LiteSpineNet Integrates Quantum Computing for Spinal Fracture Diagnosis

Daniel Alejandro Lopez , Oscar Montiel * , Oscar Castillo , Miguel Lopez-Montiel

Posted Date: 27 March 2026

doi: 10.20944/preprints202603.2213.v1

Keywords: osteoporosis; spine fractures; computer aided diagnosis; image classification; deep learning; quantum computing; quantum deep learning

 doi:10.20944/preprints202603.2213.v1

Article

LiteSpineNet Integrates Quantum Computing for Spinal Fracture Diagnosis

Daniel Alejandro Lopez [1], Oscar Montiel [1,*], Oscar Castillo [2] and Miguel Lopez-Montiel [3]

[1] Instituto Politécnico Nacional, CITEDI, Ave. Instituto Politécnico Nacional No. 1310, Col. Nueva Tijuana, 22435, Tijuana, B.C., México
[2] TecNM, Technological Institute of Tijuana; Calz. del Tecnologico No. 12950, Fracc. Tomas Aquino, 22414,Tijuana. B. C. México
[3] CETYS Universidad, Sv. Cetys Universidad 4, El Lago, Lago Sur, 22217, Tijuana, B.C., México
* Correspondence: oross@ipn.mx

Abstract

In medical diagnostics, integrating quantum computing with deep learning offers potential improvements in sensitivity and computational efficiency. This study introduces two models: LiteSpineNet, a streamlined CNN, and HQ-LiteSpineNet, its hybrid quantum counterpart that incorporates a novel seven-qubit variational feature extraction block for multiclass spine lesion classification. Using the VinDr-SpineXR dataset of 12,000 X-ray images across eight classes, we show that both models achieve strong performance while requiring substantially fewer parameters than state-of-the-art methods. LiteSpineNet attains 76.81% test accuracy, 79.69% precision, 76.81% recall, 77.92% F1, and 89.75% AUROC. HQ-LiteSpineNet achieves 72.43% accuracy, 80.40% precision, 72.43% recall, 75.46% F1, and 89.95% AUROC. Importantly, HQ-LiteSpineNet improves minority class detection, raising recall by +6.7% and F1 by +4.4% compared to the classical baseline. Statistical significance was confirmed through paired t-tests ($p < 0.0001$), supported by 99% confidence intervals, external validation, and decision curve analysis, which showed net benefit gains in minority classes. Computational cost analysis highlights the exponential overhead of quantum simulators ($O(2^n)$) but emphasizes the linear scalability of real quantum hardware ($O(n)$). These findings establish LiteSpineNet as an efficient baseline and HQ-LiteSpineNet as a promising quantum-enhanced alternative, advancing AI-driven diagnostics by combining efficiency with improved sensitivity in underrepresented lesion categories.

Keywords: osteoporosis; spine fractures; computer aided diagnosis; image classification; deep learning; quantum computing; quantum deep learning

1. Introduction

Fractures of the hip and spine in older adults represent not only medical emergencies but also major public health crises with far-reaching societal and economic consequences. In adults aged 50 years and above, such fractures pose substantial health challenges, with wide-ranging effects on life expectancy, social well-being, and national economies [1]. Globally, vertebral fractures are among the most common osteoporotic fractures, affecting more than 30% of women and 20% of men over the age of 50, with incidence expected to rise sharply as populations age [2,3]. These fractures frequently remain undiagnosed, yet they are associated with subsequent hip fractures, chronic pain, and progressive spinal deformity, underscoring their clinical importance [4]. The consequences are severe: patients often face prolonged recovery periods, recurrent hospitalizations, and substantially increased healthcare costs [5]. Mortality within the first year is notably higher, largely due to surgical complications, extended immobility, and exacerbation of underlying health conditions [6]. Socially, affected individuals experience a marked decline in their ability to perform daily activities, increased dependence on family members, and the need for long-term care, which frequently results in frustration, depression, diminished quality of life, and social isolation [7]. Economically, the combined

 doi:10.20944/preprints202603.2213.v1

burden of medical care, rehabilitation, and institutionalized care imposes heavy strain on healthcare systems worldwide, with direct costs amounting to billions annually and indirect costs such as lost productivity further amplifying the impact [2,5]. Although both men and women are affected, women are disproportionately impacted due to higher rates of osteoporosis, while men often experience worse outcomes, including higher mortality rates [1,8,9].

Accurate and early detection of fractures and bone damage is a critical challenge in modern medicine, with significant implications for treatment and recovery [3,10,11]. Undiagnosed or improperly treated fractures often lead to severe complications, prolonged recovery, and increased healthcare costs [2,5]. Conventional approaches such as radiographic assessment and specialist interpretation remain effective, but they are limited by time demands, uneven availability of equipment, and observer variability [12,13]. These limitations have motivated the rapid growth of computer-aided diagnosis (CAD), where artificial intelligence methods are applied to medical imaging. Deep learning (DL), in particular, has demonstrated strong performance in tasks such as image classification, detection, and segmentation, achieving substantial improvements in diagnostic accuracy across multiple domains [14]. In the context of osteoporosis and spinal conditions, DL has shown promise for lesion detection and fracture identification. Nevertheless, most existing approaches are restricted to binary or one-vs-all formulations, which simplify deployment but fail to capture the full spectrum of clinically relevant lesion types. Multi-class classification, though more representative of real diagnostic needs, remains underexplored and is further complicated by data imbalance and computational constraints. A more detailed review of prior works is provided in Section 2.

However, despite these advances, DL methods still face important limitations when applied to complex medical imaging tasks [15]. Challenges such as class imbalance, limited annotated datasets, and reduced generalization across institutions often restrict their clinical adoption. Quantum computing offers a new computational paradigm that leverages principles of quantum mechanics to enable richer data representations and novel optimization strategies [16–18]. Although today's noisy intermediate-scale quantum (NISQ) devices remain constrained by qubit counts and circuit depth, shallow variational circuits have already demonstrated competitive performance in simulation and small-scale hybrid applications. The integration of quantum computing with artificial intelligence, commonly referred to as quantum machine learning (QML), has shown promising results in data science, optimization, and computer vision tasks [19–21]. Because quantum algorithms are particularly well suited to high-dimensional data, they can reveal subtle correlations that may remain hidden to classical models [18]. This capability is highly relevant to medical imaging, where improving the sensitivity and robustness of diagnostic systems could contribute to earlier and more accurate identification of pathological changes.

Given these properties, quantum computing emerges as a promising complement to medical imaging for fracture detection [18,22]. Building on this motivation, we propose two models: (i) LiteSpineNet, a parameter-efficient convolutional neural network tailored for eight-class spine lesion classification; and (ii) HQ-LiteSpineNet, a hybrid quantum–classical extension that integrates a seven-qubit variational block to enrich feature extraction. We evaluate both models under identical experimental conditions to assess accuracy, robustness, and reproducibility, with particular attention to minority-class sensitivity.

All experiments are conducted in accordance with CLAIM and TRIPOD+AI guidelines to ensure transparent and reproducible reporting of AI-based diagnostic models in medical imaging. Within this framework, we introduce our proposed architectures and outline their contributions, which are summarized below.

- We propose an eight-class multiclass framework on VinDr-SpineXR, moving beyond binary or multi-binary schemes.
- We design LiteSpineNet, a compact CNN with approximately 3.4× fewer trainable parameters than MobileNetV3, tailored for eight-class spine lesion classification in resource-constrained clinical environments.

 doi:10.20944/preprints202603.2213.v1

- We introduce a seven-qubit variational block to extract high-order features, improving sensitivity to minority classes.
- We propose HQ-LiteSpineNet, a novel hybrid quantum–classical variant of LiteSpineNet designed for multiclass spine lesion classification.
- We evaluate robustness and reproducibility under CLAIM and TRIPOD+AI guidelines, using fixed seeds, class-wise metrics, decision-curve analysis, and external validation.
- We release all code, trained weights, and logs openly to ensure transparency and facilitate replication.

In summary, this work presents a framework that integrates a parameter-efficient convolutional network with a hybrid quantum extension for the classification of spine lesions. The proposed approach is developed within a rigorous experimental design and adheres to established guidelines for transparent and reproducible reporting in medical imaging. By situating our contribution at the intersection of clinical need, deep learning, and quantum computing, the present work contributes to advancing the development of more reliable and accessible diagnostic tools.

This manuscript is structured as follows. Section 1 introduces the clinical context of osteoporosis and spine lesion diagnosis, emphasizing current challenges in computer-aided diagnosis and the potential advantages of integrating artificial intelligence with quantum computing, and concludes with a summary of the key contributions. Section 2 reviews recent advances in spine lesion diagnosis using deep learning approaches for classification, detection, and segmentation, and highlights the absence of established hybrid quantum methods in this field, motivating their inclusion in medical imaging tasks. Section 3 details the proposed methodology, including data preprocessing, the LiteSpineNet architecture, the quantum feature extraction block, and the hybrid model (HQ-LiteSpineNet). Section 4 presents the experimental results, covering hyperparameter selection, sensitivity analysis, hardware/-software configuration, and the evaluation protocol. A total of eight experiments were conducted, encompassing classical model training with K-fold cross-validation, hybrid quantum model training with K-fold cross-validation, statistical analysis, ablation studies, external dataset validation, and computational cost and complexity analysis. Section 5 discusses the findings in relation to the state of the art, addresses the main limitations of the study, and outlines directions for future research. Finally, Section 6 also provides the work's conclusions, summarizing the main results, highlighting their practical significance, and pointing to avenues for further exploration.

2. Literature Review

The dynamic field of medical imaging research and CAD has advanced through diverse methodologies, including improved data preparation, conventional data science workflows, and novel paradigms such as artificial intelligence and quantum computing. This review highlights key contributions, focusing on computer vision tasks, spine lesion classification and detection, and the use of hybrid quantum models in medical image diagnostics.

Accurate image segmentation is essential for automated systems that support clinicians with reliable predictions. Recent advances in AI for clinical computer vision tasks have produced notable improvements in this area. Sun et al. [23] introduced DA-TransUNet, a dual attention-based transformer that integrates position and channel attention blocks within a U-Net, achieving higher segmentation accuracy, parameter efficiency, and data privacy across six public medical datasets. Similarly, Sun et al. [24] proposed MIPC modules that combine spatial and channel-level information through skip-residual connections, improving boundary delineation and segmentation accuracy on the Synapse, ISCI 2018, and SegPC datasets.

Beyond segmentation performance, CAD systems face persistent challenges related to data sufficiency and quality. Limited annotated datasets and high communication overhead in distributed settings remain major obstacles for clinical deployment. To address these issues, Sun et al. [25] proposed FKD-Med, a model that integrates federated learning with knowledge distillation for medical image segmentation. FKD-Med reduces student model size while maintaining or improving accuracy,

 doi:10.20944/preprints202603.2213.v1

enabling privacy-preserving training, lowering communication costs, and supporting collaborative learning across institutions without compromising data confidentiality.

Similar to other medical imaging tasks, DL has significantly advanced spine lesion diagnosis, often reaching performance comparable to medical experts when supported by classical models and statistical validation [4,12]. For example, Chen et al. [13] used residual neural networks on frontal radiographs, achieving 73% accuracy by leveraging Grad-CAM to identify and visualize regions of interest. Yabu et al. [26] reported an AUROC of 94% with ensemble CNN architectures, while Dong et al. [27] combined datasets and annotations with a modified algorithm to build a screening tool for osteoporotic fractures, attaining an AUROC of 93%. Karno et al. [28] achieved 99% accuracy and a 96% F1 score using different EfficientNet variants, despite higher loss rates.

