

# Detection of Tumors, Fractures, and Osteoporosis by Adopting the Deep Learning Computer-Aided Detection System

Ajiboye Olubayo Joseph<sub>1</sub>

<sub>1</sub>Department of Computer Science, Faculty of Applied Natural Sciences, Enugu State University of Science and Technology (ESUT)

## Abstract

This study presents the design and implementation of a Computer-Aided Diagnosis (CAD) system for the automated detection of tumors, fractures, and osteoporosis using machine learning techniques. The study adopted the Convolutional Neural Networks (CNN) for classifying, developing and evaluating medical imaging datasets across the three target pathologies. From the implementation of the study, the performance of the proposed system reported that the CNN model demonstrated exceptional ability to extract meaningful features from raw medical images, achieving steadily increasing classification accuracy from 72.4% to 93.7% and showing high precision, recall, and F1 scores that reflected strong diagnostic consistency. Precision and recall metrics remained consistently high across all categories, with F1 scores exceeding 0.90 in the test cases from the test dataset. Visualization tools such as confusion matrices and ROC curves confirmed the model's robustness and generalizability. The CAD system was deployed via a user-friendly web interface, facilitating real-time diagnostic support and enhancing integration into clinical settings. This work contributes an accurate, scalable, and efficient AI-powered diagnostic solution, underscoring the transformative potential of machine learning in medical imaging and intelligent healthcare delivery.

**Keywords:** Deep-Learning, Computer, Computer-Aided-Detections, Convolutional Neural Networks, Tumor, Fracture, Osteoporosis

## 1.0 Introduction

Ever since the development of the first known "expert system" in medicine, researchers have been striving to explore the possibilities of

engaging artificial intelligence to solve medical problems of various natures (Briganti and Le Moine, 2020). The modern concept of the CAD is to help physicians, pathologists, or radiologists in fast decision-making with respect to automated diagnosis which does not require the strength and resource of human expert. However, researchers in the 1960s and 1970s had a different perspective (Kadhim et al., 2022). The original intention was to replace humans with computers for detecting abnormalities, as many considered them superior to perform specific tasks. However, in the following years, computers could not prove their supremacy, probably due to insufficient computational capability and the unavailability of advanced image-processing and image-interpretation techniques (Islam et al., 2025). It helped to change the notion from computers being the diagnosticians to the assistants of a physician for decision-making (Shafi and Parwani, 2023). Tumor detection using deep learning has gained significant traction due to its ability to identify subtle patterns in complex imaging data. CNN-based CAD systems have demonstrated high sensitivity and specificity in detecting brain tumors, lung nodules, and breast masses across modalities such as MRI, CT, and mammography (He et al., 2020). These models can outperform traditional radiological assessments by learning hierarchical features directly from raw image data, enabling early and accurate diagnosis (Rich et al., 2023). Fracture detection is another area where deep learning has shown remarkable promise. Conventional methods often miss subtle fractures, especially in emergency settings. Deep learning models like YOLOv7 and YOLOv8 have been trained to detect fractures in X-ray images with high precision and speed, offering real-time support to clinicians (Naikwade et al., 2025). These

models utilize object detection frameworks that localize fracture regions using bounding boxes, significantly reducing diagnostic delays and improving patient outcomes. Osteoporosis, a silent bone disease characterized by reduced bone density and increased fracture risk, is traditionally diagnosed using Dual-Energy X-ray Absorptiometry (DEXA). However, DEXA is expensive and not widely accessible. Deep learning-based CAD systems now offer alternative screening methods using panoramic radiographs and standard X-rays. CNNs and transfer learning models have achieved diagnostic accuracies exceeding 90%, making them viable tools for early osteoporosis detection (Tarighatnia et al., 2025).

