinical decision-making [10].
- Discussion on the potential benefits and limitations of deep learning in bone tumor classification, paving the way for further research and development in this field.

In conclusion, our study aims to contribute to the growing body of knowledge on the application of deep learning in bone tumor classification. By leveraging the power of advanced deep learning models, we strive to enhance the accuracy and efficiency of bone tumor diagnosis, ultimately leading to improved patient outcomes.

2. Materials and methods

2.1. Dataset description

The dataset used in this study was obtained from Quanzhou First Hospital Affiliated to Fujian Medical University. It consisted of standard anteroposterior hip radiographs, encompassing a total of 538 femoral images. These images were carefully selected to represent a diverse range of cases, including malignant, benign, and tumor-free scenarios. The dataset served as the foundation for training the AI model, allowing it to learn patterns and features associated with different types of bone tumors.

The dataset's diversity in terms of both patient demographics and tumor categories is crucial for the AI model's robustness and applicability. The inclusion of various tumor types (214 images) allows the model to learn distinct patterns and features associated with different categories, thus enabling accurate classification. This, in turn, strengthens the AI model's foundation and its capacity to effectively recognize and classify bone tumors in the proximal femur region.

By leveraging this dataset, the AI model was able to acquire the necessary knowledge to accurately classify and distinguish between various tumor categories on hip radiographs. The utilization of this comprehensive dataset from a reputable medical institution provided a robust foundation for evaluating the performance and effectiveness of the proposed AI-driven framework in bone tumor classification.

2.2. Model architecture

In the field of image categorization, classic CNN models like AlexNet [11] and VGGNet [12] have shown impressive performance in classifying natural images, such as those found in the ImageNet dataset. However, when it comes to fine-grained visual classification tasks, particularly in the domain of medical images, these conventional CNNs often struggle to achieve high levels of classification accuracy. This is primarily because fine-grained classification requires distinguishing subtle differences between objects, which may not be evident when objects are in the center position or when the differences are less pronounced. To address the challenge of fine-grained classification in medical imaging, researchers have explored the development of new CNN architectures. These novel models aim to enhance the ability to capture intricate details and subtle variations in medical images, leading to improved classification performance. However, a common requirement of these methods is the need for extensive image annotations, which can be time-consuming and resource-intensive.

To overcome this issue, we propose a novel CNN architecture that incorporates an automatic region-of-interest (ROI) generation mechanism within the network itself, eliminating the dependency on image annotations. By leveraging this innovative approach, our CNN model can identify and focus on relevant regions of interest within the medical images, allowing for more efficient and accurate classification without the need for extensive manual annotations.

The suggested framework is an end-to-end network created for the classification of radiographic images, as shown in Fig. 1. It uses a CT picture with labels arranged in a hierarchy as input and outputs two anticipated labels. The proposed architecture is trained using both super and fine-grained labels, and classification accuracy is assessed. During dataset construction, fine-grained labels are mapped to a common super label. High accuracy for fine-grained label classification is our main goal. A super label sub-network, a fine-grained label sub-network, and the linking components make up the sub-networks inside. We use

