Bio-Design and Manufacturing (2023) 6:464–477
https://doi.org/10.1007/s42242-023-00244-4

RESEARCH ARTICLE

The use of machine learning to predict the effects of cryoprotective agents on the GelMA-based bioinks used in extrusion cryobioprinting

Qian Qiao[1,5,6] · Xiang Zhang[1,2,5,6] · Zhenhao Yan[1,5,6] · Chuanyu Hou[1,5,6] · Juanli Zhang[3] · Yong He[2] · Na Zhao[1,5,6] · Shujie Yan[1,5,6] · Youping Gong[4] · Qian Li[1,5,6]

Received: 2 December 2022 / Accepted: 7 May 2023 / Published online: 14 June 2023

Abstract

Cryobioprinting has tremendous potential to solve problems to do with lack of shelf availability in traditional bioprinting by combining extrusion bioprinting and cryopreservation. In order to ensure the viability of cells in the frozen state and avoid the possible toxicity of dimethyl sulfoxide (DMSO), DMSO-free bioink design is critical for achieving successful cryobioprinting. A nontoxic gelatin methacryloyl-based bioink used in cryobioprinting is composed of cryoprotective agents (CPAs) and a buffer solution. The selection and ratio of CPAs in the bioink directly affect the survival of cells in the frozen state. However, the development of universal and efficient cryoprotective bioinks requires extensive experimentation. We first compared two commonly used CPA formulations via experiments in this study. Results show that the effect of using ethylene glycol as the permeable CPA was 6.07% better than that of glycerol. Two datasets were obtained and four machine-learning models were established to predict experimental outcomes. The predictive powers of multiple linear regression (MLR), decision tree (DT), random forest (RF), and artificial neural network (ANN) approaches were compared, suggesting an order of ANN>RF>DT>MLR. The final selected ANN model was then applied to another dataset. Results reveal that this machine-learning method can accurately predict the effects of cryoprotective bioinks composed of different CPAs. Outcomes also suggest that the formulations presented here have universality. Our findings are likely to greatly accelerate research and development on the use of bioinks for cryobioprinting.

✉ Xiang Zhang
zhangxiang@zzu.edu.cn

1 School of Mechanics and Safety Engineering, Zhengzhou University, Zhengzhou 450001, China

2 State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China

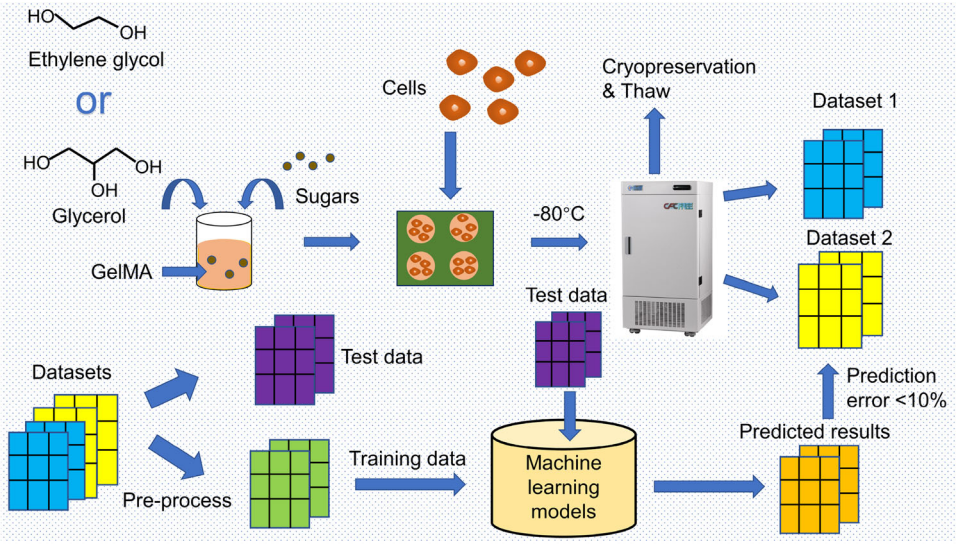
3 NMPA Key Laboratory for Quality Control of In Vitro Diagnostics, Henan Institute of Medical Device Inspections, Zhengzhou 450003, China

4 School of Mechanical Engineering, Hangzhou Dianzi University, Hangzhou 310018, China

5 National Center for International Joint Research of Micro-Nano Molding Technology, Zhengzhou University, Zhengzhou 450001, China

6 Key Laboratory for Micro Molding Technology of Henan Province, Zhengzhou University, Zhengzhou 450001, China

Graphic abstract



Keywords Cryobioprinting · Cryoprotective bioink · 3D bioprinting · Machine learning · Artificial intelligence · Prediction model

Introduction

As one example of an additive biomanufacturing technology, three-dimensional (3D) bioprinting is frequently used to manufacture engineered tissues because of its favorable characteristics and ability to provide in vitro models for therapeutic screening. This approach is also able to meet the needs of tissue and organ shortages for in vivo transplantation [1]. Numerous 3D bioprinting techniques have been developed over the last decades utilizing inkjet bioprinting [2], extrusion bioprinting [3], laser-assisted bioprinting [4], and acoustic bioprinting [5] and enable cells to be patterned in a layer-by-layer manner [6]. Compared to traditional techniques, the major benefits of 3D bioprinting are the ability to digitally define the tissue structures of interest and reproduce the physical 3D constructs through automated operations [7].

In order to generate a 3D structure that is biologically functional, cell-laden bioinks must be optimized to meet certain key characteristics such as rheological and physicomechanical properties as well as biofunctionality [8]. Bioinks are usually composed of two or more components mixed together, which can form synergistic effects. Different components of bioinks have different applications [9]. It is often required to tailor the bioink formulation for a specific print with desired tissue-level functions [10]. However, due to complex factors in the manufacturing and storage processes, traditional bioprinting usually lacks shelf availability. Our previous works in this area have shown that cryobioprinting can make up for this shortcoming by seamlessly integrating extrusion bioprinting and cryopreservation [11, 12]. Traditional extrusion bioprinting uses a bioink which is deposited layer-by-layer into a preset shape through a nozzle. This bioink is usually composed of a hydrogel laden with living cells. The working principle of cryobioprinting is similar to its conventional counterpart, except that bioprinting is carried out directly on a cold plate set at a subzero temperature. Once a tissue has been bioprinted in the frozen state, it is immediately transferred to low-temperature conditions for long-term storage. Thus, to survive at low temperatures, cells must rely on one or more cryoprotective agents (CPAs) that prevent permeable shock and limit the formation of ice crystals that destroy cellular integrity.

In cryobioprinting, cell-laden bioink needs to go through three stages: the cooling process, the storage process, and the thawing process [13]. CPAs are involved in the entire cryopreservation procedure, which helps improve the viability of bioink-embedded cells during cryobioprinting procedures, making it critical to select appropriate CPAs [14]. In general, CPAs are mainly divided into permeable CPAs and impermeable CPAs. Permeable CPAs such as dimethyl sulfoxide (DMSO), ethylene glycol, glycerol, methanol, and propylene glycol are characterized by stable infiltration into cells [15]. During cooling, they can enter cells through the membrane, improve the permeability of the cell membrane to water, reduce the freezing point of the cryopreservation solution, and minimize the formation of ice crystals by water within the cells, preventing inactivation and degeneration of intracellular proteins and rupture of cells [14]. Impermeable CPAs

include polyvinylpyrrolidone (PVP), hydroxyethyl starch (HES), polyethylene glycol (PEG), and sugars, among others [15]. In most cases, in the absence of permeable CPAs, impermeable CPAs do not have the protective effect of cell cryopreservation; however, in the presence of permeable CPAs, they will enhance the effect of impermeable CPAs. Impermeable CPAs help accelerate the vitrification of the solution, stabilize the proteins and membrane of the cells, and prevent the formation of ice crystals by water within the cells [15, 16].

However, the properties of bioinks that contain different components are difficult to predict, which nonetheless are critical since they will affect the performance of tissue constructs formed by subsequent 3D bioprinting. In addition, the relationship between the CPAs and the survival of cells recovered from cryopreservation can be hard to correlate directly. Traditional experimental explorations usually need experienced operators to select suitable components with specific characteristics according to previous experience with experimental verifications, to adjust the new formula according to the experimental results and to carry out experiments repeatedly to optimize, i.e., the trial-and-error method that is oftentimes time-consuming, laborious, and costly.

