A Survey on ML-Optimized Scaffolds in Tissue Engineering

Daniela Nikoloska
Electrical & Computer Engineering Department
University of Nevada, Las Vegas
nikolosk@unlv.nevada.edu

***Abstract*—Machine learning adoption in 3D bioprinting has reshaped scaffold optimization methods in tissue engineering. This survey tracks the progression of scaffold design, from basic fabrication methods to artificial intelligence applications: bone, skin, and general tissue engineering. Through analysis of mathematical models, cellular simulations, and neural networks, the review evaluates how computational techniques improve the prediction of mechanical properties and biological responses. According to recent studies, supervised learning models achieve F1 scores of 0.96 for printability and 0.88 for scaffold quality metrics, indicating substantial gains in prediction accuracy. Insights from three case studies reveal clear trends in algorithm efficacy: XGBoost and gradient boosting consistently excel across tissue types. For bone tissue, non-linear models provide 99% accuracy in geometry classification with R^2 exceeding 0.86, while AdaBoost achieves 98.58–99.6% accuracy in skin tissue datasets. Despite the availability of extensive datasets with over 1,000 scaffold designs, challenges persist in mechanical performance and data accessibility. Research indicates only 25% of studies provide public data access, revealing a need for standardized sharing protocols. Looking ahead, 4D printing with responsive biomaterials and volumetric fabrication methods suggest promising paths for machine learning integration in scaffold development.**

***Index Terms*—3D bioprinting, artificial intelligence, computer-aided design, machine learning, regenerative medicine, scaffolds, tissue engineering**

I. Introduction

Organ failure and tissue loss present formidable medical hurdles that dramatically affect patient survival rates and life quality. It's a stark clinical reality – organ transplantation, while theoretically promising, falls devastatingly short of demand. Due to the shortage of the organ pool and restricted access to the transplant waiting list, more than 100,000 people die in the US each year. [1]. These limitations, paired with the persistent challenge of immune rejection, have initiated innovations in tissue engineering. Then comes 3D bioprinting, a development that has entirely altered the way scaffolds are made. Yet optimization proves elusive. The calibration of bioprinting parameters remains complex, with interdependent variables requiring precise control. Material compositions demonstrate non-linear behaviors. Cell densities convey temporal variations. Printing conditions demand continuous refinement.

In this work, how machine learning techniques manage multifarious challenges is reviewed. The analysis spans recent developments, from fundamental predictive models to advanced neural networks. The paper addresses gaps in current methodologies, offering fresh perspectives on parameter optimization. The data reveals compelling patterns. Machine learning algorithms demonstrate unprecedented capabilities in scaffold design prediction.

This paper examines machine learning methodologies in bone, skin, and general tissue engineering and evaluates their efficacy in addressing the various challenges of scaffold optimization.

II. Evolution of Scaffold Design

a. Traditional Approaches

The genesis of scaffold engineering presented remarkable simplicity: fundamental support matrices for cellular growth. [2] Such straightforward beginnings, while foundational, revealed substantial limitations. Pore design control, weak mechanical properties, and limited cell infiltration helped early studies. [3] During the 1990s and 2000s, the field evolved quickly through the development of various fabrication techniques such as freeze-drying, textile techniques, and particulate leaching. [4] Federal investments exceeding $4 billion over two decades fueled this progress. [5]

b. Material and Cellular Integration

Since scaffolds for tissue engineering are ultimately destined for implantation in the human body, they must meet strict biological safety criteria and compatibility standards. They must be free of toxic, carcinogenic, or mutagenic properties to avoid adverse reactions while exhibiting excellent biocompatibility with cells and tissues. [4] Synthetic polymers and natural polymers represent the primary materials used for scaffold construction. Composite scaffolds, blending these materials, are becoming more prevalent in addressing the limitations of individual types. [3]

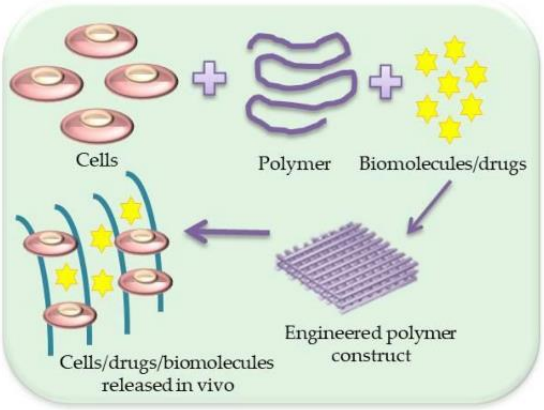

Fig 1. Schematic Representation of Tissue Engineering Components [6]

Fig 1. exemplifies the components and their integration into functional constructs. The integration process combines three essential elements: cells, structural polymers, and bioactive molecules. Synthetic polymers offer mechanical strength and controlled degradation rates. Popular options include polylactic acid (PLA), polyglycolic acid (PGA), and their copolymers, which break down into natural metabolic products [4]. Natural polymers complement synthetic materials by providing biological recognition signals. Collagen mimics native extracellular matrix components, facilitating cell attachment. [3] Hyaluronic acid supports cell migration and tissue hydration. [6] Chitosan adds antimicrobial properties while promoting wound healing. These biomolecules create an environment that supports cellular function and tissue development.

As shown in Fig. 1, the engineered construct releases its cellular and molecular cargo in controlled phases. Matrix degradation rates match tissue formation speeds. Bioactive molecules—growth factors, cytokines, and small molecules—exit the structure through diffusion or matric breakdown. This temporal control guides tissue development.