More recently, attention mechanisms have been applied to spine lesion classification. Dastgir et al. [29] proposed MAFMv3, a MobileNetV3-based model with multiscale attention and atrous spatial pyramid pooling to enhance focus on lesion regions. The model incorporates normalized, histogram-equalized, and raw images to construct comprehensive 3D feature maps. On the VinDr-SpineXR dataset, MAFMv3 achieved 96.81% accuracy, 98.38% precision, 97.95% recall, 98.15% F1 score, and 99.98% AUROC.

Object detection has also been applied to vertebral body and spine lesion analysis. Namki et al. [30] proposed a scoring-based method for detecting osteoporosis and vertebral fractures, achieving 91% and 94% accuracy for vertebral fracture detection on internal and external test sets, respectively. When extended to osteoporosis identification, the model reached 77% and 72% accuracy on the same datasets.

Iyer et al. [31] developed a vertebral compression fracture detection method that combines imitation learning with majority voting. Using 3D CT abdominal and chest scans, the system localizes spine regions through deep reinforcement learning, generates 3D bounding boxes, and converts them into 2D sagittal slices and patches for input into CNN detectors. Final predictions are obtained by majority voting, achieving an average threefold cross-validation accuracy of 85.95% and an F1 score of 85.94%.

Chlad et al. [32] evaluated the use of Vision Transformers for cervical spine fracture detection, reporting an accuracy of 98%. This work highlights the potential of transformer-based models in medical imaging tasks beyond traditional CNN approaches.

Kim et al. [33] explored classification of spine lesions to distinguish lumbosacral radicular pain from spinal stenosis using anteroposterior, lateral, and bilateral lumbar radiographs. Predictive performance was assessed by therapeutic outcomes after transforaminal epidural steroid injection (TFESI), where favorable outcomes were defined as stenosis reductions greater than 50% and poor outcomes as reductions below this threshold. The model achieved an AUROC of 92% and an accuracy of 87.2%.

Classical machine learning approaches have also achieved favorable results in spine lesion diagnosis. Nicolaes et al. [34] developed algorithms for vertebral fracture identification on abdominal and chest CT scans, reporting an AUROC of 88%, an accuracy of 92%, a sensitivity of 81%, and a specificity of 95%.

Hybrid deep learning strategies have likewise been explored. Polzer et al. [35] combined a hierarchical CNN with a fracture classification network for automated detection and stability analysis of vertebral bodies on CT scans, achieving 75% classification accuracy and an AUROC of 87% for fracture detection. Liawrungrueang et al. [36] analyzed more than 500 lateral cervical spine radiographs, obtaining 88.6% recall, 95.4% precision, 95.7% specificity, 92.1% accuracy, and a 92.1% AUROC.

More recently, custom detection models have been proposed. Sutradhar et al. [37] introduced VertNet-10, a modified YOLOv8 model that integrates object detection with a novel attention mechanism for cervical spine fracture detection and vertebral classification in CT images. VertNet-10 achieved a mean average precision (mAP) of 93% for fracture detection and 99.5% accuracy in vertebral classification.

Hybrid quantum methodologies for medical imaging remain in early development, yet initial studies show promising results. Houssein et al. [19] implemented quantum convolution for COVID-19 classification, reporting 98% accuracy in both binary and multiclass tasks and outperforming state-of-the-

 doi:10.20944/preprints202603.2213.v1

art methods. Amin et al. [38] further demonstrated that quantum convolutional neural networks can enhance performance on generated COVID-19 data, achieving 94% accuracy, precision, and F1 score.

Quantum circuits have also been applied beyond convolution and feature extraction. Azevedo et al. [39] used a quantum classifying circuit after transfer learning to improve binary and multiclass breast cancer classification, obtaining 84% accuracy.

In the context of bone fractures and spine lesions, hybrid quantum methods have shown advantages compared to classical approaches. Gadu et al. [40] proposed a hybrid GAN with a quantum U-Net for 3D spine X-ray registration, which enhanced capture range by providing richer semantic information. More recently, Mewada et al. [41] integrated a quantum convolutional layer at the input of a CNN for bone fracture classification in X-ray images, achieving 96% accuracy and a 95% F1 score.

Despite progress in both classical and quantum approaches for binary and multiclass tasks, no prior study has presented an end-to-end hybrid quantum–classical model capable of discriminating among eight clinically relevant spine lesion categories within a single network using X-ray data. Table 1 summarizes the reviewed literature on spine lesions, outlining the approach, data type, computer vision task, and results, while highlighting the gap addressed by this work.

To bridge this gap, we introduce an efficient lightweight neural network tailored for spine lesion classification, along with its quantum-enhanced counterpart, HQ-LiteSpineNet. The latter integrates a quantum feature extraction block to improve complex pattern learning and minority-class generalization. Inspired by MobileNetV3, LiteSpineNet comprises 760,292 parameters—substantially fewer than most deep learning models—yet effectively addresses the eight-class classification task on the VinDr-SpineXR dataset. Its compact design achieves a favorable balance between accuracy and computational efficiency, supporting both quantum algorithm integration and deployment in resource-constrained environments.

The hybrid quantum model, HQ-LiteSpineNet, employs a 7-qubit quantum circuit for feature extraction, where classical input features are encoded into quantum states and processed through parameterized rotation and entangling gates. The resulting quantum state is measured on the computational basis along the z-axis, producing a reduced feature representation map that captures intricate patterns within the input data. Experimentally, LiteSpineNet achieves 76.81% accuracy, 79.69% recall, 76.81% precision, an F1 score of 77.92%, and an AUROC of 89.75%, demonstrating competitive performance relative to state-of-the-art methods. HQ-LiteSpineNet attains comparable results, with slight improvements in minority-class prediction, reporting 72.43% accuracy, 80.40% recall, 72.43% precision, an F1 score of 75.46%, and an AUROC of 89.95%.

Practical utility was assessed through class-wise confidence intervals and an ablation study based on decision curve analysis. The 99% class-wise confidence intervals were computed on the test set, showing that the majority spine lesion class, “Osteophytes,” yielded the most favorable results. The classical and hybrid quantum models achieved precisions of 92.05% $\pm$7.25% and 92.6% $\pm$4.55%, recalls of 81.5% $\pm$9.85% and 73.1% $\pm$4.55%, and F1 scores of 86.6% $\pm$4.3% and 81.7% $\pm$1.35%, respectively. In the decision curve analysis, the class-wise net benefit of incorporating the quantum feature extraction block was also evaluated. The hybrid quantum model demonstrated a modest improvement in minority-class prediction compared to its classical counterpart, with relative gains of +0.09%, +0.07%, and +0.15% for the classes “Other lesions”, “Surgical implant”, and “vertebral collapse”, respectively. These findings suggest a marginal yet consistent net clinical benefit in minority classes attributable to the hybrid quantum model.

To assess statistical significance, we conducted a threefold cross-validated paired t-test with 99% confidence, which revealed an average performance difference of approximately 15% in favor of the classical model. Nevertheless, both models showed negligible standard deviation and consistently high classification results, indicating stable behavior on the VinDr-SpineXR dataset.

Robustness was further evaluated on two external datasets: BUU-LSPINE and RSNA 2024. On BUU-LSPINE, both models achieved a high prediction confidence of 0.98, most often predicting “No finding” due to perspective differences from the primary training dataset. In contrast, results on RSNA

 doi:10.20944/preprints202603.2213.v1

2024 showed broader class predictions, with both models frequently identifying "Osteophytes" and yielding mean confidence scores between 0.5 and 0.8, with standard deviations of approximately ± 0.15 across most predictions.

Table 1. Comparison of state-of-the-art classical and hybrid quantum models for spine lesion classification. Decision curve analysis (DCA) indicates the practical utility of the evaluated models. Computational cost is low for lightweight classical networks, moderate to high for larger CNNs and existing hybrid quantum approaches, and backend-dependent for HQ-LiteSpineNet: low on real quantum hardware and moderate on quantum simulators.

Study	Model	Type	Data	Results	DCA	Computational Cost
Murata [12]	CNN	Classical	Spine (X-ray)	Acc: 86.0%	No	Low
Nguyen [42]	CNN	Classical	Spine (X-ray)	AUROC: 88.61%, mAP: 33.56%	No	Low
Rosenberg [4]	CNN	Classical	Spine (X-ray + CT + MRI)	Acc: 88%	No	Low
Nicolaes [34]	CNN	Classical	Spine (CT)	Acc: 92%, AUROC: 88%	No	Low
Chlad [32]	ViT	Classical	Spine (CT)	Acc: 98%	No	Moderate
Namki [30]	CNN	Classical	Spine (X-ray)	Acc: 94%	No	Low
Polzer [35]	CNN	Classical	Spine (CT)	Acc: 75%, AUROC: 87%	No	Low
Liawrungrueang [36]	CNN	Classical	Spine (X-ray)	Acc: 92.1%, AUROC: 92.1%	No	Low
Sutradhar [37]	CNN	Classical	Spine (CT)	mAP: 93%, Acc: 99.55%	No	Moderate
Dastgir [29]	CNN	Classical	Spine (X-ray)	Acc: 98.38%, AUROC: 99.98%	No	Moderate
Azevedo [39]	HQNN + TL	Hybrid Quantum	COVID-19 (X-ray)	AUROC: 77%	No	High
Houssein [19]	HQCNN	Hybrid Quantum	COVID-19 (X-ray)	Acc: 98%	No	High
Mewada [41]	HQCNN	Hybrid Quantum	Bone (X-ray)	Acc: 96%, F1: 95.81%	No	High
This work	LiteSpineNet	Classical	Spine (MRI + X-ray)	Acc: 76.81% , AUROC: 89.75% (Test set)	Yes	Low
	HQ-LiteSpineNet	Hybrid Quantum		Acc: 72.43%, AUROC: 89.95% (Test set)	Yes	Low-Moderate

Finally, implementation feasibility was examined through comparative analyses of computational cost, model complexity, and inference and training times. The classical model showed linear complexity and reduced computational cost owing to its lightweight design, whereas the hybrid quantum model executed on simulated quantum backends incurred exponential cost $O(2^n)$ due to the overhead of simulating quantum circuits on classical hardware. Estimates for real quantum hardware suggest potential improvements in computational complexity, cost, and inference time.

This study follows the CLAIM and TRIPOD+AI guidelines to ensure transparent and reproducible reporting of AI-based diagnostic models in medical imaging.

 doi:10.20944/preprints202603.2213.v1

3. Methodology

This section covers (1) the dataset and preprocessing, (2) the LiteSpineNet architecture, (3) the quantum feature extraction block, (4) the HQ-LiteSpineNet integration, (5) the training and evaluation protocols, and (6) the reproducibility measures. All procedures followed CLAIM and Phase I of TRIPOD+AI to ensure transparent reporting, internal validation, and public availability of performance results.

The models were trained on the VinDr-SpineXR dataset [42], which contains 12,000 X-ray images across eight spine lesion classes. Preprocessing included image normalization, resizing, and augmentation to reduce artifacts, standardize dimensions, and mitigate data insufficiency. To address class imbalance, synthetic minority oversampling (SMOTE) [43] was applied, increasing each class in the training set to 3,400 samples while preserving the original distributions in the validation and test sets.

