Deep Learning-based Surrogate Model of Subject-Specific Finite-Element Analysis for Vertebrae

Yuanrui Cai, Enrico Dall'Ara, Damien Lacroix, and Lingzhong Guo, *Member, IEEE*

***Abstract*— Subject-specific finite-element analysis (FEA) models enable accurate simulation of vertebral biomechanics but are often time-consuming to construct and solve under varying conditions. This study presents a novel deep learning (DL)/machine learning (ML)-based surrogate model that predicts stress distributions in vertebral bodies with high efficiency. The model integrates vertebral shape encoding and employs separate decoding branches for surface and internal nodes. It was trained on 3,960 synthetic L1 vertebrae generated via data augmentation from 42 real computed tomography (CT) scans. Evaluation on independent test samples yielded a mean absolute error (MAE) of 0.0596 MPa and an R^2 of 0.864 for von Mises stress. Visualization results confirm strong agreement between predicted and FEA-computed stress patterns, with localized discrepancies observed at the anteroinferior margin and pedicles. Moreover, an end-to-end automated pipeline was established based on the developed model, reducing the total processing time from 90–120 min to approximately 134–154 s per subject. These findings highlight the potential of the proposed surrogate model to facilitate rapid, subject-specific biomechanical assessments in clinical workflows.**

***Index Terms*— Biomechanics, Deep learning, Finite element analysis, Shape encoding, Stress prediction, Surrogate modeling, Vertebra**

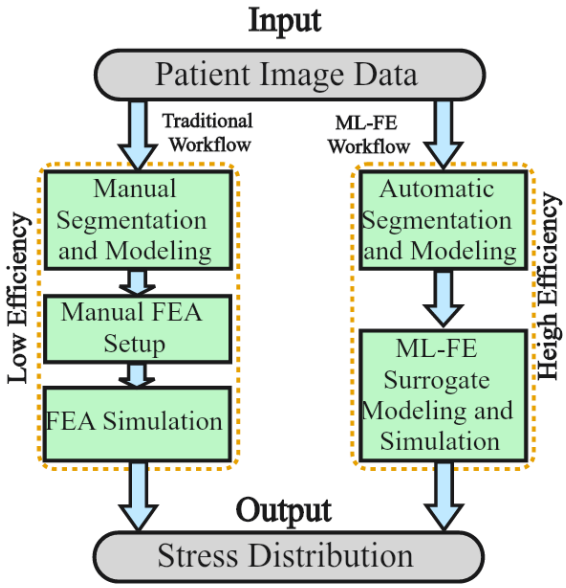

Fig. 1. Comparison between traditional workflow and current workflow with machine learning that we proposed for patient-specific computational modeling.

I. Introduction

STRESS analysis was central to investigations conducted in the discipline of mechanics, within which finite element analysis (FEA) has been extensively used to study the biomechanics of human vertebrae [1], [2]. Accurate prediction of spinal stress distributions was crucial to understand various pathological conditions, such as low back pain and metastatic tumors [3]–[5].

Fig. 1 (left) illustrates the conventional workflow of FEA. After acquiring subject-specific medical images, manual spine segmentation was performed to construct a subject-specific finite element model (FEM). The material properties were assigned, and the relevant boundary conditions and loading conditions were defined. Subsequently, the FEM was submitted to a finite element solver for numerical stress analysis. This traditional FEA workflow is both complex and time-consuming, making it impractical for clinical scenarios requiring rapid diagnosis and treatment planning [6].

Although several studies have explored the combination of FEM and DL/ML to improve the efficiency of FE simulations in general [7], similar studies specifically focusing on vertebra or spine FEA are still very limited. Ahmadi et al. [8] integrated FEA with a physics-informed neural network (PINN) to inversely estimate material properties of the lumbar spine by solving a physics-constrained optimization problem, achieving an accuracy of 94.30%. In 2024 [9], a reduced order model (ROM) was proposed by using the Cartesian grid FEM (cgFEM) in combination with the Sparse Subspace Learning (SSL) technique. The method has been used for the assessment of the risk of failure of metastatic vertebrae. Kok et al. [10] presented a DL-based approach to automatic generation of subject-specific finite element models of the spine from magnetic resonance images (MRI). Furthermore,

Yuanrui Cai and Lingzhong Guo are with the School of Electrical and Electronic Engineering, The University of Sheffield, Sheffield S10 2TN, U.K. (e-mail: {ycai50,l.guo}@sheffield.ac.uk).

Enrico Dall'Ara is with the School of Division of Clinical Medicine, The University of Sheffield, Sheffield S10 2TN, U.K. (e-mail: {e.dallara}@sheffield.ac.uk).

Damien Lacroix is with the School of Mechanical, Aerospace and Civil Engineering, The University of Sheffield, Sheffield S10 2TN, U.K. (e-mail: {d.lacroix}@sheffield.ac.uk).

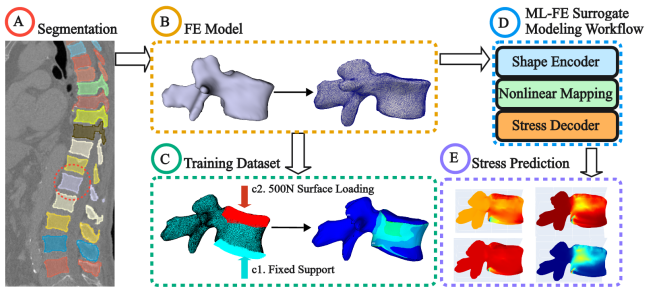

Fig. 2. ML-FE surrogate modeling pipeline for biomechanics prediction from clinical CT: (A) Anatomical structure segmentation from medical imaging; (B) Mesh morphing by Deformetrica and Tetramesh by Hypermesh; (C) Biomechanics simulation by PyAnsys (Training dataset generation); (D) ML-FE surrogate modeling; (E) Biomechanics prediction outcomes.

machine learning-based surrogates of quantitative computed tomography finite element analysis (QCT-FEA) have been developed for predicting fracture risk and estimating strains in various anatomical regions [11], [12]. These methodologies are predicated on manually designed anatomical and clinical features, executing pointwise predictions by treating each node autonomously. While these pointwise strategies manifest efficacy within their specific domains, they may not adequately encapsulate the complex biomechanical fields necessary for more intricate anatomical structures. Despite these advancements, the application of advanced AI techniques, including DL and ML in the study of biomechanical behavior of vertebrae remains underexplored. Therefore, we propose a DL-based surrogate model that replaces traditional FEA, enabling the automatic generation of subject-specific FEM in approximately 3 seconds and rapidly providing stress analysis results for time-sensitive clinical scenarios. The workflow consists mainly of two components: first, an in-house DL-based automatic segmentation model processes the patient's original clinical CT images to directly obtain the anatomical geometry of the patient's vertebrae; second, a DL-based mechanical analysis model performs FEA directly on the anatomical model, quickly releasing the corresponding stress analysis results. This workflow circumvents the need for manual tasks and conventional finite element modeling, thus substantially reducing processing time without compromising accuracy. Moreover, we integrate the DL-based segmentation network with the ML-FE surrogate model into an end-to-end pipeline that directly produces stress distribution results from the patient's clinical CT images (Fig. 1 (right)). This end-to-end automation workflow demonstrates the feasibility of accelerating biomechanical stress analysis, substantially reducing the time and manual effort required for subject-specific modeling, and may benefit clinical applications in the future.

II. Materials and Methods

A. Dataset

This study used publicly available spinal CT images from the Verse2020 dataset [13]–[15], focusing on the first lumbar vertebra (L1) of 42 healthy subjects to ensure biomechanical analysis under normal physiological conditions. Three-dimensional L1 models were reconstructed for all subjects, with 33 cases used for training and 9 for independent testing.