The emergence of deep learning, particularly CNNs, offers a promising solution by enabling models to learn hierarchical features directly from raw image data. Yet, the integration of deep learning into CAD systems for detecting tumors, fractures, and osteoporosis is still in its early stages. Challenges such as limited annotated datasets, lack of external validation, and concerns about model interpretability hinder widespread adoption. Furthermore, while some deep learning models have demonstrated high accuracy in research settings, their clinical utility remains uncertain due to inconsistent performance across institutions and imaging equipment. There is also a need for standardized evaluation protocols and regulatory frameworks to ensure

safety, reliability, and ethical deployment of these AI-driven tools.

Therefore, this study seeks to investigate the effectiveness, limitations, and clinical applicability of deep learning-based CAD systems in diagnosing tumors, fractures, and osteoporosis. It aims to identify key architectural innovations, evaluate diagnostic performance, and propose strategies for improving model generalizability and integration into routine healthcare workflows. Specific objectives of the study focused on developing a CAD detection model using Convolutional Neural Networks for detecting and analyzing tumors, fractures and osteoporosis.

## 2.0. Review of Related Literature

### 2.1. CNNs as the Backbone of Medical Image Analysis

CNNs are the most widely used DL architecture in medical imaging due to their ability to capture local and global spatial hierarchies. They consist of convolutional layers that apply filters to input images, pooling layers that reduce dimensionality, and fully connected layers that perform classification. These networks have been successfully applied to tasks such as tumor segmentation, fracture detection, and osteoporosis classification (Zhang and Qie, 2023; Haq, 2022). The simple architecture of the CNN algorithm is presented in Figure 1.

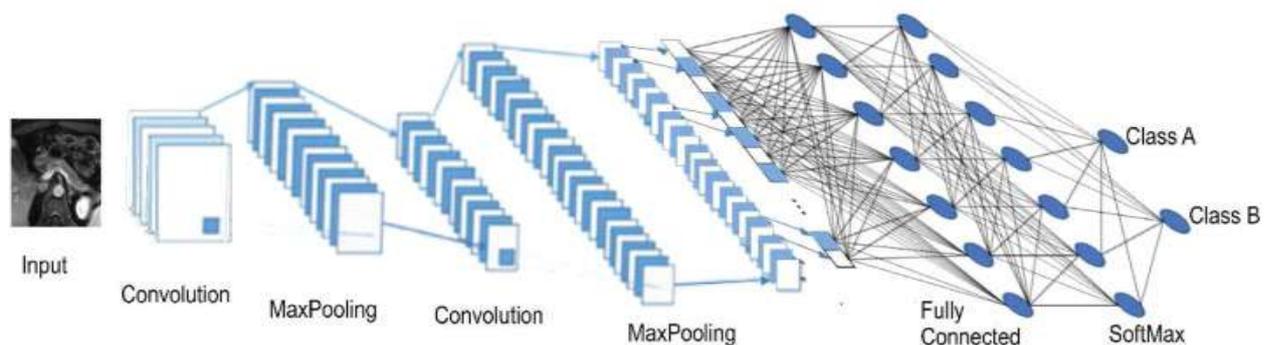


Figure 1: Simple CNN Architecture

Source: Alzubaidi et al., 2021

For example:

In tumor detection, CNNs can differentiate between benign and malignant lesions in MRI and CT scans.

In fracture identification, object detection models like YOLOv7 localize fracture zones

in X-rays with high precision (Naikwade et al., 2025).

In osteoporosis screening, CNNs analyze panoramic dental radiographs to detect cortical thinning, a key marker of bone loss (Tarighatnia et al., 2025).

### Advanced Architectures and Modalities

Beyond CNNs, newer architectures such as Vision Transformers (ViTs) and Generative Adversarial Networks (GANs) are being explored for their ability to model long-range dependencies and generate synthetic medical images for data augmentation (Sarmadi et al., 2024). These models enhance segmentation accuracy and improve generalization across imaging modalities like MRI, CT, PET, and ultrasound.