Integration challenges persist. Achieving uniform cell distribution requires meticulous consideration of pore architecture, and maintaining consistent material properties across large structures demands precise process control. Traditional fabrication methods struggled to meet these demanding specifications, pushing the industry toward computer-aided techniques.

c. Transition to Computer-Aided Design

Between 2000-2010, scaffold fabrication transitioned from conventional methods to computer-aided (CAD) manufacturing, driven by the need for more precise control over internal architecture. Solid freeform fabrication (SFF) technologies emerged, using layer-by-layer manufacturing strategies guided by CAD models. [7]

Three main additive manufacturing methods transpired – photo resin-based, nozzle-based, and powder-based- [4]. This understanding led to the establishment of fundamental design principles. Contemporary scaffold design follows the 4F principle – form, function, formation, and fixation [4].

Photo resin-based techniques, such as stereolithography, employ UV light or lasers to selectively cure liquid photopolymers layer by layer in a resin vat. [8] This approach achieves feature resolution at the 100-micrometer scale. Nozzle-based systems deposit materials through controlled extrusion, melting polymers for fused deposition or dispensing cell-containing bio inks at physiological temperatures. [4] Precise pressure and temperature regulation ensure consistent material deposition and maintain cellular viability. Multi-material printing capabilities enable the creation of gradient structures mimicking natural tissue interfaces.

Powder-based methods, illustrated in Fig 2, demonstrate exceptional spatial control. Print systems deposit binding agents onto powder substrates, while alternative approaches employ laser sintering for particle fusion. The process repeats as the build platform descends incrementally. Powder composition ranges from biodegradable polymers to bioactive ceramics. [8]

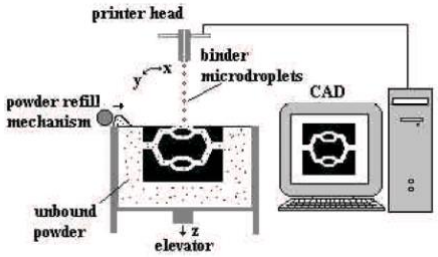

Fig 2. Schematic Diagram of the 3D Printing System [E.Sachlos]

The convergence of these manufacturing methods and computational design marked a defining shift in tissue engineering. Optimization, however, proved complex—interactions between parameters triggered ripple effects, where even small changes led to unpredictable results. Machine learning rendered the answer, providing a means to manage multiple variables concurrently.

III. Core ML Concepts

The following are the most favorable ML models/algorithms for tissue engineering. [9] The first three are traditional mathematical models:

Traditional Mathematical Models

a. Finite Element Analysis (FEA)

A numerical technique for simulating deformation and mechanical properties under load, this method subdivides complex structures into more minor elements, enabling precise calculations of stress distribution and mechanical behavior. [9]

b. Cellular Automata Models

A mathematical framework simulating the behavior and migration of cells and how tissue grows. These models incorporate fundamental biological processes through defined rule sets, predicting developmental patterns by analyzing cell-cell interactions and environmental responses.

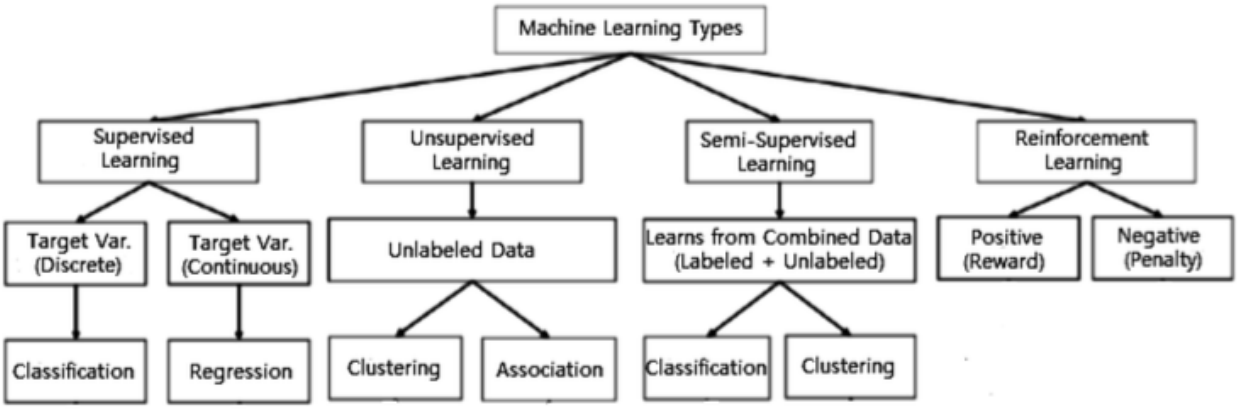

Fig 3. Four Categories of Machine Learning Techniques [Iqbal Sarker]

c. Partial Differential Equations (PDEs)

Mathematical models simulating the distribution and exchange of biochemical factors within tissue. These equations characterize how molecules move and interact, enabling precise predictions of concentration gradients and transport phenomena.

The remaining machine learning models mentioned can be categorized into four categories based on how signals and feedback are processed: supervised ML, unsupervised ML, semi-supervised ML, and reinforcement ML. [6] In Fig. 3, evident pathways appear – supervised learning splits into classification and regression based on target variable types, while unsupervised learning focuses on pattern discovery through clustering and association. [10] Deep learning architectures operate across multiple categories, spanning supervised classification, unsupervised feature extraction, and reinforcement learning implementations. Classification and clustering methods also cross categorical boundaries – classification functions in both supervised and semi-supervised domains, while clustering applies to both unsupervised and semi-supervised approaches.