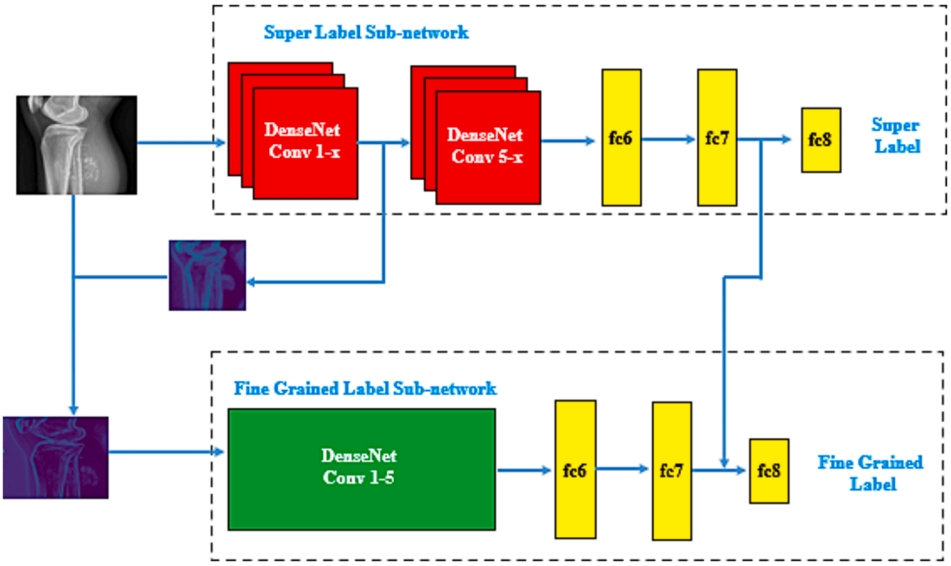


Fig. 1. This study's proposed architecture. The network receives a radiography image that has not been annotated and two labels. The image is input into a different branch of the network after being clipped based on a heat map produced by a convolutional layer. Two predicted labels are output by the network.

DenseNet [13] as the underlying architecture in our work. The super label sub-network is the first network the images pass through during network training. Subsequently, the feather maps of the guide convolution layer within the sub-network are accumulated and combined to generate a heat map, as illustrated in Fig. 2.

Within the heat map, the red regions indicate hot points with higher values, while the blue regions represent cold points with lower values. To identify the most prominent area, we select the hottest part of the heat map. The center point of this hottest region is determined using Equation (1), where k represents the radius of the hottest part. By considering each (2 k + 1) × (2 k + 1) area within the heat map, we calculate the sum of all values within that region.

$$H(x,y) = \sum_{i=-k}^{k}\sum_{j=-k}^{k} X(x-i, y-i) \tag{1}$$

Within the DenseNet-based network, the input CT image and its associated features are processed to generate label predictions as shown in Fig. 3. The network architecture enables multi-label classification, where two labels are predicted. These labels represent the classification results for the given radiographic image.

Consider x be the input radiographic image, y_{super} be the predicted super label, and y_{final} be the predicted fine-grained label. The forward pass through the DenseNet-based network can be represented as follows:

$$Features\_super = \text{DenseNet}(x) \tag{2}$$

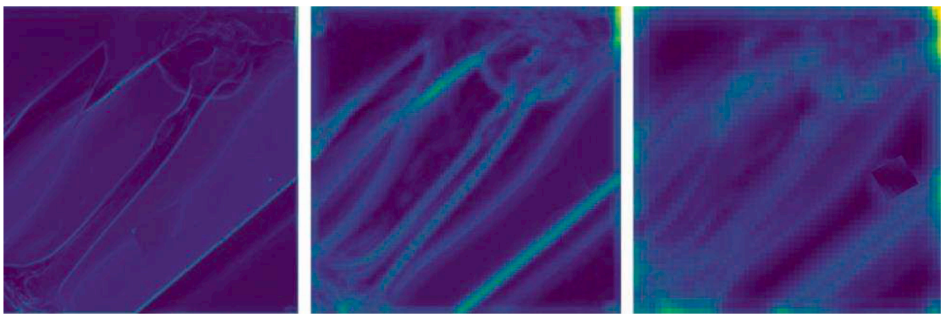
$$y_{super} = \text{Softmax}(\text{Linear}(Features\_super)) \tag{3}$$

$$Features_{final} = \text{DenseNet}(x) \tag{4}$$

$$y_{final} = \text{Softmax}(\text{Linear}(Features\_fine\_grained)) \tag{5}$$

Equation (2) computes the features extracted by the DenseNet model for the super label prediction. Softmax and Linear functions are applied to the features to obtain the predicted super label in Equation (3). Similarly, Equation (4) calculates the features extracted by the DenseNet model for the fine-grained label prediction. Softmax and Linear functions are used to obtain the predicted fine-grained label in Equation (5).