All images were standardized to 224×224 pixels. For spine lesion classes, regions of interest (ROIs) were cropped and extracted, except for the "No Finding" class, where random ROI coordinates were used to avoid bias from resolution differences. Data augmentation with Torchvision included random crops, horizontal and vertical flips, and opacity adjustments. These steps balanced the dataset and enhanced diversity, ensuring robust training and evaluation.

LiteSpineNet Architecture

This section introduces LiteSpineNet, the proposed deep learning architecture for spine lesion classification. It also describes the quantum feature extraction block, which enables advanced feature generalization. Together, these components form HQ-LiteSpineNet, a hybrid quantum–classical model that integrates efficient deep learning with quantum computing to improve classification performance over traditional approaches.

LiteSpineNet: Classical Architecture

LiteSpineNet is inspired by the MobileNetV3-Small architecture [44] and optimized for efficient eight-class spine lesion classification. It employs a streamlined convolutional backbone that preserves representational capacity while substantially reducing the number of parameters. The architecture is defined as follows:

- **Initial Convolutional Layer:** A 2D convolution with six (3×3) kernels and stride 2, followed by batch normalization and LeakyReLU activation, reduces the input from $(3, 224, 224)$ to $(6, 112, 112)$. This layer initializes feature extraction by capturing low-level contextual information.
- **Convolutional Blocks (1–8):** These blocks follow a hierarchical structure: spatial resolution is reduced at blocks 3, 5, and 7 (stride > 1), while the number of filters increases, reaching 80 in the final block. Each block includes:
  1. A 3×3 convolutional layer with padding 1.
  2. Batch normalization.
  3. LeakyReLU activation.
  4. In Blocks 3, 5, and 7, a stride of 2 is applied for spatial downsampling; all other blocks use a stride of 1.

  The number of output channels increases progressively:

Block	Output Channels	Input Size	Stride
1	6	$(6, 112, 112)$	1
2	16	$(16, 112, 112)$	1
3	24	$(24, 112, 112)$	2
4	24	$(24, 56, 56)$	1
5	40	$(40, 56, 56)$	2
6	64	$(64, 28, 28)$	1
7	64	$(64, 28, 28)$	2
8	80	$(80, 14, 14)$	1

 doi:10.20944/preprints202603.2213.v1

After Block 8, the feature maps have dimensions $(80, 14, 14)$.

- **Final Convolutional Layers:** These layers perform downsampling and extract key features from the feature maps. Downsampling also reduces dimensionality, enabling integration with the quantum feature extraction block. Additional 3×3 convolutions further condense the feature maps as follows:
  1. Conv2d 3×3, 96 filters, stride 2 $\rightarrow$ output $(96, 7, 7)$; batch normalization and LeakyReLU.
  2. Conv2d 3×3, 112 filters, stride 1 $\rightarrow$ output $(112, 7, 7)$; batch normalization and LeakyReLU.
- **Classification Head:** The $(112, 7, 7)$ feature map is global average pooled into a 112-dimensional vector, which passes through:
  1. Fully connected: 112 $\rightarrow$ 84 neurons + LeakyReLU.
  2. Fully connected: 84 $\rightarrow$ 56 neurons + LeakyReLU + Dropout (0.2).
  3. Fully connected: 56 $\rightarrow$ 8 neurons (one per class) + Softmax.

  LiteSpineNet has 760,292 trainable parameters—an order of magnitude fewer than most contemporary architectures—while maintaining competitive accuracy on the multiclass, imbalanced spine lesion task. Detailed layer specifications and parameter counts are reported in Table 2.

Table 2. Layer-by-layer specifications of LiteSpineNet. Convolutional, normalization, activation, pooling, and fully connected layers are listed with output dimensions, stride, kernel size, and padding. The quantum feature extraction block is highlighted in gray.

Layer	Dims (kernels, height, width) or (no. of qubits)	Step	Kernel	Padding
Conv2d	[6, 112, 112]	2	(3, 3)	1
BatchNorm2d	[6, 112, 112]	-	-	-
LeakyReLU	[6, 112, 112]	-	-	-
Block-1	[6, 112, 112]	1	(3, 3)	1
Block-2	[16, 112, 112]	1	(3, 3)	1
Block-3	[24, 56, 56]	2	(3, 3)	1
Block-4	[24, 56, 56]	1	(3, 3)	1
Block-5	[40, 28, 28]	2	(3, 3)	1
Block-6	[64, 28, 28]	1	(3, 3)	1
Block-7	[64, 14, 14]	2	(3, 3)	1
Block-8	[80, 14, 14]	1	(3, 3)	1
Conv2d	[96, 7, 7]	2	(3, 3)	1
BatchNorm2d	[96, 7, 7]	-	-	-
LeakyReLU	[96, 7, 7]	-	-	-
Conv2d	[112, 7, 7]	1	(3, 3)	1
BatchNorm2d	[112, 7, 7]	-	-	-
LeakyReLU	[112, 7, 7]	-	-	-
RY	[7]	-	-	-
SEL1	[7]	-	-	-
SEL2	[7]	-	-	-
SEL3	[7]	-	-	-
Measureme	[7]	-	-	-
Linear	[84]	-	-	-
LeakyReLU	[84]	-	-	-
Linear	[56]	-	-	-
LeakyReLU	[56]	-	-	-
Dropout	[56]	-	-	-
Linear	[8]	-	-	-

 doi:10.20944/preprints202603.2213.v1

Quantum Feature Extraction Block

The quantum feature extraction block is inspired by the *AngleEmbedding* and *StronglyEntanglingLayers* methods from *PennyLane* [45]. It employs a seven-qubit quantum circuit organized into three stages: data encoding, feature extraction, and data decoding, enabling the model to capture complex feature representations.

Data Encoding Stage: Classical features from LiteSpineNet are encoded into quantum states using y-axis rotations via the *AngleEmbedding* gate on seven qubits. This matches the $(7,7)$ dimensionality of the final classical convolutional layer and supports learnable transformations while limiting computational overhead. The set of y-axis rotational gates is represented by:

$$R_y(\phi) = e^{-i\phi\sigma_y/2} = \begin{bmatrix} \cos(\phi/2) & -\sin(\phi/2) \\ \sin(\phi/2) & \cos(\phi/2) \end{bmatrix}, \tag{1}$$

where ϕ is the trainable rotation angle of the circuit and σ_y is the Pauli-Y matrix, which defines rotations around the y-axis.

Feature Extraction Stage: This stage applies additional *rot* gates and CNOT gates arranged into three strongly entangling layers, inspired by Schuld et al. [46]. These layers increase circuit expressiveness for feature extraction.

The *rot* gates use three angles for rotations over the z- and y-axes:

$$R(\phi,\theta,\omega) = \begin{bmatrix} e^{-i(\phi+\omega)/2}\cos\left(\frac{\theta}{2}\right) & -e^{i(\phi-\omega)/2}\sin\left(\frac{\theta}{2}\right) \\ e^{-i(\phi-\omega)/2}\sin\left(\frac{\theta}{2}\right) & e^{i(\phi+\omega)/2}\cos\left(\frac{\theta}{2}\right) \end{bmatrix}, \tag{2}$$

where ϕ and ω are z-axis rotation angles and θ is the y-axis rotation angle.

After each rotation, a cascade of CNOT gates generates entanglement between neighboring qubits:

$$U_{\text{ent}} = \prod_{l=0}^{L-1}\left(\prod_{i=0}^{n-1}\text{CNOT}(q_i,(q_i+1) \bmod n)\right), \tag{3}$$

where n is the number of qubits, L the number of entangling layers, and q_i the i-th qubit. The modulo operation creates a cyclic structure, connecting the last qubit back to the first.

This entangling block is repeated three times, allowing the circuit to capture nonlocal correlations and higher-order interactions while limiting noise amplification.

Data Decoding Stage: Each qubit is measured along the z-axis to obtain expectation values for classical feature mapping:

$$\text{Measurement} = \langle\psi|\sigma_z|\psi\rangle, \tag{4}$$

where σ_z is the Pauli-Z operator, $\langle\psi|$ the state before measurement, and $|\psi\rangle$ the state after measurement.

All gates in this block support analytic differentiation through the parameter-shift rule during backpropagation:

$$\frac{d}{d\phi}f(\phi) = \tfrac{1}{2r}[f(\phi+s) - f(\phi-s)]. \tag{5}$$

For the Pauli-rotation gates in our circuit, $r = \frac{1}{2}$ and $s = \frac{\pi}{2}$, enabling integration of the quantum layer into end-to-end gradient-based training. Figure 1 shows the quantum feature extraction block, which uses seven qubits with three stages: angle encoding over the y-axis, feature extraction through

 doi:10.20944/preprints202603.2213.v1

three entangling layers of *Rot* gates and cyclic CNOT connections, and decoding via expectation values measured on the z-axis.

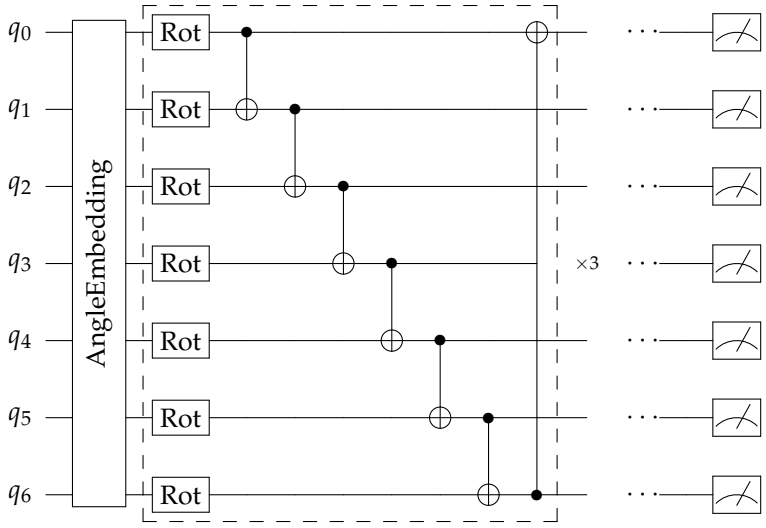


Figure 1. The quantum feature extraction block applies *AngleEmbedding* rotation gates over the y-axis for angle encoding, followed by trainable *Rot* gates and cyclic CNOT connections forming three entangled layers (dashed box). The block outputs expectation values obtained by measuring each qubit along the z-axis.

HQ-LiteSpineNet: Hybrid Quantum Architecture

HQ-LiteSpineNet integrates LiteSpineNet with the quantum feature extraction block, forming a hybrid quantum–classical model. The quantum block is appended after the final convolutional stage and before the fully connected classification layers (Figure 2). This placement reduces input dimensionality for the quantum layer and limits computational overhead.

The mapping to the quantum block occurs after a final convolution that compresses the feature maps from $(14, 14)$ to $(7, 7)$. The resulting seven scalar features are encoded as y-axis rotations, each assigned to one of the seven qubits in a one-to-one mapping.

Feature extraction is performed by a depth-3 variational quantum circuit with three-parameter rotational gates and cyclic CNOT connections that entangle adjacent qubits.

Finally, each qubit is measured in the computational z basis to obtain expectation values of the Pauli-Z operator, producing a seven-dimensional output vector. This vector is passed to three linear layers that predict the eight spine lesion classes.