Supervised Learning

Pattern recognition algorithms are trained on labeled datasets. These establish predictive relationships between input parameters and biological outcomes.

a. Neural Networks

Computational architectures that mirror biological neural processing. These systems stack multiple layers to decode complex parameter relationships, enabling pattern recognition across diverse biological datasets. The processing power grows with architectural depth.

b. Supper Vector Machines (SVM)

Classification systems that optimize hyperplanes in n-dimensional feature spaces. SVMs strength lies in handling high-dimensional data with clear decision boundaries.

c. Decision Trees

Hierarchical algorithms that refine parameters through sequential binary decisions. Branching logic structures that optimize processes by incorporating quantitative measurements and qualitative criteria are created.

d. Random Forests

Algorithms that combine multiple decision trees to improve prediction accuracy. These methods leverage the power of numerous trees operating in parallel, each analyzing different subsets of data and features.

e. K-Nearest Neighbors (KNN)

Classifying data involves analyzing patterns relative to known reference points. KNN determines outcomes by identifying the closest matches within the training set. This method excels at handling non-linear relationships in biological data.

f. Naïve Bayes

Probabilistic algorithms applying Bayesian statistics to classification tasks. These methods calculate the likelihood of outcomes based on input features, assuming independence between variables.

g. Regression

A modeling technique for predicting continuous numerical outcomes with high precision. These algorithms establish mathematical correlations between input variables and quantitative metrics.

h. Classification

A categorical prediction method that assigns outcomes to distinct groups based on input features. These algorithms simultaneously analyze multiple input features, creating clear boundaries between categories.

Unsupervised Learning

Algorithms that discover hidden patterns without predefined labels.

a. Clustering

Grouping similar data points based on inherent patterns and relationships. Such methods identify natural categories within the dataset, revealing underlying structures that are not immediately apparent.

b. Dimensionality Reduction

Methods that compress high-dimensional data while preserving essential information. These methods transform complex datasets into more straightforward representations.

Semi-Supervised Learning

Hybrid approaches combining labeled and unlabeled data analysis. These methods leverage small amounts of labeled data to guide the interpretation of larger unlabeled datasets,

offering a practical balance between supervised and unsupervised techniques. [10]

Reinforcement Learning

Dynamic algorithms that learn optimal strategies through environmental interaction continuously adjust parameters based on outcome measurements.

a. Agent-Based Models

Computational entities that learn through action-consequence relationships. These models make decisions based on defined reward systems, adapting their behavior to maximize positive outcomes.

I. Case Studies

Predicting the Quality of 3D Bioprinted Scaffolds

a. Dataset Design and Characteristics

Rafieyan et al.'s study in 2024 reshaped tissue engineering by introducing an unprecedented open-source dataset encompassing 1,171 scaffolds. The input parameters contain 60 biomaterial compositions, including natural polymers (alginate, gelatin, GelMA), synthetic polymers (PCL, PLGA), crosslinkers (CaCl2, Irgacure 2959), and cell-related parameters across 49 cell lines. The cell populations, including six co-cultured systems, introduced biological variables. [11]

The dataset, featuring high-dimensional data and significant sparsity, limits the applicability of traditional mathematical models. As a result, techniques such as dimensionality reduction, regularization, or machine learning algorithms are necessary. Each scaffold entry contains material concentrations (w/v%), crosslinking parameters (duration, type), printing conditions (pressure, temperature, nozzle specifications), and biological outcomes. Material concentrations range from 0.25-20 w/v% for alginate, 1-50 w/v% for gelatin, and 3-30 w/v% for GelMA.

b. Scaffold Quality Assessment Methodology

The authors established a systematic quantification method using two primary metrics: cell response (1-5 scale) and printability (0-3 scale). These metrics combine multiplicatively to generate a scaffold quality score, as shown in the equation below.

$$\text{Scaffold Quality} = \text{Cell Response} \times \text{Printability}.$$

c. Preprocessing and Data Handling

The preprocessing pipeline employs XGBoost regression to calculate missing values. The authors implemented SMOTE oversampling to address the class imbalance, which is particularly important given the uneven distribution of scaffold quality scores. The authors employed standardized Euclidean distance calculations in the normalized 75-dimensional feature space. The optimal number of clusters (k=5) appeared from a systematic evaluation using three complementary metrics: Silhouette Score for cluster cohesion and separation, Davies-Bouldin Score for cluster compactness, and Calinski-Harabasz Score for intercluster variance. Centroid initialization utilized the k-means++ algorithm to ensure optimal cluster distribution.

The neural network implementation utilized a symmetric architecture:

(input → 128 → 256 → 512 → 256 → 128 → output)

with dropout regularization and adaptive learning rates.

d. Model Performance

The comparative performance analysis of 29 machine learning algorithms revealed patterns in predictive capability, as shown in Fig. 4. XGBoost and LightGBM demonstrated dominance in algorithmic performance, achieving precision scores of 0.88 and 0.87, respectively, for scaffold quality prediction. Most notably, GradientBoostingClassifier maintained consistent scores above 0.85 across all evaluation metrics – precision, recall, accuracy, and F1-score.