With the development of computational technologies such as artificial intelligence in recent years, machine learning has become the preferred method to address this type of multivariable and complex problem. Ruberu et al. [17] explored machine learning as a new tool for quantitative evaluations of printing performance and constructed a machine learning black box based on the Bayesian method to optimize the printing performance of gelatin methacryloyl (GelMA) and hyaluronic acid-methacrylate (HAMA) bioinks. Lee et al. [18] developed a method based on machine learning to design bioinks for 3D printing, and they found a general relationship between the mechanical properties and printing adaptability of the bioinks. Chen et al. [19] proposed an artificial-intelligence-assisted high-throughput printing-condition-screening system (AI-HTPCSS) for rapid selection of the optimal printing conditions of a given combination of (bio)printer and biomaterial ink. Machine learning is a subset of artificial intelligence, but currently, there are few studies, if any, on the combination of machine learning and CPA-containing bioinks geared towards cryobioprinting.

In addition, there are many researchers who use machine-learning methods to advance other materials research fields. Ueki et al. [20] succeeded in predicting the grafting yields for the radiation-induced graft polymerization of a methacrylate ester monomer to produce a polyethylene-coated polypropylene nonwoven fabric. Yu et al. [21] reported the development of the biosynthesis of inorganic nanomaterials into a simple, stable, and accurate strategy for distinguishing microorganisms from multiple classification levels without gene amplification, biochemical testing, or target recognition through machine learning. Liu et al. [22] took the first step towards enabling the data-driven discovery of innovative quasicrystals by machine learning. Beckham et al. [23] constructed machine-learning models to explore the factors that drove the transformation of amorphous carbon into graphene nanocrystals during flash Joule heating. Hence, we hypothesized that if we can take advantage of the use of an advanced machine-learning strategy to determine the appropriate proportions of CPAs in bioinks, it might be possible to greatly reduce the workload and cost of cryobioprinting to accelerate further developments and applications in this field.

Accordingly, in this study, we used GelMA with a content of 5% (0.05 g/mL) as the base bioink and then explored ethylene glycol, glycerol, raffinose, trehalose, and lactose as the various CPAs. Four characteristic concentrations of each CPA were selected to prepare bioinks with different compositions for cell cryopreservation experiments and to record cell viability. Subsequently, two datasets (dataset 1 for the ethylene glycol group and dataset 2 for the glycerol group) with a total sample size of 96 were established. Afterwards, four different machine-learning models were established to predict the performances of the bioinks with different CPAs. We obtained the optimal prediction model on the ethylene glycol group dataset and migrated it to the glycerol group dataset. The experimental results suggested that the method that we developed could accurately predict the effects of bioinks with different CPAs on cell survival, and the model had good transferability. The flowchart of this study is shown in Fig. 1.

Experiments and data

Data preparation

The bioinks were designed by mixing in the form of A+B+C. A represents GelMA solution with 5% content, and B represents ethylene glycol or glycerol, with four different concentrations. C stands for lactose, raffinose, or trehalose, and there are also four different concentrations. Meanwhile, to verify the migration performance of the machine-learning model, two different datasets of the bioink system were established (Table 1). In order to ensure the accuracy of experimental data, four identical samples were prepared for each component and the same experiments were carried out independently. Average values were taken to eliminate errors as much as possible. The samples produced from each bioink with a different composition were frozen separately for 1 day, 7 days, and 15 days, and experimental results were recorded to increase the breadth of the dataset for model optimization and verification in the future.

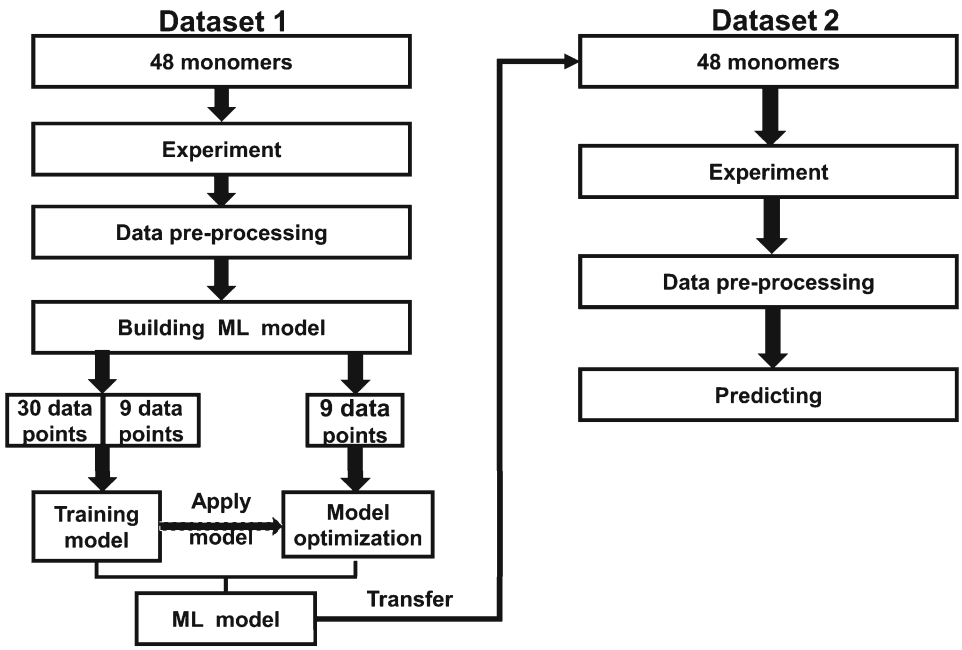


Fig. 1 Diagram showing the cell viability prediction model development process employed. ML: machine learning

Table 1 Two datasets established based on experimental design (GelMA: gelatin methacryloyl)

	Dataset 1		Dataset 2	
	Composition	Concentration (g/mL)	Composition	Concentration (g/mL)
A	GelMA	0.05	GelMA	0.05
B	Ethylene glycol	0.05, 0.10, 0.15, 0.20	Glycerol	0.05, 0.10, 0.15, 0.20
C	Lactose	0.05, 0.10, 0.15, 0.20	Lactose	0.05, 0.10, 0.15, 0.20
	Raffinose	0.05, 0.10, 0.15, 0.20	Raffinose	0.05, 0.10, 0.15, 0.20
	Trehalose	0.05, 0.10, 0.15, 0.20	Trehalose	0.05, 0.10, 0.15, 0.20

Cell cryopreservation experiments

In this study, polydimethylsiloxane (PDMS) was used to fabricate molds for bioink cryopreservation experiments. Ethylene glycol, glycerol, gelatin, raffinose, trehalose, lactose and photoinitiator Irgacure 2959 were purchased from Aladdin Reagent (Shanghai, China). Human umbilical vein endothelial cells (HUVECs), phosphate-buffered saline (PBS), and live/dead cell viability kits were purchased from Oligobio (Beijing, China), BI Company (Shanghai, China), and Biovision (Shanghai, China), respectively. More detailed information on the experimental materials and instruments can be found in Supplementary Information. GelMA was synthesized in-house according to our published protocols [24, 25], and the preparation method is briefly described as follows. First, 10 g of gelatin was dissolved in 100 mL of deionized water and stirred at 50 °C for 1 h. Then, 6 mL of methacrylic anhydride was added dropwise into the solution at a ratio of 0.5 mL/min. The reaction solution was placed into a dialysis bag and dialyzed against deionized water at 50 °C for 10 days. The deionized water was changed three times a day to dialyze out the unreacted substances. Finally, the prepared solution was freeze-dried to obtain GelMA as a raw material for the bioinks used in cryobioprinting.

Experimental method

The experimental methods were as follows: first, 0.5% photoinitiator Irgacure 2959 was added to 5% GelMA solution under dark conditions to avoid light. Then, 5%, 10%, 15%, and 20% ethylene glycol solutions were added to the GelMA solution in separate groups and thoroughly mixed. Subsequently, lactose in the mass ratio of 5%, 10%, 15%, or 20% was added to the respective group. Finally, 0.22-μm filters were used to filter the prepared solutions in the biosafety cabinet to filter out impurities and sterilize.