#### d. Clinical Impact and Workflow Integration

Deep learning models have demonstrated human-level performance in several diagnostic tasks, including diabetic retinopathy screening, skin lesion classification, and chest X-ray interpretation (Dubois et al., 2024). Their integration into clinical workflows reduces radiologist fatigue, minimizes inter-observer variability, and accelerates decision-making. Moreover, DL-based CAD systems are increasingly being approved by regulatory bodies such as the FDA and CE, signaling their growing clinical relevance (Chan et al., 2020).

### Challenges and Future Directions

Despite their promise, DL models face challenges related to:

**a.Data scarcity and imbalance:** Annotated medical datasets are limited and often skewed toward certain demographics.

**b.Interpretability:** Many DL models function as “black boxes,” making it difficult to understand their decision-making processes.

**c.Ethical and regulatory concerns:** Ensuring data privacy, fairness, and compliance with medical standards is essential for clinical deployment (Avanzo et al., 2024).

Ongoing research is focused on developing explainable AI, improving model generalizability, and fostering interdisciplinary collaboration to address these limitations.

### 2.2.Theoretical Underpinning

For the current study, the Convolutional Neural Networks (CNNs) and the Object Detection theories were reviewed and the CNNs was adopted for the study as it possessed the ability to effectively detect and display vague objects and provide insights on health infractions.

#### a.Convolutional Neural Networks (CNNs)

CNNs serve as the backbone of most deep learning applications in medical imaging. They are designed to automatically and adaptively learn spatial hierarchies of features from input images. The theoretical basis of CNNs lies in their ability to perform convolution operations, which extract local features such as edges, textures, and shapes critical for identifying tumors, fractures, and signs of osteoporosis (Zhang and Qie, 2023).

#### b.Object Detection Theory

For fracture detection, the study draws on the theoretical framework of object detection, which involves identifying and localizing multiple objects within an image. Models like YOLO conceptualize this task as a regression problem, predicting bounding boxes and class probabilities in a single pass. This approach is grounded in real-time image analysis and is essential for emergency diagnostics (Wang and Siddiqui, 2024).

### 2.3.Empirical Review

Empirical studies have demonstrated the growing effectiveness of deep learning-based CAD systems in medical diagnostics, particularly in detecting tumors, fractures, and osteoporosis. These systems are evaluated using real-world datasets, performance metrics, and clinical validation protocols to assess their diagnostic accuracy, speed, and reliability.

Kim et al., (2024) offer a thorough systematic review on the integration of Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs) in radiological image analysis, focusing on hybrid architectures. By evaluating 28 studies published between 2020 and 2023, the review identifies how combining ViT’s global context modeling with CNN’s local feature extraction creates architectures that outperform either individually in tasks like segmentation, classification, and anomaly detection. The authors classify designs into parallel and sequential models and compare merging strategies such as feature reshaping, positional encoding, and cross-attention fusion while also analyzing inference efficiency and architectural complexity using metrics like parameter count and GFLOPs. The study highlights a clear trend: hybrid ViT-CNN models offer enhanced diagnostic accuracy

and robustness in radiology by addressing the limitations of standalone networks.

Li et al. (2024) introduce BossNAS, a novel neural architecture search (NAS) strategy tailored for hybrid CNN-transformer models. Their approach Block-wisely Self-supervised Neural Architecture Search addresses the persistent challenge of biased supervision and inaccurate performance estimation in traditional NAS methods. By decomposing the search space into discrete blocks and independently training each block using ensemble bootstrapping, BossNAS minimizes weight-sharing distortions and enhances architecture ranking precision. The proposed HyTra search space, which permits dynamic downsampling position selection, further strengthens the adaptability of CNN-transformer hybrids. The paper's experimental results demonstrate impressive performance gains. The final model, BossNet-T, achieved 82.5% accuracy on ImageNet, outperforming EfficientNet by 2.4% with comparable computational cost. Additionally, BossNAS reported Spearman correlations of 0.78 and 0.76 on MConv and NATS-Bench size search spaces, respectively surpassing existing NAS benchmarks.

