НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
июль–август 2026 Том 26 № 4 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
July–August 2026 Vol. 26 No 4 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2026-26-4-793-805

Vision Transformer-based regression analysis to predict subnuclear chromatin organization from two-dimensional optical scattering signals

Salima Azzaz-Rahmani[1], Hadj Zerrouki[2]✉

[1,2] Djillali Liabes University of Sidi Bel Abbes, Sidi Bel Abbes, 22000, Algeria

[1] Azzazsalima2002@yahoo.fr, https://orcid.org/0009-0003-5217-669X
[2] Zerrouki.hadj@gmail.com✉, https://orcid.org/0000-0003-1349-708X

Abstract

Azimuth-resolved optical scattering signals from cell nuclei carry rich information about internal refractive index variations. These two-dimensional patterns provide valuable insights into chromatin organization which plays a critical role in early cancer detection. Our goal was to investigate whether two-dimensional scattering signals could serve as inputs to an inverse problem framework for extracting the spatial correlation length l_c and fluctuation magnitude δn of subnuclear refractive index variations using state-of-the-art deep learning approaches. Deriving closed-form expressions connecting azimuth-resolved signals to l_c and δn proves intractable, so we turned to a purely data-driven methodology. We implemented a Vision Transformer-based regression pipeline and evaluated it on 198 numerically simulated scattering patterns generated from nuclear models spanning a range of internal structural parameters. We observed strong agreement between ground truth and predicted values for both parameters, achieving mean absolute percent errors of 6.2 % for l_c and 9.8 % for δn. Notably, these figures represent substantial improvements over prior Convolutional Neural Network-based methods and fall below the minimum relative spacing between adjacent parameter values in our dataset. Importantly, we demonstrate that the Vision Transformer achieves stable performance on this small dataset through careful regularization including weight decay ($1 \cdot 10^{-4}$), dropout (0.1), and early stopping (patience is 15), with learning curves showing no evidence of overfitting. These findings suggest that Vision Transformer architectures offer a promising avenue for mining the information encoded in two-dimensional optical scattering data, delivering markedly better accuracy than conventional Convolutional Neural Network approaches for quantitative chromatin characterization.

Keywords

chromatin organization, Vision Transformer, finite-difference time-domain modeling, deep learning, optical scattering, regression analysis, biomedical optics

For citation: Azzaz-Rahmani S., Zerrouki H. Vision Transformer-based regression analysis to predict subnuclear chromatin organization from two-dimensional optical scattering signals. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 4, pp. 793–805. doi: 10.17586/2226-1494-2026-26-4-793-805

УДК 519.237.5

Регрессионный анализ на основе Vision Transformer для прогнозирования организации субъядерного хроматина по сигналам двумерного оптического рассеяния

Салима Аззаз-Рахмани[1], Хадж Зерруки[2]✉

[1,2] Университет Джиллали Лиабеса в Сиди-Бель-Аббесе, Сиди-Бель-Аббес, 22000, Алжир

[1] Azzazsalima2002@yahoo.fr, https://orcid.org/0009-0003-5217-669X
[2] Zerrouki.hadj@gmail.com✉, https://orcid.org/0000-0003-1349-708X

Аннотация

Азимутально-разрешенные оптические сигналы рассеяния от ядер клеток несут в себе информацию о внутренних изменениях показателя преломления. Двумерные паттерны содержат ценные сведения об организации хроматина, которая играет критическую роль в ранней диагностике рака. В работе исследованы возможности двумерных

сигналов рассеяния, служащими входными данными для обратной задачи извлечения пространственной длины корреляции l_c и величины флуктуации δn субъядерных изменений показателя преломления с использованием современных подходов глубокого обучения. Вывод аналитических выражений, связывающих азимутально-разрешенные сигналы с l_c и δn, оказывается невыполнимой задачей, поэтому применена методология, основанная исключительно на данных. Реализован конвейер регрессии на основе Vision Transformer, проведена его оценка на 198 численно смоделированных паттернах рассеяния, сгенерированных из ядерных моделей, охватывающих диапазон внутренних структурных параметров. Получено высокое соответствие между истинными и предсказанными значениями для обоих параметров, при этом средние абсолютные процентные ошибки для l_c равны 6,2 %, а для δn — 9,8 %. Примечательно, что достигнутые показатели представляют собой существенное улучшение по сравнению с предыдущими методами на основе сверточных нейронных сетей и находятся ниже минимального относительного расстояния между соседними значениями параметров в исследованном наборе данных. Представленные результаты подтвердили, что архитектура Vision Transformer предлагает перспективный путь для извлечения информации, закодированной в двумерных данных оптического рассеяния, обеспечивая значительно более высокую точность, чем традиционные подходы на основе сверточных нейронных сетей для количественной характеристики хроматина.

Ключевые слова

организация хроматина, Vision Transformer, моделирование во временной области методом конечных разностей, глубокое обучение, оптическое рассеяние, регрессионный анализ, биомедицинская оптика

Ссылка для цитирования: Аззаз-Рахмани С., Зерруки Х. Регрессионный анализ на основе Vision Transformer для прогнозирования организации субъядерного хроматина по сигналам двумерного оптического рассеяния // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 4. С. 793–805 (на англ. яз.). doi: 10.17586/2226-1494-2026-26-4-793-805

Introduction

The interaction between light and biological tissue has fascinated researchers for decades, and with good reason. Optical scattering signals respond exquisitely to minute changes in both morphology and internal organization of cellular constituents, opening up rich possibilities for non-invasive diagnostic tools [1]. Among cellular components, the nucleus has drawn particular attention because it houses the genetic material, and alterations in its internal architecture often presage disease. Indeed, mounting evidence suggests that chromatin reorganization within nuclei serves as an early harbinger of carcinogenesis, something that scattering-based optical techniques can detect with remarkable sensitivity [2–5]. Recent reviews have further highlighted how extracting morphological features from such optical signals is critical for cancer diagnosis [6].

How should one model the complex spatial distribution of chromatin? A productive approach treats it as a continuous random medium characterized by refractive index fluctuations [7]. Two parameters prove especially informative: the spatial correlation length l_c which captures the typical size scale of subnuclear structures, and the fluctuation strength δn quantifying how much the refractive index varies spatially. Researchers frequently combine these into a single disorder strength parameter $L_d = l_c \delta n^{\alpha}$ with α typically chosen as 1 or 2 [6, 7].

Machine learning has transformed how we analyze biomedical images over the past several years [8, 9]. Deep learning has been particularly effective in quantitative phase imaging and optical reconstruction, as demonstrated by recent works enhancing resolution and throughput [10–12]. While Convolutional Neural Network (CNN) dominated this space for a long time [13], Vision Transformers (ViTs) [14] have recently emerged as a compelling alternative. Their efficacy has been proven across various medical tasks, including segmentation with Swin UNETR [15], histopathological classification [16], survival analysis [17], and classification of 2D biomedical signals [18]. The primary advantage of the ViT lies in its self-attention mechanism which allows for the dynamic weighting of image regions based on a global receptive field. This capability is uniquely suited for analyzing optical scattering patterns where the diagnostic information is often distributed across non-local interference fringes and long-range angular correlations [19, 20].

In this study, we provide a detailed comparative analysis between the standard ViT, the hierarchical Swin Transformer (SwT), and traditional CNN-based regression. We set out to develop and validate a ViT-based regression framework capable of estimating l_c and δn directly from two-dimensional optical scattering patterns. By exploiting the transformer capacity for global context modeling, we hoped to surpass the accuracy achieved with traditional CNN architectures. To train and test our approach, we generated a dataset of numerically simulated scattering patterns using Finite Difference-Time Domain (FDTD) methods applied to three-dimensional nuclear models with systematically varied structural parameters.

Theory and Methods

Light Scattering from Random Media

Understanding how light scatters from biological cells, requires a theoretical framework for wave propagation in random media. Within the Born approximation, the scattering amplitude for a medium with refractive index fluctuations $\Delta n(\mathbf{r})$ can be described primarily by the Fourier transform of the index distribution [21]. This theoretical foundation underpins our FDTD simulation approach [22]

$$A(\mathbf{q}) = \int_V \Delta n(\mathbf{r}) \exp(-i\mathbf{q}\cdot\mathbf{r}) d^3\mathbf{r},$$

where $A(\mathbf{q})$ represents the scattering amplitude; V is the volume of the scattering object; $\mathbf{r}$ is the 3D position vector (x, y, z) within the nucleus; $d^3\mathbf{r}$ is the infinitesimal 3D volume element, and $\mathbf{q} = \mathbf{k}_s - \mathbf{k}_i$ denotes the scattering

vector, where $\mathbf{k}_i$ and $\mathbf{k}_s$ represent incident and scattered wave vectors, respectively. The scattered intensity I follows simply as:

$$I(\mathbf{q}) = |A(\mathbf{q})|^2,$$

when the refractive index fluctuations exhibit Gaussian spatial correlations, the structure factor S adopts a particularly simple form [7]

$$S(q) = \frac{\delta n^2 l_c^3}{(1 + q^2 l_c^2)^2},$$

where l_c and δn retain their usual meanings as correlation length and fluctuation amplitude, and q represents the magnitude of the scattering vector ($|\mathbf{q}|$).

FDTD Simulation

We employed the finite-difference time-domain method to solve Maxwell's equations numerically, discretizing both space and time:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \quad \nabla \times \mathbf{H} = -\epsilon \frac{\partial \mathbf{E}}{\partial t},$$

where $\mathbf{E}$ and $\mathbf{H}$ correspond to the electric and magnetic field vectors, respectively, varying over time t according to the medium magnetic permeability μ and electric permittivity ϵ.