In our experiments, HUVECs were used as the exemplary cell type for cryopreservation, where the cells after cultivation, harvesting, and counting were added to the solutions with the density of 500,000 cells/mL. First, 100 μL of the above homogeneously blown cell solution was added to each cavity of the PDMS mold with a pipette gun. Then, the PDMS

mold along with the petri dish it was housed was placed on a cooling platform at a precooled temperature of −20 °C, and a 6 W ultraviolet (UV) lamp was used for UV crosslinking for 120 s. The light source was 10 cm away from the samples. The samples were transferred to a refrigerator (Form 900 Series, Thermo Fisher Technology Company, Shanghai, China) at −80 °C for short-term storage. After storage for 1 day, 7 days, and 15 days, the samples were removed and thawed in an incubator at 37 °C. Then, a live/dead cell staining kit was used to stain the samples at room temperature for 30 min in the dark. After staining, the samples were rinsed with PBS, and an inverted fluorescence microscope (DMI 3000 M, Leica Company, Shanghai, China) was used for observation. Finally, the cell viability of each subset of samples was calculated by ImageJ software: the numbers of green cells (i.e., living cells) and red cells (i.e., dead cells) were counted and recorded, and then the total number of cells was obtained by adding them up. Cell viability is equal to the number of living cells divided by the total number of cells.

The trehalose group and the raffinose group were both experimented according to the above method. The bioinks were prepared by replacing ethylene glycol with glycerol for dataset 2. Four samples were repeated for each concentration, and we took the average as the cell viability value for each experimental subset/group.

Ethylene glycol group experiment

Trehalose group experiment The experiments were carried out according to the compositions of dataset 1 in Table 1 and the experimental method described in Sect. "Experimental method." The cell viability results of cryopreservation in each group are shown in Fig. 2. Figures 2a–2d show experiments for 1, 7, and 15 days with ethylene glycol concentrations of 5%, 10%, 15%, and 20%, respectively. Different colors in the figure represent different concentrations of trehalose added to the bioinks.

As shown in Fig. 2, the lowest cell viability of bioink with added trehalose after 15 days of cryopreservation was 52.54%, and the highest cell viability was 73.16%. The cell viability increased with an increase in trehalose mass fraction, and the highest cell viability was reached when the volume fraction of ethylene glycol was 15%.

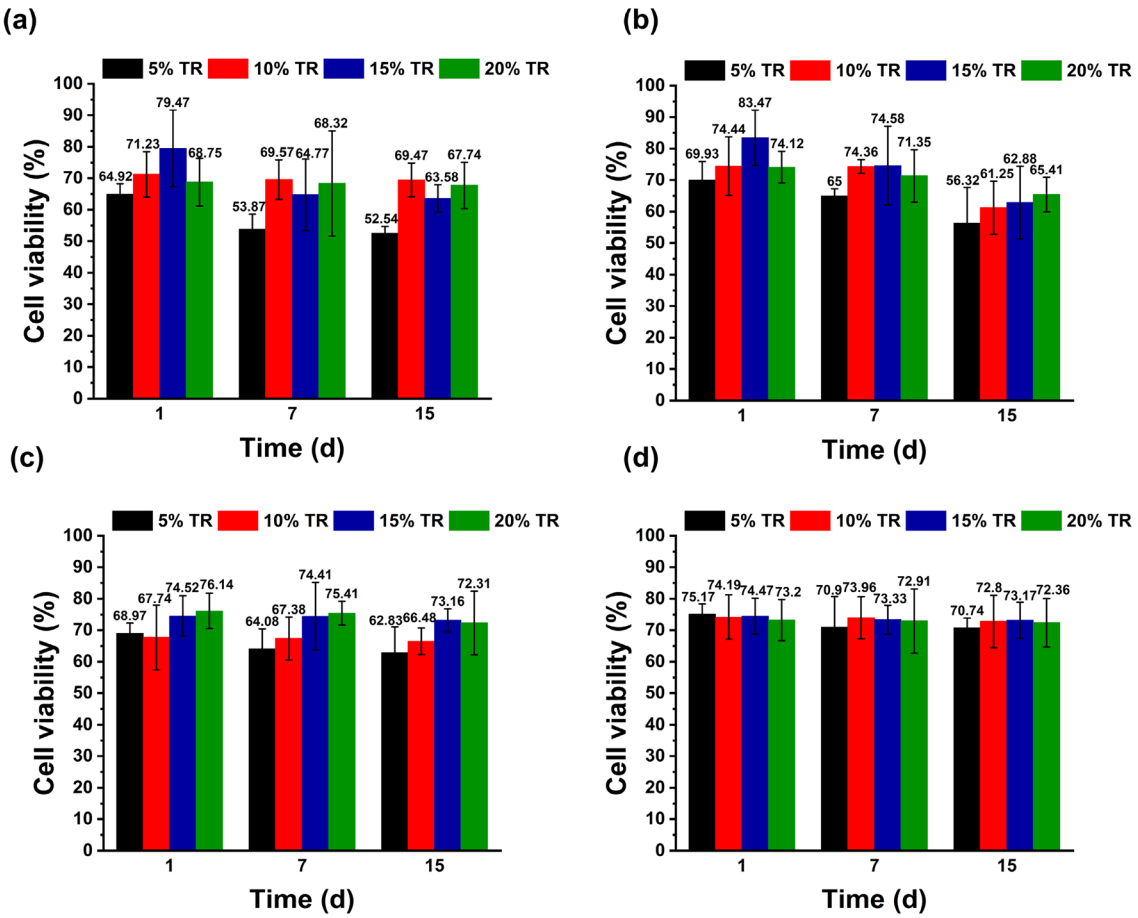


Fig. 2 Comparison of cell viability values following cryopreservation after 1, 7, and 15 days for the trehalose group experiment (n=4): **a** 5% EG; **b** 10% EG; **c** 15% EG; **d** 20% EG. EG: ethylene glycol; TR: trehalose

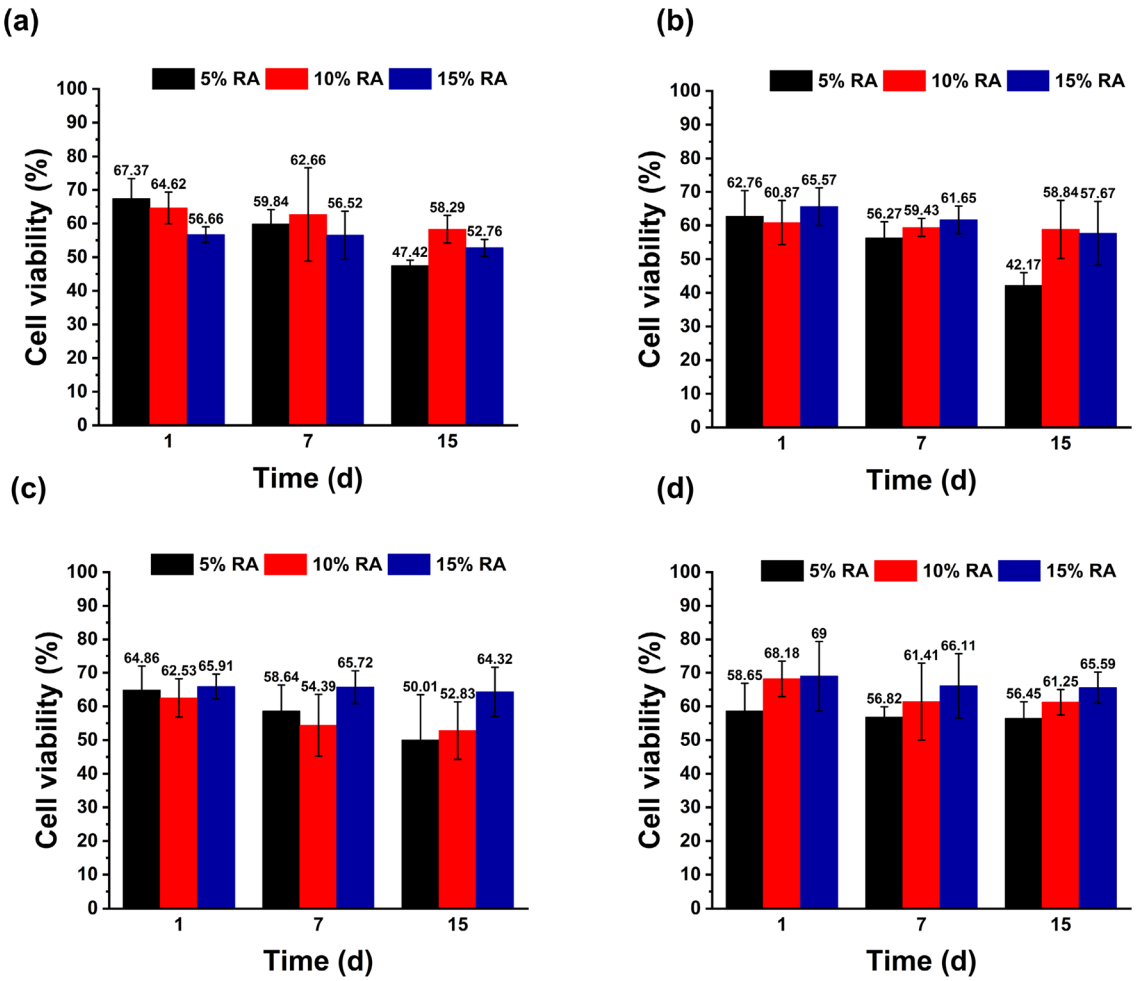


Fig. 3 Comparison of cell viability values following cryopreservation after 1, 7, and 15 days for the raffinose group experiment (n=4): **a** 5% EG; **b** 10% EG; **c** 15% EG; **d** 20% EG. EG: ethylene glycol; RA: raffinose

Raffinose group experiment The experiments were carried out according to the compositions of dataset 1 in Table 1 and the experimental method in Sect. "Experimental method." The cell viability values after cryopreservation of each group are shown in Fig. 3. Figures 3a–3d show experiments for 1, 7, and 15 days with ethylene glycol concentrations of 5%, 10%, 15%, and 20%, respectively. Different colors in the figure represent different concentrations of raffinose added to the bioinks.