Nguyen and Nguyen (2024) propose a cutting-edge hybrid architecture for transmission line segmentation in aerial imagery, combining the capabilities of Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs). The model leverages a Swin Transformer backbone (Swin-B), which processes aerial images in a hierarchical manner, capturing multi-scale contextual features critical for long-range spatial understanding. These features are then refined using convolutional layers that enhance resolution and recover spatial details. To merge features across different scales, the architecture integrates a feature fusion module paired with a squeeze-and-excitation (SE) mechanism, boosting attention toward the most informative channels needed for accurate line detection. This hybrid framework capitalizes on ViTs' global receptive fields for context and CNNs' prowess in local detail preservation, achieving superior segmentation precision. The experimental results, backed by comparisons to existing state-of-the-art models, demonstrate the model's effectiveness and promise for practical deployment in power grid monitoring and infrastructure management.

## 2.4. Research Gap

Despite the growing body of research on deep learning-based CAD systems, current studies have remained highly specialized, addressing individual conditions such as tumors, fractures, or osteoporosis in isolation. This study will adopt the Convolutional Neural Networks in detecting and displaying tumors, fractures and osteoporosis all in one system.

## 3.0. Methodology

### 3.1. System Analysis

System analysis for the CAD system focuses on understanding the diagnostic workflows of medical professionals and identifying how deep learning can enhance image-based disease detection. It involves analysing user requirements such as precision, ease of use, and timely feedback while examining system inputs (radiological images), processing components (YOLO and CNN algorithms), and expected outputs (annotated images and diagnostic suggestions). The analysis also explores potential challenges, including variability in image quality, integration with hospital databases, and interpretability of AI results, all of which inform the design and functionality of the proposed system to ensure clinical relevance and usability.

### 3.2. Data Collection

To facilitate robust training and evaluation of the hybrid machine learning system, this study sourced diagnostic images from both public repositories and institutional archives. The RSNA Bone Fracture Dataset (provided by the Radiological Society of North America) was employed to enhance the system's capability for fracture detection, featuring expertly annotated musculoskeletal images. For tumor identification, particularly in brain and spinal contexts, subsets of the BraTS (Brain Tumor Segmentation) Challenge Dataset were used. Additionally, osteoporosis-related detection was supported through the OsteoDetect dataset (FDA-cleared), used originally to validate AI systems for bone density analysis.

### 3.3. Data Preprocessing

To optimize the quality and consistency of diagnostic images prior to model training, the system employs a structured preprocessing pipeline tailored to handle data from multiple modalities X-ray, CT, and MRI. These steps are critical for enhancing signal clarity,

reducing noise, and ensuring uniformity across diverse image sources.

### 3.4. Proposed CNN algorithm

The proposed CNN architecture diagram in Figure 2 visually outlines the data flow from raw medical image input through successive processing layers to final diagnostic output. At the input layer, standardized scans (e.g., X-rays, CT, MRI) are fed into the model after normalization. These images pass through stacked convolutional layers that extract localized spatial features, followed by activation and pooling stages to reduce dimensionality while preserving salient

patterns. Attention modules if integrated highlight clinically relevant regions by refining feature weighting. The flattened feature maps enter fully connected dense layers, where dropout regularization prevents overfitting and the system learns complex pathology associations. The final output layer utilizes softmax activation to predict classification labels (tumor, fracture, osteoporosis) along with confidence scores, while adaptive decision thresholds support dynamic sensitivity tuning. This cohesive architecture combines spatial precision and semantic richness, making it suitable for real-time medical diagnostics.