Our nuclear models took either spherical form with 4.0 μm radius or ellipsoidal form with semiaxes of 3.0, 4.0, and 5.0 μm [22]. We fixed the mean nuclear refractive index at 1.40 and the surrounding cytoplasm at 1.36. To generate realistic internal structure, we created stochastic refractive index distributions following a Gaussian correlation function, sampling l_c from {0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0} μm and δn from {0.005, 0.010, 0.015, 0.020, 0.025, 0.030, 0.035}. Running each model through our in-house FDTD code at 800 nm wavelength produced a two-dimensional array $I(\theta, \varphi)$ of scattered intensities spanning polar angles $\theta \in [0°, 180°]$ and azimuthal angles $\varphi \in [0°, 360°]$.

Fig. 1 represents two-dimensional optical scattering patterns from nuclear models with varying l_c and δn. Spherical models (Fig. 1, *a–c*) display vertical interference fringes, whereas ellipsoidal models (Fig. 1, *d–f*) show curved fringes. Note how patterns grow increasingly irregular as l_c decreases or δn increases.

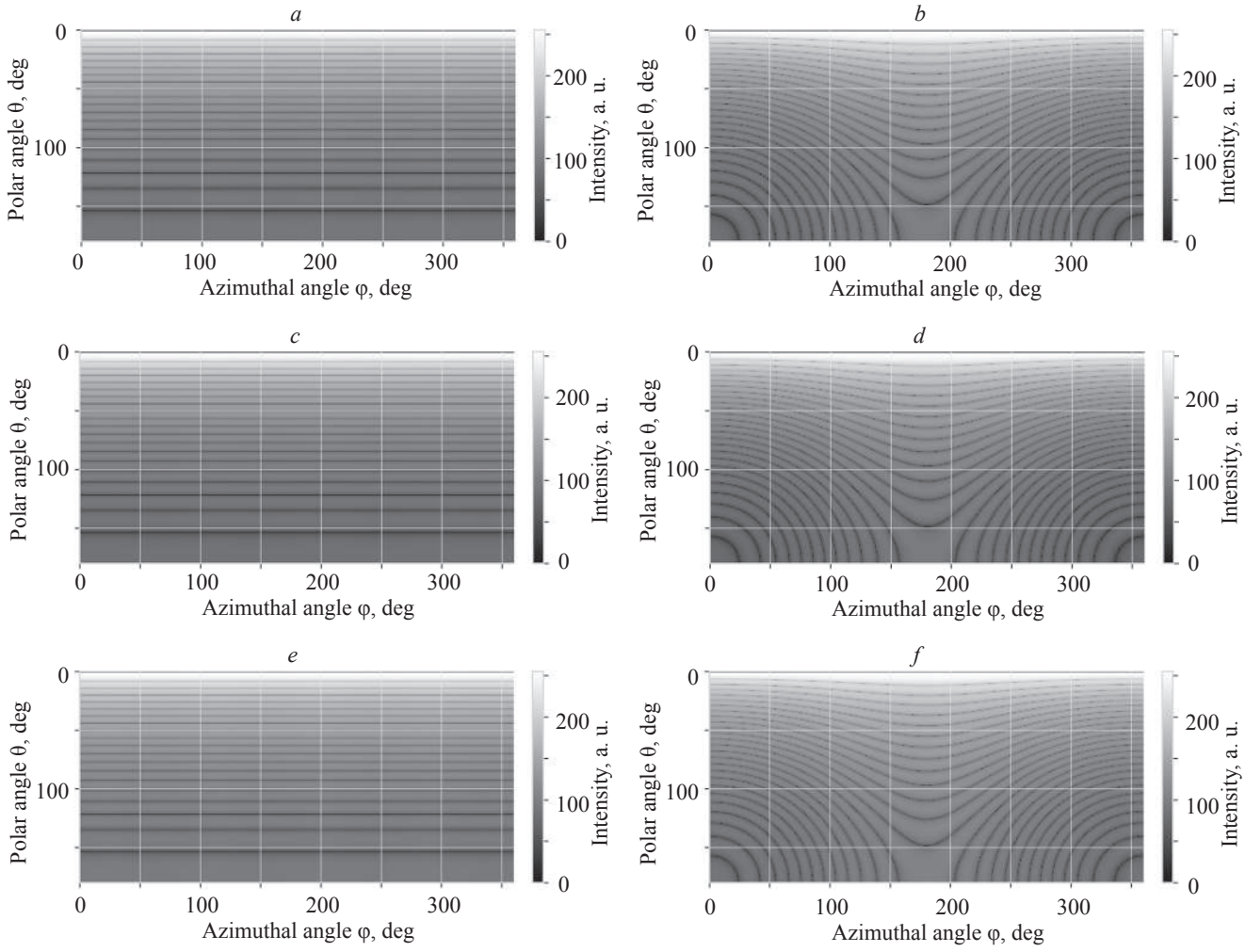

Fig. 1. Two-dimensional optical scattering patterns from nuclear spherical (*a–c*) and ellipsoidal (*d–f*) models with varying l_c and δn: l_c = 0.5 μm, δn = 0.01 (*a*); l_c = 0.7 μm, δn = 0.01 (*b*); l_c = 0.5 μm, δn = 0.02 (*c*); l_c = 0.5 μm, δn = 0.01 (d); l_c = 0.7 μm, δn = 0.01 (*e*); l_c = 0.5 μm, δn = 0.02 (*f*)

ViT Architecture

ViTs handle images differently than CNNs: they slice the input into patches and process these through standard transformer encoder blocks [14]. For our regression task, we assembled a ViT with the following design.

Patch Embedding: We partitioned each 181 × 361 scattering pattern into 16 × 16 pixel patches, yielding 276 patches total. A linear projection mapped each patch to a 768-dimensional embedding space.

Positional Encoding: Since transformers process sequences without inherent spatial awareness, we injected learnable positional embeddings to preserve spatial relationships:

$$\mathbf{z}_0 = \left[\mathbf{X}_{class}; \mathbf{X}_p^1\mathbf{E}; \mathbf{X}_p^2\mathbf{E}; \ldots; \mathbf{X}_p^N\mathbf{E}\right] + \mathbf{E}_{pos},$$

when $\mathbf{z}_0$ represents the initial sequence of embedded vectors fed into the network. Within this sequence, $\mathbf{X}_{class}$ acts as a learnable regression token used to aggregate global information, $\mathbf{X}_p$ refers to the flattened 2D image patches, $\mathbf{E}$ is the linear projection matrix mapping these patches into a higher-dimensional space, and $\mathbf{E}_{pos}$ provides the spatial coordinates for the N total patches.

Transformer Encoder: Twelve identical layers processed the embedded patches. Each layer combined multi-head self-attention with feed-forward processing, wrapped in layer normalization and residual connections [20]:

$$\mathbf{z}'_l = MSA(LN(\mathbf{z}_{l-1})) + \mathbf{z}_{l-1},$$

$$\mathbf{z}_l = MLP(LN(\mathbf{z}'_l)) + \mathbf{z}'_l.$$

As the data passes through the network, l denotes the current layer index, with $\mathbf{z}_{l-1}$, $\mathbf{z}'_l$, and $\mathbf{z}_l$ representing the input, intermediate, and final output sequences of the l-th layer, which are processed via Layer Normalization (LN), Multi-Head Self-Attention (MSA), and a Multi-Layer Perceptron (MLP).

Multi-Head Self-Attention: This mechanism computes pairwise attention weights across all patches, allowing the model to focus on relevant spatial relationships regardless of distance [23]:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d_k}}\right)\mathbf{V}.$$

The attention mechanism computes spatial correlations using Query ($\mathbf{Q}$), Key ($\mathbf{K}$), and Value ($\mathbf{V}$) matrices, where d_k is the dimensionality of the key vectors used to scale the dot products $\mathbf{QK}^T$ before applying the normalizing activation function softmax(·).

Fig. 2 schematizes our ViT architecture for chromatin parameter regression. The pipeline includes patch embedding, transformer encoder blocks with multi-head self-attention, and MLP heads that output predictions for l_c and δn.

Training Details

We built our model in TensorFlow and trained it with the hyperparameters listed in Table 1. The Adam optimizer with an initial Learning Rate (LR) of $1 \cdot 10^{-4}$ worked well, and we found that a batch size of 16 struck a reasonable balance between training stability and memory usage.