For ML-FE model training, geometric models were obtained by manual segmentation in 3D Slicer (version 5.6.1) (Fig. 2. A) [16]. In the end-to-end pipeline, vertebral models were generated by automatic segmentation from clinical CT images. To ensure consistent mesh topology, mesh morphing with Deformetrica (Version 4.3.0) [17], [18] was applied to deform a reference FEM to each subject (Fig. 2. B), and 120 intermediate meshes were generated per sample for augmentation. All meshes were solved using PyAnsys (Fig. 2. C) [19] with linear tetrahedral elements (1 mm max size) and homogeneous isotropic bone properties (Young's modulus: 500 MPa [20]; Poisson's ratio: 0.3 [3]). The inferior surface was fully constrained, and a 500 N compressive load applied to the superior surface.

B. Automatic Segmentation: BiFusion-UNet

To ensure accurate and robust vertebral segmentation from clinical CT images, we employed a dual-encoder neural network, termed BiFusion-UNet. which serves as the first step in our end-to-end workflow. The structure consists of two complementary encoders: a VGG16-based encoder [21], which preserves low-level texture information, and a ResNet-based encoder [22] with embedded Squeeze-and-Excitation (SE) [23] modules that emphasize informative channels through learned attention.

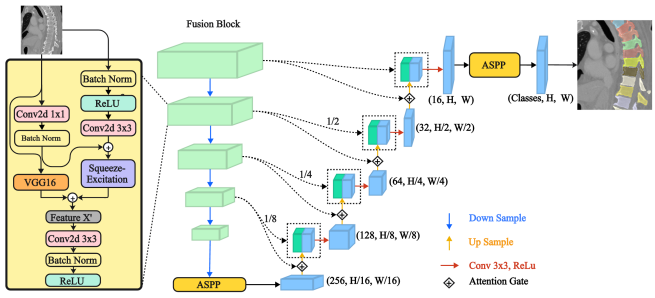

Fig. 3. BiFusion-UNet: Combination of dual-encoders (VGG16 and ResNet-SE) with multi-scale fusion, attention-guided decoding, and an Atrous Spatial Pyramid Pooling (ASPP) module [24]. The ASPP module uses parallel dilated convolutions at multiple dilation rates with multiple dilation rates (1, 3, 6, 9) to effectively capture multi-scale context. The attention gate where decoder features g guide the modulation of encoder features x via $\psi(g_{\text{pool}}, x)$.

At each resolution level, features from both encoders are adaptively fused through learnable fusion blocks, enabling the network to automatically balance complementary information and improve generalization across diverse anatomical variations.

In order to remove noise, fill internal voids, and enhance anatomical consistency, post-processing was applied for each CT image segmentation result to refine the predicted mask, including morphological denoising, hole filling, and largest component extraction.

C. Subject-specific FEM Generation

After segmentation, the L1 vertebrae of 33 healthy subjects were spatially normalized and converted to point clouds with consistent topology. A mean shape was computed and meshed in ANSYS^{TM} (version 2024 R2) to create the reference FEM. Subject-specific FEMs were generated by applying mesh morphing with Deformetrica to deform the reference mesh to each subject, ensuring consistent node correspondence. Since Deformetrica only provides surface displacements, a HyperMesh (`.tcl`) script was used to regenerate internal tetrahedral elements, maintaining volumetric consistency. For data augmentation, Deformetrica also produced 120 intermediate shapes per subject along the geodesic deformation path. These intermediate meshes introduce gradual morphological variations between the reference and subject geometries, thereby augmenting the dataset with anatomical diversity while preserving mesh topology consistency. All augmented shapes were further masked and included in the training set (Fig. 4).

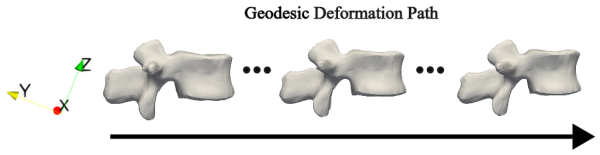

Fig. 4. Intermediate shapes (middle) were generated along the geodesic deformation path between the reference mesh (left) and subject mesh (right), providing anatomically diverse but topologically consistent training samples.

D. ML-FE Surrogate Modeling

In this study, a dataset of 3,960 L1 vertebrae samples was constructed from data augmentation on 33 subjects, and used to train the ML-FE surrogate model. In addition, 9 independent L1 vertebrae samples were used to evaluate the performance of the model.

1) Model Structure and Data Representation: Fig. 6 illustrates the proposed ML-FE surrogate modeling pipeline. Each vertebral shape was represented by two point clouds: surface nodes (with consistent number and order across all samples via prior mesh morphing) and internal nodes (generated by tetrahedral meshing in HyperMesh (version 2022) [25] based on the morphed surfaces, allowing anatomical variability).

The shape encoder extracts a global latent vector from the surface nodes, which serves as a shared feature for both the surface and internal branches. This latent representation was projected by two separate nonlinear mapping modules into stress latent codes for the surface and internal regions, respectively. Each branch uses its own decoder to reconstruct nodal stress distributions: the surface decoder operates on the stress latent, whereas the internal decoder additionally incorporates the resampled internal node coordinates using farthest point sampling (FPS) [26] for spatial conditioning. The predicted stress values from both branches were concatenated to form the complete stress field of the vertebra.

To address potential discontinuities at the interface, a surface-constrained Laplacian smoothing was applied as post-processing. Surface stress values were held fixed, whereas internal stresses were iteratively refined over 50 iterations using Gaussian-weighted averaging of their $k = 10$ nearest neighbors, with the Gaussian scale parameter σ set to the mean neighborhood distance.

2) Data Pre-processing and Augmentation: All point cloud data were normalized upon extraction to ensure consistent scale and centering. Since the number of internal nodes varied across FEMs, farthest point sampling (FPS) was applied to obtain a fixed set of 23,528 points, matching the surface resolution and ensuring a consistent input size across all subjects. To further alleviate the limited number of patient scans and enrich the dataset, five independent FPS operations with different random seeds were performed for each sample. While each subset contained the same number of points, the specific spatial distributions differed across seeds, as demonstrated through principal component analysis (PCA) [27] analysis (Fig.5), thereby introducing structural diversity while maintaining geometric consistency. As a result, each original internal sample yielded five augmented versions, leading to a total of 19,800 training samples. The corresponding coordinates and stress values were simultaneously extracted from the FEA result files (.rst output files). Importantly, all pre-processing procedures were applied on geometric information (xyz coordinates), without involving stress or strain values. Therefore, these steps were entirely independent of the outputs.

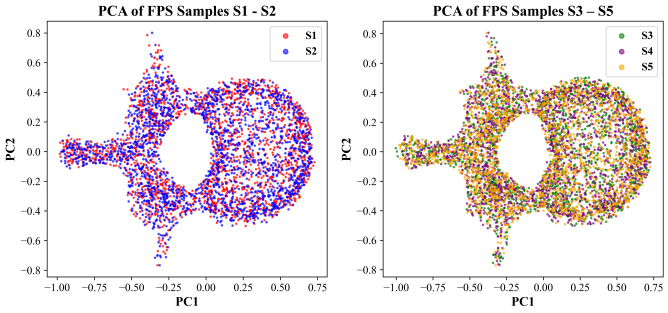

Fig. 5. PCA visualization of FPS augmentation. Different random seeds (S1–S5) yield consistent global distributions but exhibit local variations in point arrangement, illustrating structural diversity while preserving geometric consistency.

However, during the testing stage, only a single instance derived from FPS was used per original sample for prediction, specifically, the variant generated using the first random seed. This choice ensured consistency and avoided redundancy, allowing for an unbiased assessment of model performance without over-representing structurally similar samples.

3) Shape Encoding Stage: The shape encoder was a PointNet-based [28] autoencoder network (Fig. 7). Let the input surface point cloud be denoted as $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N} \subset \mathbb{R}^3$, where each point $\{\mathbf{x}_i\} \in \mathbb{R}^3$ represents the 3D coordinates of a surface node. The shape encoder was designed to extract a global latent vector $\mathbf{z} \in \mathbb{R}^{512}$ that compactly represents the overall geometry of the vertebrae.