As shown in Fig. 3, the lowest cell viability of bioink supplemented with raffinose after 15 days of cryopreservation was 42.17%. The highest cell viability was 65.59%. The cell viability increased with an increase in raffinose mass fraction and reached the highest cell viability when the volume fraction of ethylene glycol was 20%.

Lactose group experiment The experiments were carried out according to the compositions of dataset 1 in Table 1 and the experimental method in Sect. "Experimental method." The cell viability values of cryopreservation in each group are shown in Fig. 4. Figures 4a–4d show experiments for 1, 7, and 15 days with ethylene glycol concentrations of 5%, 10%, 15%, and 20%, respectively. Different colors in the figure represent different concentrations of lactose added to the bioinks.

As shown in Fig. 4, the lowest cell viability of bioink with lactose after 15 days of cryopreservation was 55.56%. The highest viability was 73.61%. The cell viability increased with increasing lactose mass fraction and reached the highest cell viability when the volume fraction of ethylene glycol was 15%.

The results of this section suggested that adding a certain amount (volume fraction) of ethylene glycol and a certain amount (weight/volume) of trehalose, raffinose, or lactose to GelMA bioinks could play a role in cryopreservation protection. The specific conclusions are as follows:

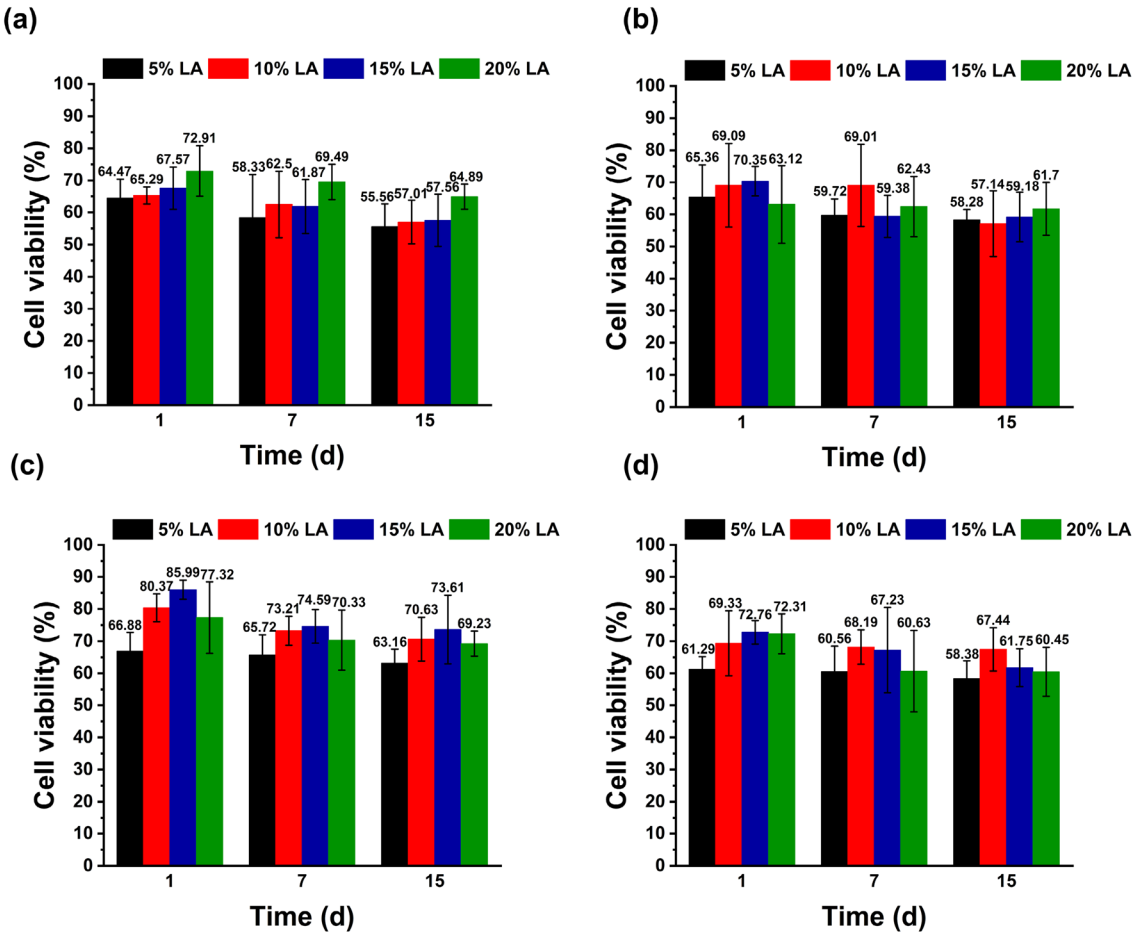


Fig. 4 Comparison of cell viability values following cryopreservation after 1, 7, and 15 days for the lactose group experiment (n=4): **a** 5% EG; **b** 10% EG; **c** 15% EG; **d** 20% EG. EG: ethylene glycol; LA: lactose

(1) 5% (0.05 g/mL) GelMA + ethylene glycol (volume fraction) bioink could protect cells during freezing, among which 5% (0.05 g/mL) GelMA + 15% (volume fraction) ethylene glycol bioink had the best efficacy, and the cell viability after 15 days of cryopreservation was 37.51%.
(2) Among the trehalose groups, 5% (0.05 g/mL) GelMA + 20% (volume fraction) ethylene glycol + 20% (0.20 g/mL) trehalose bioink had the best cell protection efficacy. The cell viability after 15 days of cryopreservation was 72.36%.
(3) Among the raffinose groups, 5% (0.05 g/mL) GelMA + 15% (volume fraction) ethylene glycol + 15% (0.15 g/mL) raffinose bioink had the best cell protection efficacy. The cell viability after 15 days of cryopreservation was 64.32%.
(4) Among the lactose groups, 5% (0.05 g/mL) GelMA + 15% (volume fraction) ethylene glycol + 15% (0.15 g/mL) lactose bioink had the best cell protection efficacy. The cell viability after 15 days of cryopreservation was 73.61%.

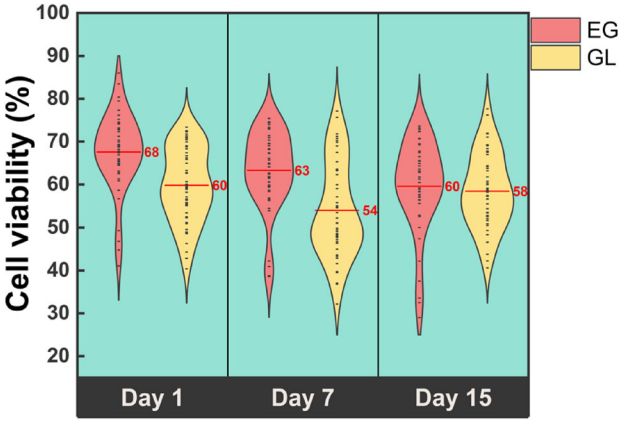


Fig. 5 Distributions of cell viability values for ethylene glycol and glycerol groups. Each group has 48 samples with different compositions and sugar concentrations (5%, 10%, 15%, 20%) for Days 1, 7, and 15. EG: ethylene glycol; GL: glycerol

Based on the outcomes above, it was confirmed that bioinks with different compositions of CPAs had different

protective effects on cell cryopreservation, but the relationship between them was difficult to explicitly describe through experiments; therefore, it became apparent to use machine-learning algorithms to further mine the correlations between them.