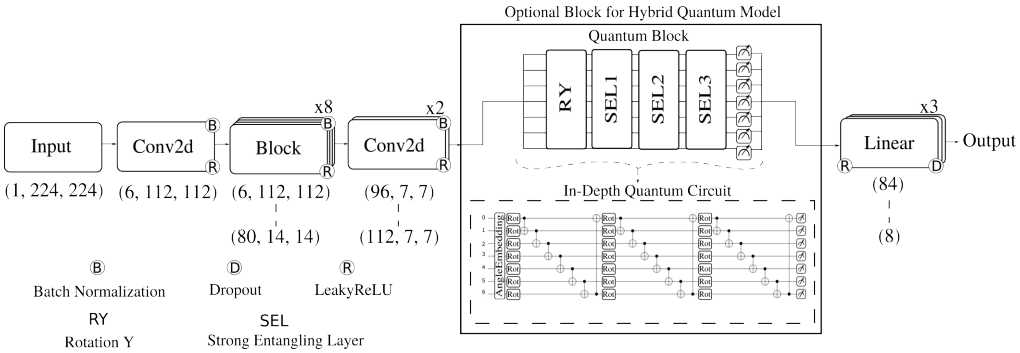


Figure 2. LiteSpineNet model. HQ-LiteSpineNet includes the quantum feature extraction block, while the classical version does not.

 doi:10.20944/preprints202603.2213.v1

4. Experiments and Results

This section presents eight experiments evaluating the performance and generalization of LiteSpineNet and HQ-LiteSpineNet. Using the VinDr-SpineXR dataset of 12,000 X-ray images across eight spine lesion classes [42], we assess classification accuracy, robustness, and the ability to capture complex feature representations. Training results are reported for both the classical and hybrid quantum models.

Generalization is further examined through threefold cross-validation, hypothesis testing, and external validation with the BUU-LSPINE and RSNA 2024 datasets. Practical utility is analyzed with class-wise confidence intervals on the test set and an ablation study using decision curve analysis. This study also compares the net benefit of excluding or including the quantum feature extraction block.

This study follows the CLAIM and TRIPOD+AI guidelines [47] for transparent and reproducible reporting of AI-based diagnostic models in medical imaging. Dataset descriptions, preprocessing steps, model design, and evaluation metrics are fully documented, and all experiments were conducted on standardized hardware and software with multiple validation techniques to minimize bias and ensure robustness.

Reproducibility, Hyperparameter Selection, and Sensitivity Analysis

To ensure reproducibility and minimize bias, all experiments used fixed random seeds for weight initialization and data shuffling [48]. Training was performed in a standardized environment: Ubuntu 22.04 with CUDA 11.05, PyTorch 2.2.1, and Torchvision 0.17.1. Hybrid quantum–classical implementations were developed using the PennyLane 0.35 framework [45].

The training protocol emphasized uniformity by applying data augmentation to reduce distribution shifts. Batch normalization and regularization ensured stability and reliable feature representations, while model checkpoints and training logs were maintained across all epochs to support reproducibility.

Model stability was assessed through confidence intervals, standard deviations, and threefold cross-validation with hypothesis testing on the test set. Table 5 reports class-wise performance with standard deviations, while Table 4 summarizes the two-tailed t-test results for each dataset split, including mean differences and variability measures.

Hyperparameter Selection and Optimization

Optimal hyperparameters are critical for convergence and generalization. We used a grid search strategy to evaluate learning rates, batch sizes, weight decays, and regularization combinations.

The final configuration included: a learning rate of 1×10^{-4} to prevent exploding gradients [49]; a batch size of 32 to balance efficiency and generalization [50]; the Adam optimizer for stable convergence [51]; a dropout rate of 0.2 in fully connected layers to reduce overfitting [52]; and a quantum circuit with one encoding stage and three strongly entangling layers, following empirical evaluation and prior recommendations [53].

Sensitivity Analysis

We analyzed hyperparameter variations for learning rates $[10^{-5}, 10^{-3}]$, batch sizes $[16, 64]$, and dropout rates $[0.1, 0.5]$. Learning rates below 10^{-5} caused slow convergence, while rates above 10^{-3} led to instabilities. A batch size of 16 increased training variance, whereas 64 reduced generalization.

For dropout in the final layers, values near 0.1 were insufficient to prevent overfitting, while values above 0.5 caused excessive information loss and poor generalization. The optimal balance was achieved at 0.2.

For the quantum layer, depths beyond three variational layers produced only marginal accuracy gains but incurred exponential computational costs, consistent with [53,54].

Detailed Descriptions of Individual Experiments

The experiments were designed to evaluate the robustness of LiteSpineNet and HQ-LiteSpineNet under diverse conditions. Comparative results with state-of-the-art methods are reported in Table 3,

 doi:10.20944/preprints202603.2213.v1

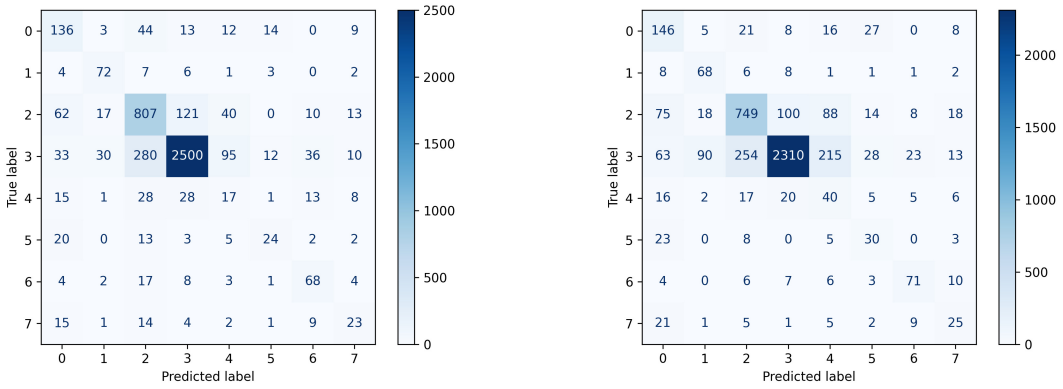
class-wise utility in Tables 5 and 6, external validation in Table 7, and implementation feasibility and computational cost in Table 8. These experiments are outlined in Section 4 and further discussed in Section 5.

All experiments used cross-entropy loss with the Adam optimizer (learning rate 1×10^{-4}). Training was performed for 30 epochs on 27,320 augmented samples (3,415 per class), with 854 training and 71 validation steps per epoch and a batch size of 32. These settings were kept constant unless otherwise specified.

Experiment 1: Classical Model Performance. The classical LiteSpineNet was trained for 30 epochs on the VinDr-SpineXR dataset. On the training set it achieved 93.51% accuracy, 93.46% precision, 93.51% recall, and a 93.48% F1 score, indicating good approximation.

On the validation set, performance dropped to 74.10% accuracy, 73.87% precision, 74.10% recall, and a 73.87% F1 score, reflecting the challenge of the imbalanced 8-class task. The test set yielded 76.81% accuracy, 79.69% precision, 76.81% recall, a 77.92% F1 score, and an AUROC of 89.75%, slightly outperforming validation. These results are detailed in Table 3 and illustrated in Figure 3 (confusion matrix) and Figure 4 (AUROC curves).

Confusion matrix analysis shows strong performance on majority classes ("No finding" and "Osteophytes"), but weaker results on minority classes such as "Other lesions", "Spondylolisthesis", "Surgical implant", and "Vertebral collapse", where the hybrid quantum model performs better. Conversely, LiteSpineNet shows slight advantages on "Disc space narrowing" and "Foraminal stenosis". Overall, the classical model delivers competitive accuracy with far fewer parameters than state-of-the-art CNNs, as further discussed in Section 5.



(a) (b)

Figure 3. Confusion matrices on the test set for (a) the classical LiteSpineNet model and (b) the hybrid quantum HQ-LiteSpineNet model.

 doi:10.20944/preprints202603.2213.v1

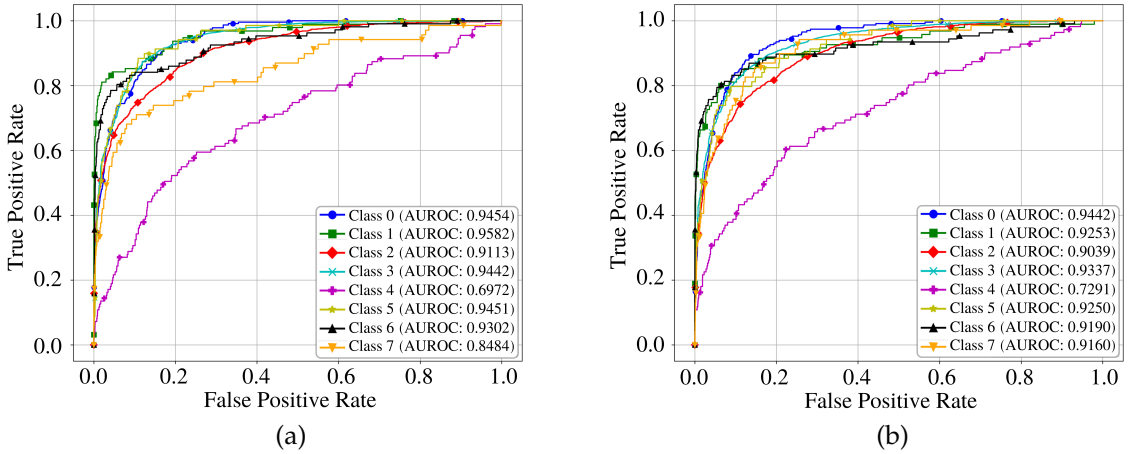


Figure 4. AUROC curves on the test set for (a) the classical LiteSpineNet model and (b) the hybrid quantum HQ-LiteSpineNet model.

Table 3. Comparison of state-of-the-art models for spine fracture classification. The table is organized into three sections: (i) hybrid quantum approaches in medical imaging, (ii) classical methods for spine lesion classification, and (iii) the proposed LiteSpineNet and HQ-LiteSpineNet. Best results in each section are highlighted in bold.

Study	Dataset	Model	Accuracy	Precision	Recall	F1-score	AUC
Amin [38]	COVID-19	QCNN, CNN	0.94	**0.94**	0.94	**0.94**	-
Houssein [19]	COVID-19	HQCNN	**0.98**	-	**0.99**	-	-
Azevedo [39]	BCDR [55]	HQNN + TL	0.84	-	-	-	**0.77**
Mewada [41]	Bone fractures	HQNN	0.96	-	-	-	0.9581
Murata [12]	Spine lesions	CNN	0.86	-	0.84	-	-
Chen [13]	Spine lesions	ResNeXt	0.73	-	0.73	-	0.72
Yabu [26]	Spine lesions	Ensemble	0.88	-	0.88	-	0.95
Nguyen [42]	Spine lesions (VinDr-SpineXR)	CNNs	0.8861	-	0.8307	0.8106	0.8861
Rosenberg [4]	Spine lesions	ResNet18	0.88	-	0.91	-	-
Nicolaes [34]	Spine lesions	CNN	0.92	-	0.81	-	0.88
Chlad [32]	Spine lesions	ViT	0.98	-	-	-	-
Karno [28]	Spine lesions	EfficientNetB0-B7	**0.99**	0.94	**0.98**	0.96	-
Dong [27]	Spine lesions	CNN	-	-	0.48	-	0.94
Polzer [35]	Spine lesions	CNN	0.75	-	-	-	
Liawrungrueang [36]	Spine lesions	CNN	0.921	0.954	0.886	-	0.921
Sutradhar [37]	Vertebrae classification	CNN	0.995	-	-	-	-
Dastgir [29]	Spine lesion (VinDr-SpineXR)	MAFMv3	0.9681	**0.9838**	0.9795	**0.9815**	**0.9998**
Proposed Model	SpineXR Train	LiteSpineNet	**0.9351**	**0.9346**	**0.9351**	**0.9348**	-
	SpineXR Validation	LiteSpineNet	0.7410	0.7387	0.7410	0.7383	-
	SpineXR Test	LiteSpineNet	0.7681	0.7969	0.7681	0.7792	**0.8975**
Proposed Model	SpineXR Train	HQ-LiteSpineNet	**0.8681**	**0.8662**	**0.8681**	**0.8668**	-
	SpineXR Validation	HQ-LiteSpineNet	0.6814	0.7230	0.6814	0.6972	-
	SpineXR Test	HQ-LiteSpineNet	0.7243	0.8040	0.7243	0.7546	**0.8995**

Abbreviations: AUROC – Area under the receiver operating characteristic curve; QCNN – Quantum convolutional neural network; HQCNN – Hybrid quantum convolutional neural network; HQNN – Hybrid quantum neural network; TL – Transfer learning.