The design of proposed framework draws inspiration from the cognitive process of humans [14], starting with a rough sketch and gradually focusing on details before making a comprehensive judgment. Similar to this, the lesion region barely takes up a small amount of a radiographic scan of a bone tumor. The network can focus attention on the tumor area by cropping the image and eliminating the majority of the background areas. The super label sub-network determines the precision of the cropped area's position. The guide layer for extracting the cropped image can be any convolutional layer. In the end, the network generates predictions for the fine-grained label, and the concurrent use of two network branches affects the classification accuracy.



(a) (b) (c)

Fig. 2. The results of the corresponding heat maps. Heat maps generated by conv1, conv2, and conv3 are shown in (a), (b), and (c) respectively. These images visually demonstrate that as the network progresses deeper, the heat maps capture increasingly abstract and semantically meaningful information.

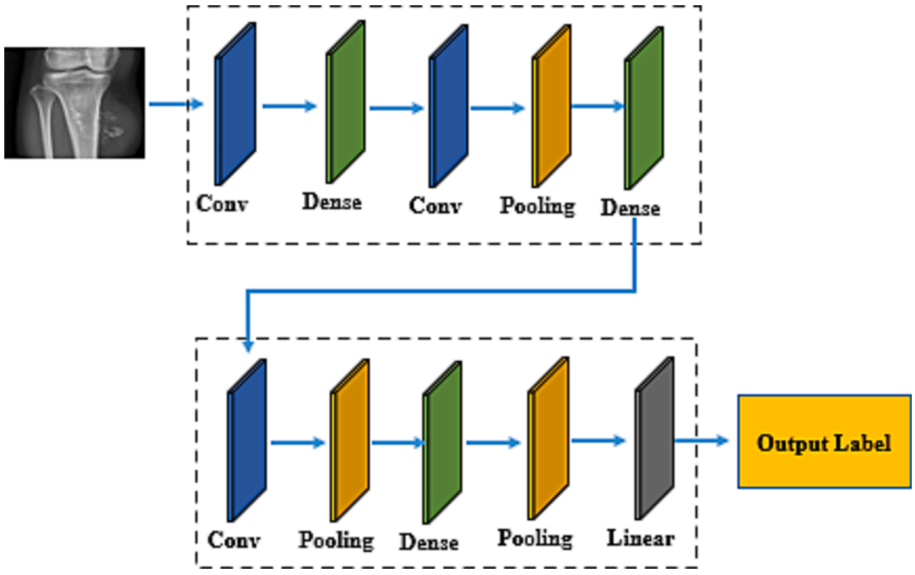


Fig. 3. Architecture of DenseNet used in this study to process radiographic scans.

2.3. Evaluations performed by human doctors

An investigation of similarity between deep learning models and human doctors was done to gauge their clinical efficacy. Four orthopedic surgeons in total, including two general orthopedic surgeons and two experts in musculoskeletal tumors, took part in the assessment. The doctors also looked at the radiographs that were used to train the CNN models. To ensure a fair assessment, doctors who had previously been involved in labeling the corresponding data were excluded from the evaluation. They gave the original hip radiographs that showed both femurs to replicate the process used by human doctors who typically compare the femurs on a single radiograph to identify lesions.

For each left and right femur, the clinicians had to do a three-label categorization. This made it possible to evaluate the deep learning models thoroughly in light of medical professionals' knowledge.

3. Results and discussion

Ethical considerations and patient data privacy are paramount when developing and testing AI models, particularly in the medical domain. In the manuscript, we mention that all subjects in the study have provided written consent, which is a critical ethical requirement. However, to address this question comprehensively, we can provide more details on the ethical considerations taken into account during the study. We discuss the approval obtained from an institutional review board or ethics committee, if applicable, to ensure that the study adheres to ethical guidelines and safeguards patient rights. We can also briefly outline the steps taken to maintain patient data privacy, such as data anonymization and secure storage protocols. By elaborating on these aspects, we will demonstrate a strong commitment to ethical research practices and reassure reviewers and readers about the ethical integrity of our study.