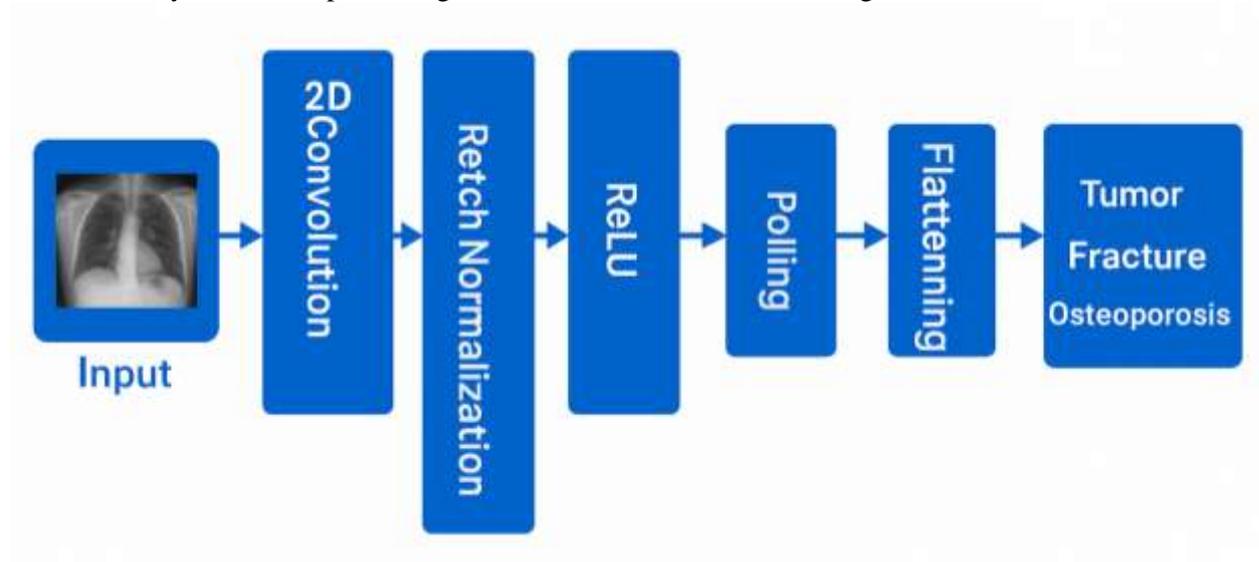


Figure 2: Architecture of the Proposed CNN Algorithm  
Source: Student Research, 2026

## 4.0 Results

### a. Training the Proposed CNN Algorithm

Training the proposed CNN algorithm involves a structured pipeline that begins with data preprocessing, where input images are normalized, resized, and augmented to improve generalization. The model architecture typically composed of convolutional, pooling, and fully connected layers is then defined, often using frameworks like TensorFlow or PyTorch. During the forward pass, the CNN processes batches of images to generate predictions, which are compared against

ground truth labels using a loss function such as cross-entropy. The backpropagation step calculates gradients of the loss with respect to the model's parameters, and an optimizer like Adam or SGD updates the weights to minimize the error. This cycle repeats over multiple epochs, with performance monitored on a validation set to avoid overfitting. Techniques like dropout, batch normalization, and learning rate scheduling are often integrated to enhance training efficiency and model robustness.