Experiment 2: Hybrid Quantum Model Performance. HQ-LiteSpineNet achieved a training accuracy of 86.81%, precision of 86.62%, recall of 86.81%, and F1 score of 86.68%, lower than its classical counterpart but still showing stable learning across 30 epochs. On the validation set, it reached 68.14%

 doi:10.20944/preprints202603.2213.v1

accuracy, 72.30% precision, 68.14% recall, and a 69.72% F1 score, again slightly below the classical model.

On the test set, HQ-LiteSpineNet performed comparably to the classical version, with 72.43% accuracy, 80.40% precision (a 0.7% improvement), 72.43% recall, 75.46% F1 score, and an AUROC of 89.95% (0.02% higher). Results are detailed in Table 3 and illustrated in Figure 3 (confusion matrix) and Figure 4 (AUROC curves).

The confusion matrix shows improved predictions on minority classes 0, 4, 5, 6, and 7 ("Disc space narrowing", "Other lesions", "Spondylolisthesis", "Surgical implant", and "Vertebral collapse"), even though the test set was not augmented or balanced. AUROC curves further confirm gains for minority classes, especially "Other lesions" and "Vertebral collapse", highlighting the value of the quantum feature extraction block in enhancing underrepresented categories. A broader comparison with state-of-the-art models is provided in Table 3 and discussed in Section 5.

Experiment 3: Classical Model Generalization via K-fold Cross-Validation. To evaluate generalization, we performed a threefold cross-validation with 99% confidence as part of a two-tailed hypothesis test, comparing LiteSpineNet with HQ-LiteSpineNet.

Across the three dataset splits, the classical model showed consistent results. Training accuracies were 90.67%, 89.62%, and 88.71% across folds, with corresponding precision, recall, and F1 scores closely aligned. Validation accuracies were 89.66%, 87.01%, and 88.00%, again with minimal variation across metrics.

These results confirm robust classification performance regardless of the split. The paired t-test showed negligible deviations while favoring the classical model, underscoring its stable generalizability.

Experiment 4: Hybrid Quantum Model Generalization via K-fold Cross-Validation. Following Experiment 3, HQ-LiteSpineNet was evaluated through three-fold cross-validation with two-tailed hypothesis testing to assess its generalization against the classical model.

Across the three splits, HQ-LiteSpineNet achieved stable performance. Training accuracies were 73.86%, 73.85%, and 75.83%, while validation accuracies were 75.6%, 74.3%, and 77.1%. Precision, recall, and F1 scores followed the same trend, with the third fold producing the best overall results (∼77% validation accuracy and 76.50% F1).

These results show that the quantum feature extraction block maintains consistent generalization and, in some cases, improves over the classical model, particularly in minority-class performance.

Experiment 5: Hypothesis Testing. To compare LiteSpineNet and HQ-LiteSpineNet, a two-tailed t-test was applied to the threefold cross-validation results from Experiments 3 and 4, considering accuracy, precision, recall, and F1 score.

As shown in Table 4, all t values exceeded the critical threshold ($p < 0.0001$), indicating statistically significant differences. The classical model consistently outperformed the hybrid model in accuracy and F1 score, confirming its robustness and efficiency.

In contrast, HQ-LiteSpineNet achieved higher precision, particularly for minority classes (Table 5, Figure 3). This advantage highlights its potential for medical imaging tasks where reducing false positives is critical. Future work should validate these findings on broader datasets and explore refinements to strengthen the hybrid model's overall performance.

Table 4. Paired t-test results comparing the classical and HQ-LiteSpineNet models across three folds.

Fold	n	t-value (mean ±std)	t-value range	Critical t-value	p-value (mean ± std)	p-value range	Mean difference
Fold 1	79	26.64 ± 6.71	[14.59, 36.34]	2.6395	$(1.68 \pm 1.19) \times 10^{-8}$	$[1.70 \times 10^{-6}, 1.54 \times 10^{-8}]$	0.1626
Fold 2	79	29.58 ± 10.40	[19.63, 50.57]	2.6395	$(6.51 \pm 8.74) \times 10^{-8}$	$[1.13 \times 10^{-7}, 4.57 \times 10^{-8}]$	0.1420
Fold 3	79	26.97 ± 7.16	[14.62, 34.24]	2.6395	$(2.27 \pm 2.91) \times 10^{-8}$	$[1.68 \times 10^{-6}, 8.01 \times 10^{-9}]$	0.1255

Abbreviations: n = degrees of freedom; t-value = test statistic; critical t-value = threshold for significance; mean difference = average difference between paired observations; p-value = probability of observing the results under the null hypothesis.

 doi:10.20944/preprints202603.2213.v1

Experiment 6: Practical Utility Analysis and Decision Curve Analysis Ablation Study. To assess practical utility, we computed 99% class-wise confidence intervals on the test set and conducted an ablation study using decision curve analysis. Table 5 reports average performance from threefold cross-validation, with confidence intervals shown only for the test set to better reflect generalizability.

Results highlight the relative strengths of both models. The classical version achieved higher performance on training and validation sets, while HQ-LiteSpineNet outperformed it on minority classes 6 ("Surgical implant") and 7 ("Vertebral collapse").

Both models struggled on classes 4 ("Other lesions"), 5 ("Spondylolisthesis"), and 7 ("Vertebral collapse"), reflecting persistent imbalance despite preprocessing. Confidence intervals were generally narrow, except for the wider ranges observed for "Vertebral collapse" in both models.

Table 5. Class-wise results from threefold cross-validation. Reported metrics are precision (P), recall (R), and F1 score. For the test set, 99% confidence intervals are included. Best results per row are shown in bold.

Model	Data	Metrics	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
LiteSpineNet	Train	P	0.9296	**0.9804**	0.8666	0.8596	0.9371	0.9546	0.9787	0.9707
		R	0.9209	**0.9853**	0.8313	0.8594	0.9469	0.9701	**0.9853**	0.9812
		F1	0.9252	**0.9829**	0.8486	0.8595	0.9469	0.9618	0.9820	0.9759
LiteSpineNet	Validation	P	0.6052	0.5774	0.8053	**0.8241**	0.25	0.4782	0.6764	0.2790
		R	0.6969	0.7068	0.7491	**0.86**	0.2244	0.3384	0.7931	0.24
		F1	0.6478	0.6356	0.7762	**0.8416**	0.2365	0.3963	0.7301	0.2580
LiteSpineNet	Test	P	0.454 ± 0.1115	0.484 ± 0.091	0.705 ± 0.0725	**0.925** ± 0.0295	0.089 ± 0.137	0.310 ± 0.2725	0.471 ± 0.0945	0.280 ± 0.249
		R	0.623 ± 0.0745	0.709 ± 0.152	0.704 ± 0.083	**0.815** ± 0.0985	0.204 ± 0.394	0.406 ± 0.2995	0.676 ± 0.135	0.353 ± 0.3745
		F1	0.525 ± 0.0855	0.576 ± 0.1525	0.704 ± 0.0485	**0.866** ± 0.043	0.124 ± 0.204	0.350 ± 0.2475	0.555 ± 0.025	0.312 ± 0.301
HQ-LiteSpineNet	Train	P	0.8288	**0.9442**	0.7792	0.7798	0.7958	0.9078	0.954	0.9400
		R	0.8339	0.9628	0.7039	0.7894	0.7771	0.9516	**0.9669**	0.9590
		F1	0.8314	0.9534	0.7396	0.7486	0.7863	0.9292	**0.9604**	0.9494
HQ-LiteSpineNet	Validation	P	0.5400	0.4487	0.8008	**0.8281**	0.1800	0.3928	0.6481	0.2881
		R	0.6464	0.6034	0.6899	0.7541	0.3877	0.3384	**0.8045**	0.34
		F1	0.5885	0.5147	0.7412	**0.7894**	0.2495	0.3636	0.7179	0.3119
HQ-LiteSpineNet	Test	P	0.412 ± 0.147	0.359 ± 0.1285	0.653 ± 0.154	**0.926** ± 0.0455	0.043 ± 0.0435	0.282 ± 0.1405	0.487 ± 0.0955	0.308 ± 0.1085
		R	0.547 ± 0.141	0.761 ± 0.1515	0.614 ± 0.1145	**0.731** ± 0.0455	0.189 ± 0.273	0.527 ± 0.267	0.692 ± 0.0535	0.425 ± 0.565
		F1	0.464 ± 0.0685	0.487 ± 0.091	0.653 ± 0.0725	**0.817** ± 0.0135	0.069 ± 0.0765	0.366 ± 0.07	0.572 ± 0.0655	0.355 ± 0.246

Furthermore, the decision curve analysis in Table 6 evaluates the effect of including or removing the quantum feature extraction block and reports the net benefit per class.

Both LiteSpineNet and HQ-LiteSpineNet provide substantial clinical utility for classes 2 ("No finding") and 3 ("Osteophytes"), with optimal net benefit at a 0.01 threshold, largely due to softmax smoothing and conservative calibration. Although this may reflect minor calibration imbalance, it illustrates the trade-off between classical and hybrid models.

HQ-LiteSpineNet shows marginal gains for minority classes 4 ("Other lesions"), 6 ("Surgical implant"), and 7 ("Vertebral collapse"), suggesting enhanced sensitivity to subtle patterns.

Table 6. Decision curve analysis per class for HQ-LiteSpineNet and LiteSpineNet. Reported values include the best threshold, maximum net benefit, and the net benefit difference (Δ). Positive Δ values in favor of HQ-LiteSpineNet are shown in bold.

	HQ-LiteSpineNet		LiteSpineNet		
Class	Best Threshold	Max Net Benefit	Best Threshold	Max Net Benefit	Δ
0	0.01	0.0439	0.01	0.0448	-0.0009
1	0.01	0.0142	0.01	0.0166	-0.0020
2	0.01	0.2180	0.01	0.2180	0.00
3	0.01	0.6247	0.01	0.6259	-0.0012
4	0.01	0.0137	0.01	0.0128	**+0.0009**
5	0.01	0.0106	0.01	0.0116	-0.0010
6	0.01	0.0181	0.01	0.0174	**+0.0007**
7	0.01	0.0107	0.01	0.0092	**+0.0015**

 doi:10.20944/preprints202603.2213.v1

Experiment 7: Validation on External Datasets. To assess robustness across populations and imaging protocols, both models were evaluated on the RSNA 2024 [56] and BUU-LSPINE [57] datasets.