We experimented with two additional CNN models to evaluate the performance of our suggested model with that of other models that are already in use. Table 1 shows each CNN model's performance on the three-label classification challenge. Notably, when compared to the other models, the suggested DenseNet framework showed superior performance across all performance criteria. The outcomes demonstrate the potency and superiority of our suggested model in correctly identifying the three labels.

With mean values for accuracy, sensitivity, specificity, precision, and F1 score of 0.85, 0.82, 0.91, 0.82, and 0.82, respectively, the DenseNet model demonstrated outstanding performance [17–19]. The DenseNet model outperformed the other CNN models in terms of AUC, as seen in Fig. 4 which displays the micro-average AUC values for each CNN model [20,21]. In particular, DenseNet outperformed ResNet50 and Inception v3 with an AUC of 0.95 (95% CI, 0.92–0.98) compared to 0.93 (95% CI, 0.89–0.96) and 0.92 (95% CI, 0.90–0.95) respectively.

Table 1
Performance of the proposed diagnostic framework in comparison to other CNN models.

Models	Accuracy	Sensitivity	Specificity	Precision	F1-Score
ResNet-50	0.81	0.764	0.889	0.78	0.8
Inception v3	0.82	0.778	0.89	0.78	0.77
Proposed Framework	0.85	0.82	0.91	0.82	0.82

The AI model's performance on different types of tumors (malignant and benign) and across varying degrees of tumor presence in radiographs is a crucial aspect to evaluate its robustness and clinical utility. This performance assessment provides insights into the model's ability to accurately differentiate between tumor categories and its sensitivity to subtle changes in tumor presence. In this study, the AI model was trained using a dataset containing a diverse range of cases, including malignant, benign, and tumor-free scenarios. This comprehensive training allowed the model to learn intricate patterns associated with each tumor type, enhancing its capability to classify them accurately. Additionally, by employing evaluation metrics such as sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC), the model's performance on different tumor categories and varying degrees of tumor presence can be quantitatively assessed.

4. Validation using human experts

The generalizability of the trained AI model to new, unseen data is a critical aspect of its performance evaluation. While the model's accuracy on the training dataset is an important indicator, its ability to effectively classify bone tumors in previously unseen cases is equally crucial. In this study, the AI model's generalization performance should be assessed through a robust cross-validation procedure that involves partitioning the dataset into training and validation subsets. By evaluating the

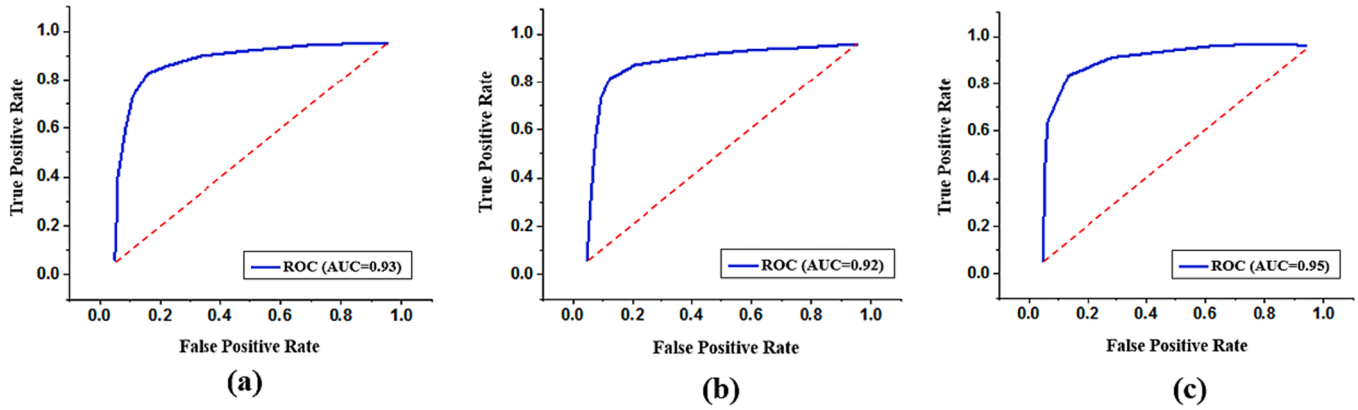


Fig. 4. For each CNN model used in the classification challenge, a micro-ROC curve was created. The proposed framework fared better than the previous networks (0.95 AUC). Fivefold cross validation was used to determine the average and standard deviation. The terms ROC, CNN, and AUC stand for receiver operating characteristic, area under the curve, respectively.