Glycerol group experiment

The glycerol group experiments were performed in the same manner as the ethylene glycol group experiments. The detailed experimental results can be found in Supplementary Information.

Comparison of short-term cryopreservation effects

The violin plots in Fig. 5 show the probability distributions and the mean values of cell viability of the ethylene glycol and glycerol groups for short-term cryopreservation experiments. Each distribution was constructed from 48 data points. The experimental results of the ethylene glycol group were better than those of the glycerol group. It can be seen from the figure that the mean values of the ethylene glycol group after cryopreservation for 1, 7, and 15 days were 68%, 63%, and 60%, respectively. The mean values of the glycerol group after cryopreservation for 1, 7, and 15 days were 60%, 54%, and 58%, respectively. After one day of freezing, the mean cell viability of the ethylene glycol group was 8% higher than that of the glycerol group. By the 7th day, the means of both the ethylene glycol and glycerol groups decreased, but the mean of the ethylene glycol group was still 9% higher than that of the glycerol group. After 15 days of cryopreservation, the data distribution of the two groups tended to be consistent, while the mean cell viability of the ethylene glycol group was still 2% higher than that of the glycerol group.

The above results suggested that the protective efficacy of the ethylene glycol group was better than that of glycerol for the short-term cryopreservation that we evaluated. The main trend of the two groups was that the cell viability decreased gradually with increasing cryopreservation time.

Establishment of the dataset

Two datasets were established with a total sample number of 96 according to Table 1, and each dataset contained 48 samples, as shown in Table 2. Each 48 samples with different CPA compositions were separated into 39 samples as training data and 9 samples as test data. The composition of CPA was a subjective variable, which was manually set according to our prior experience. Cell viability was an objective variable, which was obtained by cellular experiments.

Data pre-processing

In the multi-index evaluation system, due to the different dimensions of each effect factor, they usually have large differences in terms of values. When the level of each index varies greatly, if the original index value is directly used for analysis, it will highlight the role of the index with the higher value in the comprehensive analysis and comparatively weaken the role of the index with the lower value level [26]. Therefore, to ensure the reliability of the results, it is necessary to standardize the original index data [27]. In this study, we used min–max normalization to process the input sample features. The mathematical definition of this method is as described in Eq. (1):

$$y_i = \frac{x_i - \min_{1\le j\le n}\{x_j\}}{\max_{1\le j\le n}\{x_j\} - \min_{1\le j\le n}\{x_j\}}, \tag{1}$$

where y_i represents the processed data, x_i represents the original data, max represents the maximum value of the original data sequence, and min represents the minimum value of the original data sequence.

Building the machine-learning model

Four different machine-learning models were established in the study, namely, multiple linear regression (MLR), decision tree (DT), random forest (RF), and artificial neural network (ANN). Table 3 details the experimental setup used in developing the cell viability prediction models.

To evaluate the prediction performances of the different models, the mean absolute error (MAE) calculated by Eq. (2) was used to measure the accuracy of a model [28]:

$$\mathrm{MAE} = \frac{1}{n} \cdot \left[\sum_{i=1}^{n} |y_i - \widehat{y}_i| \right], \tag{2}$$

where y_i is the real value, $\hat{y}_i$ is the predicted value, and n is the sample size.

Because the cell viability was measured as a percentage and the experimental error was also in the form of a percentage, using MAE as the evaluation index of a prediction model is very intuitive and efficient. The MAE calculates the average absolute difference between the predicted and real values; hence, a lower MAE is desired. In the previous cryopreservation experiments in Sect. "Experiments and data," there was an experimental error of 5%–10% between the average value and the maximum/minimum values. Therefore, we chose 5% as the minimum experimental error, which means that the prediction error was 5% as the prediction baseline of a model.

Table 2 The format of the dataset created

Sample	EG	EGP (%)	RA	RAP (%)	LA	LAP (%)	TR	TRP (%)	Cell viability (%)		
									Day 1	Day 7	Day 15
1	1	5	0	0	0	0	0	0	41.07	38.77	28.99
2	1	10	0	0	0	0	0	0	44.77	38.57	33.58
3	1	15	0	0	0	0	0	0	49.31	42.2	37.51
…											
46	1	20	0	0	1	10	0	0	69.33	68.19	67.44
47	1	20	0	0	1	15	0	0	72.26	67.23	61.75
48	1	20	0	0	1	20	0	0	72.31	60.63	60.45

In dataset 1, EG represents ethylene glycol, RA represents raffinose, LA represents lactose, and TR represents trehalose; (0,1) indicates whether to add the substance, 0: not added, 1: added; P indicates the percentage content. In dataset 2, GL presents glycerol (see Supplementary Information), and the rest is the same as in dataset 1

Table 3 Experimental setup

Parameter	Specification
Programming language	Python 3.8.12
Machine learning (ML) libraries	Scikit-learn 1.0.2 (for DT, RF) PyTorch 1.10.0 (for MLR, ANN)
Environment	PyCharm 2021.2.3 (Community Edition)
Central processing unit (CPU)	Intel® Core™ i7-10750H CPU @2.60 GHz
Graphics processing unit (GPU)	NVIDIA GeForce GTX 1660 Ti
Random access memory (RAM)	16.0 GB
System type	64-bit operating system (×64) based processor

DT: decision tree; RF: random forest; MLR: multiple linear regression; ANN: artificial neural network

Construction of the multiple linear regression (MLR) prediction model

Figure 6 shows the results of constructing an MLR model using dataset 1 established in Sect. "Establishment of the dataset." Figures 6a–6c show the experimental results and the predicted results using the constructed MLR model. The blue and red dots represent the results of the training and test data, respectively. The horizontal value of each dot is the observed cell viability obtained from the experiments, and the vertical value is the predicted cell viability based on the trained MLR model. The MAE results, which are indicators of the prediction accuracies of the model formula, are also shown. It can be observed that the MLR model could accurately predict the cell viability outcomes of cells frozen in bioinks with different compositions. Whether frozen for 1 day, 7 days, or 15 days, the results of the MLR model in the training data and test data were relatively accurate, and most data points fell within the acceptable error range.

The MAE decreased from 3.70% on the 1st day to 3.24% on the 7th day, and the accuracy was still stable at this value when predicting the cell viability after freezing for 15 days. Therefore, it was believed that the MLR model was suitable for predicting cell viability after cryopreservation.

Figure 6d shows the standard partial regression coefficient (β) of the MLR model. The value of β reflects the importance of each variable to the predicted result [29], i.e., the larger the β value, the higher the impact of this factor on the results. It was obvious from Fig. 6d that the order of importance (ethylene glycol > trehalose > lactose > raffinose) of whether to add ethylene glycol, raffinose, lactose, or trehalose was consistent with the order of importance of the concentration of the corresponding component. Among them, ethylene glycol had the greatest impact on improving the viability of frozen cells because in the bioink formulations, ethylene glycol was the only permeable CPA, while the other three sugars were impermeable CPAs. Figure 6d reveals that among the sugars that we evaluated, trehalose had the greatest enhancement effect on permeable CPAs, and raffinose had the least enhancement effect.

The above results showed that the prediction model of CPAs could be constructed according to the composition of different CPAs by machine learning. In addition, we successfully quantified the importance of different CPA components on cell viability after cryopreservation.