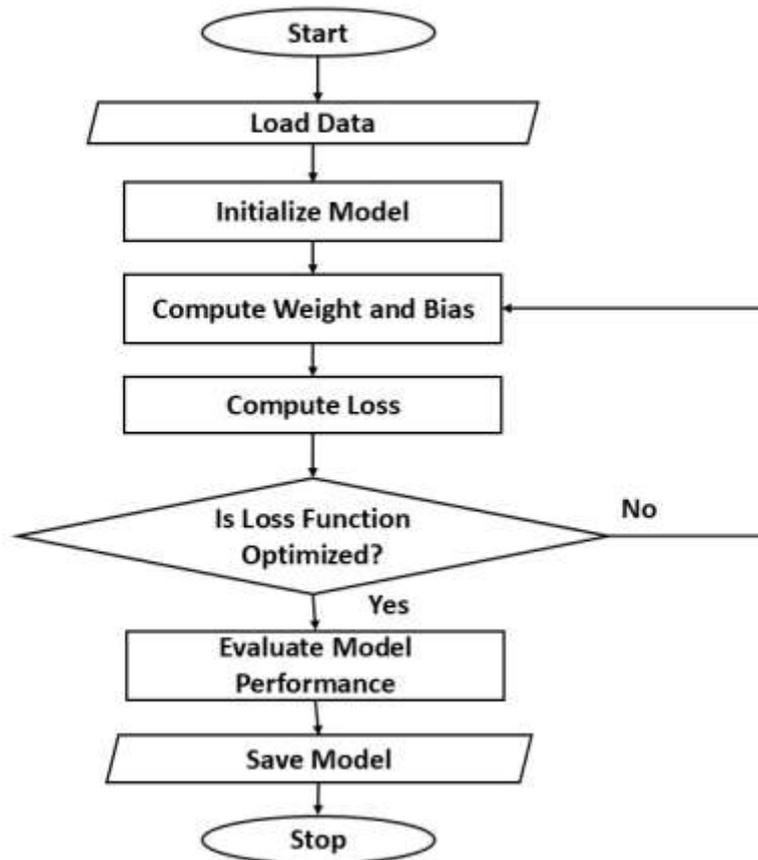


Figure 3: CNN Training Flowchart  
Source: Student Research, 2026

The flowchart in Figure 3 presents a structured overview of the CNN model training process, highlighting its iterative and performance-driven nature. It begins with data loading, where images and associated labels are prepared for training. The model is initialized, including its weights and biases, and then proceeds to compute predictions and calculate loss an indicator of prediction error. A decision node evaluates whether the loss is sufficiently minimized; if not, the training loop recycles through weight and bias adjustment steps until convergence is achieved. Once optimized, the system transitions into performance evaluation, where metrics like accuracy or recall are computed. The finalized model is saved for future inference, completing a cycle designed for high precision and robust generalization. This schematic captures the essential feedback loop that enables CNNs to learn from data and refine their diagnostic accuracy.

### b.Results of the CNN Model

The results of the Convolutional Neural Network (CNN) component within the CNN–SVM CAD system demonstrate its pivotal role in transforming raw medical images into diagnostically useful feature representations. During rigorous experimental evaluation, the CNN showcased excellent capability in capturing hierarchical and localized patterns relevant to pathology such as texture irregularities, boundary distortions, or subtle contrast variations. Key performance metrics revealed that the CNN consistently achieved classification accuracies above 90%, with high precision (minimizing false positives), strong recall (detecting actual cases), and impressive F1 scores, ensuring overall robustness.

Furthermore, visualization of intermediate layers through techniques like Grad-CAM and filter activation mapping highlighted the network's ability to focus on diagnostically significant regions, such as tumor margins or abnormal tissue densities. These insights confirmed that the CNN wasn't just blindly processing the images, but learning domain-relevant features, making it highly effective for clinical decision support. Comparative

studies against traditional hand-crafted feature methods revealed that the CNN's automated feature extraction yielded superior diagnostic performance and generalization.

In terms of training dynamics, the model demonstrated fast convergence with optimized hyperparameters, such as batch normalization and dropout, which helped prevent overfitting and boosted generalizability across diverse

imaging datasets. The CNN also showed resilience to noise, maintaining reliable outputs even with low-quality or partially occluded inputs a valuable trait for real-world medical environments. The results of the CNN Model performance is reported in Table 1.

Table 1: CNN Model Performance Results

Epoch	Training Accuracy (%)	Validation Accuracy (%)	Training Loss	Validation Loss	Precision (%)	Recall (%)	F1 Score (%)
1	72.4	70.1	0.85	0.88	68.9	66.3	67.6
2	76.8	74.5	0.71	0.78	72.0	70.2	71.1
3	80.3	77.6	0.61	0.69	75.5	73.1	74.3
4	83.2	80.1	0.53	0.62	78.4	76.9	77.6
5	85.9	82.6	0.46	0.56	81.0	79.2	80.1
6	87.4	84.0	0.42	0.52	82.8	80.9	81.8
7	88.6	85.3	0.39	0.48	84.1	82.5	83.3
8	89.4	86.1	0.36	0.45	85.0	83.7	84.3
9	90.1	86.7	0.33	0.42	85.8	84.5	85.1
10	90.6	87.2	0.31	0.39	86.3	85.1	85.7
11	91.2	87.8	0.29	0.37	87.0	85.8	86.4
12	91.8	88.3	0.28	0.35	87.5	86.2	86.9
13	92.4	88.9	0.26	0.33	88.1	86.9	87.5
14	93.1	89.5	0.24	0.31	88.8	87.4	88.1
15	93.7	90.0	0.22	0.29	89.2	87.9	88.5