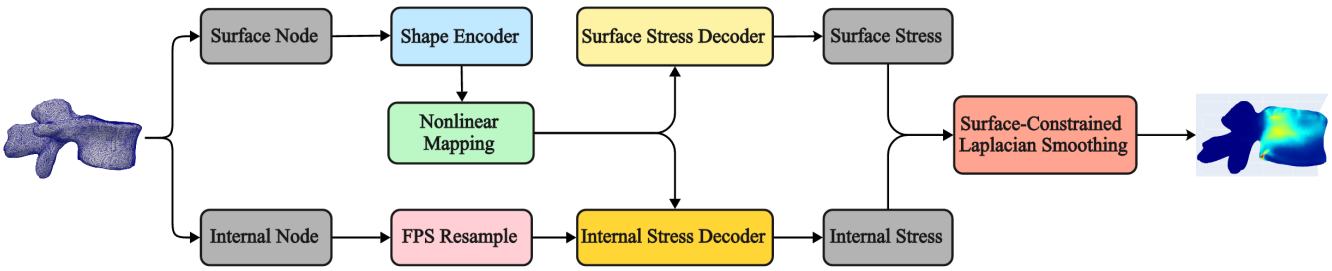

Fig. 6. ML-FE surrogate modeling overall pipeline.

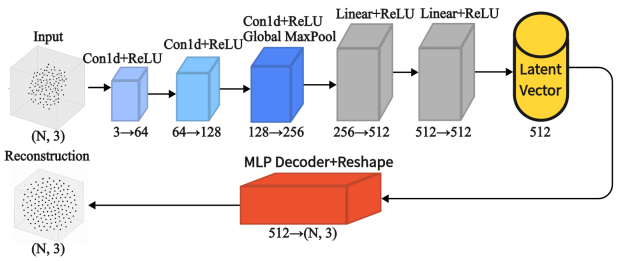

Fig. 7. PointNet-based shape encoder that extracts a 512 dimensional latent vector from surface point clouds.

The encoding process begins with a sequence of points-wise feature extraction layers $\mathbf{h} : \mathbb{R}^{N\times 3} \rightarrow \mathbb{R}^{N\times 256}$, implemented as 1D convolutions with ReLU activation. These features are aggregated by the global max pooling operator $\mathcal{P} : \mathbb{R}^{N\times 256} \rightarrow \mathbb{R}^{256}$. The pooled feature was subsequently transformed by multilayer perceptron (MLP) $\gamma_{enc} : \mathbb{R}^{256} \rightarrow \mathbb{R}^{512}$ to produce the final latent shape: $\mathbf{z} = \gamma_{enc}(\mathcal{P}(\mathbf{h}(\mathbf{X})))$. The latent vector serves as a compact shape representation and was shared across both surface and internal stress prediction branches. Since all input point clouds have the same node ordering across samples, to assess the information capacity of the shape latent vector, the reconstruction error of the autoencoder was additionally evaluated. Specifically, a decoder network with an MLP layer was built that $\gamma_{dec} : \mathbb{R}^{512} \rightarrow \mathbb{R}^{N\times 3}$ was used to reconstruct the original point cloud $\mathbf{X}$ from its latent representation $\mathbf{z}$.

The reconstruction loss which equivalent to point-wise MSE was defined as

$$\mathcal{L}_{shape} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2 \tag{1}$$

where $\hat{\mathbf{x}} = \gamma_{dec}(\mathbf{z}) \in \mathbb{R}^{N\times 3}$ denotes the reconstructed point cloud. Because the point indices were consistent across all samples, this point-wise MSE directly reflects how well the latent code retains geometric fidelity.

4) Shape to Stress Latent Mapping: To predict stress-specific latent representations from the encoded vertebral shape, two separate Random Forest (RF) regressors [29] were employed (Fig. 8). One for surface and the other one for internal stress latent mapping. Each RF takes the global shape latent vector $\mathbf{z} \in \mathbb{R}^{512}$ as input and output a corresponding stress latent vector $\tilde{\mathbf{z}}_{surf} \in \mathbb{R}^{512}$ or $\tilde{\mathbf{z}}_{internal} \in \mathbb{R}^{512}$. The target stress latent vectors used to supervise RF training were generated from the stress autoencoder model, which will be described in detail in the next section. To further enhance

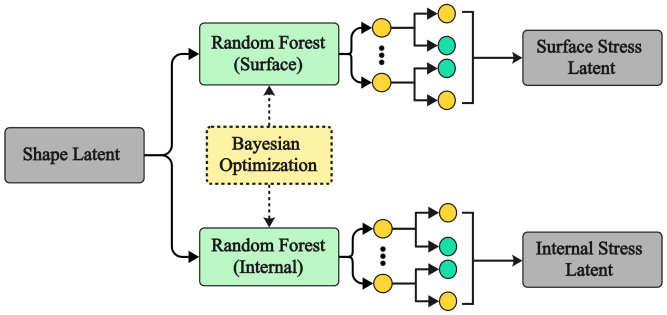

Fig. 8. Bayesian optimizer-Random Forest model with the input shape latent $\mathbf{z}$ and mapping to $\tilde{\mathbf{z}}_{surf}$ and $\tilde{\mathbf{z}}_{internal}$.

the performance of the RF model, the hyperparameters in RF could be optimized using Bayesian optimization [30], which minimizes the cross-validated mean squared error (MSE) across the entire training set:

$$\theta^* = \arg\min_{\theta} \frac{1}{K}\sum_{k=1}^{K} \mathcal{L}_{\text{MSE}}^{(k)}(\theta) \tag{2}$$

where θ denotes the hyperparameter vector, and $\mathcal{L}_{\text{MSE}}^{(k)}$ is the MSE for the k-fold cross-validation. In this study, $k = 4$ was selected to balance the reliability of the estimation and computational cost. This strategy improves generalization whereas avoiding the need for manual hyperparameter tuning. The outputs of the mapping functions, $\mathcal{F}_{surf}(\mathbf{z})$ and $\mathcal{F}_{Internal}(\mathbf{z})$, serve as inputs to the corresponding stress decoder to reconstruct the stress distributions.

5) Stress Autoencoder: To reconstruct both surface and internal stress distributions, two separate autoencoder (AE) architectures were employed. The encoder was used to generate latent vectors as a training supervisor for the mapping module. The decoder was reused during inference to reconstruct stress from predicted stress latent codes. The input for surface stress AE was 1D stress sequences. In contrast, the internal stress AE takes the stress sequences as input and uses the corresponding 3D point cloud information as a conditional feature.

The surface stress AE (Fig. 9) was designed to learn a compact latent representation of the high-dimensional stress field

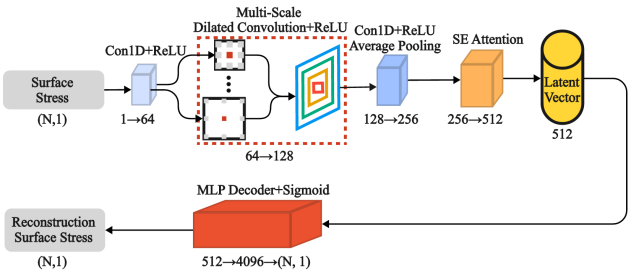

Fig. 9. Surface stress autoencoder structure: Reconstructing nodal surface stress, incorporating multi-scale dilated convolutions and SE attention.

defined over the surface nodes. Given an input stress vector $\mathbf{s} \in \mathbb{R}^N$ where N is the number of nodes. The encoder: $\phi_{surf} : \mathbb{R}^{N\times 1} \rightarrow \mathbb{R}^{512}$ transforms the normalized stress sequence into a latent vector $\tilde{\mathbf{z}}_{surf} = \phi(s)$. The encoder consists of a 1D convolution layer followed by a multi-scale dilated convolution block with dilation rates $\{1, 2, 4, 8\}$, enabling multi-resolution feature extraction. The fused features were further processed by another 1D convolution and adaptive average pooling. To enhance the quality of the representation, a SE attention module was applied. Whereas the dilated convolutions capture spatial patterns in different receptive fields, the SE module adaptively recalibrates channel-wise responses, emphasizing more informative features and suppressing irrelevant ones. This combination improves both the diversity and selectivity of the learned representation. The MLP layers-based decoder network $\gamma_{surf} : \mathbb{R}^{512} \rightarrow \mathbb{R}^{N\times 1}$ was built as a two-layer MLP with ReLU and sigmoid activations, reconstructing the stress sequence as $\hat{\mathbf{s}} = \gamma_{surf}(\tilde{\mathbf{z}}_{surf})$.