Construction of the decision tree (DT) prediction model

To explore the possibility of other nonlinear models, some more complex models were further utilized. We established a DT model because it is suitable for small datasets and efficient and insensitive to missing values [30]. To find the optimal model, we divided the 48 samples in dataset 1 into a training set, validation set, and test set, with 30, 9, and

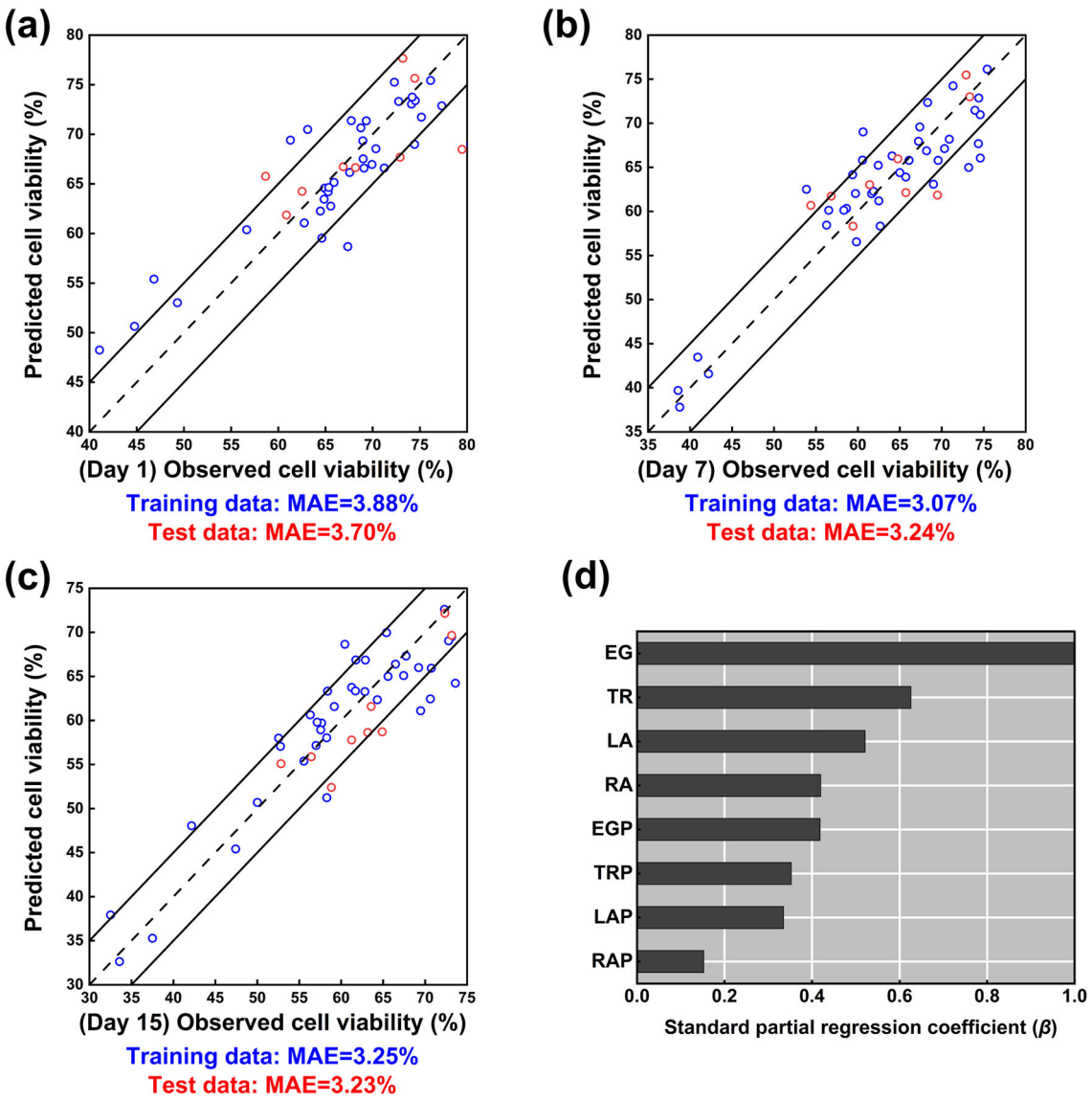


Fig. 6 MLR results: **a** predicted versus observed cell viability on Day 1; **b** predicted versus observed cell viability on Day 7; **c** predicted versus observed cell viability on Day 15; **d** the importance of variables in the MLR prediction model. MLR: multiple linear regression; MAE: mean absolute error; EG: ethylene glycol; RA: raffinose; LA: lactose; TR: trehalose. P indicates the percentage content

9 samples, respectively. The training set and validation set were used for cross validation when establishing the model. The test set did not participate in the model training process but only in model tuning. The final DT model and superparameter search space are shown in Table 4.

Figures 7a–7c show the predicted results of cell viability using the established DT models. Although the prediction errors of a few test samples are outside the minimum experimental error range, most of them are still within the acceptable range. However, it is easy to see from the figures that the MAE of the DT model on the test set is significantly higher than that on the training set, which indicates that the

Table 4 Hyperparameter tuning of the decision tree (DT) algorithm

Hyperparameter	Selected	Search space
Max_features	Log2	Auto, Sqrt, Log2
Max_depth	40	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Min_samples_split	2	2, 5, 10
Min_samples_leaf	1	1, 2, 4

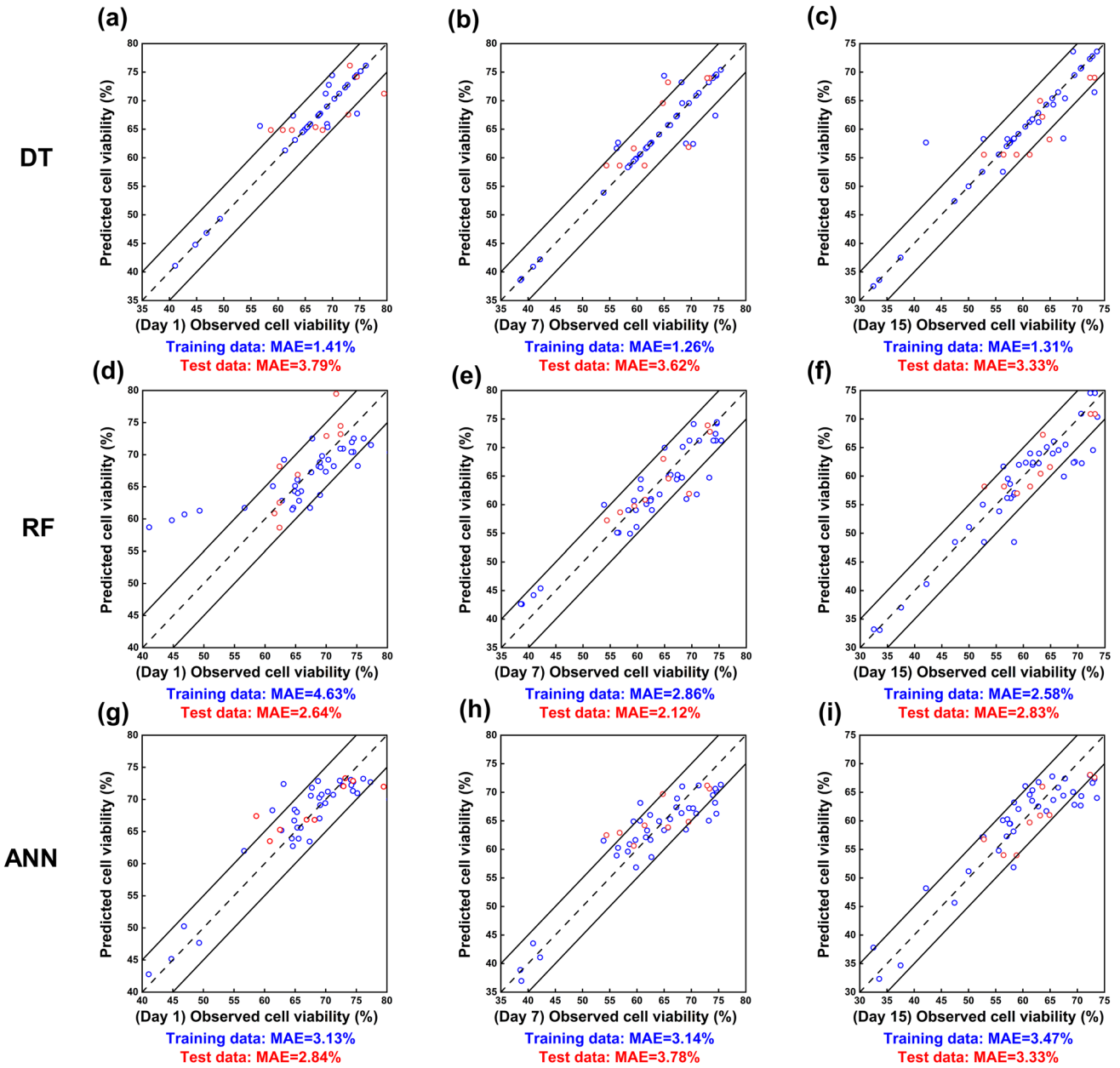


Fig. 7 Prediction results of the DT, RF, and ANN prediction models: the DT model predicts the experimental data on **a** Day 1, **b** Day 7, and **c** Day 15; the RF model predicts the experimental data on **d** Day 1, **e** Day 7, and **f** Day 15; the ANN model predicts the experimental data on **g** Day 1, **h** Day 7, and **i** Day 15; DT: decision tree; RF: random forest; ANN: artificial neural network; MAE: mean absolute error

DT model has overfitting. Overfitting is one of the shortcomings of the DT model. Therefore, the DT model is not recommended in this study.