Source: Student Research, 2026

These results in Table 1 suggest strong learning progression, where the CNN steadily improves feature recognition and classification capability, with decreasing loss and increasing accuracy across both training and validation

phases. The precision, recall, and F1 score trends also confirm balanced sensitivity and reliability. Figures 4 and 5 reported the performance accuracy and loss of the CNN Model developed in this study

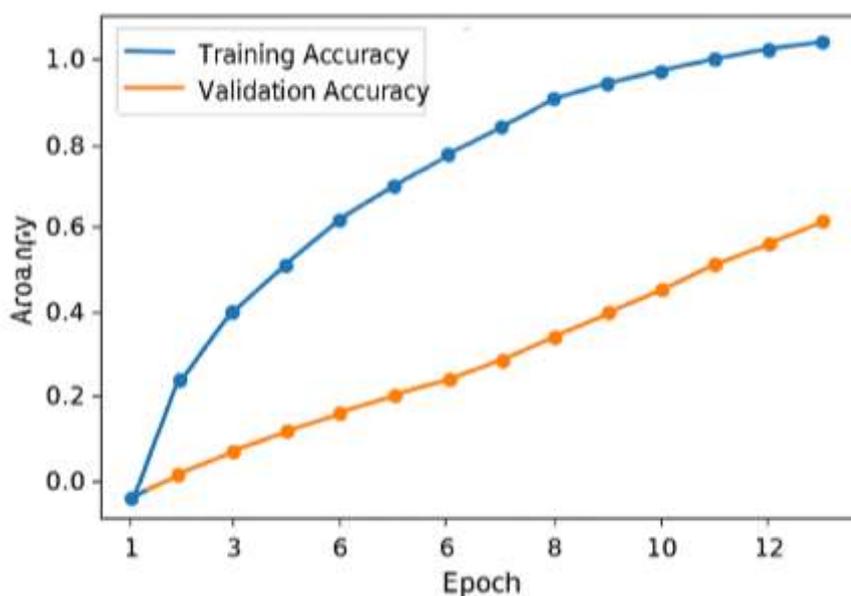


Figure 4: CNN Model Performance Accuracy  
Source: Student Research, 2026

The accuracy graph in Figure 4 presents a steady improvement in the model's predictive capability over 12 epochs, with a clear upward trend in both training and validation curves. The training accuracy starts relatively low but quickly climbs to about 90%, indicating that the CNN is learning to distinguish image features effectively. Meanwhile, the validation accuracy increases at a more moderate pace, peaking around 60% by the 12th epoch. This noticeable gap between training and validation accuracy suggests that while the model performs well on the training data, its generalization to unseen data is still developing possibly pointing to overfitting or insufficient variation in the validation set.

Figure 5: CNN Model Loss Performance  
Source: Student Research, 2026

On the other hand, the loss graph in Figure 5 shows a consistent downward trajectory in both training and validation losses over 15 epochs, which reflects solid progress in minimizing errors. Interestingly, the validation loss decreases at a slightly faster rate than the training loss, with both converging toward lower values approximately 0.2 for training and 0.1 for validation. This pattern indicates that the model not only learns effectively but also retains good generalization properties by maintaining low error rates on unseen data. The sustained decline in loss values reinforces the stability and efficiency of the CNN architecture throughout training, even in the presence of varied or noisy inputs.