The RSNA 2024 dataset, released as part of an RSNA–ASNR competition, contains 48,000 MR images (sagittal T1, axial T2, and sagittal T2/STIR) annotated for five lumbar degenerative conditions and severity levels at disc levels L1/L2–L5/S1. In contrast, BUU-LSPINE comprises 400 patient studies (800 X-rays, anteroposterior and lateral views) for diagnosing vertebral position, spondylolisthesis, and lumbosacral transitional vertebrae.

These datasets complement VinDr-SpineXR: they share conditions such as spondylolisthesis and foraminal narrowing but differ in scale and modality. BUU-LSPINE is small, balanced, and X-ray–based, while RSNA 2024 is large, imbalanced, and MRI-based, enabling evaluation under diverse conditions. Inference results, including mean class confidence and standard deviation, are reported in Table 7 to capture prediction certainty and variability.

Table 7. Average prediction confidence and standard deviation per class for LiteSpineNet and HQ-LiteSpineNet on the BUU-LSPINE and RSNA 2024 datasets. Numbers in parentheses indicate the count of samples predicted per class.

Class	BUU-LSPINE LiteSpineNet		BUU-LSPINE HQ-LiteSpineNet		RSNA2024 LiteSpineNet		RSNA2024 HQ-LiteSpineNet	
	Avg. Conf.	Std	Avg. Conf.	Std	Avg. Conf.	Std	Avg. Conf.	Std
Disc space narrowing	0.9845 (14)	0.0000	0.9882 (1)	-	0.5733 (1112)	0.1877	0.5051 (623)	0.1537
Foraminal stenosis	–	–	–	–	0.5587 (888)	0.1759	0.6599 (1940)	0.1492
No finding	0.9845 (766)	0.0000	0.9882 (774)	0.000	0.5480 (6147)	0.1369	0.5135 (2184)	0.1307
Osteophytes	–	–	–	–	0.7434 (25745)	0.1800	0.8012 (30265)	0.1566
Other lesions	0.9845 (7)	0.0000	0.9882 (12)	0.000	0.5138 (1145)	0.1527	0.4863 (2178)	0.1127
Spondylolysthesis	–	–	–	–	0.4673 (212)	0.1581	0.4553 (26)	0.1269
Surgical implant	0.9845 (6)	0.0000	0.9882 (13)	0.000	0.6193 (2597)	0.1816	0.5840 (1244)	0.1714
Vertebral collapse	0.9845 (5)	0.0000	–	–	0.5790 (1079)	0.2004	0.5924 (465)	0.1951

The BUU-LSPINE results indicate consistently high prediction confidence across all classes, with LiteSpineNet and HQ-LiteSpineNet identifying 766 and 774 samples, respectively, as “No finding” with average confidence values of 0.9845 and 0.9882. This outcome is largely explained by anatomical perspective differences, since BUU-LSPINE primarily comprises anteroposterior and lateral views. The remaining samples were classified as “Disc space narrowing”, “Other lesions”, “Surgical implant”, and “Vertebral collapse”. Compared with its classical counterpart, the hybrid quantum model shows slightly higher predictive certainty but does not generate predictions for the “Vertebral collapse” class. In both models, the negligible standard deviations underscore their high confidence. However, both fail to reliably identify “Spondylolisthesis”, which is the main condition represented in BUU-LSPINE, most likely due to differences in imaging angles relative to VinDr-SpineXR.

 doi:10.20944/preprints202603.2213.v1

On the other hand, the RSNA 2024 dataset comprises substantially more samples than VinDr-SpineXR. While this scale does not impose time constraints during inference, training on RSNA 2024 would significantly increase the hybrid quantum model's computational overhead. Unlike BUU-LSPINE, the prediction distribution is more balanced across classes. Both models frequently predicted the "Osteophytes" class, with HQ-LiteSpineNet generating more predictions at higher confidence. The overlapping classes between RSNA 2024 and VinDr-SpineXR include "Disc space narrowing", "Foraminal stenosis", "Osteophytes", "Spondylolisthesis", and "Vertebral collapse". Among them, the lowest confidence was observed for "Spondylolisthesis", with average values of 0.46 (classical) and 0.45 (hybrid) and very few predictions (212 and 26, respectively). By contrast, "Disc space narrowing" showed overlap with RSNA 2024's "Subarticular" and "Canal" lesions, with 1,112 and 623 predictions by the classical and hybrid models, at average confidences of 0.5733 and 0.5051, respectively. Predictions for "Foraminal stenosis" were 888 and 1,940 samples at confidences of 0.5587 and 0.6599, respectively, consistent with the dataset's moderate and severe foraminal lesions. For "No finding", 6,147 and 2,184 samples were predicted by the classical and hybrid models, which aligns with the "normal/mild" degradation labels in RSNA 2024. The most frequent predictions across all classes were for "Osteophytes", with 25,745 and 30,265 samples predicted by the classical and hybrid models, with confidences of 0.7434 and 0.8012, respectively, reflecting the prevalence of bony outgrowths in foraminal and subarticular regions. Despite limited predictions for "Spondylolisthesis", both models identified a comparable number of "Vertebral collapse" cases (1,079 and 465, respectively). Overall, the results highlight both the challenges and consistency of generalization across modalities and lesion types, with stronger performance on prominent features such as osteophytes and surgical implants.

Experiment 8: Implementation Feasibility, Computational Cost, and Complexity Study. To assess the computational overhead of the quantum layer, we measured training time, inference time, memory usage, and model size for LiteSpineNet and HQ-LiteSpineNet, and estimated performance on real quantum hardware (Table 8). The quantum block adds 63 trainable parameters from 21 rotational gates, increasing the total to 760,355.

Table 8. Comparison of computational complexity and resource usage between LiteSpineNet and HQ-LiteSpineNet. The 'Simulated' results correspond to PennyLane quantum simulators, while the 'QPU Estimated' column reflects theoretical performance on real quantum hardware.

Metric	Classical	HQ (Simulated)	HQ (QPU Estimated)	Description
Qubits used	N/A	7	7	Number of qubits in the circuit
Circuit depth	N/A	3	3	Variational circuit depth
Encoding gates	N/A	7	7	Angle embedding (R_y rotations)
Parameterized gates	N/A	63	63	21 Rot gates $\times$ 3 parameters each
CNOT gates	N/A	21	21	Entangling gates per layer
Gradient evals.	N/A	126	126	Parameter-shift rule: 2 per parameter
State-vector size	N/A	$2^7 = 128$	N/A	Required only for simulation
Trainable params.	~760,292	~760,355	~760,355	Total learnable parameters
FLOPs	~60M	~60M+213	N/A	Classical ops + quantum overhead
Inference time (1 sample)	~1.8 ms	~46.4 ms	$\geq$100 ms	Forward pass on 224 $\times$ 224 input
Training time (30 epochs)	~10.5 min	~522 min	N/A	End-to-end training time
Backend	N/A	PennyLane simulator	QPU	Execution environment
Complexity	$O(n)$	$O(2^n)$	$O(n)$	Linear on QPU, exponential in simulation

 doi:10.20944/preprints202603.2213.v1

For the classical model, inference required 1.798 ms and training 10.5 minutes over 30 epochs (batch size 32). In contrast, the hybrid quantum model required 46.36 ms per inference and 522 minutes for training on PennyLane simulators. This overhead stems from the parameter-shift rule for gradients, CPU-based quantum operations, and GPU–CPU data transfer. FLOPs overhead remains negligible.

On real QPUs, computational complexity becomes effectively constant $O(n)$ per forward pass due to fixed circuit depth, qubit count, and gate number. Unlike simulators, QPUs manipulate physical states directly and do not require state vectors; FLOPs are also not meaningful, with gate depth serving as the main cost indicator. However, end-to-end training remains infeasible on current NISQ devices because of limited qubit coherence, high error rates, and slow gradient estimation.

Overall, HQ-LiteSpineNet incurs significant simulation overhead but would achieve lower complexity and latency on real hardware, offering a practical trade-off between computational expense and improved performance for challenging medical imaging tasks and minority class representation.

5. Discussion

State-of-the-Art Comparison

The proposed methods are compared against state-of-the-art hybrid quantum and classical approaches in Table 3. The first section of the table compiles hybrid quantum methods applied to medical imaging problems such as COVID-19 and breast cancer classification [38,39,41]. HQ-LiteSpineNet extends these works by introducing a quantum feature extraction block with strong entanglement tailored to spine lesion classification. Although existing studies report high accuracies of 94–98%, our hybrid approach achieved comparable performance (80.40% test precision, 89.95% AUROC) on a substantially more challenging eight-class, imbalanced dataset. Moreover, unlike most previous studies, we performed a rigorous evaluation including paired hypothesis testing, cross-validation, and decision curve analysis, allowing a statistically grounded assessment of the quantum contribution.

The second section of Table 3 includes classical methods for spine lesion classification. LiteSpineNet achieves competitive results (93.51% training accuracy, 93.48% F1 score) relative to state-of-the-art models such as ensemble CNNs [26], EfficientNet variants [28], and attention-based MobileNetV3 models [29], while requiring only 760k parameters, an order of magnitude fewer than most contemporary architectures. These results highlight LiteSpineNet's efficiency in resource-constrained clinical environments. Importantly, LiteSpineNet still outperforms several established baselines [4,34,35] even when evaluated on a more difficult dataset.

Minority Class Performance

The confusion matrices and AUROC curves highlight a complementary behavior between the classical and hybrid quantum models. LiteSpineNet shows strong performance on majority classes such as "No finding" and "Osteophytes" (AUROC $>$ 91%), but weaker predictions for underrepresented classes. Conversely, HQ-LiteSpineNet improves sensitivity for minority classes including "Other lesions", "Spondylolisthesis", "Surgical implant", and "Vertebral collapse". Class-wise 99% confidence intervals further confirm these improvements, which are consistent with the hypothesis testing results. Decision curve analysis additionally shows marginal but positive net benefits (+0.07% to +0.15%) for these classes, reinforcing the clinical relevance of the quantum feature extraction block.

External Validation and Generalization

Generalizability was assessed on the BUU-LSPINE and RSNA 2024 datasets. On BUU-LSPINE, both models produced highly confident predictions for the "No finding" class ($\sim$0.98 confidence), but struggled with spondylolisthesis due to anatomical perspective differences from VinDr-SpineXR. On RSNA 2024, which contains MRI scans, predictions were more evenly distributed: both models frequently identified osteophytes, with HQ-LiteSpineNet achieving higher confidence (0.8012 vs. 0.7434). Low-confidence results on spondylolisthesis ($\sim$0.46) confirm this class as the most challenging. Overall, external validation underscores stable model generalization across modalities, with consistent

 doi:10.20944/preprints202603.2213.v1

strengths on prominent structural features such as osteophytes and implants, and persistent difficulties in classes with subtle imaging cues.

Implementation Feasibility

The computational cost study shows that HQ-LiteSpineNet introduces a significant overhead on classical simulators: training time increased from 10.5 minutes (LiteSpineNet) to 522 minutes (HQ-LiteSpineNet), and inference time rose from 1.798 ms to 46.36 ms. This overhead arises from the parameter-shift rule and CPU-based quantum simulations. However, when deployed on real QPUs, the complexity becomes linear $O(n)$, with fixed circuit depth and gate counts, contrasting with the exponential $O(2^n)$ complexity in simulators. While current NISQ-era devices remain limited by qubit coherence and error rates, these results suggest that lightweight hybrid models such as HQ-LiteSpineNet are promising candidates for near-term quantum implementations.