Construction of the random forest (RF) prediction model

In machine learning algorithms, a single DT model often exists as a weak classifier [31], and its model complexity is not high, so overfitting easily occurs [32]. Therefore, the RF algorithm was used to improve the complexity of the model, reduce overfitting, and improve the accuracy of the model. RF is a kind of integrated learning algorithm. In the training process, it "votes" to select which tree nodes to activate by establishing multiple single-decision trees [33].

The RF model is established, the hyperparameter search spaces are listed in Table 5, and the experimental results are shown in Figs. 7d–7f. The established RF prediction model

Table 5 Hyperparameter tuning of the random forest (RF) algorithm

Hyperparameter	Selected	Search space
N_estimators	10	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Max_features	Auto	Auto, Sqrt
Max_depth	80	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Min_samples_split	5	2, 5, 10
Min_samples_leaf	4	1, 2, 4
Bootstrap	True	True, false

Table 6 Hyperparameter tuning of the artificial neural network (ANN) algorithm

Hyperparameter	Selected	Search space
Number of hidden layers	2	1, 2, 4
Number of neurons in each hidden layer	8	4, 6, 8, 10, 12
Number of epochs	1000	100, 200, 500, 1000, 2000
Learning rate	0.02	[0.01, 0.1]
Weight decay	0.1	0.01, 0.1, 0.5
Batch size	4	4, 8, 16
Training algorithm	Adam	Adam, RMSprop
Activation function	Sigmoid & PReLU	Sigmoid, ReLU, PReLU, Tanh

reduces the error on the test set to be less than 3% and avoids the overfitting phenomenon according to the results of the training set. However, the MAE on the test set is smaller than that on the training set when the model predicts the cell viability on Day 1 and Day 7. If the number of test samples is increased, the error will be increased. The number of training samples should also be increased.

Construction of the artificial neural network (ANN) prediction model

Through the experiments in Sects. "Construction of the multiple linear regression (MLR) prediction model"–"Construction of the random forest (RF) prediction model," it can be seen that the complex algorithm model has high accuracy, but the DT and RF models have some defects. We continued to improve the complexity of the model and used the ANN algorithm for modeling. Neural networks have become the most popular deep learning algorithm in recent years [34]. Its learnable parameters can reach trillions or even more [35, 36]. The neural network model architecture established on the experimental dataset is shown in Table 6, and the model prediction results are shown in Figs. 7g–7i.

We successfully established the ANN prediction model, which not only avoids the overfitting phenomenon but also reduces the gap between the prediction result errors in the test set and the training set. Especially when predicting the cell viability on the 15th day, the prediction errors of the test set are all within the lowest experimental error range.

Model migration

Figure 8a shows the performance of different machine learning models we established on the test set. Different colors represent different machine learning models (MLR, DT, RF, and ANN). Overall, the prediction efficiency of the models is ranked as ANN>RF>DT>MLR. When predicting the cell viability on the first day, the ANN model has the best effect, the average value of errors is the smallest, which is 2.84%, and the maximum and minimum values of errors are also the smallest, which are 5.75% and 0.03%, respectively. The effect of the MLR model is the worst, and the maximum error is more than 10%, indicating that the MLR model is not sufficiently stable. Although the effect of the RF model is better than that of the ANN model in predicting cell viability on Day 15, the model error fluctuation is greater than that of the ANN model in predicting cell viability on Day 7.

In conclusion, the ANN model has the best performance in predicting the cryopreservation efficiency of bioinks with different CPAs. Therefore, we applied the ANN model migration trained in Sect. "Construction of the artificial neural network (ANN) prediction model" to dataset 2 to further investigate the migration generalization ability of the ANN model.

To further evaluate the performance of the ANN model in predicting the effects of the bioinks, we applied the ANN model used in Sect. "Construction of the artificial neural network (ANN) prediction model" with dataset 2 for the same training process and randomly selected 10 samples as the test set. The error of the prediction results after a series of model tunings is shown in Fig. 8b.

The average error of all prediction results is within 5%, which once again proves the effectiveness of the ANN model in predicting the efficacy of CPAs, indicating that the ANN model is very suitable for solving this type of problem. Although the error distribution is gradually dispersed with the increase of time, and the maximum error and minimum error tend to increase, the whole is still within the acceptable range. Therefore, we believe that the ANN model has certain universality in finding the optimal cell CPAs, which can provide a new idea for researchers and accelerate bioink development used in tissue engineering and artificial organs.

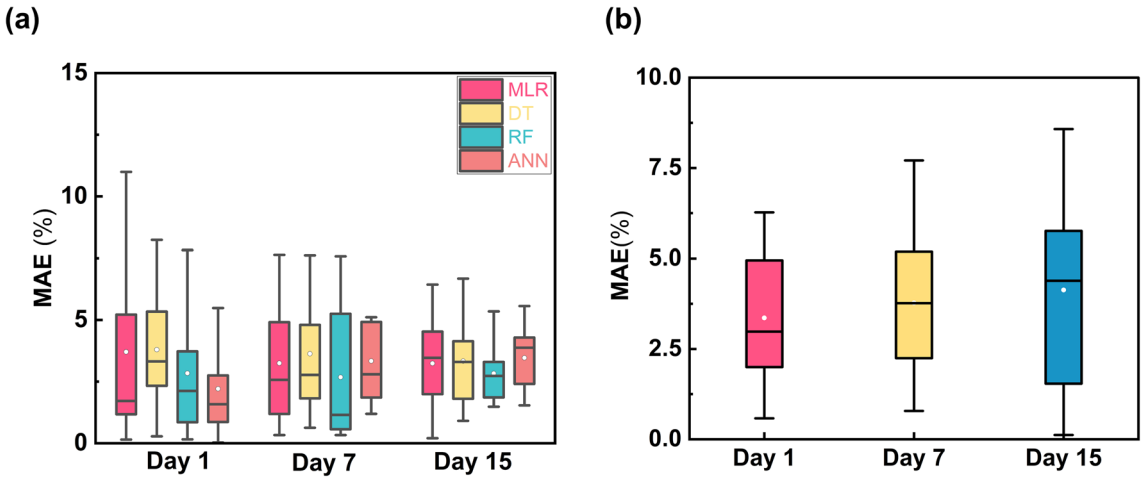


Fig. 8 Prediction error distribution of ML models: **a** mean absolute error distribution of different machine learning models on the test set on dataset 1 (n=9); **b** mean absolute error distribution of ANN on dataset 2 (n=10). ML: machine learning; ANN: artificial neural network; MAE: mean absolute error

Conclusions

In this study, we first compared the protection efficiency of two permeable CPAs through cell cryopreservation experiments. It was found that the cryopreservation efficiency of ethylene glycol as a permeable CPA was better than that of glycerol. Then, two datasets were obtained, and four machine learning prediction models were established. By comparing different machine learning models, we obtained the prediction effects of different models ranked as ANN>RF>DT>MLR. Finally, the selected optimal model was migrated and verified. The experiments showed that the ANN model was very suitable for this kind of problem, and even when used in a different bioink system, it could also achieve good prediction performance.

We successfully predicted the efficiency of cell CPAs with only eight explanatory variables and obtained the ranking of the enhancement efficiency of three sugars. It was understood that we were the first to apply machine learning to the research and development of the bioinks used in cryobioprinting, which provided a reference for follow-up researchers and could greatly improve the development efficiency of bioinks, and it was of great significance for long-term preservation of artificial tissues and organs. However, the datasets established in this study were not large enough and did not take environmental factors into account. In the future, more compositions of other CPAs will be chosen to expand the scale of the experiment and the factors that may affect the results in the experimental environment will be taken into account as independent variables to improve the quality and quantity of data for comprehensive analysis. By doing so, we expect that this will enable us to develop universally applicable cryoprotective bioinks. Combined with the cryobioprinting technology we developed before, it will greatly increase the application scope of artificial tissue and organs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42242-023-00244-4.

Acknowledgements This work was financially supported by the Major Science and Technology Special Project of Henan Province, China (No. 171100210600), the Program of China Scholarship Council (No. 201807045057), the High Level Talent Internationalization Training Program of Henan Province, China (No. 2019004), the Scientific and Technological Research Project of Henan Province, China (Nos. 212102310854 and 222102310526), and the Open Foundation of the State Key Laboratory of Fluid Power and Mechatronic Systems (No. GZKF-202105). We would like to thank Prof. Yu Shrike Zhang from Harvard Medical School and Brigham and Women's Hospital for his proofreading and suggestions.