In contrast to the surface AE, the internal stress AE (Fig. 10) operates on 3D point clouds unstructured and the corresponding internal stress values. The input consisted of 3D coordinates of nodes $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{R}^3$ and the associated normalized stress vector $\mathbf{s} \in \mathbb{R}^N$, which were concatenated to form a 4D input feature $[\mathbf{x}_i, \mathbf{s}_i] \in \mathbb{R}^4$ and were processed by the encoder $\phi_{internal} : \mathbb{R}^{N\times 4} \rightarrow \mathbb{R}^{512}$.

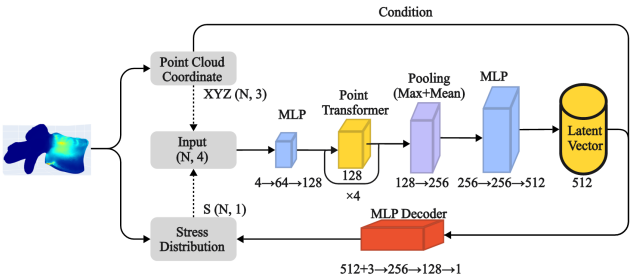

Fig. 10. Internal stress autoencoder structure.

To extract the latent representation, the concatenated input was first passed through a MLP layer, followed by a sequence of four points transformer blocks (Fig. 11).

In each block, a residual cross-head fusion block was employed after multi-head attention to explicitly enhance interactions across different attention heads. The encoded features were aggregated using global max and mean pooling, and

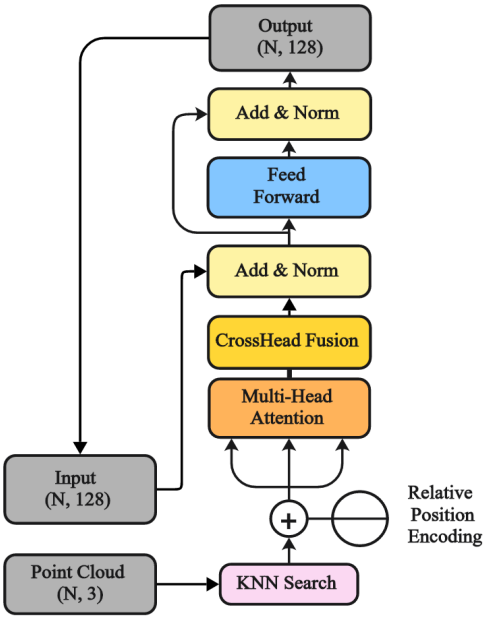

Fig. 11. PointNet transformer block with cross multi-head fusion for enhancing interactions across each heads.

subsequently mapped to a latent $\tilde{\mathbf{z}}_{internal} \in \mathbb{R}^{512}$ by a fully connected projection. The MLP layers-based decoder network $\gamma_{internal} : \mathbb{R}^{512+3} \rightarrow \mathbb{R}^N$ was employed to reconstruct the stress values by conditioning on both latent code and the point coordinates: $\mathbf{s}_i = \gamma_{internal}([\mathbf{x}_i, \tilde{\mathbf{z}}_{internal}])$.

For both the surface and internal stress AE, all stress values were normalized to the range [0, 1] using the global minimum and maximum values computed across the entire training set. The reconstruction was supervised using the MSE loss between the predicted and ground truth stress values, defined as:

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{s}_i - \hat{\mathbf{s}}_i)^2 \tag{3}$$

Beyond serving as a training objective, the reconstruction loss also reflects the extent to which the latent vector preserves essential stress information.

E. Post Processing

Since the stress values for surface and internal nodes were independently decoded and then merged, this may introduce local discontinuities or inconsistencies at the boundary between the two regions. To further improve the spatial continuity of the predicted stress field, a surface-constrained Laplacian smoothing was applied as a post processing step. Specifically, let the coordinates of all nodes be $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ and the stress values $\mathbf{s} = \{\mathbf{s}_i\}_{i=1}^N$. For each node i, its neighborhood was defined by the set of its nearest neighbors k, denoted as $\mathcal{N}(i)$. The local weighting coefficients were constructed using a Gaussian kernel:

$$w_{ij} = \exp(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}) \tag{4}$$

where σ is the Gaussian bandwidth, chosen as the average distance among the k-nearest neighbors. During each iteration, the stress value of the surface nodes was kept fixed, whereas the internal nodes i were updated as follows:

$$\mathbf{s}_i^{(t+1)} = \begin{cases} \mathbf{s}_i^{(t)}, & \text{if } i \text{ is a surface node} \\ \dfrac{\sum\limits_{j\in\mathcal{N}(i)} w_{ij}\mathbf{s}_j^{(t)}}{\sum\limits_{j\in\mathcal{N}(i)} w_{ij}}, & \text{if } i \text{ is an internal node} \end{cases} \tag{5}$$

after multiple iterations, this process effectively improved the global smoothness of the stress field and reduces discontinuities between the surface and internal regions.

III. Results

A. BiFusion-UNet Segmentation Result

To ensure anatomical accuracy of the input geometry for the ML-FE surrogate modeling, the segmentation performance of the proposed BiFusion-UNet should be evaluated. The training and validation loss and F1 curves are shown in Fig. 12, indicating stable convergence and superior performance of BiFusion-UNet compared with the baseline models.

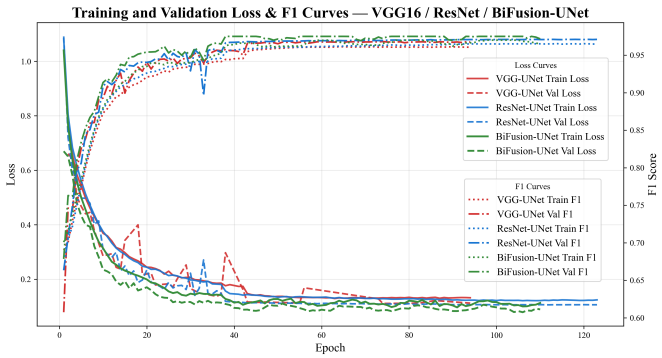

Fig. 12. Training and validation loss and F1 curves of VGG16-UNet, ResNet-UNet, and BiFusion-UNet. The results show that BiFusion-UNet achieves higher stability during training compared with the baseline models.

The proposed dual encoder fusion mechanism in BiFusion-UNet dynamically learns the contribution of VGG and ResNet features across five resolution levels (f1–f5). Specifically, the VGG encoder contributes more in the shallowest (f1: 64.4%), intermediate (f2: 60.2%, f3: 57.7%), and deepest layers (f5: 69.7%), whereas ResNet becomes dominant at the mid-resolution layer (f4: 62.7%). This adaptive fusion enables the model to generalize across anatomical variability by automatically adjusting feature emphasis during training.

TABLE I
QUANTITATIVE EVALUATION OF DIFFERENT METHODS.