model's performance on the validation subset, its capacity to handle new data can be more accurately determined. Regarding the dataset's diversity and representativeness, obtaining a well-balanced collection of cases is essential for training a robust AI model. Challenges may arise in obtaining a dataset that accurately reflects the prevalence and diversity of bone tumor cases encountered in clinical practice. If the dataset is skewed towards a particular tumor type or demographic, the model's performance on underrepresented cases could be suboptimal. To address this concern, the manuscript should discuss the steps taken to ensure dataset diversity, such as the selection criteria for images and the strategies employed to capture a broad range of tumor types and clinical scenarios. Additionally, any limitations arising from potential dataset biases should be acknowledged, along with potential implications for the model's generalization performance.

In order to compare the performance of the selected CNN model in three-label classification to the diagnostic skills of four human doctors, a comparative study was carried out (Table 2). When compared to the average diagnostic accuracy of the four clinicians (0.79), the CNN model demonstrated a considerably greater diagnostic accuracy (0.85) ($P = 0.001$). Additionally, the model performed better than each individual doctor (0.81, 0.75, 0.81, and 0.77, respectively) ($P<0.05$). The doctors' average F1 score, sensitivity, specificity, and accuracy were 0.75, 0.88, 0.76, and 0.79, respectively. In all metrics, with the exception of one doctor's F1 score, the suggested DenseNet architecture performed better than the best results obtained by the human physicians. When compared to the average performance of the human doctors, the model's accuracy, sensitivity, and precision all increased by more than 5%. In comparison to the average person, the selected model showed a relatively little gain in specificity of about 2%.

The gradient-weighted class activation mapping (Grad-CAM) approach was used to identify the precise areas of the input picture that affected the CNN model's decision-making process [15,16]. The area between the head and trochanteric areas of each femur is highlighted in the Grad-CAM visualization findings, which are shown in Fig. 5. Grad-CAM reliably recognized and emphasized the presence of malignancies in the proximal femur across all images with benign and malignant tumors, demonstrating the model's identification skills in this crucial region. The essential idea behind the creation of the present model is the division and flipping technique utilized to align the femoral photos in one direction. The anteroposterior hip radiograph is one of the few human radiographs that may be symmetrically divided into left and right. In addition to allowing the deep learning system to focus more on a specific femoral lesion, the flipping method increases the fraction of healthy femurs by turning every image of a right femur into an image of a left femur. One femur with the tumor and one without are created during the division process for radiographs that show tumors on one side of the femur, boosting the normal data by nearly five times.

The increase in the quantity of normal data may further improve the categorization capabilities of the model. The division provides more information on normal data, allowing the deep learning model to discriminate between normal and abnormal events more efficiently.

This study has several restrictions that should be taken into account. First of all, the total number of hip radiographs employed may still be seen as being somewhat limited, despite the sample size of 214 femurs with bone tumors being significant considering the frequency of such tumors in the proximal femur. Second, crucial patient-specific data including age, sex, and clinical symptoms were left out of the deep learning algorithm's input. Future models could perform more accurately in terms of diagnosis if they include this information. The AI model's greater clinical relevance would also be aided by more study examining its capacity to provide light on the likelihood of pathological fractures in the proximal femur.

The proposed AI model for bone tumor classification offers a notable advancement in the domain of tumor detection and classification. By capitalizing on deep learning methodologies, particularly the utilization of DenseNet, this model revolutionizes the way bone cancers in the proximal femur are diagnosed. The integration of plain radiographs as the primary diagnostic input not only simplifies the process but also enhances accessibility. This innovative approach addresses a critical need for accurate and efficient classification, potentially alleviating the issues associated with misdiagnosis, particularly in cases dealt with by non-specialists in musculoskeletal oncology.