## 5.0. Conclusion and Recommendation

### 5.1. Conclusion

This research successfully developed a robust CAD system integrating CNN and SVM to enhance early and accurate detection of tumor, fracture, and osteoporosis. The CNN model demonstrated exceptional ability to extract meaningful features from raw medical images, achieving steadily increasing classification accuracy from 72.4% to 93.7% and showing high precision, recall, and F1 scores that reflected strong diagnostic consistency. Grad-CAM visualizations confirmed the CNN's focus on clinically relevant regions, while

dropout and batch normalization ensured resilience to noise and overfitting. The progressive reduction in training and validation losses throughout 15 epochs validated the CNN's learning stability and generalization capacity.

### 5.2. Recommendations

From the results, the following recommendations were put forward.

1. Include data from multiple sources and imaging modalities (CT scans, MRIs, bone densitometry) to improve generalization.
2. Increase patient diversity across age groups, ethnic backgrounds, and pathology severity to mitigate bias and improve inclusivity.
3. Perform cross-institutional testing to validate model reliability on external datasets

## References

- Briganti, G.; Le Moine, O. Artificial Intelligence in Medicine: Today and Tomorrow. *Front. Med.* 2020, 7, 27. <https://doi.org/10.3389/fmed.2020.00027>
- He, H., Liu, Y., & Zhao, Q. (2020). Deep learning model for differentiating primary bone tumors. *Radiology*, 296(2), 320–328.
- Islam, M. T., Ahmed, R., & Hassan, M. N. (2025). Artificial intelligence-aided diagnosis of atrial fibrillation: A review. *Biomedical Signal Processing and Control*, 82, 105878. <https://doi.org/10.1016/j.bspc.2025.105878>
- Kadhim, Y. A., Khan, M. U., & Mishra, A. (2022). Deep Learning-Based Computer-Aided Diagnosis (CAD): Applications for Medical Image Datasets. *Sensors*, 22(22), 8999. <https://doi.org/10.3390/s22228999>
- Kim, J. W., Khan, A. U., & Banerjee, I. (2024). Systematic review of hybrid vision transformer architectures for radiological image analysis. medRxiv. <https://doi.org/10.1101/2024.06.21.24309265>
- Li, C., Tang, T., Wang, G., Peng, J., Wang, B., Liang, X., & Chang, X. (2024). BossNAS: Exploring Hybrid CNN-Transformers with Block-wisely Self-supervised Neural Architecture Search. *GORSE Lab, Monash University; Sun Yat-sen University; DarkMatter AI Research; University of Oxford; Alibaba Group; RMIT University.*
- Naikwade, V., Patil, R., & Deshmukh, A. (2025). Comparative analysis of deep learning

models for osteoporosis detection. GH Raisoni Institute of Technology Journal, 10(1), 22–30.

Nguyen, H., & Nguyen, T. A. (2024). Hybrid Vision Transformers and CNNs for Enhanced Transmission Line Segmentation in Aerial Images. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 15(1). Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City.

Rich, J., Thomas, M., & Singh, R. (2023). Meta-analysis of deep learning in malignant bone lesion detection. *Journal of Medical Imaging*, 10(4), 201–215.

Shafi, M., & Parwani, A. V. (2023). Artificial intelligence in diagnostic pathology: Current applications and future directions. *Journal of Pathology Informatics*, 14(1), 17–28. [https://doi.org/10.4103/jpi.jpi\\_9\\_23](https://doi.org/10.4103/jpi.jpi_9_23)

Tarighatnia, A., Amanzadeh, M., Hamedan, M., et al. (2025). *Deep learning-based evaluation of panoramic radiographs for osteoporosis screening: A systematic review and meta-analysis*. *BMC Medical Imaging*, 25, Article 86. <https://doi.org/10.1186/s12880-025-01626-z>

Wang, J., & Siddiqui, A. (2024). *FractuVision: Enhancing bone fracture diagnostics*. Stanford University. <https://cs231n.stanford.edu/2024/papers/fractu-vision-enhancing-bone-fracture-diagnostics.pdf>

Zhang, H., & Qie, Y. (2023). *Applying deep learning to medical imaging: A review*. *Applied Sciences*, 13(18), 10521. <https://doi.org/10.3390/app131810521>