Methods	DSC	FND	FPD	Recall
BiFusion-UNet	0.866	0.080	0.182	0.920
VGG16-UNet	0.859	0.078	0.194	0.910
ResNet-UNet	0.817	0.147	0.216	0.853

DSC: Dice similarity coefficient; FND: False negative dice; FPD: False positive dice; Recall: Sensitivity

The segmentation performance was quantitatively evaluated using the Dice Similarity Coefficient (DSC), False Negative Dice (FND), False positive Dice (FPD) and sensitivity (recall). These metrics are derived from the confusion matrix, where True Positives (TP) denote correctly segmented voxels belonging to the target class, False Negatives (FN) correspond to target voxels missed by the model, and False Positives (FP) indicate voxels incorrectly labeled as the target. The metrics are formally defined by the following equations:

$$DSC = \frac{2TP}{2TP + FN + FP} \tag{6}$$

$$FND = \frac{2FN}{2TP + FN + FP} \tag{7}$$

$$FPD = \frac{2FP}{2TP + FN + FP} \tag{8}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

DSC measures the overall overlap between prediction and ground truth, while FND and FPD quantify under-segmentation and over-segmentation, respectively, and sensitivity evaluates the ability to capture ground-truth positives.

As summarized in Table I for the L1 vertebra (class 20). BiFusion-UNet outperformed the other methods with the highest DSC (0.866) and sensitivity (0.920), together with the lowest FPD (0.182). VGG16-UNet achieved comparable DSC and sensitivity but presented a higher FPD (0.194), indicating a greater tendency toward over-segmentation despite its reduced FND. In contrast, ResNet-UNet exhibited the lowest DSC and sensitivity with both higher FND and FPD, reflecting inferior segmentation performance. These findings demonstrate that the dual-encoder design provides a more balanced and robust segmentation outcome (Fig. 13).

B. ML-FE Surrogate Model Result

To comprehensively evaluate the predictive performance of the proposed ML-FE surrogate model, the convergence of all sub-networks was examined based on their training and validation losses. The shape encoder converged rapidly and stably within approximately 30 epochs (train: 0.99×10^{-3}; val: 9.01×10^{-4}). The surface and internal stress encoders for the principal and von Mises stresses exhibited smooth loss decay and reached convergence within 100-150 epochs. The final losses were 2.01×10^{-4} (train) and 1.79×10^{-4} (val) for the surface encoder, and 1.61×10^{-2} (train) and 1.02×10^{-2} (val) for the internal encoder, indicating stable optimization without overfitting and good generalization across all components.

Furthermore, three metrics were used: the coefficient of determination (R^2), mean absolute error (MAE) and normalized mean absolute error (NMAE). The definitions were as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\mathbf{s}_i - \hat{\mathbf{s}}_i)^2}{\sum_{i=1}^{N}(\mathbf{s}_i - \bar{\mathbf{s}})^2} \tag{10}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\mathbf{s}_i - \hat{\mathbf{s}}_i| \tag{11}$$

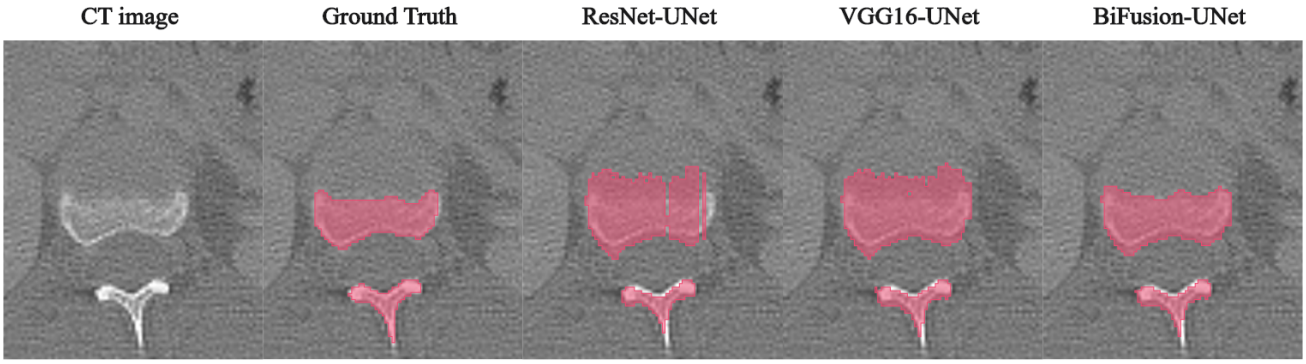

Fig. 13. Horizontal visualization of L1 lumbar vertebrae segmentation results

$$\mathrm{NMAE} = \frac{\mathrm{MAE}}{\max(|\mathbf{s}_i|)} \times 100\% \tag{12}$$

where $\mathbf{s}_i$ and $\hat{\mathbf{s}}_i$ denote the ground truth and the predicted stress value at node i, respectively. $\bar{\mathbf{s}}$ was the mean stress values of the ground truth, and N represents the total number of nodes (including both surface and internal nodes) within a given 9 test samples. All metrics were computed on each sample basis by aggregating stress values from all nodal locations in each sample. The resulting distribution of these metrics across the entire test set was summarized using boxplots for comparative analysis. R^2 reflects the overall goodness of fit, MAE quantifies the average prediction error, and NMAE enables direct comparison between samples with different stress scales. Visualization results are included to illustrate the complexity of prediction errors.

TABLE II
PERFORMANCE METRICS (MEAN ± STD) FOR THE PREDICTION.

Metric	s_1	s_2	s_3	s_{vm}
MAE (MPa)	0.0304 ± 0.0154	0.0196 ± 0.0107	0.0775 ± 0.0405	0.0663 ± 0.0384
NMAE (%)	2.62 ± 2.05	1.66 ± 1.51	2.42 ± 1.72	2.67 ± 1.95
R^2	0.6855 ± 0.2305	0.7405 ± 0.2642	0.8097 ± 0.1523	0.8635 ± 0.1376

Prediction metrics were reported for the first (s_1), second (s_2), third (s_3) principal stresses, and von Mises stress (s_{vm}).

The predictive performance of the ML-FE surrogate model for s_1, s_2, and s_3 as well as for s_{vm} was summarized in Table II and Fig. 16 -17.

The results indicate that for all stress types in whole vertebrae, the majority of $\mathbf{R}^2$ values exceeded 0.7, confirming the overall reliability of the surrogate model, as shown in Fig. 15, the post-processing procedure enhances spatial continuity.

Given the superior quantitative performance observed for $\mathbf{s}_{vm}$, more analysis was performed using spatial visualizations (Fig. 18) to assess the predictive fidelity of the model at the voxel level. Specifically, three representative samples (Samples A, B) from the test set were selected for detailed visualization including relative error computed as $\hat{s} - s$.

The results indicated that the surrogate model accurately reconstructed the von Mises stress distribution across diverse vertebral anatomies. Higher errors were mainly observed at the anteroinferior margin and pedicles, where the model tended to underestimate reference stresses. As shown in Fig. 17, the boundary and loading areas exhibited larger MAE and NMAE values, which can be attributed to pronounced stress gradients and complex contact conditions that increase the sensitivity of local predictions to geometric and morphological variations. In contrast, the central and posterior vertebral body, as well as the lamina, consistently exhibited lower prediction errors, reflecting the robust performance of the model in anatomically less variable regions.

TABLE III
MAX STRESS VALUE AND LOCATION COMPARISON

Subject	GT Max (MPa)	Pred Max (MPa)	Positional deviation (mm)	GT Index	Pred Index
A	2.5523	2.1331	0.0000	23410	23410
B	2.3477	1.7246	17.7170	10559	23410
C	1.9988	1.7469	0.0344	16377	16387
D	2.7295	2.4886	0.7357	21333	21310
E	3.4839	2.7550	0.0000	23410	23410
F	3.0709	2.5747	0.0587	23409	23410
G	2.3222	2.3749	0.0000	16377	16377
H	1.7564	2.7134	13.0736	10473	23408
I	1.7804	2.6673	14.0510	10473	23218

GT Max(MPa): The peak stress from FEA; Pred Max(MPa): The peak stress predict from ML-FE surrogate model; Positional deviation (mm): Spatial distance between GT and prediction stress; GT Index: Node index of peak stress from FEA; Pred Index: Node index of peak stress from ML-FE surrogate model.