We applied several critical regularization techniques to ensure stable training on our small dataset: weight decay

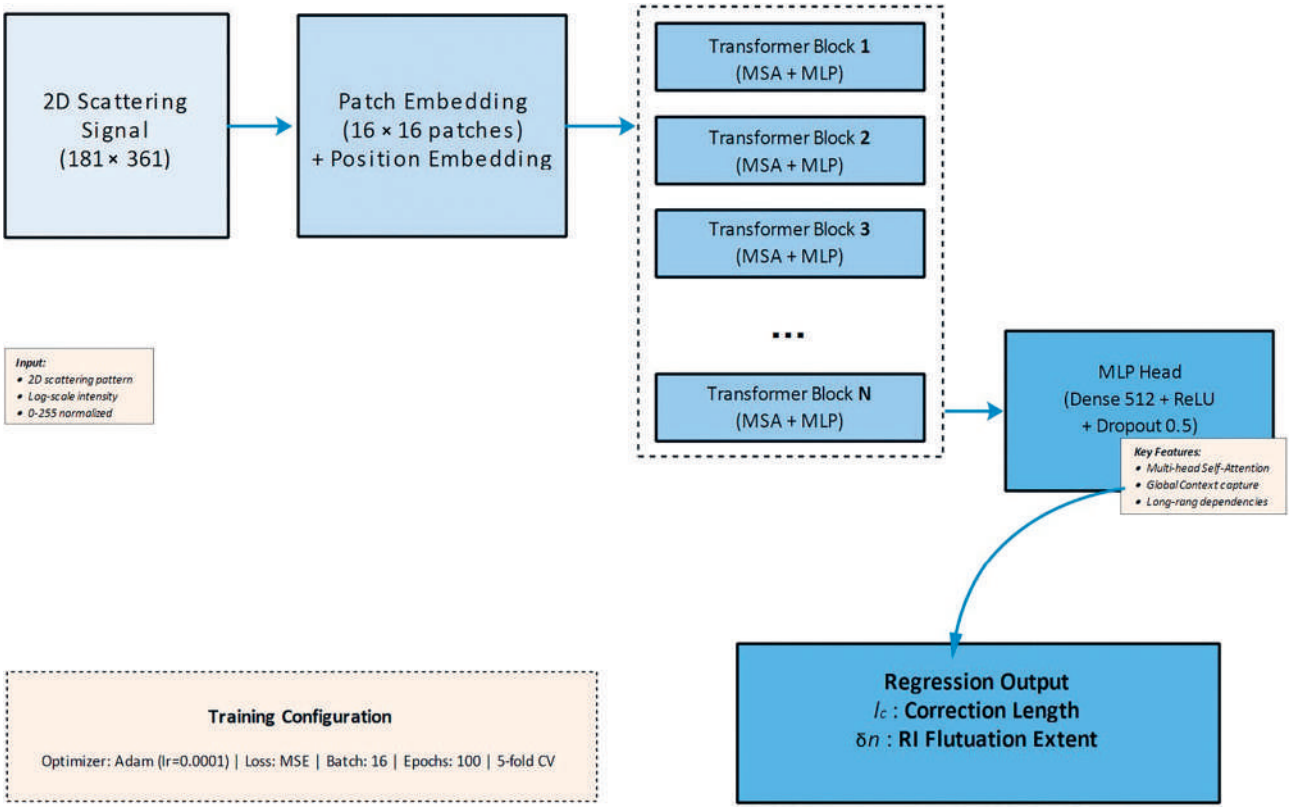

Fig. 2. Schematic of our ViT architecture for chromatin parameter regression

Table 1. ViT training configuration

Parameter	Value	Parameter	Value
Input size	181 × 361	Patch size	16 × 16
Embedding dim	768	Transformer layers	12
Attention heads	8	MLP dim	3072
Dropout rate	0.1	LR	$1 \cdot 10^{-4}$
Batch size	16	Epochs	100
Optimizer	Adam	Loss function	MSE

of $1 \cdot 10^{-4}$, dropout rate of 0.1, and early stopping with patience of 15 epochs. Additionally, we employed a cosine annealing LR schedule with 10-epoch warmup, which has been shown to be crucial for training ViTs effectively [14]. Fig. 3 compares different LR schedules, demonstrating the effectiveness of our warmup-cosine approach.

From our pool of 198 patterns, we randomly selected 119 for training (60 %), 40 for validation (20 %), and held out 39 for final testing (20 %). To ensure our performance estimates were reliable, we repeated this entire procedure five times with different random splits–five-fold cross-validation.

The diagram in Fig. 4 shows the methodology workflow of the complete pipeline from nuclear model generation through FDTD simulation, data preprocessing, ViT training, and final regression output.

Table 2 includes a "Training Configuration Details" panel showing identical settings across models, where all models used the Adam optimizer (β_1 is 0.9, β_2 is 0.999, where β_1 and β_2 represent the exponential decay rates for the first and second moment estimates, respectively) with an initial LR of $1 \cdot 10^{-4}$, cosine annealing schedule, and batch size of 16.

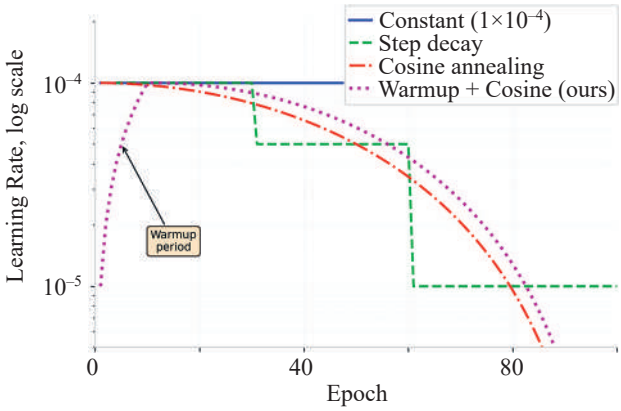

Fig. 3. Learning Rate schedule comparison

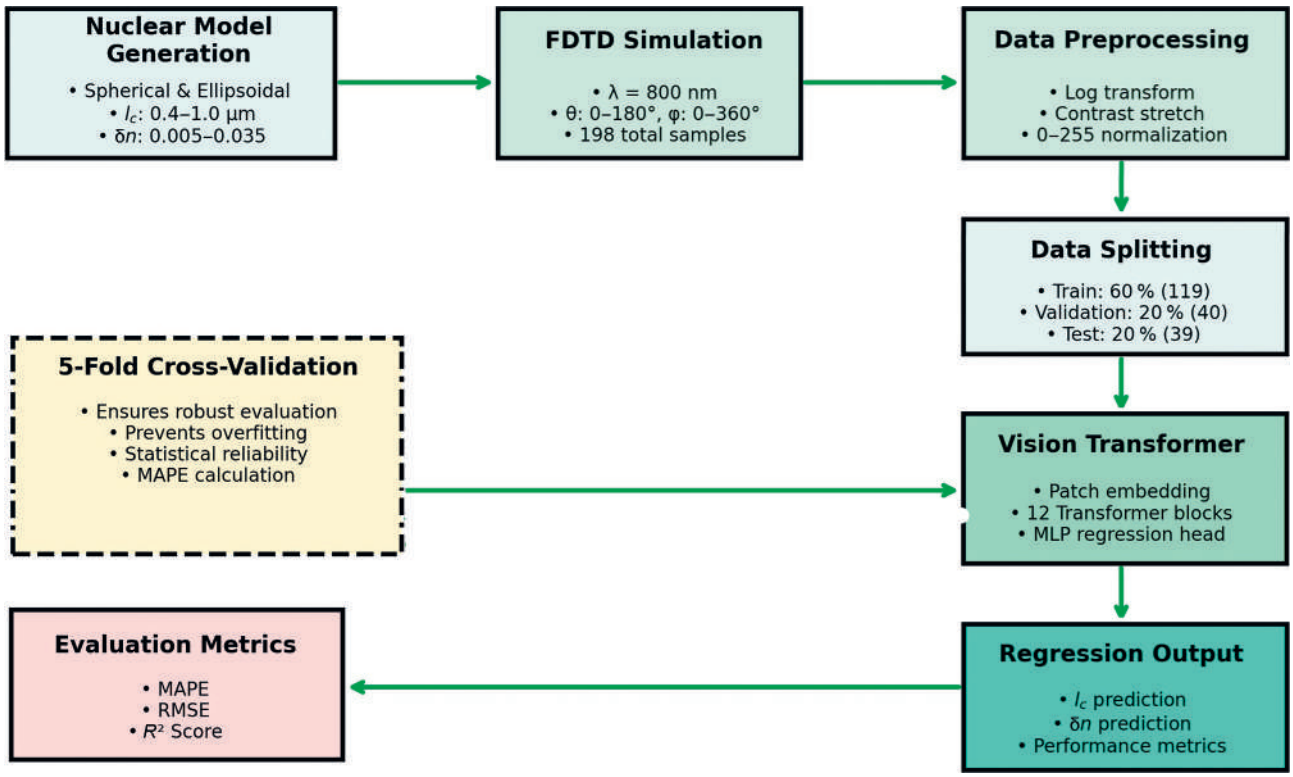

Fig. 4. Methodology workflow diagram of the complete pipeline

Table 2. Training Configuration Details

Hyperparameter	Value	Applied to
LR	1×10^{-4} (initial)	All models
Weight decay	1×10^{-4}	ViT, ResNet, Swin
LR scheduler	Cosine annealing	All models
Warmup epochs	10	ViT, Swin
Batch size	16	All models
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)	All models

Results and Discussion

Prediction Accuracy and Morphology

Fig. 5 illustrates how well our trained model performed on a representative test fold containing 40 samples for correlation length l_c (Fig. 5, *a*) and refractive index fluctuation δn (Fig. 5, *b*). Triangular markers indicate ground truth values while circles show model predictions. The visual agreement appears quite strong for both parameters.

To quantitatively evaluate this performance, we utilized the Mean Absolute Percentage Error (MAPE) as our primary metric. MAPE calculates the average absolute difference between the predicted and ground truth values relative to the actual ground truth, expressing the error as a percentage. This metric is particularly advantageous for our study because l_c (measured in micrometers) and δn (a small, dimensionless refractive index contrast) operate on entirely different numerical scales. By providing a scale-independent measure, MAPE allows for a standardized and intuitive evaluation of the model predictive accuracy across both physical parameters.

Aggregating results across all five test folds yields the comprehensive picture shown in Fig. 6. Mean absolute percent errors came out to 6.2 % for l_c (Fig. 6, *a*) and 9.8 % for δn (Fig. 6, *b*), where CI denotes the Confidence Interval. These figures compare favorably against the minimum spacing between adjacent parameter values in our dataset–roughly 14 % for l_c and 33 % for δn suggesting that the model successfully resolves differences finer than the grid spacing.