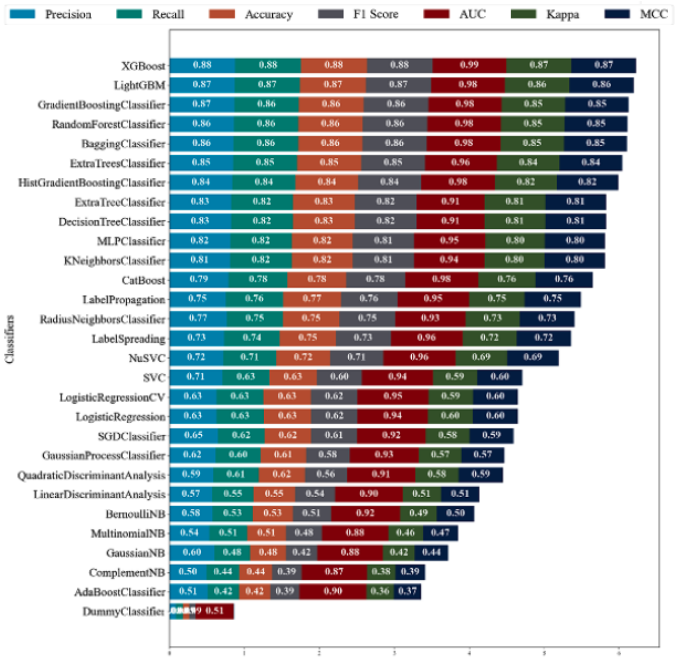

Fig 4. Comparative Analysis of 29 ML Algorithms for Scaffold Quality Prediction [11]

K-means clustering (k=5) revealed material-property relationships. Critical hyperparameters varied by prediction task: dropout rate and optimizer selection dominated scaffold quality prediction, while learning rate influenced printability outcomes. The neural network achieved F1 scores of 0.80, 0.79, and 0.93 for scaffold quality, cell response, and printability, respectively.

The neural network architecture required specific optimizations for high-dimensional biomaterial data processing. Hyperparameters surfaced from Optuna-based optimization: dropout rates between 0.1-0.3 prevented overfitting, particularly decisive given the limited dataset size relative to feature dimensionality. Batch normalization between dense layers facilitated stable training across the deep architecture. The learning rate schedule, implemented through Adam optimizer with initial rates between 10^{-3} and 10^{-4}, balanced convergence stability with model plasticity.

The dimensionality reduction route incorporated principal component analysis followed by t-SNE visualization, preserving 97.3% of variance while enabling interpretable 2D projections. Data augmentation through SMOTE yielded synthetic samples with realistic parameter combinations. The

resampling ratio optimization revealed a 75% synthetic data inclusion.

The hyperparameter search space spanned 156 distinct combinations across learning rate schedules, architecture depths, and regularization strategies. Bayesian optimization with Gaussian processes converged rapidly, identifying optimal architecture search revealed that skip connections between symmetric layers

(128→512, 512→128)

improved gradient flow while preserving spatial hierarchy in the learned scaffold representations.

Tree-based models demonstrated complementarity: XGBoost captured non-linear interactions between material parameters (partial dependence plots revealed threshold effects in crosslinker concentrations). In contrast, LightGBM revealed exceptional handling of categorical features like cell types.

e. Results

Through rigorous receiver operating characteristic (ROC) curve analysis, model performance evaluation reveals distinctive patterns across multiple algorithmic frameworks. The validation results demonstrate nuanced performance characteristics across three critical prediction tasks: scaffold quality, cell response, and printability assessment.

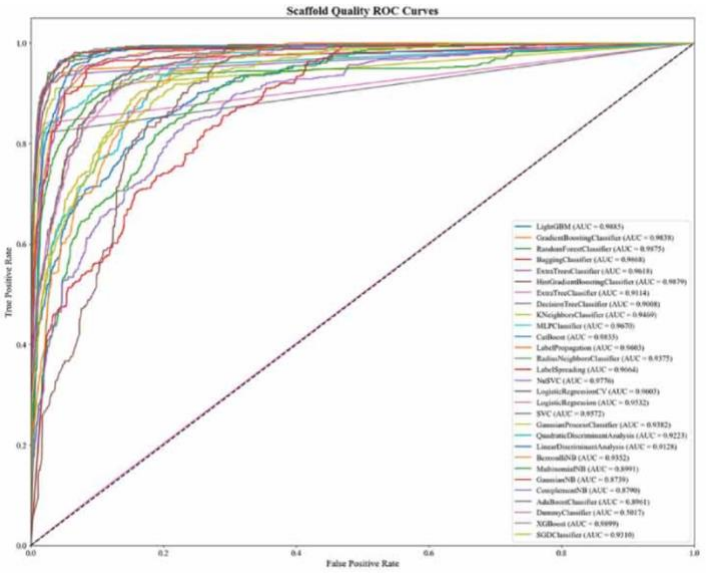

Fig 5. Comparative Analysis of 29 ML Algorithms for Scaffold Quality Prediction [11]

Fig. 5 provides quantitative validation through ROC curve analysis, revealing the superior discriminative capabilities of ensemble-based methods. The area under the curve (AUC) metrics demonstrate powerful performance in printability prediction, with several algorithms achieving AUC values exceeding 0.95. The scaffold quality prediction tasks exhibit more varied performance characteristics, with AUC values ranging from 0.82 to 0.97, indicating algorithm-specific sensitivity to feature complexity.

The ROC analysis stratifies algorithmic performance across three distinct tiers, each characterized by specific discriminative capabilities and robustness profiles. In the high-performance tier, algorithms achieving AUC values exceeding 0.95 demonstrate exceptional predictive power. XGBoost, LightGBM, and Gradient Boosting variants exhibit advanced discriminative capabilities, especially in printability prediction tasks. These algorithms maintain consistent performance across heterogeneous data distributions, suggesting robust generalization capabilities across scaffold characteristics.

The intermediate performance tier, characterized by AUC values between 0.90 and 0.95, comprises algorithms such as Random Forest and Extra Trees Classifier. While these methods demonstrate strong discriminative power, their performance exhibits slight reductions compared to the high-performance tier. Nevertheless, they maintain notable consistency across multiple prediction tasks and demonstrate particular resilience to class imbalance challenges.