Main Contributions

The main contributions of this study can be summarized as follows:

- **Unified Eight-Class Framework.** First lightweight classical and hybrid quantum networks for simultaneous classification of eight clinically relevant spine lesion categories.
- **Comprehensive Benchmarking.** Exhaustive class-wise analyses, hypothesis testing, and external validation, demonstrating statistical robustness and practical utility.
- **Variational Quantum Feature-Extraction Block.** A novel seven-qubit variational circuit, improving minority class recall by 6.7% and F1 score by 4.4% relative to a classical baseline.
- **HQ-LiteSpineNet Hybrid Architecture.** Integration of this quantum block into a compact CNN yields statistically significant improvements in precision (+0.7%) and AUROC (+0.2%), as confirmed by paired t-tests ($p < 0.0001$).
- **Cross-Validation and Statistical Rigor.** Robust validation across threefold cross-validation and significance testing ensures results are not artifacts of random seeds or data splits.
- **Transparency and Reproducibility.** Compliance with CLAIM and TRIPOD+AI standards: fixed seeds, public release of code and pretrained weights, complete environment specification, and reporting of 99% confidence intervals.

6. Conclusions

This research introduced LiteSpineNet, a lightweight CNN, and HQ-LiteSpineNet, its hybrid quantum variant with a seven-qubit variational feature extraction block, for multiclass spine lesion classification. The conclusions derive directly from the statistical evidence obtained through cross-validation and hypothesis testing. Specifically, we found that:

- Paired two-tailed t-tests across three folds confirmed statistically significant differences between the models ($p < 0.0001$). LiteSpineNet consistently achieved higher accuracy and F1 scores, with a mean performance advantage of $\sim$15%.
- HQ-LiteSpineNet achieved comparable overall results but outperformed LiteSpineNet in minority classes such as “Surgical implant” and “Vertebral collapse”, improving recall by $+6.7\%$ and F1 score by $+4.4\%$, supported by 99% confidence intervals.
- Both models demonstrated stable behavior across threefold cross-validation, with low standard deviations, confirming reproducibility and reliability.
- External validation on BUU-LSPINE and RSNA 2024 underscored generalization capabilities across modalities, although spondylolisthesis detection remained challenging.
- Computational cost analysis showed that HQ-LiteSpineNet incurs exponential overhead on simulators ($O(2^n)$) but would scale linearly on real QPUs ($O(n)$), highlighting feasibility in future hardware implementations.

In conclusion, LiteSpineNet provides an efficient and robust baseline for multiclass spine lesion classification, while HQ-LiteSpineNet demonstrates the potential of quantum-enhanced models for

 doi:10.20944/preprints202603.2213.v1

improving sensitivity in underrepresented classes. Together, these findings highlight the relevance of our contributions to both clinical practice and the emerging field of quantum medical AI.

Future Work

Future work should aim to improve model stability and robustness, especially under highly imbalanced data and limited training samples. Approaches such as adaptive loss weighting and uncertainty-aware training could mitigate these challenges. On the quantum side, more expressive variational circuits with optimized qubit connectivity and alternative encodings may enhance feature extraction for complex medical patterns. Hybrid learning frameworks should also be optimized through improved classical–quantum design and efficient parameter tuning to better support diverse imaging modalities and lesion types. Finally, clinical validation remains essential: future studies should include prospective evaluations with real patient data, workflow integration in hospital settings, and telemedicine deployment to assess practical utility and impact.

Author Contributions: Conceptualization; methodology; software validation; formal analysis investigation; data curation; writing—original draft preparation; writing—review and editing; Daniel López Montiel. Conceptualization; methodology; software validation; formal analysis investigation; writing—review and editing; supervision; project administration; funding acquisition; Oscar Montiel. Methodology; formal analysis investigation; supervision; writing—review and editing; Oscar Castillo. Methodology, formal analysis, data curation, software, Writing—review and editing; Miguel Lopez-Montiel

Funding: Instituto Politécnico Nacional SIP20253768/SIP2026, Secretaría de Ciencia, Humanidades, Tecnología e Innovación CF-2023-I-108.

Data Availability Statement: The VinDr-SpineXR dataset used to conduct this research is publicly available. It can be accessed from its official website: `https://vindr.ai/datasets/spinexr` and from the repository `https://physionet.org/content/vindr-spinexr/1.0.0/`. For long-term accessibility, the dataset DOI is `10.1007/978-3-030-87240-3_28`. The code used to conduct this research is protected under the MIT license and is available in the following Github repository: `https://mmontielpz.github.io/LiteSpineNet/login.html`.

Conflicts of Interest: The authors declare that they have no competing interests or conflicts of interest related to this work.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
HQ	Hybrid Quantum
CAD	Computer-Aided Diagnosis
DL	Deep Learning
QML	Quantum Machine Learning
NISQ	Noisy Intermediate-Scale Quantum
CT	Computed Tomography
DCA	Decision Curve Analysis
ACC	Accuracy
R	Recall
P	Precision
F1	F1 Score
AUROC	Area Under the Receiving Operating Characteristic
ROI	Region of Interest
QPU	Quantum Processing Unit
FLOPs	Floating-point Operations Per Second

 doi:10.20944/preprints202603.2213.v1

References

1. Clark, P.; Caló, M.; Torres-Naranjo, J.F.; Cisneros-Dreinhofer, F.; Silveira-Torre, L.H.; Tapia-Hernández, M.; Medina-Chávez, J.H.; Gutiérrez-Robledo, L.M.; Reza-Albarrán, A.A.; Coronado-Zarco, R.; et al. Osteoporosis and Fragility Fractures in Mexico: A Call to Action. *Archives of Medical Research* **2024**, *55*, 103062. https://doi.org/10.1016/J.ARCMED.2024.103062.
2. Ebeling, P. Osteoporosis in men: Why change needs to happen. https://www.osteoporosis.foundation/educational-hub/material/thematic-reports, 2014. Accessed: 2023-02-20.
3. Kanis, J.A.; McCloskey, E.V.; Johansson, H.; Oden, A.; Melton, L.J.; Khaltaev, N. A reference standard for the description of osteoporosis. *Bone* **2008**, *42*, 467–475. https://doi.org/10.1016/J.BONE.2007.11.001.
4. Rosenberg, G.S.; Cina, A.; Schiró, G.R.; Giorgi, P.D.; Gueorguiev, B.; Alini, M.; Varga, P.; Galbusera, F.; Gallazzi, E. Artificial Intelligence Accurately Detects Traumatic Thoracolumbar Fractures on Sagittal Radiographs. *Medicina 2022, Vol. 58, Page 998* **2022**, *58*, 998. https://doi.org/10.3390/MEDICINA58080998.
5. Aziziyeh, R.; Amin, M.; Habib, M.; Perlaza, J.G.; Szafranski, K.; McTavish, R.K.; Disher, T.; Lüdke, A.; Cameron, C. The burden of osteoporosis in four Latin American countries: Brazil, Mexico, Colombia, and Argentina. *Journal of medical economics* **2019**, *22*, 638–644. https://doi.org/10.1080/13696998.2019.1590843.
6. Fuggle, N.R.; Beaudart, C.; Bruyère, O.; Abrahamsen, B.; Al-Daghri, N.; Burlet, N.; Chandran, M.; Rosa, M.M.; Cortet, B.; Demonceau, C.; et al. Evidence-Based Guideline for the management of osteoporosis in men. *Nature Reviews Rheumatology 2024 20:4* **2024**, *20*, 241–251. https://doi.org/10.1038/s41584-024-01094-9.
7. Kashfi, S.S.; Abdollahi, G.; Hassanzadeh, J.; Mokarami, H.; Jeihooni, A.K. The relationship between osteoporosis and depression. *Sci. Rep. 2022 12:1* **2022**, *12*, 1–7. https://doi.org/10.1038/s41598-022-15248-w.
8. Clark, P.; Carlos, F.; Martínez, J.L.V. Epidemiology, costs and burden of osteoporosis in Mexico. *Archives of Osteoporosis* **2010**, *5*, 9–17. https://doi.org/10.1007/S11657-010-0042-8/TABLES/6.
9. Wu, F.; Chen, X.; Jiang, R.; Li, L.; Qin, L.; Qi, W.; Hao, C.; Hao, C.; Tang, J. Risk factor analysis of adjacent vertebral compression fracture following the surgery of percutaneous kyphoplasty in postmenopausal women. *Scientific Reports* **2025**, p. 5772. https://doi.org/10.1038/s41598-025-85381-9.
10. Kanis, J.A.; Glüer, C.C. An update on the diagnosis and assessment of osteoporosis with densitometry. *Osteoporosis International* **2000**, *11*, 192–202. https://doi.org/10.1007/S001980050281/METRICS.
11. Deng, L.; Yao, Y.; Shang, A.L.; Du, T.; Zhang, J.; Yang, Q.; Li, J.; Wang, Q.; Li, X. Opportunistic screening for osteoporosis using hydroxyapatite measurements of the vertebral by thorax dual-energy spectral CT in postmenopausal females. *Sci. Rep. 2022 12:1* **2022**, *12*, 1–8. https://doi.org/10.1038/s41598-022-26237-4.
12. Murata, K.; Endo, K.; Aihara, T.; Suzuki, H.; Sawaji, Y.; Matsuoka, Y.; Nishimura, H.; Takamatsu, T.; Konishi, T.; Maekawa, A.; et al. Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. *Scientific Reports* **2020**, *10*. https://doi.org/10.1038/S41598-020-76866-W.
13. Chen, H.Y.; Hsu, B.W.Y.; Yin, Y.K.; Lin, F.H.; Yang, T.H.; Yang, R.S.; Lee, C.K.; Tseng, V.S. Application of deep learning algorithm to detect and visualize vertebral fractures on plain frontal radiographs. *PLoS ONE* **2021**, *16*. https://doi.org/10.1371/JOURNAL.PONE.0245992.
14. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. https://doi.org/10.1038/NATURE21056.
15. Schuld, M.; Killoran, N. Is Quantum Advantage the Right Goal for Quantum Machine Learning? *PRX Quantum* **2022**, *3*. https://doi.org/10.1103/PRXQUANTUM.3.030101.
16. Montiel Ross, O.H. A Review of Quantum-Inspired Metaheuristics: Going From Classical Computers to Real Quantum Computers. *IEEE Access* **2020**, *8*, 814–838. https://doi.org/10.1109/ACCESS.2019.2962155.
17. Liu, Y.; Arunachalam, S.; Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics* **2021**, *17*, 1013–1017. https://doi.org/10.1038/S41567-021-01287-Z.
18. Rasool, R.U.; Ahmad, H.F.; Rafique, W.; Qayyum, A.; Qadir, J.; Anwar, Z. Quantum Computing for Healthcare: A Review. *Future Internet* **2023**, *15*. https://doi.org/10.3390/FI15030094.
19. Houssein, E.H.; Abohashima, Z.; Elhoseny, M.; Mohamed, W.M. Hybrid quantum-classical convolutional neural network model for COVID-19 prediction using chest X-ray images. *Journal of Computational Design and Engineering* **2022**, *9*, 343–363. https://doi.org/10.1093/JCDE/QWAC003.
20. Wei, L.; Liu, H.; Xu, J.; Shi, L.; Shan, Z.; Zhao, B.; Gao, Y. Quantum machine learning in medical image analysis: A survey. *Neurocomputing* **2023**, *525*, 42–53. https://doi.org/10.1016/J.NEUCOM.2023.01.049.
21. Ahalya, R.K.; Nkondo, G.F.; Snekhalatha, U. Automated detection of parkinson's disease based on Hybrid CNN and quantum machine learning techniques in MRI images. *Biomedical Engineering - Applications, Basis and Communications* **2024**, *36*. https://doi.org/10.4015/S1016237224500054.