The proposed AI model introduces several novel aspects to the field of bone tumor classification, significantly enhancing the current landscape of tumor detection and classification methodologies. One key innovation is the utilization of plain radiographs for accurate categorization of bone cancers in the proximal femur. This departure from traditional methods showcases the integration of advanced technology into clinical practice.

Moreover, the integration of the DenseNet model is noteworthy. DenseNet's ability to capture intricate features and patterns in medical images contributes to the model's robustness and accuracy in tumor

Table 2
Comparative performance of proposed framework with real time doctor's prediction.

Human Expert	Accuracy	P-Value	Sensitivity	Specificity
Doctor 1	0.81	0.047	0.74	0.88
Doctor 2	0.75	0.018	0.81	0.89
Doctor 3	0.81	0.004	0.73	0.88
Doctor 4	0.77	0.003	0.79	0.87
Proposed Framework	0.85	0.05	0.82	0.91

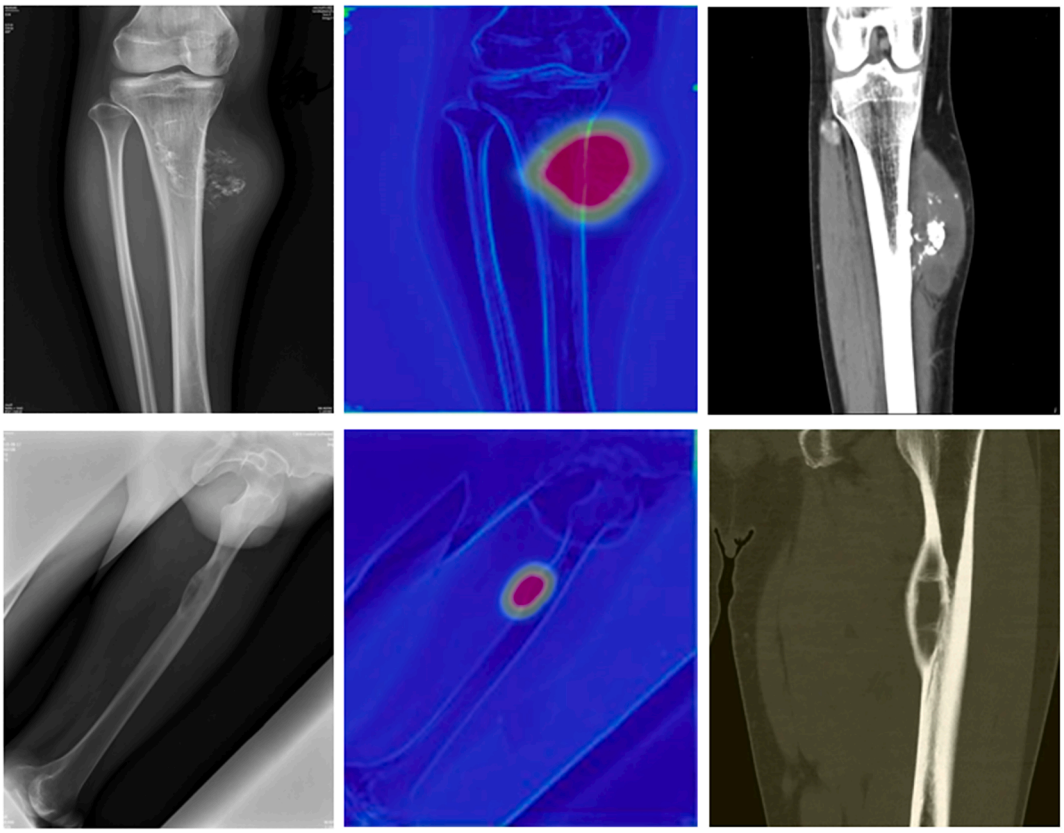


Fig. 5. Grad-CAM showed the position of a tumor. Whereby (a) is original image by radiography, (b) is the position of a tumor highlighted by Grad-Cam, and (c) is the final verified CT image for the position of a tumor.

classification. This highlights the potential of deep learning approaches to revolutionize diagnostic accuracy and clinical decision-making.