Furthermore, the predicted peak stress magnitudes are highly consistent with the FEA results and the nodes of peak stress also exhibit strong spatial correspondence, with only small deviations observed in most samples. For vertebrae with more pronounced morphological variations (e.g., Subjects B, H and I), larger discrepancies were noted (Table III). However, the results still demonstrate the potential of the surrogate model to capture critical stress responses even in geometrically complex cases.

To further elucidate these findings, a comparison between the predicted and FEA-derived stress distributions for each sample was presented in Fig. 14. Sample A, E, G represent a case whose morphology closely corresponds to the mean shape within the training dataset and the predicted stress field is highly consistent with the ground truth in both magnitude and spatial pattern. Whereas samples B, H, I exhibit a more pronounced geometric deviation and two high-stress regions were observed on opposite sides of the inferior endplate in the FEA results, while the ML-FE model predicted only one of them. This led to a large index difference but reflects that

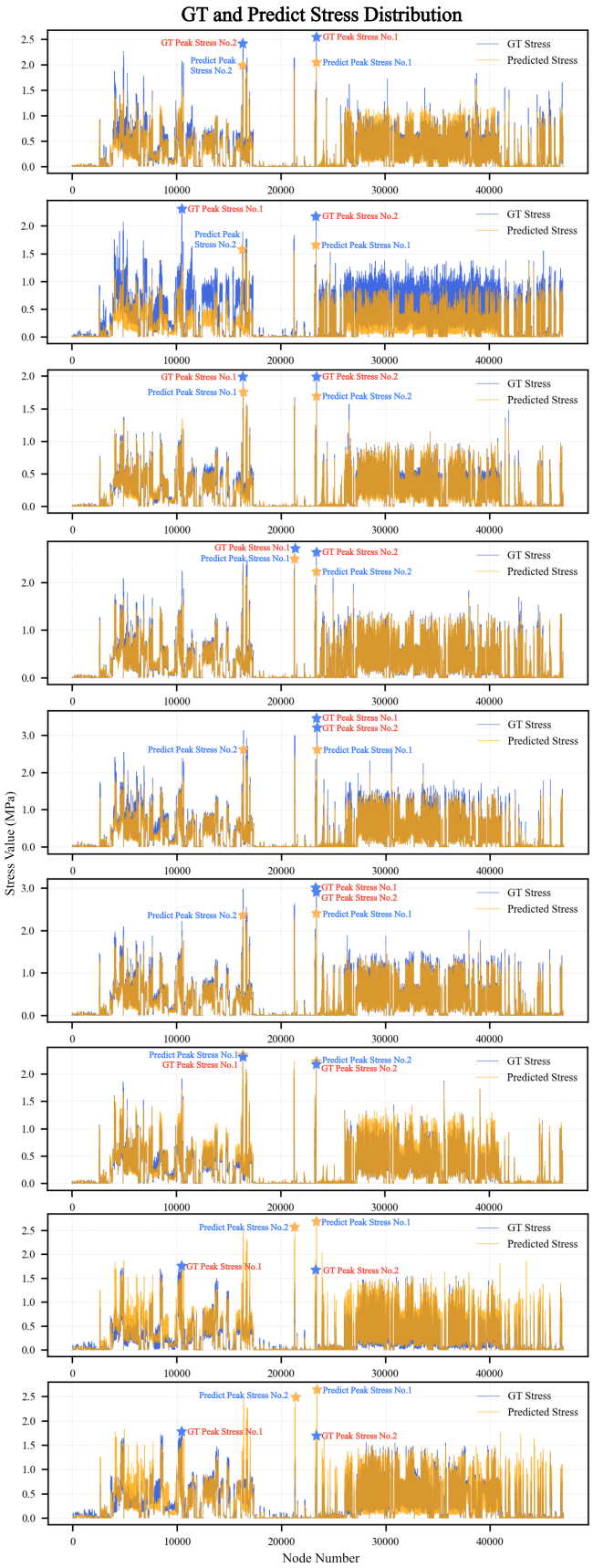

Fig. 14. Comparison between the ground truth (GT) and predicted stress distributions for each subject.

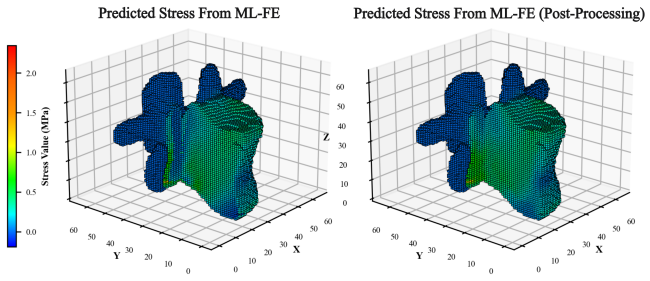

Fig. 15. Visualization of Von Mises stress distributions from FEA, ML-FE prediction, and the post-processed prediction.

the model captured one of the genuine high-stress zones rather than an incorrect location, reflecting its limited capability in capturing multiple coexisting high-stress regions. The results shown in Fig. 14 further demonstrate that the proposed ML-FE model could reliably reconstruct the overall stress patterns and magnitudes even in morphologically distinct or locally complex cases, while exhibiting potential for further improvement in multi-peak stress region recognition.

To further investigate the sources of these prediction errors, PCA was used to quantify the morphological variability. After spatial normalization, each vertebral shape was converted into a vector and the principal components of the variation were extracted by PCA. Furthermore, its morphological variability was measured by the Euclidean distance between its PCA coordinates and those of the mean shape from the training set (Fig. 19). The mean shape of the training set was defined as:

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{13}$$

where x_i was the shape vector of the i-th training sample and N was the number of training samples.

In the visualized results (Fig. 19), sample A was identified as the closest to the mean shape, with a distance of 4.2128mm, whereas sample B was the farthest, with a distance of 25.5502 mm. The results indicated that samples with higher prediction errors are typically located farther away from the mean shape in the PCA space, reflecting a greater anatomical deviation, such as increased curvature or abnormal pedicle anatomy, which poses additional challenges for the model in accurately capturing local stress distributions. In contrast, samples closer to the mean shape tend to exhibit lower prediction errors, demonstrating robust generalization for vertebral anatomies that are well represented in the training distribution. These findings underscore the importance of incorporating sufficient morphological diversity in the training data to further enhance the generalizability of the model across the full range of anatomical variations.

IV. Discussion

This study presented a deep learning-based model for predicting the distribution of vertebral stress, which is very good

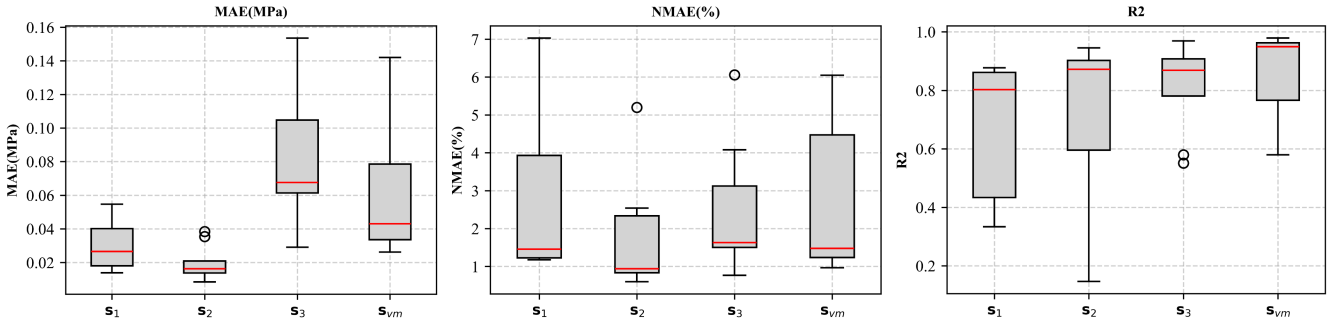

Fig. 16. Box plots of MAE, NMAE, and $\mathbf{R^2}$ for $\mathbf{s_1}$, $\mathbf{s_2}$, $\mathbf{s_3}$, and $\mathbf{s_{vm}}$ predictions in whole vertebrae.