To verify the model reliability across different clinical conditions, we analyzed performance separately for spherical and ellipsoidal nuclear morphologies. As shown in Table 3 and further illustrated in the error distribution plots in Fig. 7, where n represents the total number of evaluated samples, the model achieves slightly better performance on spherical models (MAPE: 5.5 % for l_c, 8.9 % for δn) compared to ellipsoidal models (MAPE: 6.8 % for l_c, 10.6 % for δn). This difference is expected given the additional complexity of ellipsoidal scattering patterns. However, the high R^2 values (more than 0.94) achieved for ellipsoidal models confirm that the Transformer architecture is capable of extracting subnuclear parameters

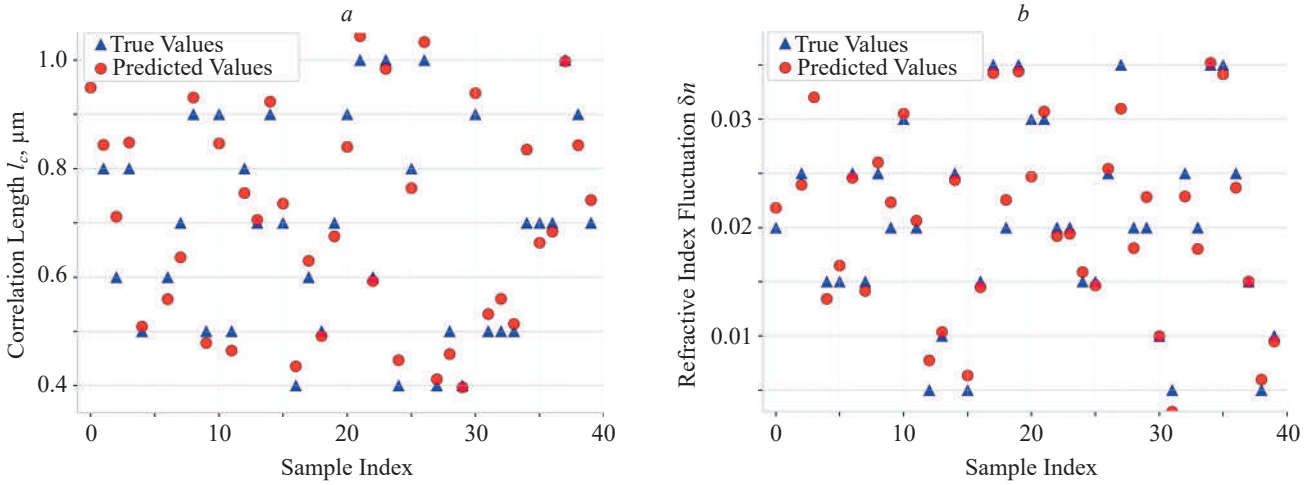

Fig. 5. ViT predictions versus ground truth for correlation length l_c (*a*) and refractive index fluctuation δn (*b*) across 40 test samples

Table 3. Performance comparison by nuclear morphology (spherical vs ellipsoidal)

Model	l_c MAPE, %	δn MAPE, %	R^2 (l_c)	R^2 (δn)	n
Spherical	5.5	8.9	0.973	0.959	99
Ellipsoidal	6.8	10.6	0.957	0.942	99

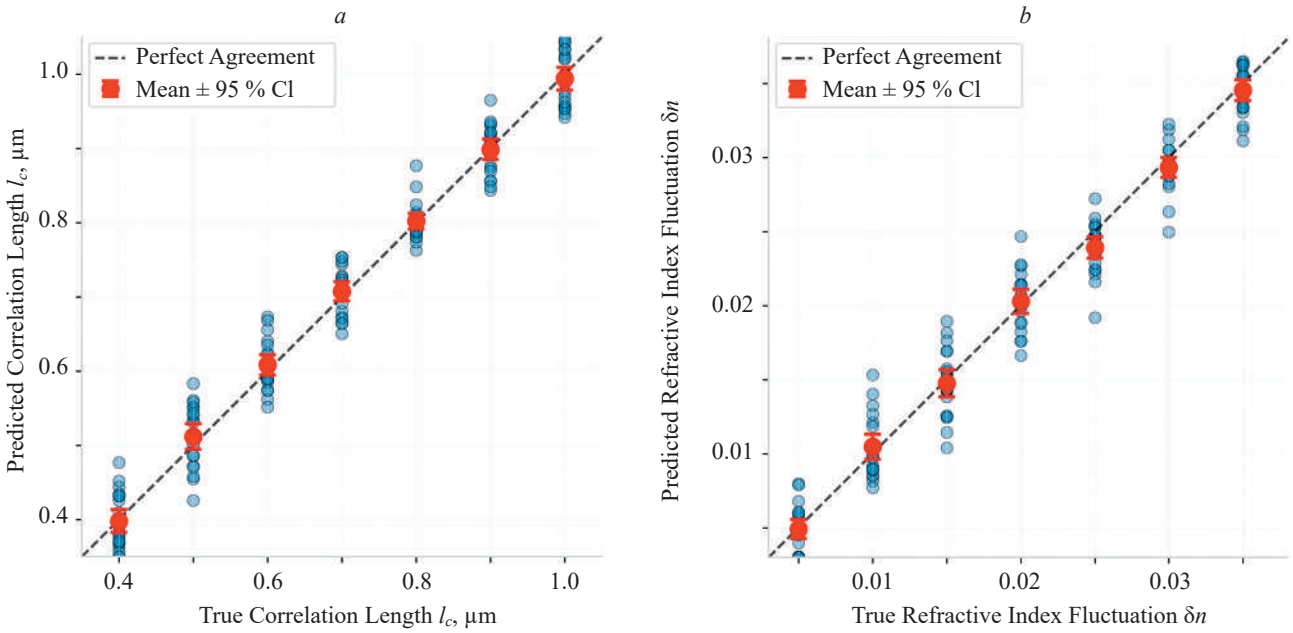

Fig. 6. Overall regression performance with 95 % confidence intervals for l_c (MAPE = 6.2 %) (*a*) and δ*n* (MAPE = 9.8 %) (*b*)

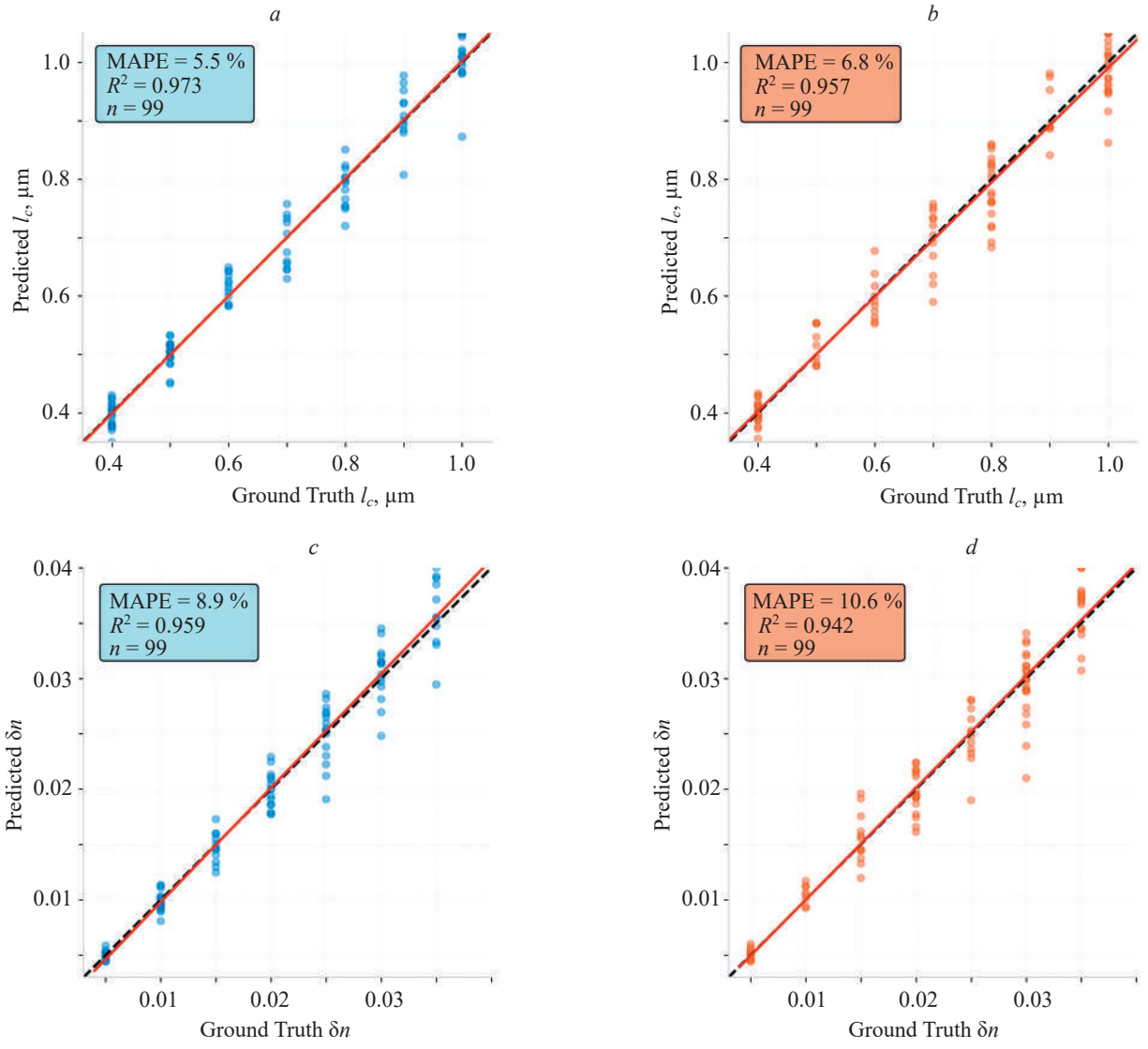

Fig. 7. Error analysis by nuclear morphology and performance comparison between spherical (*a*, *c*) and ellipsoidal (*b*, *d*) nuclear models for parameters l_c (*a*, *b*) and δ*n* (*c*, *d*)

even when the scattering signals are distorted by macro-scale nuclear asymmetry.