The proposed AI model also addresses the issue of misdiagnosis, especially among non-specialists in musculoskeletal oncology. By providing a non-invasive, AI-driven preliminary screening process, the model can significantly reduce the risk of misclassification and guide medical professionals towards more accurate diagnostic pathways. This fusion of medical expertise with advanced AI technology reflects a pioneering approach in enhancing the efficacy of bone tumor detection.

5. Limitations and future work

Several limitations should be acknowledged in this study. Firstly, the sample size of hip radiographs used may be considered relatively small, although it included 214 of the femurs had bone cancers, which is a significant number given the prevalence of proximal femur tumors. Second, the deep learning algorithm's lack of input for patient-specific data including age, sex, and clinical symptoms may have limited the diagnostic potential of the model. Incorporating such information in future models could enhance diagnostic performance.

The proposed AI-driven bone tumor classification approach exhibits several potential limitations that deserve consideration. Firstly, the model's performance may be influenced by the quality and diversity of the training dataset. If the training dataset is not sufficiently comprehensive, the model's ability to generalize to different tumor types and clinical scenarios could be compromised. Additionally, the model's performance might be affected by the rarity of certain tumor types, leading to challenges in accurately classifying them. In future implementations, for the development of more advanced diagnosis of the orthopedics [28], algorithms in the cybernetical intelligence context [29] can be computationally powerful and may bring about viable diagnostics techniques based on medical image processing [30].

6. Conclusion

The suggested AI-driven approach outperformed orthopedic doctors' diagnostic performance in identifying bone cancers in the proximal femur based on plain hip radiographs. Utilizing advanced deep learning models like DenseNet, the framework demonstrated superior accuracy, sensitivity, and precision. By automatically generating regions of interest (ROIs) without annotations, it improved efficiency and reduced dependence on manual labeling. When compared to human physicians, DenseNet fared better than previous models, exhibiting considerable gains of over 5% in accuracy, sensitio vity, and precision. The effectiveness of the framework was further supported by gradient-weighted class activation mapping (Grad-CAM), which revealed tumor locations. However, limitations include a smaller sample size and the absence of patient-specific information. Overall, the framework's success in bone tumor classification promises reduced misdiagnosis rates and improved accuracy in musculoskeletal oncology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W. M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.