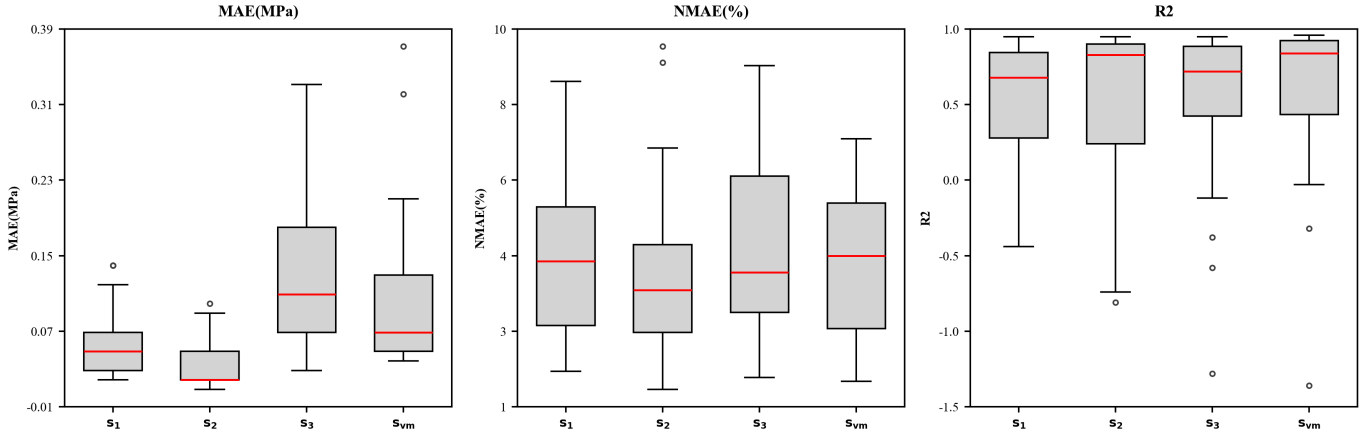

Fig. 17. Box plots of MAE, NMAE, and $\mathbf{R^2}$ for $\mathbf{s_1}$, $\mathbf{s_2}$, $\mathbf{s_3}$, and $\mathbf{s_{vm}}$ predictions in the boundary and loading area.

overall in a range of vertebral anatomical variations and could effectively reconstruct stress patterns in major anatomical regions. However, spatial visualizations and further shape analysis revealed that the model exhibits relatively higher prediction errors in anatomically complex or atypical regions such as the anteroinferior margin of the vertebral body and areas adjacent to the pedicles.

In addition to its accuracy, the proposed end-to-end pipeline offers significant improvements in processing efficiency. As summarized in Table IV, the entire workflow from automated segmentation to stress prediction takes approximately 134 to 154 seconds per subject, whereas conventional FEA requires over 90 to 120 minutes, even with homogeneous isotropic properties. For stress computation alone, the surrogate model infers results in about 3 seconds on an NVIDIA RTX 4090 GPU, compared to 11 seconds for a traditional FEM solver (ANSYS) under similar conditions. This efficiency highlights the potential for scalable, real-time application in clinical biomechanics, where rapid and automated stress analysis is essential. The present comparison was conducted under homogeneous and linear elastic assumptions, but the computational burden of traditional FEA will increase substantially when nonlinearities (e.g., nonlinear constitutive behavior, large deformations, geometrical nonlinearities, contact in implant–bone interfaces) are incorporated. In such cases, the proposed ML-FE surrogate would offer even greater advantages, as its inference time remains nearly constant regardless of model complexity, enabling rapid biomechanical evaluation of subject-specific FEMs.

TABLE IV
COMPARISON OF PROCESSING TIME BETWEEN THE ENTIRE END-TO-END PIPELINE AND CONVENTIONAL FEA.

Step	**End-to-End Pipeline**	**Conventional FEA**
CT segmentation	25 $\sim$ 45 s	30 $\sim$ 60 min
FE modeling	96 $\sim$ 106 s	60 $\sim$ 90 min
Stress computation	3 s	11 s
Total time per subject	**134 $\sim$ 154 s**	**90 $\sim$ 120 min**

Other limitations of this study include the relatively small data set and the assumption of the same homogenous material properties within and across all samples. This approach will improve model generalization and workflow integration to strengthen clinical applicability. Moreover, further investigations are needed to examine whether the surrogate model can effectively capture such nonlinear mechanical behaviors within patient-specific FEMs, thereby improving its physiological realism and clinical relevance. Overall, the present findings demonstrate the feasibility of transforming the traditionally offline, expert-dependent FEA process into a fully automated and near–real-time biomechanical analysis pipeline, capable of supporting rapid subject-specific evaluation and personalized

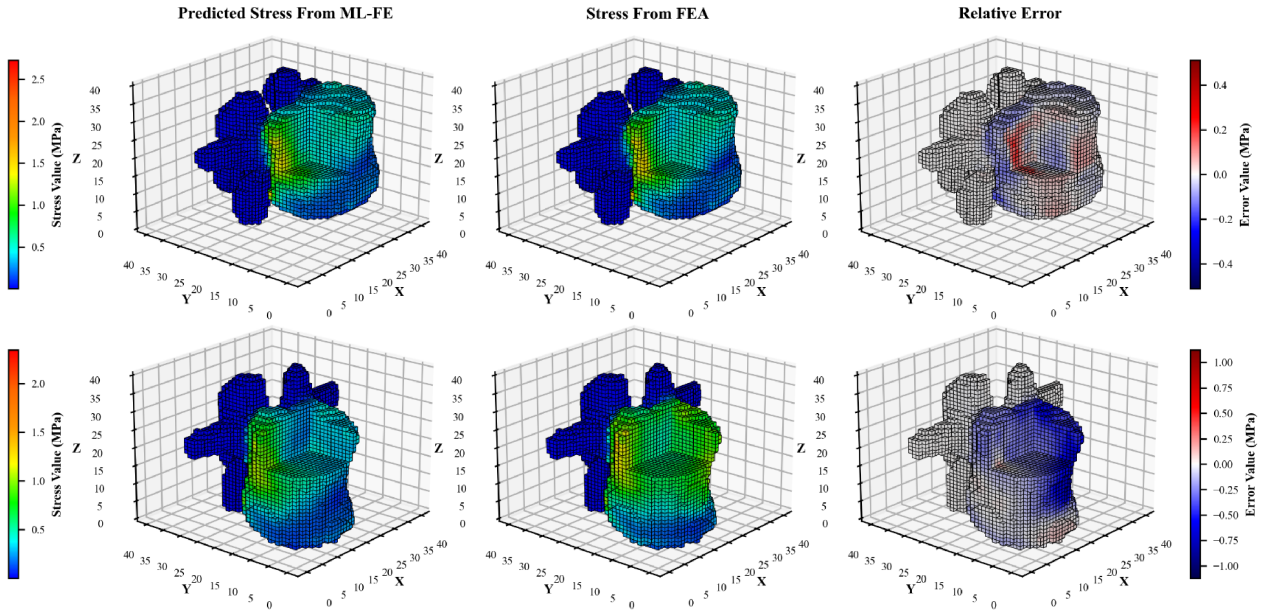

Fig. 18. Visualization of predicted stress, reference FEA stress, and relative error for three representative samples (A, B). Each row corresponds to one sample.