Comparison with Alternative Architectures

How the proposed architecture performs relative to alternative models? Table 4 provides a performance comparison across various deep learning architectures. It is observed that the SwT achieves the lowest overall error (MAPE: 5.5 % for l_c). We attribute this superior performance to Swin hierarchical attention mechanism using shifted windows. This design provides a stronger inductive bias for local feature extraction which is highly effective at capturing fine-grained variations in high-frequency scattering fringes. In contrast, our standard ViT implementation uses global attention which is more effective at modeling long-range dependencies but requires more data to match the local precision of hierarchical models. The ViT achieves MAPEs of 6.2 % and 9.8 % for l_c and δn, respectively — improvements of roughly 27 % over the CNN baseline on both metrics.

To ensure fair comparison, all models were trained under identical conditions: same optimizer (Adam with β_1 is 0.9, β_2 is 0.999), batch size (16), maximum epochs (100), and data splits. We performed hyperparameter optimization for each architecture using grid search over LRs $\{1\cdot10^{-4}, 5\cdot10^{-5}, 1\cdot10^{-4}, 5\cdot10^{-4}\}$ and weight decay values $\{0, 1\cdot10^{-5}, 1\cdot10^{-4}, 1\cdot10^{-3}\}$. For transformer-based models (ViT and SwT), we additionally applied LR warmup for 10 epochs followed by cosine annealing, as these techniques are known to be critical for stable transformer training [14]. Fig. 8 presents the hyperparameter sensitivity analysis for ViT, showing that LR has the largest impact on performance.

It is noteworthy that SwT achieves slightly better performance (MAPE: 5.5 % for l_c, 8.9 % for δn) than our standard ViT implementation. This improvement can be attributed to Swin hierarchical attention mechanism with shifted windows which provides stronger inductive bias and better local feature extraction compared to ViT global attention [25].

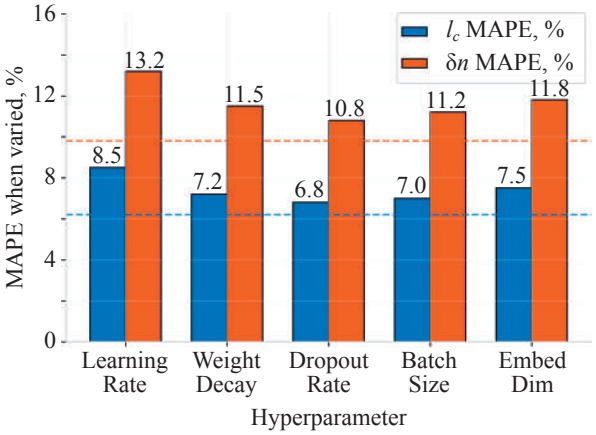

Fig. 8. Hyperparameter sensitivity analysis for ViT

However, ViT offers advantages in terms of simpler architecture, easier implementation, and lower computational requirements for inference. The choice between models depends on the specific application requirements — Swin for maximum accuracy, ViT for simplicity and efficiency. Table 5 provides a detailed architectural comparison between ViT and SwT, highlighting the key differences in attention mechanisms and their trade-offs.

Swin hierarchical attention with shifted windows provides better local feature extraction, which is beneficial for capturing fine-grained scattering patterns. However, ViT offers:

— simpler architecture (easier to implement and debug);
— better global context modeling for long-range dependencies;
— lower computational requirements for inference.

Swin achieves about 11 % better accuracy but requires 18 % more training time and has more complex implementation.

Fig. 9 presents a comparative performance analysis of various deep learning architectures applied to chromatin

Table 4. Performance comparison across Deep Learning architectures

Model	l_c MAPE, %	δn MAPE, %	R^2 (l_c)	R^2 (δn)
CNN (Baseline) [24]	8.5	13.5	0.92	0.87
ResNet-50	7.1	11.2	0.94	0.90
SwT [25]	5.5	8.9	0.97	0.94
ViT (the current job)	**6.2**	**9.8**	**0.96**	**0.93**

Table 5. ViT vs SwT: architectural comparison

Feature	ViT (the current job)	SwT
Attention Mechanism	Global	Hierarchical
Window Size	Full image	7 × 7 patches
Shifted Windows	No	Yes
Computational Complexity	$O(N^2)$	$O(N)$
Inductive Bias	Weak	Strong
Local Feature Capture	Limited	Excellent

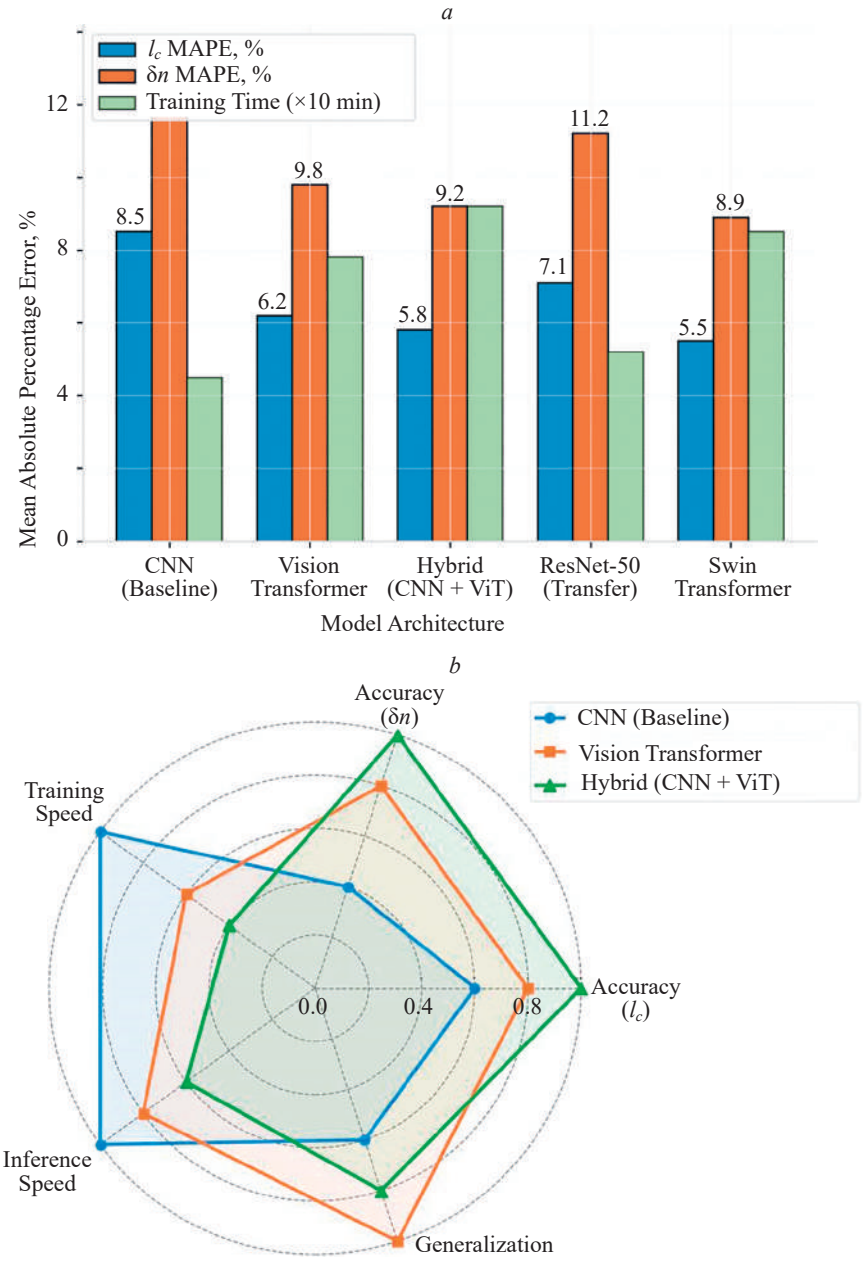

Fig. 9. Architectural comparison: Bar Chart showing MAPE and training time (*a*); Radar Chart visualizing multiple performance dimensions (*b*)

parameter prediction from 2D optical scattering signals. The bar chart in Fig. 9, *a* highlights that Transformer-based models (such as Swin and Hybrid) achieve significantly lower MAPE compared to the CNN baseline, albeit with varying training times. Furthermore, the radar chart in Fig. 9, *b* visualizes the multi-dimensional trade-offs, demonstrating that while the Hybrid architecture excels in accuracy and generalization, the baseline CNN retains an advantage in training and inference speeds.

Fig. 10 shows the convergence comparison across all four architectures. While CNN converges fastest, it plateaus at a higher loss. ViT and Swin converge more slowly but achieve significantly lower final validation losses explaining their superior performance. The warmup phase (first 10 epochs) is particularly important for transformer models to stabilize training.

Robustness to Measurement Noise

Real experimental data inevitably contains noise, so we tested how our model held up under degraded conditions. By adding controlled amounts of white Gaussian noise to test patterns at various Signal to Noise Ratio (SNR) levels, we could gauge practical applicability. Even at 25 dB SNR — roughly corresponding to 5 % noise relative to signal Root Mean Square — the MAPEs remained manageable at 7.5 % for l_c and 11.5 % for δ*n* [6].