[2] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118.
[3] O. Ronneberger, et al., U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[4] K. Simonyan, et al. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
[5] K. He, et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE International Conference on Computer Vision. 2015;1026-1034.
[6] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng. 19 (1) (2017) 221–248.
[7] J. Zhang, et al., Bone tumor classification in extremities with multiplanar MRI scans using convolutional neural networks, Comput. Methods Programs Biomed. 197 (2020), 105732.
[8] P. Rajpurkar, et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225. 2017.
[9] B.Q. Huynh, et al., Deep learning for image-based cancer detection and diagnosis–A survey, Pattern Recogn. 83 (2018) 134–149.
[10] H.C. Shin, et al., Deep convolutional neural networks for computer-aided detection: CNN9 architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.
[11] H. Choi, et al., Extending convolutional neural network for hemodynamic classification of intracranial aneurysms using color maps from 3D angiograms, Med. Phys. 47 (4) (2020) 1363–1373.
[12] L. Zhu, Z. Li, C. Li, J. Wu, J. Yue, High performance vegetable classification from images based on alexnet deep learning model, Int. J. Agric. Biol. Eng. 11 (4) (2018) 190–196.
[13] N. Abbassi, R. Helaly, M.A. Hajjaji, A. Mtibaa. A deep learning facial emotion classification system: a VGGNet-19 based approach. In 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2020 (pp. 271-276).
[14] Y. Tao, M. Xu, Z. Lu, Y. Zhong, DenseNet-based depth-width double reinforced deep learning neural network for high-resolution remote sensing image per-pixel classification, Remote Sens. (Basel) 10 (5) (2018) 779.
[15] A. Szulewski, H. Braund, R. Egan, A. Gegenfurtner, A.K. Hall, D. Howes, D. Dagnone, J.J.G. van Merrienboer, Starting to think like an expert: an analysis of resident cognitive processes during simulation-based resuscitation examinations, Ann. Emerg. Med. 74 (5) (2019) 647–659.
[16] K. Vinogradova, A. Dibrov, G. Myers. Towards interpretable semantic segmentation via gradient-weighted class activation mapping (student abstract). In Proceedings of the AAAI conference on artificial intelligence, 2020 (Vol. 34, No. 10, pp. 13943-13944).
[17] P. Das, A. Ortega. Gradient-Weighted Class Activation Mapping for Spatio Temporal Graph Convolutional Network. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022 (pp. 4043-4047). IEEE.
[18] J. Loo, T.E. Clemons, E.Y. Chew, M. Friedlander, G.J. Jaffe, S. Farsiu, Beyond performance metrics: automatic deep learning retinal OCT analysis reproduces clinical trial outcome, Ophthalmology 127 (6) (2020) 793–801.
[19] D. Bae, J. Ha, Performance metric for differential deep learning analysis, J. Internet Serv. Inf. Secur. 11 (2) (2021) 22–33.
[20] A. Korotcov, V. Tkachenko, D.P. Russo, S. Ekins, Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets, Mol. Pharm. 14 (12) (2017) 4462–4475.
[21] X.L. Zhang, M. Xu, AUC optimization for deep learning-based voice activity detection, EURASIP J. Audio, Speech, Music Process. 2022 (1) (2022) 1–12.
[22] J. Sulam, R. Ben-Ari, P. Kisilev. Maximizing AUC with Deep Learning for Classification of Imbalanced Mammogram Datasets. In VCBM, 2017 (pp. 131-135).
[23] C. Errani, D. Vanel, M. Gambarotti, M. Alberghini, P. Picci, C. Faldini, Vascular bone tumors: a proposal of a classification based on clinicopathological, radiographic and genetic features, Skeletal Radiol. 41 (12) (2012) 1495–1507.
[24] A.R. Vaccaro, R.J. Hulbert, A.A. Patel, C. Fisher, M. Dvorak, R.A. Lehman, P. Anderson, J. Harrop, F.C. Oner, P. Arnold, M. Fehlings, R. Hedlund, I. Madrazo, G. Rechtine, B. Aarabi, M. Shainline, The subaxial cervical spine injury classification system: a novel approach to recognize the importance of morphology, neurology, and integrity of the disco-ligamentous complex, Spine 32 (21) (2007) 2365–2374.
[25] S.S. Yadav, S.M. Jadhav, Deep convolutional neural network based medical image classification for disease diagnosis, J. Big Data 6 (1) (2019) 1–18.
[26] E.K. Wai, A.M. Davis, A. Griffin, R.S. Bell, J.S. Wunder, Pathologic fractures of the proximal femur secondary to benign bone tumors, Clin. Orthop. Relat. Res. 1976–2007 (393) (2001) 279–286.
[27] T. Triwiyanto, I.P.A. Pawana, M.H. Purnomo, An improved performance of deep learning based on convolution neural network to classify the hand motion by evaluating hyper parameter, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (7) (2020) 1678–1688.
[28] X. Deng, Y. Zhu, S. Wang, Y. Zhang, H. Han, D. Zheng, Z. Ding, K.K.L. Wong, CT and MRI determination of intermuscular space within lumbar paraspinal muscles at different intervertebral disc level, PLoS One 10 (10) (2015) e0140315.
[29] K.K.L. Wong, Cybernetical Intelligence: Engineering Cybernetics with Machine Intelligence, First Edition. The Institute of Electrical and Electronics Engineers, Inc., John Wiley & Sons, Inc., England, U.K., ISBN: 9781394217489, 2024.
[30] K.K.L. Wong, R.M. Kelso, S.G. Worthley, P. Sanders, J. Mazumdar, D. Abbott, Medical imaging and processing methods for cardiac flow reconstruction, J. Mech. Med. Biol. 9 (1) (2009) 1–20.