The baseline performance tier, containing algorithms with AUC values between 0.85 and 0.90, integrates traditional approaches such as Decision Trees and k-NN classifiers. While these methods provide acceptable discrimination capabilities, they exhibit reduced sensitivity compared to more complex ensemble approaches. Their performance characteristics reveal greater susceptibility to variations in data distribution, suggesting limitations in generalizability across diverse scaffold configurations.

Bone Tissue Engineering Scaffolds

a. Dataset Design and Characteristics

Ibrahim et al.'s study in 2024 leveraged a dataset encompassing 1,279 Triply Periodic Minimal Surace (TPMS) scaffolds. The parameter space spans two critical variables: offset parameter b (0-2, $\Delta b=0.1$) and thickness parameter δ (0.4-1.5, $\Delta\delta=0.05$). The architectural distribution comprises Diamond (n=321), Gyroid (n=475), and IWP (n=483) topologies. Each scaffold undergoes multi-parametric characterization, generating a 37-dimensional feature vector. The feature space incorporates quantitative descriptors across mechanical (normalized elastic moduli, shear coefficients), morphological (volumetric fraction ρ^*, trabecular parameters), and transport (effective pore connectivity) domains. A resolution of 300 voxels per edge was applied in spatial discretization across three repeating units, ensuring efficient computation and accurate microstructure representation. [12]

b. Scaffold Quality Assessment Methodology

Quality evaluation implements a tripartite analytical framework. Mechanical characterization employs voxel-based finite element analysis through a ParOSol parallel solver, computing the full elastic tensor under confined tension and pure shear conditions. The strength assessment incorporates probabilistic mechanics, utilizing Weibull statistics (m=6.6) to predict failure indices at 95% confidence thresholds.

Morphometric analysis deploys the BoneJ algorithmic suite, extracting critical parameters such as trabecular thickness (Tb. Th), spacing (Tb. Sp), connectivity density, and anisotropy tensors. Physical transport characterization implements a novel permeability index (EPCI), quantifying three-dimensional pore networks through image-based analysis under saturated conditions. Integrating these metrics provides a comprehensive quality assessment framework spanning multiple physical domains.

c. Preprocessing and Data Handling

Geometric validation excludes architectures exhibiting topological discontinuities (pinching surfaces) or incomplete

connectivity. Feature extraction standardizes measurements through pixel-wise operations, maintaining consistent spatial resolution (300 voxels/edge). The normalized dataset undergoes correlation analysis, eliminating features with Pearson coefficients >0.6 to minimize collinearity.

The Greedy Feedforward Feature Selection algorithm identifies optimal feature subsets and evaluates performance through geometric mean R-squared metrics on validation data. This process culminates in selecting five critical features, balancing predictive power with practical accessibility. The framework demonstrates particular sensitivity to mechanical property normalization, utilizing reference elastic modulus (E_0=1 GPa) and Poisson ratio (ν=0.3) for standardization.

d. Model Performance

The machine learning framework demonstrates distinctive performance characteristics across three architectural paradigms: linear (L-MOD), non-linear (NL-MOD), and constrained non-linear (NL-MOD-CON) models. Performance metrics reveal R^2 values ranging from 0.62 to 0.86 for parameter b prediction, with notably lower values (0.32-0.64) for parameter δ estimation. The non-linear model performs better, achieving median prediction errors below 3% and determination coefficients exceeding 0.89 for selected features.

Cross-validation analysis reveals robust geometry classification accuracy (≥90%) across all models. The NL-MOD with Data-Driven Feature Selection (DDFS) achieves peak accuracy (99%) while maintaining computational efficiency. Feature importance analysis identifies five critical parameters sufficient for optimal prediction, with diminishing returns observed beyond this threshold. The model demonstrates particular sensitivity to mechanical properties, specifically normalized stiffness ($\bar{E}$) and strength index ($I\sigma$).

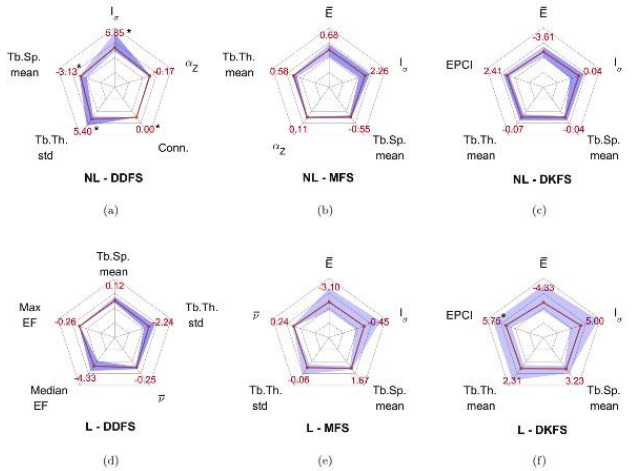

Fig 6. Spider plots representing median, 25th, and 75th percentiles of percentage errors for each model and feature selection approach [12]

As shown in Fig. 6, the radial arrangement effectively exemplifies the comparative performance between linear and nonlinear models. Each axis represents a distinct feature spanning the -50% to +50% error range, providing immediate visual insight into prediction accuracy and variability. The plot reveals that the nonlinear model consistently achieves tighter error distributions, particularly in mechanical property predictions. This visualization demonstrates the exceptional performance of the NL-MOD architecture across multiple feature domains.

e. Results

The experimental validation, conducted across 50 randomly selected test scaffolds, reveals compelling insights into model performance and practical applicability. The non-linear model (NL-MOD) with Data-Driven Feature Selection demonstrates exceptional geometry identification accuracy, achieving 100% correct classification across all test cases. This markedly outperforms the linear model (L-MOD), which achieves 88% accuracy in geometry classification. Notably, IWP structures consistently exhibit the highest classification reliability across all model variants, suggesting inherent geometric characteristics that facilitate accurate identification.