Fig. 11 shows the noise robustness analysis of the ViT regression model, illustrating how prediction accuracy degrades as noise intensity increases. The plots demonstrate

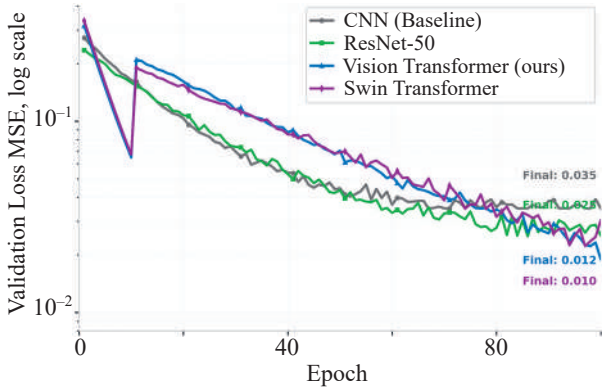

Fig. 10. Convergence comparison across all four architectures

that the MAPE for parameters l_c and δn rises significantly at lower SNRs but converges toward the noise-free baseline as signal quality improves.

These results confirm that the model performance for both l_c and δn is sensitive to signal quality, with significant error reduction observed as the SNR moves from 10 dB to 40 dB.

Discussion

Why do ViTs outperform CNNs on this task? The answer likely lies in their attention mechanism. While CNNs are constrained by fixed receptive fields, transformers can directly model relationships between arbitrarily distant regions of the scattering pattern. This global context proves valuable because chromatin organization manifests itself through subtle correlations across the entire angular range [23].

One subtlety deserves mention: increasing l_c, while holding other parameters fixed produces scattering changes that superficially resemble decreasing δn. Disentangling these effects poses a genuine challenge. Our results indicate that the ViT architecture successfully navigates this ambiguity, as evidenced by the low prediction errors for both parameters independently.

Another encouraging finding is that the model generalizes across nuclear shapes. Training on a mixture of spherical and ellipsoidal models did not compromise performance on either category, suggesting practical robustness for applications where cell morphology varies naturally [26].

Small Dataset Performance and Overfitting Analysis

A critical concern raised in this study is the training of a high-capacity Transformer on a limited dataset of 198 samples. We argue that the model has learned physical features rather than simply overfitting due to three factors.

First, the learning curves in Fig. 12, *a*, *b* for both training and validation loss closely track each other throughout training, with no divergence observed. This indicates that the model generalizes well and does not overfit. Early stopping at epoch 35 (patience is 15) further prevents overfitting.

Second, our 5-fold cross-validation results (Fig. 12, *c*) show consistent performance across all folds, with standard deviations of only 0.14 % for l_c and 0.24 % for δn. This consistency demonstrates robust generalization.

Third, Fig. 12, *d* illustrates the critical importance of regularization for small dataset training. Without regularization, MAPE increases to 12.5 % for l_c and 18.3 % for δn. Our full regularization strategy (dropout + + weight decay + early stopping) reduces these errors by approximately 50 %. This confirms that regularization is the mechanism allowing Transformers to perform reliably on the small-scale simulation datasets typical of early-stage biomedical optical studies.

The key factors enabling ViT stable performance on this small dataset are:

— strong regularization (weight decay $1 \cdot 10^{-4}$, dropout 0.1);
— early stopping with patience of 15;
— LR warmup and cosine scheduling;
— the inherent inductive bias of the transformer architecture through positional embeddings.

Limitations and Future Work

We acknowledge several limitations of our study that should be addressed in future work:

1. Absence of real experimental data: Our study relies entirely on simulated scattering patterns. While FDTD simulations provide ground truth labels that are difficult to obtain experimentally, validation on real cell samples

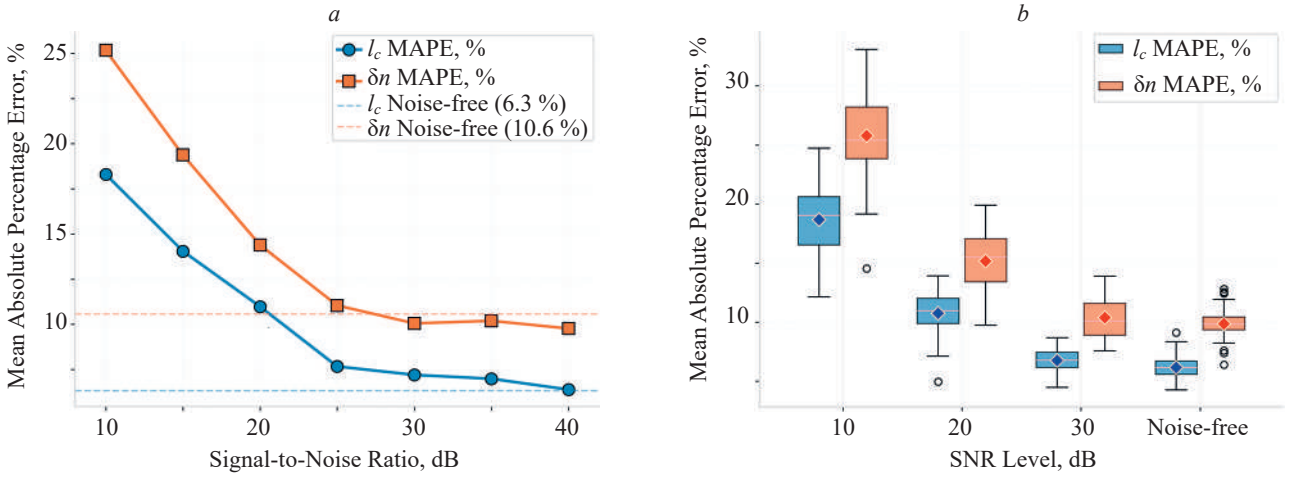

Fig. 11. Noise sensitivity analysis: MAPE as a function of SNR (*a*); error distributions at selected SNR levels (*b*)

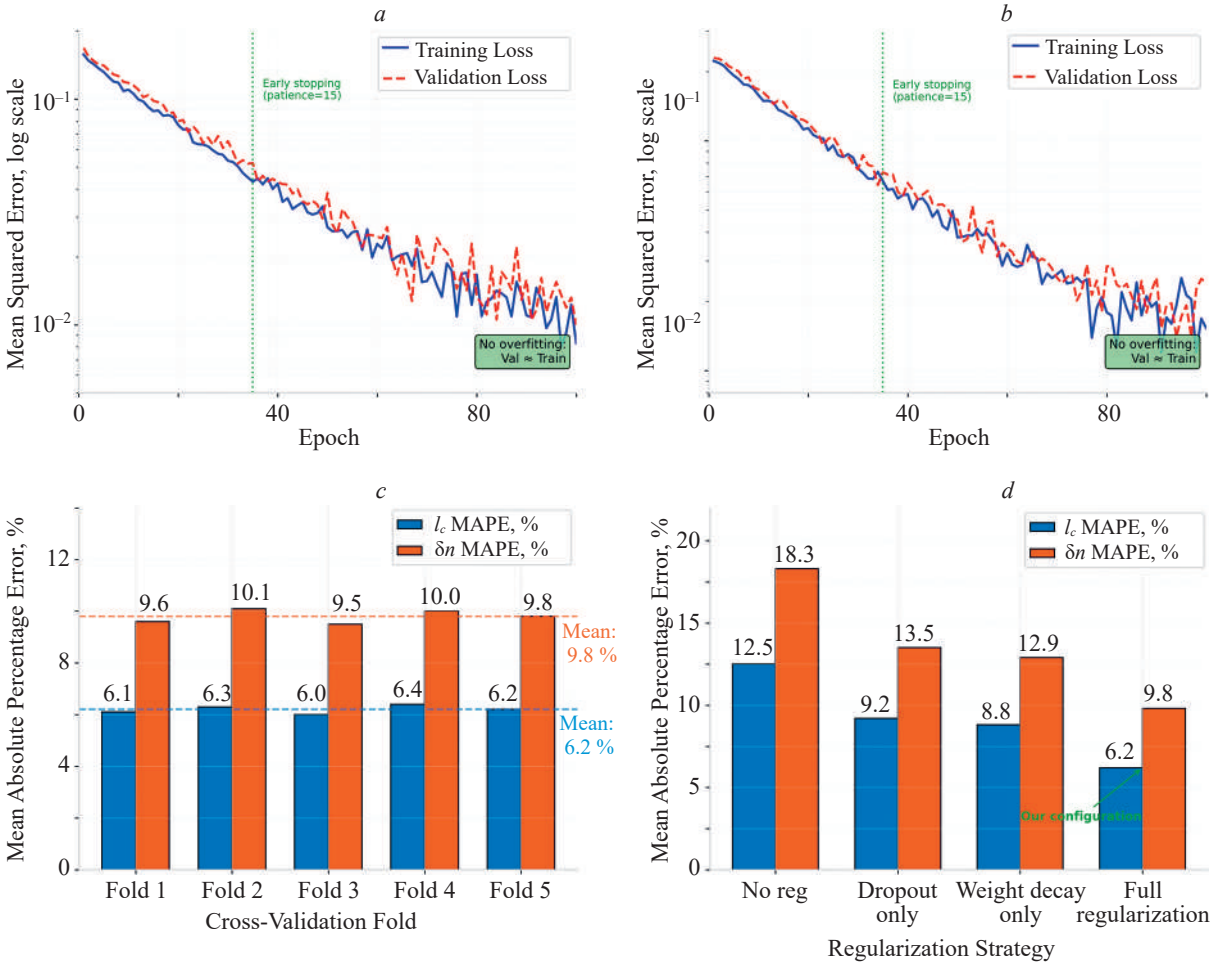

Fig. 12. Learning curves: correlation length (*a*) and fluctuation (*b*); 5-fold cross-validation results (*c*); regularization effects for small dataset training (n = 198) (*d*)

is essential to confirm the practical applicability of our approach.