Author contributions QQ: methodology, investigation, visualization, and writing—original draft—revision. XZ: conceptualization, methodology, investigation, writing—review & editing, supervision, and funding acquisition. ZHY: investigation and formal analysis. CYH: investigation and validation. JLZ: resources. YH: software and writing—review & editing. NZ: writing—review & editing. SJY: writing—review & editing and funding acquisition. YPG: software. QL: resources.

Declarations

Conflict of interest YH is an Associate Editor of *Bio-Design and Manufacturing*. The authors declare that they have no conflict of interest.

Ethical approval This study was approved by the Adminstraion Committee of Experimental Animals, Zhengzhou University, China.

References

1. Murphy SV, Atala A (2014) 3D bioprinting of tissues and organs. Nat Biotechnol 32:773–785. https://doi.org/10.1038/nbt.2958

2. Ranjit E, Hamlet S, George R et al (2022) Biofunctional approaches of wool-based keratin for tissue engineering. J Sci Adv Mater Dev 7:100398. https://doi.org/10.1016/j.jsamd.2021.10.001
3. Tian S, Stevens R, McInnes B et al (2021) Machine assisted experimentation of extrusion-based bioprinting systems. Micromachines 12(7):780. https://doi.org/10.3390/mi12070780
4. Ali M, Pages E, Ducom A et al (2014) Controlling laser-induced jet formation for bioprinting mesenchymal stem cells with high viability and high resolution. Biofabrication 6(4):045001. https://doi.org/10.1088/1758-5082/6/4/045001
5. Hu X, Zhu J, Zuo Y et al (2020) Versatile biomimetic array assembly by phase modulation of coherent acoustic waves. Lab Chip 20:3515–3523. https://doi.org/10.1039/d0lc00779j
6. Planchette C, Pichler H, Wimmer-Teubenbacher M et al (2016) Printing medicines as orodispersible dosage forms: effect of substrate on the printed micro-structure. Int J Pharm 509:518–527. https://doi.org/10.1016/j.ijpharm.2015.10.054
7. Thomas DJ (2021) Engineering regenerative tissue systems using 3D bioprinting technology. A golden era for reconstructive surgery. Int J Surg 90:105982. https://doi.org/10.1016/j.ijsu.2021.105982
8. Cui X, Li J, Hartanto Y et al (2020) Advances in extrusion 3D bioprinting: a focus on multicomponent hydrogel-based bioinks. Adv Healthc Mater 9(15):1901648. https://doi.org/10.1002/adhm.201901648
9. Ravanbakhsh H, Bao GY, Luo ZY et al (2021) Composite inks for extrusion printing of biological and biomedical constructs. ACS Biomater Sci Eng 7(9):4009–4026. https://doi.org/10.1021/acsbiomaterials.0c01158
10. Luo Z, Mu X, Zhang YS (2022) Biomaterials for bioprinting. In: Conti M, Marino M (Eds.), Bioprinting. Academic Press, Elsevier, p.51–86. https://doi.org/10.1016/B978-0-323-85430-6.00001-7
11. Ravanbakhsh H, Luo Z, Zhang X et al (2022) Freeform cell-laden cryobioprinting for shelf-ready tissue fabrication and storage. Matter 5:573–593. https://doi.org/10.1016/j.matt.2021.11.020
12. Luo Z, Tang G, Ravanbakhsh H et al (2022) Vertical extrusion cryo(bio)printing for anisotropic tissue manufacturing. Adv Mater 34(12):2108931. https://doi.org/10.1002/adma.202108931
13. El Assal R, Guven S, Gurkan UA et al (2014) Bio-inspired cryo-ink preserves red blood cell phenotype and function during nanoliter vitrification. Adv Mater 26:5815–5822. https://doi.org/10.1002/adma.201400941
14. Shi Y, Zeng Y, Yu H et al (2006) The processes and the recently progression of the cryopreservation of biological materials. Cryogen Supercond 34:141–144
15. Koebe HG, Muhling B, Deglmann CJ et al (1999) Cryopreserved porcine hepatocyte cultures. Chem Biol Interact 121:99–115. https://doi.org/10.1016/S0009-2797(99)00093-9
16. Sharma B, Fermanian S, Gibson M et al (2013) Human cartilage repair with a photoreactive adhesive-hydrogel composite. Sci Transl Med 5(167):167ra61. https://doi.org/10.1126/scitranslmed.3004838
17. Ruberu K, Senadeera M, Rana S et al (2021) Coupling machine learning with 3D bioprinting to fast track optimisation of extrusion printing. Appl Mater Today 22:100914. https://doi.org/10.1016/j.apmt.2020.100914
18. Lee J, Oh SJ, An SH et al (2020) Machine learning-based design strategy for 3D printable bioink: elastic modulus and yield stress determine printability. Biofabrication 12(3):035018. https://doi.org/10.1088/1758-5090/ab8707
19. Chen B, Dong J, Ruelas M et al (2022) Artificial intelligence-assisted high-throughput screening of printing conditions of hydrogel architectures for accelerated diabetic wound healing. Adv Funct Mater 32(38):2201843. https://doi.org/10.1002/adfm.202201843
20. Ueki Y, Seko N, Maekawa Y (2021) Machine learning approach for prediction of the grafting yield in radiation-induced graft polymerization. Appl Mater Today 25:101158. https://doi.org/10.1016/j.apmt.2021.101158
21. Yu T, Su S, Hu J et al (2022) A new strategy for microbial taxonomic identification through micro-biosynthetic gold nanoparticles and machine learning. Adv Mater 34(11):2109365. https://doi.org/10.1002/adma.202109365
22. Liu C, Fujita E, Katsura Y et al (2021) Machine learning to predict quasicrystals from chemical compositions. Adv Mater 33(36):2102507. https://doi.org/10.1002/adma.202102507
23. Beckham JL, Wyss KM, Xie Y et al (2022) Machine learning guided synthesis of flash graphene. Adv Mater 34(12):2106506. https://doi.org/10.1002/adma.202106506
24. Maharjan S, Bonilla D, Sindurakar P et al (2021) 3D human non-alcoholic hepatic steatosis and fibrosis models. Bio-Des Manuf 4:157–170. https://doi.org/10.1007/s42242-020-00121-4
25. Gong J, Schuurmans CCL, van Genderen AM et al (2020) Complexation-induced resolution enhancement of 3D-printed hydrogel constructs. Nat Commun 11(1):1267. https://doi.org/10.1038/s41467-020-14997-4
26. Yu F, Liu L, Yu N et al (2020) A method of L1-norm principal component analysis for functional data. Symmetry 12(1):182. https://doi.org/10.3390/sym12010182
27. Bowen TP (1984) Standardized data collection for software engineering. In: Third Software Engineering Standards Application Workshop (Cat. No. 84CH2071–9)
28. Mendo L (2009) Estimation of a probability with guaranteed normalized mean absolute error. IEEE Commun Lett 13:817–819
29. Schielzeth H (2010) Simple means to improve the interpretability of regression coefficients. Methods Ecol Evol 1:103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x
30. Li DC, Wu CS, Tsai TI et al (2006) Using mega-fuzzification and data trend estimation in small data set learning for early FMS scheduling knowledge. Comput Oper Res 33:1857–1869. https://doi.org/10.1016/j.cor.2004.11.022
31. Liu YL, Wang YR, Zhang J (2012) New machine learning algorithm: random forest. In: Third International Conference on Information Computing and Applications, p.246-252
32. Saleh H, Bassily N, Hammoud J (2009) Utility of a liquid-based, monolayer preparation in the evaluation of thyroid lesions by fine needle aspiration biopsy comparison with the conventional smear method. Acta Cytol 53:130–136. https://doi.org/10.1159/000325113
33. El Habib Daho M, Settouti N, El Amine Lazouni M et al (2014) Weighted vote for trees aggregation in random forest. In: International Conference on Multimedia Computing and Systems, p.1–6. https://doi.org/10.1109/ICMCS.2014.6911187
34. Zhang C, Guo Y, Li M (2021) Review of development and application of artificial neural network models. Comput Eng Appl 57:57–69
35. Bi JR, Zhu ZL, Meng QL (2021) Transformer in computer vision. In: IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology, p.178–188
36. He KM, Zhang XY, Ren SQ et al (2015) Deep residual learning for image recognition. https://arxiv.org/abs/1512.03385