The predictive accuracy for mechanical and morphological features varies significantly across different model architectures. The NL-MOD implementation with Mixed Feature Selection yields median percentage errors that remain statistically insignificant from zero, indicating robust predictive capability. Wall thickness predictions maintain exceptionally high accuracy, falling within manufacturing tolerances of ±10%, a critical factor for practical fabrication considerations. The mechanical property predictions exhibit the highest fidelity for normalized stiffness parameters, with correlation coefficients exceeding 0.89 for selected features.

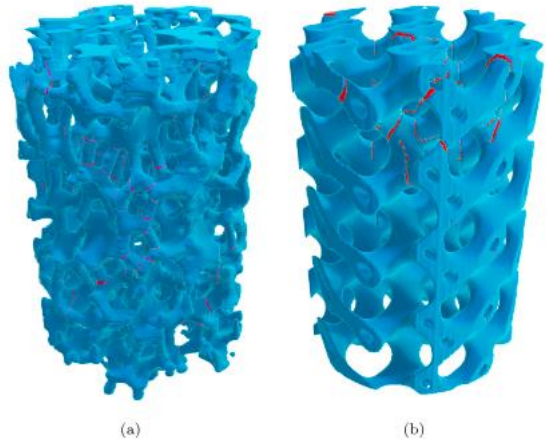

Fig 7. Finite element analysis comparing fracture patterns between (a) trabecular-like scaffold and (b) TPMS gyroid scaffold under uniaxial compression [12]

Finite element analysis of the scaffold structures reveals distinctive mechanical behavior patterns under compression loading conditions. Fig. 7 illustrates that fracture propagation mechanisms differ substantially between traditional trabecular-like scaffolds and TPMS structures. The trabecular scaffolds exhibit complete cross-sectional fractures near structural interconnections, indicated by the purple regions of damage concentration. In contrast, the TPMS gyroid scaffolds demonstrate a more distributed pattern of partial fractures, shown in red, spanning multiple locations across the structure. This variation in failure mechanisms explains the superior strength characteristics observed in TPMS structures despite comparable porosity levels. The finite element results corroborate the model's predictions regarding mechanical performance and validate its utility in scaffold design optimization.

The model's practical application to hydroxyapatite scaffold design demonstrates its clinical relevance and manufacturing feasibility. The algorithm successfully generated designs that achieved target mechanical properties within 10% of specified values while maintaining wall thickness predictions within the capabilities of VAT photopolymerization manufacturing processes. This validation confirms the model's potential for deployment in tissue engineering practices.

Skin Tissue Engineering Scaffolds

a. Dataset Design and Characteristics

Salem et al. 's experimental framework contains a dataset of 182 nanofibrous scaffold observations characterized through four pivotal numerical features: water contact angle, fiber diameter, Young's modulus, and pore diameter. [13] This multi-parametric characterization establishes a robust foundation for machine learning model development. The dataset's architecture integrates physicochemical parameters that exhibit complex interdependencies. The correlation analysis notably demonstrated this, which revealed significant relationships between contact angle and Young's modulus (r = +0.425) and fiber diameter with pore diameter (r = -0.206).

The feature space demonstrates notable distributional characteristics requiring sophisticated scaling approaches. Each scaffold entry contains material-specific attributes that span multiple orders of magnitude, necessitating careful normalization protocols. The water contact angle measurements capture surface wettability variations, while fiber diameter distributions reflect the architectural heterogeneity inherent in electrospun constructs.

b. Scaffold Quality Assessment Methodology

Quality evaluation implements a multi-tiered analytical framework integrating clustering techniques with regression modeling. The assessment protocol employs k-means clustering to partition the feature space into three distinct clusters, enabling more nuanced performance analysis. Silhouette scoring quantifies clustering efficacy, revealing intricate patterns in data point distribution and cluster cohesion. The methodology incorporates supervised and unsupervised learning approaches—decision trees, random forests, AdaBoost, and gradient boosting algorithms evaluate scaffold characteristics with varying degrees of sophistication.

The assessment framework is susceptible to feature importance, with fiber diameter emerging as the dominant predictor (R^2 score = 0.78). This hierarchical feature evaluation enables targeted optimization of scaffold design parameters while maintaining manufacturing feasibility constraints. Integrating Manhattan distance metrics in clustering analysis provides robust scaffold classification capabilities.

c. Preprocessing and Data Handling

The data preprocessing pipeline implements sophisticated cleaning and normalization protocols through the MinMaxScaler technique, ensuring consistent feature scaling across heterogeneous parameter spaces. Feature selection employs a rigorous correlation analysis using Pearson coefficients, identifying and mitigating multicollinearity effects among predictive variables. The framework demonstrates particular sensitivity to outlier detection and handling, implementing statistical filters for low-variance and high-correlation scenarios.

Cross-validation protocols utilize a 5-fold methodology, ensuring robust model performance assessment across different data partitions. The preprocessing workflow incorporates automated feature importance analysis, facilitating dynamic feature selection based on predictive power. This systematic approach enables consistent cross-comparison of scaffold architectures while maintaining the physical interpretability of the extracted features. The data handling protocol's sophistication ensures reliable model training while preserving the intrinsic characteristics of the nanofibrous scaffolds.

d. Model Performance

The machine learning framework demonstrates distinctive performance characteristics across multiple algorithmic paradigms. The AdaBoost (AB) algorithm exhibits superior predictive capability, achieving remarkable accuracy metrics across all dataset configurations: 98.58% for the whole dataset (WD), 99.6% for Cluster 1, 98.51% for Cluster 2, and 98.85% for Cluster 3. This exceptional performance is further validated by consistently high R^2 values (0.999-1.000), indicating robust model generalization capabilities.

The gradient boosting (GB) implementation demonstrates competitive performance, particularly in Cluster 3 analysis, achieving 99.97% accuracy with an R^2 value of 1.000. However, decision tree (DT) and random forest (RF) algorithms exhibit notably lower performance metrics, with 60.05-81% accuracy ranges and 45.72-66.185%, respectively. This performance disparity suggests inherent limitations in these traditional ensemble methods when handling complex nanofibrous scaffold characterization data.

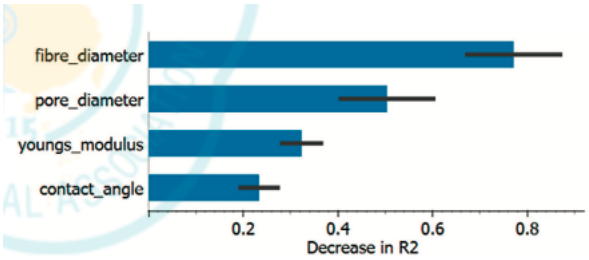

Fig 8. Feature importance graph computed using AdaBoost, ranking scaffold characteristics based on their relative contribution to model performance, with fiber diameter emerging as the dominant predictor (R^2 = 0.78) [13]

Fig. 8 visualizes the feature importance analysis, which reveals critical insights into predictor hierarchy. Fiber diameter demonstrates paramount importance, with an R^2 score of 0.78, while Young's modulus and water contact angle exhibit lower significance (R^2 = 0.32 and 0.22, respectively). This visualization effectively captures the hierarchical relationship between physicochemical parameters and scaffold performance, guiding optimization strategies.

e. Results

Cross-validation analysis reveals exceptional model stability across different data partitions. The clustering approach demonstrates particular efficacy in handling dataset heterogeneity, with silhouette analysis confirming optimal cluster separation. The AB algorithm maintains consistent performance across all clusters, suggesting robust

generalization capabilities independent of data partitioning strategies.

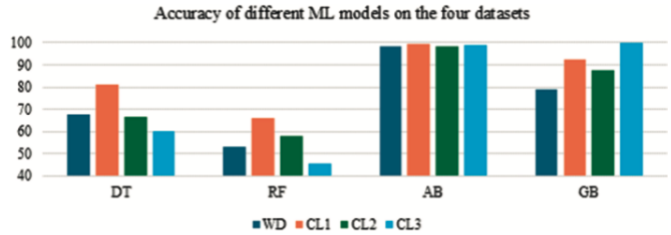

Fig 9. Performance comparison of machine learning models across clustered datasets (WD, CL1, CL2, CL3), demonstrating superior accuracy of AdaBoost and gradient boosting algorithms relative to traditional approaches [13]

Fig. 9 provides compelling evidence of the clustering strategy's impact on model performance. The visualization demonstrates how data partitioning significantly influences algorithm effectiveness, with AB and GB maintaining consistently high accuracy across all clusters. This performance improvement is particularly pronounced in Cluster 1 and Cluster 2, where both algorithms achieve accuracy exceeding 98%, substantially outperforming traditional DT and RF implementations.

The comprehensive analysis reveals that clustering-based model optimization significantly enhances predictive capabilities. The AB algorithm's superior performance across all evaluation metrics establishes it as the optimal choice for nanofibrous scaffold characterization, achieving a mean absolute percentage error of 1.41% across the entire dataset. This remarkable accuracy, coupled with consistent performance across different data partitions, validates the robustness of the proposed machine learning framework for scaffold design optimization.

II. Challenges

The application of machine learning (ML) in scaffold optimization faces four rudimentary challenges: (1) computational complexity in high-dimensional parameter spaces, (2) mechanical performance, (3) data availability, and (4) reproducibility concerns. These challenges intersect across multiple scales and stages of scaffold development, from initial design to final fabrication.

High-dimensional parameter spaces - encompassing pore architectures, material properties, and cellular responses - create computational complexity that traditional modeling needs help to handle. [6] Salem et al. identify how multiple physicochemical properties exhibit complex interdependencies. Correlation analyses reveal relationships between mechanical and structural features (e.g., Young's modulus and contact angle, r = +0.425). These intricate parameter relationships complicate feature selection and model optimization processes. [13]

Research by Drakoulas et al. demonstrates how strut bending and creeping phenomena fundamentally alter pore morphology and volume, compromising structural integrity in ways that exceed current predictive capabilities. [14]

Data availability presents a quantifiable limitation to advancement. Statistical analysis reveals that only 25% of bioprinting studies make their datasets publicly accessible, significantly impeding AI model development. [11]

Reproducibility concerns pose substantial barriers to clinical translation. Many published ML models need more documentation of training procedures and hyperparameters. The absence of standardized validation protocols and limited access to proprietary datasets hampers independent verification of results. Furthermore, laboratory variations in experimental conditions and measurement techniques introduce systematic biases affecting model generalizability. [2]

III. Future Perspectives

The evolution of scaffold fabrication technology spans multiple complementary innovations. Building on 3D printing, 4D printing enhances scaffolds by integrating innovative biomaterials, enabling 3D structures to respond to moisture, temperature, or pH stimuli. These properties allow for functions like scaffold degradation in response to tissue growth, aiding tissue regeneration and eventual restoration.

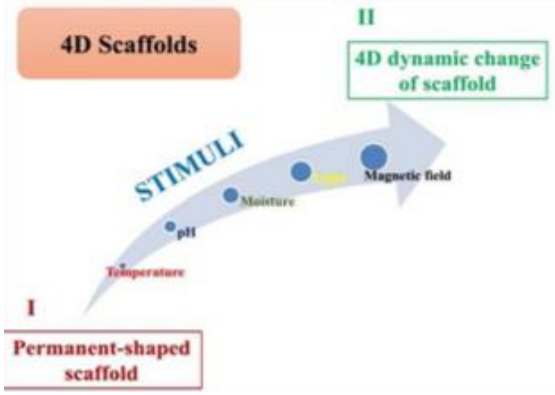

Fig 10. Schematic representation of 4D scaffold dynamics illustrating the transition from permanent-shaped scaffolds (I) to dynamically responsive structures (II) [2]

Fig. 10 illustrates the development from static 3D structures to dynamic 4D systems. In stage I, the permanent-shaped scaffold represents the traditional structural foundation. Stage II depicts the scaffold's dynamic responsiveness to multiple environmental triggers - a defining characteristic of 4D biomaterials.

This advancement couples naturally with patient customization through improved medical imaging integration. Zaszczynska et al. demonstrate how patient imaging data can be translated into bespoke scaffold architectures that precisely match individual anatomical requirements [8]. Meanwhile, volumetric fabrication techniques utilizing holographic methods have accelerated production speeds by eliminating the need for traditional layer-by-layer processing [4]

These converging technological developments - responsive materials, personalized design, and rapid fabrication - create a promising framework for next-generation tissue engineering applications.

IV. Conclusion

In this work, machine learning techniques were examined transform scaffold optimization across bone, skin, and general tissue engineering applications. Our analysis reveals that ML

implementations achieve exceptional metrics - F1 scores reach 0.96 for printability assessment and 0.88 for quality metrics. XGBoost and gradient boosting algorithms demonstrate consistently exceptional performance.

REFERENCES

[1] C. Park *et al.*, “A scoping review of inequities in access to organ transplant in the United States,” *International Journal for Equity in Health*, vol. 21, no. 1, Feb. 2022, doi: https://doi.org/10.1186/s12939-021-01616-x.

[2] I. M. Adel, M. F. ElMeligy, and N. A. Elkasabgy, "Conventional and recent trends of scaffold fabrication: A superior mode for tissue engineering," *Pharmaceutics*, vol. 14, no. 2, p. 306, Jan. 2022. [Online]. Available: https://doi.org/10.3390/pharmaceutics14020306

[3] O'Brien, F. J. (2011). Biomaterials & scaffolds for tissue engineering. Materials Today, 14(3), 88-95.

[4] D. W. Hutmacher, B. Tandon, and P. D. Dalton, "Scaffold design and fabrication," in *Tissue Engineering*, 3rd ed., ch. 11, Elsevier, 2023, pp. 355–385.

[5] S. J. Hollister, "Scaffold engineering: A bridge to where?" *Fabrication*, vol. 1, no. 1, pp. 012001, Mar. 2009, doi: 10.1088/1758-5082/1/1/012001.

[6] J. Shin, Y. Lee, Z. Li, J. Hu, S. S. Park, and K. Kim, "Optimized 3D bioprinting technology based on machine learning: A review of recent trends and advances," *Micromachines*, vol. 13, no. 3, p. 363, 2022. [Online]. Available: https://doi.org/10.3390/mi13030363

[7] E. Sachlos and J. T. Czernuszka, "Making tissue engineering scaffolds work. Review: the application of solid freeform fabrication technology to the production of tissue engineering scaffolds," *Eur. Cell Mater.*, vol. 5, pp. 29–39; discussion 39–40, Jun. 2003, doi: 10.22203/ecm.v005a03.

[8] A. Zaszczynska, M. Moczulska-Heljak, A. Gradys and P. Sajkiewicz, "Advances in 3D Printing for Tissue Engineering," Materials, vol. 14, no. 12, p. 3149, 2021.

[9] M. Gharibshahian, M. Torkashvand, M. Bavisi, N. Aldaghi, and A. Alizadeh, "Recent advances in artificial intelligent strategies for tissue engineering and regenerative medicine," *Skin Research and Technology*, vol. 30, no. 9, Art. no. e70016, Aug. 2024. doi: 10.1111/srt.70016.Med. Biol. Soc. (EMBC), Mexico, 2021, pp. 1230–1233, doi: 10.1109/EMBC46164.2021.9630587

[10] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Computer Science*, vol. 2, no. 3, pp. 1–21, Mar. 2021, doi: https://doi.org/10.1007/s42979-021-00592-x.

[11] S. Rafieyan, E. Ansari, and E. Vasheghani-Farahani, "A practical machine learning approach for predicting the quality of 3D (bio)printed scaffolds," *Biofabrication*, vol. 16, no. 4, p. 045014, July 2024. [Online]. Available: https://doi.org/10.1088/1758-5090/ad6374

[12] S. Ibrahimi, Luca D’Andrea, D. Gastaldi, M. W. Rivolta, and P. Vena, “Machine Learning approaches for the design of biomechanically compatible bone tissue engineering scaffolds,” *Computer Methods in Applied Mechanics and Engineering*, vol. 423, pp. 116842–116842, Feb. 2024, doi: https://doi.org/10.1016/j.cma.2024.116842.

[13] D. A. Salem, M. H. Moharam, and E. M. Hashem, “Development of Machine Learning Regression Models for Predicting the Performance of Nanofibrous Scaffolds for Skin Tissue Engineering,” *Journal of Bio-X Research*, vol. 7, Jan. 2024.

[14] G. Drakoulas, M. Kokkolaras, and J. Lessard, "An explainable machine learning-based probabilistic framework for the design of scaffolds in bone tissue engineering," Advanced Engineering Informatics, vol. 54, art. no. 101866, Apr. 2022, doi: 10.1016/j.aei.2022.101866.