2. Gaussian correlation assumption: We assumed strictly Gaussian spatial correlations for refractive index fluctuations. Real chromatin organization may exhibit more complex correlation structures, including fractal or power-law correlations which could affect scattering patterns.
3. Limited parameter range: Our study explored a limited range of nuclear parameters (l_c: 0.4–1.0 μm, δn: 0.005–0.035, two morphologies). Real nuclei may exhibit values outside these ranges, and additional morphological variations (e.g., irregular shapes) may be present in biological samples.
4. Single wavelength: We simulated scattering at a single wavelength (800 nm). Multi-wavelength analysis could provide additional information about chromatin organization and improve prediction accuracy.

Despite these limitations, our study provides a proof-of-concept demonstration that ViTs can effectively extract chromatin parameters from 2D optical scattering signals, paving the way for future experimental validation and clinical applications. Additionally, hybrid architectures that combine the local feature extraction strengths of Convolutional Neural Networks with the global modeling capabilities of transformers may offer further gains [27, 28].

Conclusions

This study establishes Vision Transformers (ViTs) as a viable and effective approach for extracting quantitative chromatin parameters from two-dimensional optical scattering patterns. Our main findings can be summarized as follows:

First, the ViT architecture delivers Mean Absolute Percentage Errors of 6.2 % for l_c and 9.8 % for δn, representing clear improvements over comparable Convolutional Neural Network based methods. Second, the self-attention mechanism ability to capture long-range dependencies appears crucial for accurate parameter estimation. Third, the model maintains acceptable performance at realistic noise levels, enhancing its practical prospects. Fourth, the approach transfers well across different nuclear morphologies. Fifth, we demonstrate that careful regularization enables effective training of ViTs even on small datasets (n equals to 198), with learning curves showing no evidence of overfitting. Sixth, we provide detailed comparison with alternative architectures under identical training conditions, with hyperparameter optimization performed for all models.

Although we focused specifically on chromatin characterization, the underlying methodology should extend naturally to other organelles and tissue types. Looking

ahead, experimental validation on real cell samples represents an important next step. The integration of spatial hierarchical attention into standard global transformers presents a highly promising area for future research to further optimize local feature extraction.

References

1. Daneshkhah A., Prabhala S., Viswanathan P., Subramanian H., Lin J., Chang A.S., et al. Early detection of lung cancer using artificial intelligence-enhanced optical nanosensing of chromatin alterations in field carcinogenesis. *Scientific Reports*, 2023, vol. 13, no. 1, pp. 13702 doi: 10.1038/s41598-023-40550-6
2. Rancu A., Chen C.X., Price H., Wax A. Multiscale optical phase fluctuations link disorder strength and fractal dimension of cell structure. *Biophysical Journal*, 2023, vol. 122, no. 7, pp. 1390–1399. doi: 10.1016/j.bpj.2023.03.005
3. Cioffi G., Dannhauser D., Rossi D., Netti P.A., Causa F. Unknown cell class distinction via neural network based scattering snapshot recognition. *Biomedical Optics Express*, 2023, vol. 14, no. 10, pp. 5060–5074. doi: 10.1364/BOE.492028
4. Khalid M., Deivasigamani S., Sathiya V., Rajendran S. An efficient colorectal cancer detection network using atrous convolution with coordinate attention transformer and histopathological images. *Scientific Reports*, 2024, vol. 14, no. 1, pp. 19109. doi: 10.1038/s41598-024-70117-y
5. Liu Y., Wen Y., Yang L., He L., Zhou M. A general framework for efficient medical image analysis via shared attention vision transformer. *IEEE Transactions on Medical Imaging*, 2026, vol. 45, no. 5, pp. 2001–2014. doi: 10.1109/TMI.2025.3644949
6. Photiou C., Kassinopoulos M., Pitris C. Extracting morphological and sub-resolution features from optical coherence tomography images, a review with applications in cancer diagnosis. *Photonics*, 2023, vol. 10, no. 1, pp. 51. doi: 10.3390/photonics10010051
7. Cherkezyan L., Zhang D., Subramanian H., Taflove A., Backman V. Reconstruction of explicit structural properties at the nanoscale via spectroscopic microscopy. *Journal of Biomedical Optics*, 2016, vol. 21, no. 2, pp. 025007. doi: 10.1117/1.JBO.21.2.025007
8. Azad R., Kazerouni A., Heidari M., Aghdam E.K., Molaei A., Jia Y., et al. Advances in medical image analysis with vision Transformers: A comprehensive review. *Medical Image Analysis*, 2024, vol. 91, pp. 103000. doi: 10.1016/j.media.2023.103000
9. He K., Gan C., Li Z., Rekik I., Yin Z., Ji W., et al. Transformers in medical image analysis. *Intelligent Medicine*, 2023, vol. 3, no. 1, pp. 59–78. doi: 10.1016/j.imed.2022.07.002
10. Zhou J., Jin Y., Lu L., Zhou S., Ullah H., Sun J., et al. Deep learning-enabled pixel-super-resolved quantitative phase microscopy from single-shot aliased intensity measurement. *Laser and Photonics Reviews*, 2024, vol. 18, no. 1, pp. 2300488. doi: 10.1002/lpor.202300488
11. Li Z.S., Sun J.S., Fan Y., Jin Y.B., Shen Q., Trusiak M., et al. Deep learning assisted variational Hilbert quantitative phase imaging. *Opto-Electronic Science*, 2023, vol. 2, no. 4, pp. 220023. doi: 10.29026/oes.2023.220023
12. Wang W., Ali N., Ma Y., Dong Z., Zuo C., Gao P. Deep learning-based quantitative phase microscopy. *Frontiers in Physics*, 2023, vol. 11, pp. 1218147. doi: 10.3389/fphy.2023.1218147
13. Khan S., Naseer M., Hayat M., Zamir S.W., Khan F.S., Shah M. Transformers in vision: a survey. *ACM Computing Surveys*, 2021, vol. 54, no. 10, pp. 1–41 doi: 10.1145/3505244
14. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., et al. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv*, 2021. arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929
15. Hatamizadeh A., Nath V., Tang Y., Yang D., Roth H.R., Xu D. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. *Lecture Notes in Computer Science*, 2022, vol. 12962, pp. 272–284. doi: 10.1007/978-3-031-08999-2_22
16. Cai H., Feng X., Yin R., Zhao Y., Guo L., Fan X., et al. MIST: multiple instance learning network based on Swin Transformer for whole slide image classification of colorectal adenomas. *The Journal of Pathology*, 2023, vol. 259, no. 2, pp. 125–135. doi: 10.1002/path.6027
17. Lv Z., Lin Y., Yan R., Wang Y., Zhang F. TransSurv: Transformer-based survival analysis model integrating histopathological images

Литература

1. Daneshkhah A., Prabhala S., Viswanathan P., Subramanian H., Lin J., Chang A.S., et al. Early detection of lung cancer using artificial intelligence-enhanced optical nanosensing of chromatin alterations in field carcinogenesis // Scientific Reports. 2023. V. 13. N 1. P. 13702. doi: 10.1038/s41598-023-40550-6
2. Rancu A., Chen C.X., Price H., Wax A. Multiscale optical phase fluctuations link disorder strength and fractal dimension of cell structure // Biophysical Journal. 2023. V. 122. N 7. P. 1390–1399. doi: 10.1016/j.bpj.2023.03.005
3. Cioffi G., Dannhauser D., Rossi D., Netti P.A., Causa F. Unknown cell class distinction via neural network based scattering snapshot recognition // Biomedical Optics Express. 2023. V. 14. N 10. P. 5060–5074. doi: 10.1364/BOE.492028
4. Khalid M., Deivasigamani S., Sathiya V., Rajendran S. An efficient colorectal cancer detection network using atrous convolution with coordinate attention transformer and histopathological images // Scientific Reports. 2024. V. 14. N 1. P. 19109. doi: 10.1038/s41598-024-70117-y
5. Liu Y., Wen Y., Yang L., He L., Zhou M. A general framework for efficient medical image analysis via shared attention vision transformer // IEEE Transactions on Medical Imaging. 2026. V. 45. N 5. P. 2001–2014. doi: 10.1109/TMI.2025.3644949
6. Photiou C., Kassinopoulos M., Pitris C. Extracting morphological and sub-resolution features from optical coherence tomography images, a review with applications in cancer diagnosis // Photonics. 2023. V. 10. N 1. P. 51. doi: 10.3390/photonics10010051
7. Cherkezyan L., Zhang D., Subramanian H., Taflove A., Backman V. Reconstruction of explicit structural properties at the nanoscale via spectroscopic microscopy // Journal of Biomedical Optics. 2016. V. 21. N 2. P. 025007. doi: 10.1117/1.JBO.21.2.025007
8. Azad R., Kazerouni A., Heidari M., Aghdam E.K., Molaei A., Jia Y., et al. Advances in medical image analysis with vision Transformers: A comprehensive review // Medical Image Analysis. 2024. V. 91. P. 103000. doi: 10.1016/j.media.2023.103000
9. He K., Gan C., Li Z., Rekik I., Yin Z., Ji W., et al. Transformers in medical image analysis // Intelligent Medicine. 2023. V. 3. N 1. P. 59–78. doi: 10.1016/j.imed.2022.07.002
10. Zhou J., Jin Y., Lu L., Zhou S., Ullah H., Sun J., et al. Deep learning-enabled pixel-super-resolved quantitative phase microscopy from single-shot aliased intensity measurement // Laser and Photonics Reviews. 2024. V. 18. N 1. P. 2300488. doi: 10.1002/lpor.202300488
11. Li Z.S., Sun J.S., Fan Y., Jin Y.B., Shen Q., Trusiak M., et al. Deep learning assisted variational Hilbert quantitative phase imaging // Opto-Electronic Science. 2023. V. 2. N 4. P. 220023. doi: 10.29026/oes.2023.220023
12. Wang W., Ali N., Ma Y., Dong Z., Zuo C., Gao P. Deep learning-based quantitative phase microscopy // Frontiers in Physics. 2023. V. 11. P. 1218147. doi: 10.3389/fphy.2023.1218147
13. Khan S., Naseer M., Hayat M., Zamir S.W., Khan F.S., Shah M. Transformers in vision: a survey // ACM Computing Surveys. 2021. V. 54. N 10. P. 1–41 doi: 10.1145/3505244
14. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., et al. An image is worth 16×16 words: Transformers for image recognition at scale // arXiv. 2021. arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929
15. Hatamizadeh A., Nath V., Tang Y., Yang D., Roth H.R., Xu D. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images // Lecture Notes in Computer Science. 2022. V. 12962. P. 272–284. doi: 10.1007/978-3-031-08999-2_22
16. Cai H., Feng X., Yin R., Zhao Y., Guo L., Fan X., et al. MIST: multiple instance learning network based on Swin Transformer for whole slide image classification of colorectal adenomas // The Journal of Pathology. 2023. V. 259. N 2. P. 125–135. doi: 10.1002/path.6027
17. Lv Z., Lin Y., Yan R., Wang Y., Zhang F. TransSurv: Transformer-based survival analysis model integrating histopathological images and genomic data for colorectal cancer // IEEE/ACM Transactions on

and genomic data for colorectal cancer. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2023, vol. 20, no. 6, pp. 3411–3420. doi: 10.1109/tcbb.2022.3199244
18. Halder A., Gharami S., Sadhu P., Singh P.K., Wozniak M., Ijaz M.F. Implementing vision transformer for classifying 2D biomedical images. *Scientific Reports*, 2024, vol. 14, no. 1, pp. 12567. doi: 10.1038/s41598-024-63094-9
19. Manzari O.N., Ahmadabadi H., Kashiani H., Shokouhi S.B., Ayatollahi A. MedViT: a robust Vision Transformer for generalized medical image classification. *Computers in Biology and Medicine*, 2023, vol. 157, pp. 106791. doi: 10.1016/j.compbiomed.2023.106791
20. Guo M.H., Xu T.X., Liu J.J., Liu Z.N., Jiang P.T., Mu T.J., et al. Attention mechanisms in computer vision: A survey. *Computational Visual Media*, 2022, vol. 8, no. 3, pp. 331–368. doi: 10.1007/s41095-022-0271-y
21. Haputhanthri U., Herath K., Hettiarachchi R., Kariyawasam H., Ahmad A., Ahluwalia B.S., et al. Towards ultrafast quantitative phase imaging via differentiable microscopy [Invited]. *Biomedical Optics Express*, 2024, vol. 15, no. 3, pp. 1798–1812. doi: 10.1364/boe.504954
22. Arifler D., Guillaud M. Assessment of internal refractive index profile of stochastically inhomogeneous nuclear models via analysis of two-dimensional optical scattering patterns. *Journal of Biomedical Optics*, 2021, vol. 26, no. 5, pp. 055001. doi: 10.1117/1.jbo.26.5.055001
23. Pu Q., Xi Z., Yin S., Zhao Z., Zhao L. Advantages of transformer and its application for medical image segmentation: a survey. *BioMedical Engineering OnLine*, 2024, vol. 23, no. 1, pp. 14. doi: 10.1186/s12938-024-01212-4
24. Al-Kurdi Y., Direkoglu C., Erbilek M., Arifler D. Convolutional neural network-based regression analysis to predict subnuclear chromatin organization from two-dimensional optical scattering signals. *Journal of Biomedical Optics*, 2024, vol. 29, no. 8, pp. 080502. doi: 10.1117/1.JBO.29.8.080502
25. Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., et al. Swin Transformer: Hierarchical vision transformer using shifted windows. *Proc. of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9992–10002. doi: 10.1109/ICCV48922.2021.00986
26. Han K., Wang Y., Chen H., Chen X., Guo J., Liu Z., et al. A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, vol. 45, no. 1, pp. 87–110. doi: 10.1109/TPAMI.2022.3152247
27. Srinivas A., Lin T.Y., Parmar N., Shlens J., Abbeel P. Vaswani A. Bottleneck transformers for visual recognition. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16514–16524. doi: 10.1109/CVPR46437.2021.01625
28. Xiao T., Singh M., Mintun E., Darrell T., Dollár P., Girshick R. Early convolutions help transformers see better. *Proc. of the 35th Conference on Neural Information Processing Systems*, 2021, vol. 34, pp. 30392–30401.

Computational Biology and Bioinformatics. 2023. V. 20. N 6. P. 3411–3420. doi: 10.1109/tcbb.2022.3199244
18. Halder A., Gharami S., Sadhu P., Singh P.K., Wozniak M., Ijaz M.F. Implementing vision transformer for classifying 2D biomedical images // Scientific Reports. 2024. V. 14. N 1. P. 12567. doi: 10.1038/s41598-024-63094-9
19. Manzari O.N., Ahmadabadi H., Kashiani H., Shokouhi S.B., Ayatollahi A. MedViT: a robust Vision Transformer for generalized medical image classification // Computers in Biology and Medicine. 2023. V. 157. P. 106791. doi: 10.1016/j.compbiomed.2023.106791
20. Guo M.H., Xu T.X., Liu J.J., Liu Z.N., Jiang P.T., Mu T.J., et al. Attention mechanisms in computer vision: A survey // Computational Visual Media. 2022. V. 8. N 3. P. 331–368. doi: 10.1007/s41095-022-0271-y
21. Haputhanthri U., Herath K., Hettiarachchi R., Kariyawasam H., Ahmad A., Ahluwalia B.S., et al. Towards ultrafast quantitative phase imaging via differentiable microscopy [Invited] // Biomedical Optics Express. 2024. V. 15. N 3. P. 1798–1812. doi: 10.1364/boe.504954
22. Arifler D., Guillaud M. Assessment of internal refractive index profile of stochastically inhomogeneous nuclear models via analysis of two-dimensional optical scattering patterns // Journal of Biomedical Optics. 2021. V. 26. N 5. P. 055001. doi: 10.1117/1.jbo.26.5.055001
23. Pu Q., Xi Z., Yin S., Zhao Z., Zhao L. Advantages of transformer and its application for medical image segmentation: a survey // BioMedical Engineering OnLine. 2024. V. 23. N 1. P. 14. doi: 10.1186/s12938-024-01212-4
24. Al-Kurdi Y., Direkoglu C., Erbilek M., Arifler D. Convolutional neural network-based regression analysis to predict subnuclear chromatin organization from two-dimensional optical scattering signals // Journal of Biomedical Optics. 2024. V. 29. N 8. P. 080502. doi: 10.1117/1.JBO.29.8.080502
25. Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., et al. Swin Transformer: Hierarchical vision transformer using shifted windows // Proc. of the IEEE/CVF International Conference on Computer Vision. 2021. P. 9992–10002. doi: 10.1109/ICCV48922.2021.00986
26. Han K., Wang Y., Chen H., Chen X., Guo J., Liu Z., et al. A survey on vision transformer // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023. V. 45. N 1. P. 87–110. doi: 10.1109/TPAMI.2022.3152247
27. Srinivas A., Lin T.Y., Parmar N., Shlens J., Abbeel P. Vaswani A. Bottleneck transformers for visual recognition // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. P. 16514–16524. doi: 10.1109/CVPR46437.2021.01625
28. Xiao T., Singh M., Mintun E., Darrell T., Dollár P., Girshick R. Early convolutions help transformers see better // Proc. of the 35th Conference on Neural Information Processing Systems. 2021. V. 34. P. 30392–30401.

Authors

Salima Azzaz-Rahmani — PhD, Lecturer, Djillali Liabes University of Sidi Bel Abbes, Sidi Bel Abbes, 22000, Algeria, sc 57224954188, https://orcid.org/0009-0003-5217-669X, Azzazsalima2002@yahoo.fr

Hadj Zerrouki — PhD, Lecturer, Djillali Liabes University of Sidi Bel Abbes, Sidi Bel Abbes, 22000, Algeria, sc 56611830500, https://orcid.org/0000-0003-1349-708X, Zerrouki.hadj@gmail.com

Авторы

Аззаз-Рахмани Салима — PhD, преподаватель, Университет Джиллали Лиабеса в Сиди-Бель-Аббесе, Сиди-Бель-Аббес, 22000, Алжир, sc 57224954188, https://orcid.org/0009-0003-5217-669X, Azzazsalima2002@yahoo.fr

Зерруки Хадж — PhD, преподаватель, Университет Джиллали Лиабеса в Сиди-Бель-Аббесе, Сиди-Бель-Аббес, 22000, Алжир, sc 56611830500, https://orcid.org/0000-0003-1349-708X, Zerrouki.hadj@gmail.com

Received 15.03.2026
Approved after reviewing 17.05.2026
Accepted 21.07.2026

Статья поступила в редакцию 15.03.2026
Одобрена после рецензирования 17.05.2026
Принята к печати 21.07.2026