 doi:10.20944/preprints202603.2213.v1

22. Maheshwari, D.; Garcia-Zapirain, B.; Sierra-Sosa, D. Quantum Machine Learning Applications in the Biomedical Domain: A Systematic Review. *IEEE Access* **2022**, *10*, 80463–80484. https://doi.org/10.1109/ACCESS.2022.3195044.
23. Sun, G.; Pan, Y.; Kong, W.; Xu, Z.; Ma, J.; Racharak, T.; Nguyen, L.M.; Xin, J. DA-TransUNet: integrating spatial and channel dual attention with transformer U-net for medical image segmentation. *Frontiers in Bioengineering and Biotechnology* **2024**, *12*, 1398237. https://doi.org/10.3389/FBIOE.2024.1398237/BIBTEX.
24. Pan, Y.; Xin, J.; Yang, T.; Li, S.; Nguyen, L.M.; Racharak, T.; Li, K.; Sun, G. A mutual inclusion mechanism for precise boundary segmentation in medical images. *Frontiers in Bioengineering and Biotechnology* **2024**, *12*, 1504249. https://doi.org/10.3389/FBIOE.2024.1504249/BIBTEX.
25. Sun, G.; Shu, H.; Shao, F.; Racharak, T.; Kong, W.; Pan, Y.; Dong, J.; Wang, S.; Nguyen, L.M.; Xin, J. FKD-Med: Privacy-Aware, Communication-Optimized Medical Image Segmentation via Federated Learning and Model Lightweighting Through Knowledge Distillation. *IEEE Access* **2024**, *12*, 33687–33704. https://doi.org/10.1109/ACCESS.2024.3372394.
26. Yabu, A.; Hoshino, M.; Tabuchi, H.; Takahashi, S.; Masumoto, H.; Akada, M.; Morita, S.; Maeno, T.; Iwamae, M.; Inose, H.; et al. Using artificial intelligence to diagnose fresh osteoporotic vertebral fractures on magnetic resonance images. *Spine Journal* **2021**, *21*, 1652–1658. https://doi.org/10.1016/J.SPINEE.2021.03.006.
27. Dong, Q.; Luo, G.; Lane, N.E.; Lui, L.Y.; Marshall, L.M.; Johnston, S.K.; Dabbous, H.; O'Reilly, M.; Linnau, K.F.; Perry, J.; et al. Generalizability of Deep Learning Classification of Spinal Osteoporotic Compression Fractures on Radiographs Using an Adaptation of the Modified-2 Algorithm-Based Qualitative Criteria. *Academic Radiology* **2023**. https://doi.org/10.1016/J.ACRA.2023.04.023.
28. Karno, A.S.B.; Hastomo, W.; Surawan, T.; Lamandasa, S.R.; Usuli, S.; Kapuy, H.R.; Digdoyo, A. Classification of cervical spine fractures using 8 variants EfficientNet with transfer learning. *International Journal of Electrical and Computer Engineering* **2023**, *13*, 7065–7077. https://doi.org/10.11591/IJECE.V13I6.PP7065-7077.
29. Dastgir, A.; Bin, W.; Saeed, M.U.; Sheng, J.; Saleem, S. MAFMv3: An automated Multi-Scale Attention-Based Feature Fusion MobileNetv3 for spine lesion classification. *Image and Vision Computing* **2025**, *155*, 105440. https://doi.org/10.1016/J.IMAVIS.2025.105440.
30. Hong, N.; Cho, S.W.; Shin, S.; Lee, S.; Jang, S.A.; Roh, S.; Lee, Y.H.; Rhee, Y.; Cummings, S.R.; Kim, H.; et al. Deep-Learning-Based Detection of Vertebral Fracture and Osteoporosis Using Lateral Spine X-Ray Radiography. *Journal of Bone and Mineral Research* **2023**, *38*, 887–895. https://doi.org/10.1002/JBMR.4814.
31. Iyer, S.; Blair, A.; White, C.; Dawes, L.; Moses, D.; Sowmya, A. Vertebral compression fracture detection using imitation learning, patch based convolutional neural networks and majority voting. *Informatics in Medicine Unlocked* **2023**, *38*. https://doi.org/10.1016/J.IMU.2023.101238.
32. Chłąd, P.; Ogiela, M.R. Deep Learning and Cloud-Based Computation for Cervical Spine Fracture Detection System. *Electronics (Switzerland)* **2023**, *12*. https://doi.org/10.3390/ELECTRONICS12092056.
33. Kim, J.K.; Chang, M.C. Convolutional neural network algorithm trained on lumbar spine radiographs to predict outcomes of transforaminal epidural steroid injection for lumbosacral radicular pain from spinal stenosis. *Scientific Reports 2024 14:1* **2024**, *14*, 1–7. https://doi.org/10.1038/s41598-024-59288-w.
34. Nicolaes, J.; Skjødt, M.K.; Raeymaeckers, S.; Smith, C.D.; Abrahamsen, B.; Fuerst, T.; Debois, M.; Vandermeulen, D.; Libanati, C. Towards Improved Identification of Vertebral Fractures in Routine Computed Tomography (CT) Scans: Development and External Validation of a Machine Learning Algorithm. *Journal of Bone and Mineral Research* **2023**, *38*, 1856–1866. https://doi.org/10.1002/JBMR.4916,.
35. Polzer, C.; Yilmaz, E.; Meyer, C.; Jang, H.; Jansen, O.; Lorenz, C.; Bürger, C.; Glüer, C.C.; Sedaghat, S. AI-based automated detection and stability analysis of traumatic vertebral body fractures on computed tomography. *European Journal of Radiology* **2024**, *173*. https://doi.org/10.1016/J.EJRAD.2024.111364.
36. Liawrungrueang, W.; Han, I.; Cholamjiak, W.; Sarasombath, P.; Riew, K.D. Artificial Intelligence Detection of Cervical Spine Fractures Using Convolutional Neural Network Models. *Neurospine* **2024**, *21*, 833–841. https://doi.org/10.14245/NS.2448580.290.
37. Sutradhar, D.; Fahad, N.M.; Raiaan, M.A.K.; Jonkman, M.; Azam, S. Cervical spine fracture detection utilizing YOLOv8 and deep attention-based vertebrae classification ensuring XAI. *Biomedical Signal Processing and Control* **2025**, *101*. https://doi.org/10.1016/J.BSPC.2024.107228.
38. Amin, J.; Sharif, M.; Gul, N.; Kadry, S.; Chakraborty, C. Quantum Machine Learning Architecture for COVID-19 Classification Based on Synthetic Data Generation Using Conditional Adversarial Neural Network. *Cognitive Computation* **2021**. https://doi.org/10.1007/S12559-021-09926-6.
39. Azevedo, V.; Silva, C.; Dutra, I. Quantum transfer learning for breast cancer detection. *Quantum Machine Intelligence* **2022**, *4*, 1–14. https://doi.org/10.1007/S42484-022-00062-4/FIGURES/9.

 doi:10.20944/preprints202603.2213.v1

40. Gadu, S.R.; sekhar potala, C. A Hybrid Generative Adversarial network with Quantum U-NET for 3D spine X-ray image registration. *Healthcare Analytics* **2023**, *4*, 100251. https://doi.org/10.1016/J.HEALTH.2023.100251.
41. Mewada, H.; Pires, I.M.; Rahevar, M.; Khatri, N. Quantum Convolutional Neural Network for Bone Fracture Classification from X-Ray Images. *Procedia Computer Science* **2025**, *256*, 1143–1150. https://doi.org/10.1016/J.PROCS.2025.02.222.
42. Nguyen, H.T.; Pham, H.H.; Nguyen, N.T.; Nguyen, H.Q.; Huynh, T.Q.; Dao, M.; Vu, V. VinDr-SpineXR: A Deep Learning Framework for Spinal Lesions Detection and Classification from Radiographs. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2021**, *12905 LNCS*, 291–301. https://doi.org/10.1007/978-3-030-87240-3_28/TABLES/5.
43. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* **2002**, *16*, 321–357.
44. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. *Proceedings of the IEEE International Conference on Computer Vision* **2019**, *Vol. 2019-October*, 1314–1324. https://doi.org/10.1109/ICCV.2019.00140.
45. Bergholm, V.; Izaac, J.; Schuld, M.; Gogolin, C.; Ahmed, S.; Ajith, V.; Alam, M.S.; Alonso-Linaje, G.; AkashNarayanan, B.; Asadi, A.; et al. PennyLane: Automatic differentiation of hybrid quantum-classical computations, 2022, [arXiv:quant-ph/1811.04968].
46. Schuld, M.; Bocharov, A.; Svore, K.; Wiebe, N. Circuit-centric quantum classifiers. *Physical Review A* **2018**, *101*. https://doi.org/10.1103/PhysRevA.101.032308.
47. Collins, G.S.; Moons, K.G.M.; Dhiman, P.; Riley, R.D.; Beam, A.L.; Van Calster, B.; Ghassemi, M.; Liu, X.; Reitsma, J.B.; van Smeden, M.; et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, [https://www.bmj.com/content/385/bmj-2023-078378.full.pdf]. https://doi.org/10.1136/bmj-2023-078378.
48. Pineau, J.; Pu, P.; Snoek, J.; Wilson, A.G.; et al. Improving reproducibility in machine learning research: A report from the NeurIPS 2020 reproducibility program. *Journal of Machine Learning Research* **2021**, *22*, 1–20.
49. Bengio, Y. Practical recommendations for gradient-based training of deep architectures. *arXiv preprint arXiv:1206.5533* **2012**.
50. Keskar, N.S.; Mudigere, D.; Nocedal, J.; Smelyanskiy, M.; Tang, P.T.P. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836* **2017**.
51. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR)* **2017**, [1412.6980].
52. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* **2014**, *15*, 1929–1958.
53. Schuld, M.; Killoran, N. Quantum Machine Learning in Feature Hilbert Spaces. *Phys. Rev. Lett.* **2019**, *122*, 040504. https://doi.org/10.1103/PhysRevLett.122.040504.
54. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum Machine Learning. *Nature* **2017**, *549*, 195–202.
55. Ramos-Pollán, R.; Guevara-López, M.A.; Suárez-Ortega, C.; Díaz-Herrero, G.; Franco-Valiente, J.M.; Rubio-Del-Solar, M.; González-De-Posada, N.; Vaz, M.A.P.; Loureiro, J.; Ramos, I. Discovering mammography-based machine learning classifiers for breast cancer diagnosis. *Journal of Medical Systems* **2012**, *36*, 2259–2269. https://doi.org/10.1007/S10916-011-9693-2/FIGURES/4.
56. Richards, T.; Talbott, J.; Ball, R.; Colak, E.; Flanders, A.; Kitamura, F.; Mongan, J.; Prevedello, L.; Vazirabad., M. RSNA 2024 Lumbar Spine Degenerative Classification, 2024.
57. Klinwichit, P.; Yookwan, W.; Limchareon, S.; Chinnasarn, K.; Jang, J.S.; Onuean, A. BUU-LSPINE: A Thai Open Lumbar Spine Dataset for Spondylolisthesis Detection. *Applied Sciences* **2023**, *13*. https://doi.org/10.3390/app13158646.