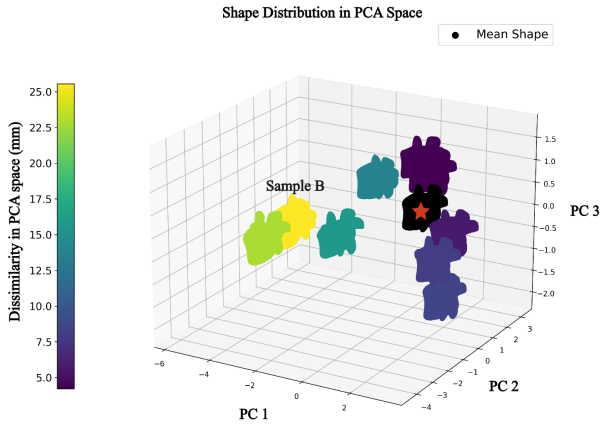

Fig. 19. Distribution of vertebral shapes from the test samples in the PCA space. Each sample was colored by its Euclidean distance to the mean shape in the normalized PCA space.

treatment planning in clinical practice.

V. Conclusion

We have developed a subject-specific surrogate model based on DL and ML that enables a rapid and accurate prediction of vertebral stress distributions in various anatomical variations. By integrating shape encoding and branch-specific decoding, the model achieved low MAE and high R^2 in independent test subjects, excelling in Von Mises stress. Although localized errors were identified in anatomically complicated locations such as the anteroinferior border and pedicles, the model generalized to be robust. Compared to conventional FEA, this approach enables inference of a single subject in 3 seconds on a NVIDIA RTX 4090 GPU, significantly improving clinical efficiency, whereas the entire end-to-end pipeline was completed in 134 to 154 seconds. This innovative approach, especially when combined with subject-specific material features, has great potential for clinical real-time vertebral stress analysis.

References

[1] M. C. Costa, P. Eltes, A. Lazary, P. P. Varga, M. Viceconti, and E. Dall'Ara., "Biomechanical assessment of vertebrae with lytic metastases with subject-specific finite element models," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 98, pp. 268–290, 2019.

[2] S. Wang, W. M. Park, H. R. Gadikota, J. Miao, Y. H. Kim, K. B. Wood, and et al., "A combined numerical and experimental technique for estimation of the forces and moments in the lumbar intervertebral disc," *Comput. Methods Biomech. Biomed. Eng.*, vol. 16, no. 12, pp. 1278–1286, 2013.

[3] A. Atkins, M. Burke, S. Samiezadeh, M. K. Akens, M. Hardisty, and C. M. Whyme., "Elevated microdamage spatially correlates with stress in metastatic vertebrae," *Ann. Biomed. Eng.*, vol. 47, pp. 980–989, 2019.

[4] F. G. Gibson, M. A. Paggiosi, C. Handforth, J. E. Brown, X. Li, E. Dall'Ara, and et al., "Altered vertebral biomechanical properties in prostate cancer patients following androgen deprivation therapy," *Bone*, vol. 195, 2025.

[5] M. Palanca, G. Cavazzoni, and E. Dall'Ara, "The role of bone metastases on the mechanical competence of human vertebrae," *Bone*, vol. 173, 2023.

[6] I. Singhal, B. Harinathan, A. Warraich, Y. Purushothaman, M. D. Budde, N. Yoganandan, and et al., "Finite element modeling of the human cervical spinal cord and its applications: A systematic review," *North Am. Spine Soc. J.*, vol. 15, 2023.

[7] D. Nath, Ankit, D. R. Neog, and S. S. Gautam., "Application of machine learning and deep learning in finite element analysis: A comprehensive review," *Arch. Comput. Methods Eng.*, vol. 31, pp. 2945–2984, 2024.

[8] M. Ahmadi, D. Biswas, R. Paul, M. Lin, Y. Tang, T. S. Cheema, and et al., "Integrating finite element analysis and physics-informed neural networks for biomechanical modeling of the human lumbar spine," *North Am. Spine Soc. J.*, vol. 22, 2025.
[9] X. Garcia-Andrés, E. Noguera, E. Arana, B. Gandía-Vañó, and J. J. Ródenas., "Methodology for the assessment of the risk of failure of metastatic vertebrae through rom-based patient-specific simulations," *Comput. Struct.*, vol. 296, p. 107298, 2024.
[10] J. Kok, Y. M. Shcherbakova, T. P. C. Schlösser, P. R. Seevinck, T. A. can der Velden, R. M. Castelein, and et al., "Automatic generation of subject-specific finite element models of the spine from magnetic resonance images," *Front. Bioeng. Biotechnol.*, vol. 11, p. 1244291, 2023.
[11] R. Awal, M. Naznin, and T. R. Faisal, "Machine learning based finite element analysis (fea) surrogate for hip fracture risk assessment and visualization," *Expert Systems with Applications*, vol. 264, p. 125916, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417424027830
[12] R. Awal, S. C. Doll, M. Naznin, and T. R. Faisal, "Interpretable machine learning classifiers for the reliable prediction of fall induced hip fracture risk," *Machine Learning for Computational Science and Engineering*, vol. 1, no. 2, p. 2, 2025. [Online]. Available: https://doi.org/10.1007/s44379-024-00004-w
[13] M. T. Löffler, A. Sekuboyina, A. Jakob, A.-L. Grau, A. Scharr, M. E. Husseini, and et al., "A vertebral segmentation dataset with fracture grading," *Radiol. Artif. Intell.*, vol. 2, 2020.
[14] H. Liebl, D. Schinz, A. Sekuboyina, L. Malagutti, M. T. Löffler, A. Bayat, and et al., "A computed tomography vertebral segmentation dataset with anatomical variations and multi-vendor scanner data," *Sci. Data*, vol. 8, no. 1, p. 284, 2021.
[15] A. Sekuboyina, M. E. Husseini, A. Bayat, M. Löffler, H. Liebl, H. Li, and et al., "Verse: A vertebrae labelling and segmentation benchmark for multi-detector ct images," *Med. Image Anal.*, vol. 73, p. 102166, 2021.
[16] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, and et al., "3d slicer as an image computing platform for the quantitative imaging network," *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.
[17] N. Toussaint, Y. Redhead, M. Vidal-Garcia, L. L. Vercio, W. Liu, E. M. C. Fisher, and et al., "A landmark-free morphometrics pipeline for high-resolution phenotyping: application to a mouse model of down syndrome," *Development*, vol. 148, no. 18, p. 188631, 2021.
[18] A. Bône, M. Louis, B. Martin, and S. Durrleman, "Deformetrica 4: An open-source software for statistical shape analysis," in *Shape in Medical Imaging*. Cham: Springer International Publishing, 2018, pp. 3–13.
[19] A. Kadantsev, "pyansys: Pythonic interface to mapdl," 2021.
[20] Y. H. Ahn, W. M. Chen, K. Y. Lee, K.-W. Park, and S.-J. Lee., "Comparison of the load-sharing characteristics between pedicle-based dynamic and rigid rod devices," *Biomed. Mater.*, vol. 3, p. 044101, 2009.
[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. Learn. Represent. (ICLR)*, 2015, arXiv:1409.1556. [Online]. Available: https://arxiv.org/abs/1409.1556
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[23] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
[24] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," 2017. [Online]. Available: https://arxiv.org/abs/1606.00915
[25] Altair Engineering Inc., "Altair hypermesh," 2022, version 2022, Troy, MI, USA. [Online]. Available: https://altair.com/hypermesh
[26] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," 2017. [Online]. Available: https://arxiv.org/abs/1706.02413
[27] I. T. Jolliffe, *Principal component analysis*, 2nd ed., ser. Springer Series in Statistics. New York: Springer, 2002.
[28] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," 2017. [Online]. Available: https://arxiv.org/abs/1612.00593
[29] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[30] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems*, vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf