JOTCS A Journal of the Turkish Chemical Society Section: A Chemistry
TURKISH CHEMICAL SOCIETY

The Effect of Different Methacrylation Amounts on Physical Properties of Gelatin Methacryloyl Biomaterials: Machine Learning Approach

Sena Çakıcı[1], Rumeysa Tutar[2*]

[1]Department of Chemical Engineering, Faculty of Engineering, Istanbul University-Cerrahpaşa, Avcılar, Istanbul, 34320, Turkey.
[2]Department of Chemistry, Faculty of Engineering, Istanbul University-Cerrahpaşa, Avcılar, Istanbul, 34320, Turkey.

Abstract: The rational design process for biomaterials is time-consuming. Machine learning (ML) is an efficient approach for reducing material synthesis and experimentation in terms of cost and time. Among the emerging biopolymers for tissue engineering applications, methacrylic anhydride (MA)-functionalized gelatin (GelMA), which was chosen as the model biomaterial for this study, has assumed a promising role owing to its excellent tunable properties and biocompatibility. The ML approach was used to determine the efficiency of the MA amounts selected for GelMA synthesis. In addition, the effect of different methacrylation amounts on the molecular structure of GelMA was indicated in terms of its physical properties. This modeling was performed to generate predictions based on 20 mL of MA. The prediction output was obtained as a result of four data models from the 20 mL MA column. First, data were collected with experimental applications for swelling and degradation ratios, and then the data processing phase was applied. The most suitable ML model, decision tree regression, was selected, and the results were interpreted graphically. The experimental results were compared with the ML results, and the efficiency of ML is shown in detail. The Mean Squared Error (MSE) value for degradation was calculated as 10.16, with a Root Mean Squared Error (RMSE) of 3.1885, Mean Absolute Error (MAE) of 2.6667, and Mean Absolute Percentage Error (MAPE) of 14.66%. For swelling, the MSE value was calculated to be 1821.25, with an RMSE of 3.1885, MAE of 2.6667, and MAPE of 14.66%. In future studies, it is anticipated that the performance of the model will improve with the expansion of the experimental dataset for swelling measurements.

Keywords: Methacrylic anhydride, Gelatin, GelMA, Physical properties, Data analysis, Machine learning.

Submitted: April 26, 2024. **Accepted:** July 6, 2024.

Cite this: Cakıcı S, Tutar R. The Effect of Different Methacrylation Amounts on Physical Properties of Gelatin Methacryloyl Biomaterials: Machine Learning Approach. JOTCSA. 2024;11(3): 1275-86.

DOI: https://doi.org/10.18596/jotcsa.1473948

***Corresponding author's E-mail:** rumeysa.tutar@iuc.edu.tr

1. INTRODUCTION

A rational design process for biomaterials can be time-consuming. The prepared biomaterial needs to have the desired functionality while mimicking human physiology. Biomaterials are used for biomedical applications; therefore, all details are important (1,2). In particular, the development of biocompatible biomaterials requires rational designs. The development of biomaterials using the traditional approach involves a long, trial-and-error cycle. It is becoming increasingly important to improve this process in terms of time and cost (3). The purpose of machine learning (ML) in biomaterial production and design is to accelerate the process and obtain accurate models.

Machine learning emulates human learning. ML is an important field of science that operates through the development of various algorithms and techniques (1). The primary objective of machine learning is to derive inferences and obtain predictions from data. ML aims to save time in data analysis and engineering. The most efficient form of the material can be achieved using ML (4). In recent years, the field of data science has seen numerous advancements in data analysis owing to the proliferation of data. Interest in this field has increased significantly to save time in concluding data as quickly as possible. The basic logic of machine learning is to learn the pattern and statistical correlation of data (5). The expected results and interpretations differ depending on the

data type. Applying an incorrect model resulted in inaccurate results and poor performance.

One machine-learning algorithm is called supervised learning. It teaches a computer to understand the relationship between input data, which are the variables provided, and output, which is the desired result (6). This approach is used in many applications, particularly for data collection and image processing (7). They can be used in language processing, medicine, finance, and many other fields. Some supervised learning methods are classified into linear regression, decision trees, random forests, and artificial neural networks. Another important ML algorithm is unsupervised learning, which is used to understand a dataset. It describes the pattern of the data in the dataset and the correlation between the data. Unsupervised learning can be classified into data collection, clustering (7,8), dimensionality reduction (9,10), and feature engineering. In particular, these two ML algorithms are most commonly used in the biological and material fields.

The input and output data are important for supervised machine learning. The model was divided into training and testing sets using the input data. It learns with the training dataset and checks it with the inputs, which it divides into a test set (11). In addition to the importance of the input and output information in supervised learning, it is recommended for situations where known data are available for the output of the data to be predicted. Supervised learning uses classification and regression techniques (12). In decision trees, a dataset can be divided into a smaller training set with a feature selected from the training data. The results are formed in this way to create branches. At this point, the results were divided into questions asked at the root nodes of the separation. Branches were created as a pure subset of the training dataset. Decision tree methods can be classified using regression analysis (13,14). It is preferable to interpret the dataset in regression according to the results of the data in the new clusters created in the regression analysis because the dataset is numerical, and the results are analyzed according to variable parameters. In the present study, the variables were the time and degree of methacrylation of the GelMA biomaterial. A dataset was created from the results of the degradation and swelling measurements based on these two variables. The model was applied using decision trees, and branching under small clusters was examined using regression analysis. In general, the aim is to use decision trees with regression analysis on the dataset.

Hydrogels are polymer materials that can absorb and retain large amounts of water without degrading their structure (15). Hydrogels are used in various applications, such as tissue engineering and environmental engineering. Many physical properties of hydrogel biomaterials, such as swelling and degradation ratios, can be experimentally determined. Gelatin methacryloyl (GelMA), often abbreviated as GelMA, is a derivative of gelatin containing primarily methacrylamide and minor methacrylate groups (2,16,17). Methacrylic groups were introduced from methacrylic anhydride (MA) onto the active amine and hydroxyl groups of GelMA. GelMA is called very high GelMA, high/medium GelMA, or low GelMA based on the addition of MA. The amount of methacrylamide-functionalization is an important parameter that affects the polymerization and crosslinking processes of GelMA hydrogel structures (16-18). The incorporation of methacrylate groups into the amine-containing side chains of gelatin enabled photopolymerization, resulting in a stable hydrogel that remained intact at 37 °C. The variation in the degree of methacrylation allows precise control over the physical properties of GelMA, offering a customizable spectrum of swelling characteristics tailored to various applications (18).

Researchers have explored hydrogel stiffness through the application of machine-learning techniques. It has been established that the stiffness of hydrogels is intricately related to their physical properties. Altering the reaction conditions induces changes in the physical, chemical, and morphological attributes of hydrogels (19,20). Researchers have mostly investigated the effect of photocrosslinking conditions on hydrogel stiffness based on the mechanical properties (21). Photocrosslinking conditions are directly related to photoinitiators. Therefore, researchers have applied artificial intelligence (AI) to overcome the limitations associated with experimental optimization. For instance, a group developed an artificial neural network (ANN) model to predict the effects of Eosin Y, triethanolamine (TEA), and N-vinyl-2-pyrrolidone (NVP) concentrations on stiffness and gelation time (22). In a previous study, GelMA hydrogels were fabricated with different DS values (49.8%, 63.8%, and 73.2%) using 1, 5, and 10 molar ratios of MA, respectively. It was observed that the average porosity of the dried hydrogels, as determined by scanning electron microscopy (SEM), decreased with higher substitutions (23,24). Considering these approaches, the present study is the first of its kind in this context. Investigating the effects of different degrees of methacrylation on GelMA biomaterials will provide researchers with different perspectives.

Our study focused on the importance of the amount of methacrylate in GelMA preparations and its physical properties, such as swelling and degradation. This is important for the application of GelMA hydrogels in tissue engineering applications. To date, 8 mL of MA has been experimentally investigated for different organs in tissue engineering applications of GelMA (2,16-18). We aimed to obtain physical property data according to varying amounts of methacrylation during the synthesis of GelMA using a machine-learning approach. The dataset was created by considering changes in the parameters as a result of the synthesis. This parameter determines the physical properties of the hydrogel structure based on the swelling and degradation ratios. In this study, there were multiple independent variables (features) in the dataset because GelMA was synthesized with different amounts of methacrylation (4, 8, 12, and 20 mL MA). Swelling and degradation data were experimentally collected for each degree of

methacrylation. For four different degrees of methacrylation (four distinct variables), 10 data points were used in the swelling studies, whereas nine data points were used in the degradation studies. Experimental results were initially obtained using a 20 mL methacrylation degree as a sample. Subsequently, the results were presented by comparing the experimental data with machine learning data for the 20 mL methacrylation degree.

Decision tree regression was chosen as the method. This model is based on the principle of continuously partitioning data fields and creating a prediction model for each partition. This partitioning creates decision trees graphically. The decision tree machine learning method was employed to perform a regression analysis on two distinct datasets and establish more specific relationships between the data. Decision tree regression analysis allows the dataset to be interpreted in smaller groups using branches. Another reason for choosing decision-tree methods is their inherent simplicity and ease of understanding.

2. EXPERIMENTAL SECTION

2.1. Materials and Instruments

Type A porcine gelatin (300 g Bloom), methacrylic anhydride (MA purity >94%), triethanolamine (TEA purity >99%), N-vinyl caprolactam (VC purity >98%), Eosin Y disodium salt, cellulose dialysis membranes (12–14 kDa molecular weight cutoff), and phosphate-buffered saline (PBS) tablets were purchased from Sigma-Aldrich. Dulbecco's phosphate-buffered saline (DPBS) was purchased from Gibco (Life Technologies). Attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectra of all studied samples were recorded between 4000 and 500 cm^{-1} using a Jasco FT/IR 6700 spectrophotometer. A JEOL ECZ500R (11.75 Tesla) spectrometer and high-performance Ultrashield TM 500 MHz superconducting magnet were used for NMR analysis. Scanning electron microscopy (SEM) micrographs were acquired at different magnifications using an SEM (JEOL JSM 5600).

2.2. GelMA Synthesis and Characterization

Gelatin was methacrylated based on previous literature (16,17). Different amounts of MA (4, 8, 12, and 20 mL) were used to prepare GelMA biomaterials. Briefly, 10 g of gelatin was dissolved in DPBS at 50 °C under constant stirring. After the gelatin was completely dissolved, MA was added dropwise to the gelatin solution. The reaction was allowed to continue for 2 h. Subsequently, a cellulose dialysis membrane was used to remove unreacted chemicals from the GelMA solution. After dialysis, the solution was placed in a glass container, frozen, and freeze-dried using a TeknoSEM brand lyophilizer device. ATR-FTIR spectroscopy and ^{1}H-NMR characterization were used to identify pristine gelatin and the prepared GelMA materials.

2.3. Hydrogel Preparation and Characterization

Hydrogels were prepared using a visible-light photoinitiator system. Visible light crosslinker solutions (PIs) were prepared using Eosin Y disodium salt (0.5 mM), VC (1.25 w/v%), and TEA (1.875 w/v%) (2,16). To determine the physical properties, 10% w/v GelMA was dissolved in a PI solution to prepare the hydrogels. A small LED light source (VALO Light Curing Device, Ultradent, USA) was used to crosslink hydrogels. Visible light was applied for 120 s to the GelMA solutions, which were then placed in a polydimethylsiloxane (PDMS) mold (diameter:9 mm and depth: 7 mm) to fabricate crosslinked hydrogel discs of the same size. The resulting GelMA hydrogel discs were frozen overnight and dried using a lyophilizer. Dry hydrogel discs were then used to determine the physical properties of GelMA.

2.4. Morphological Characterization

The SEM surface morphologies of the methacrylate-modified samples and hydrogels after crosslinking were examined using scanning electron microscopy. The samples and scaffolds were placed on a double-sided graphite tape, attached to a metal surface, and sputter-coated with gold for 10 s.

2.5. Physical Properties of GelMA Hydrogels

After preparing the hydrogel discs via visible-light crosslinking, the physical properties of the GelMA hydrogels were characterized based on their swelling and degradation ratios. The physical properties of GelMA were obtained by adding 4, 8, 12, and 20 mL of methacrylic anhydride in four different GelMA syntheses, which were tested using appropriate methods. The obtained data were used to create the dataset.

2.5.1. Swelling Ratio

The swelling capacity of the prepared GelMA hydrogels was evaluated in DPBS at 37 °C following previously reported procedures (2). The hydrogel discs were prepared using a freeze-drying method. Dry hydrogel discs were weighed and recorded to obtain their initial dry weight (W_i). Then, all weighed hydrogel samples were soaked in DPBS and placed in an incubator. They were then weighed after removing excess water at specific time points to determine the wet weight at each interval (W_s). The time intervals were set to 1 h, 3 h, 5 h, 7 h, 24 h, 168 h, and 336 h. The swelling ratios were calculated using the ratio of the increase in mass to the mass of the swollen samples. A minimum of six samples were tested for each case.

$$Swelling\ (\%) = (W_s - W_i)/Wi * 100 \quad (1)$$

W_i = initial weight before swelling.
W_s = swollen weight of the materials.

2.5.2. Degradation Ratio

The in vitro degradation behavior of the scaffolds was analyzed using collagenase type II solution at 37 °C (2). The freeze-dried hydrogel discs were weighed to determine their initial weights. The dried samples were then immersed in collagenase II solution. Samples were removed from the solution at checkpoints of 1 h, 3 h, 5 h, 7 h, 24 h, and 168 h of incubation. The hydrogel discs were then frozen and lyophilized. The dry weights were recorded at each time point. The degradation ratios of the hydrogel discs were calculated using the following equation:

$$Lost\ weight\ (\%) = (W_i - W_d)/W_i \times 100 \quad (2)$$

W_i = initial dry weight before degradation, and W_d = final dry weight after the operation.

2.6. Machine Learning (ML) Study

Visual Studio Code is a stand-alone source code editor compatible with Windows, macOS, and Linux operating systems. It is freely accessible and provides seamless internet connectivity. For our study, we utilized the Conda virtual environment, which is both free and easily integrated with the Visual Studio Code editor alongside the Python programming language. The Decision Tree Regression model was employed to demonstrate data prediction within the supervised learning domain of machine learning. Through data pre-processing, the dataset was rendered suitable and subsequently partitioned into training and testing subsets. Prior to the model implementation, normalization was applied to standardize the data, followed by the application of the model to the data frame. Hyperparameter tuning was conducted to optimize the model fit and enhance performance. Mean square error (MSE) values were computed to evaluate error performance. Finally, the predicted data were compared with the actual data displayed on the screen, and the results were analyzed.

3. RESULTS AND DISCUSSION

The bands based on methacrylation of the GelMA structures were used for characterization, as shown in Figure 1. FTIR spectra were used to identify the chemical composition of the fabricated samples. To conduct this analysis, pristine gelatin and GelMA were directly used. They were exposed to infrared waves in the range 4000–400 cm^{-1} using a Jasco FT/IR-6700 spectrometer. The band at 1210 cm^{-1} was assigned to the secondary (amide) group, v(C-N), in the spectra of Gelatin and GelMA. The spectra of GelMA with different methacrylated ratios showed peak shifts. For GelMA containing 8 mL MA, a shift in the C-N band occurred. The bands at 1640 cm^{-1} and 1540 cm^{-1} correspond to C=0 and N-H bending, respectively. While there is C=O stretching (amide I) in the gelatin structure, C=C groups are formed in the GelMA structure owing to the methacrylated anhydride (25,26). The peaks of both groups were very close to each other. The $-NH_2$ groups in the gelatin structure exist as -NH groups in GelMA after methacrylation. The broad amide A band observed at approximately 3300 cm^{-1} as a characteristic medium absorption peak can be attributed to the stretching vibrations of O–H and N–H. The $-NH_2$ bands in the gelatin structure are shadowed by the -OH band. A widespread peak was observed (2,26).

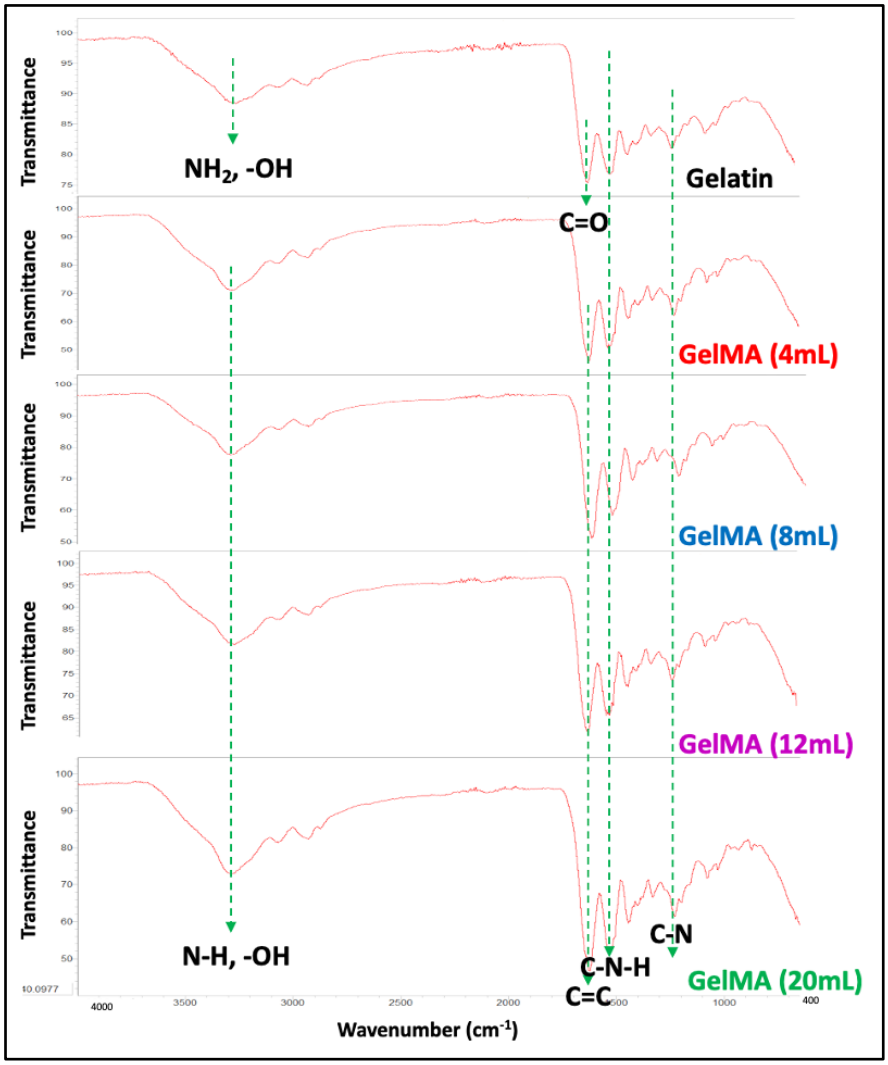


Figure 1: The ATR-FTIR spectra of pristine Gelatin and GelMA samples.

In ^{1}H-NMR spectra of pristine gelatin and four different methacrylated GelMA samples are shown in Figure 2. Deuterium oxide (D_2O) was used to prepare solutions for the measurements at room temperature. Because of the methacrylation reaction, two main peaks (belonging to the magnetically different protons at vinyl groups, CH_2=C) were observed at 5.4-5.6 ppm in the GelMA spectrum (2,27). This indicated that the methacrylic group was successfully connected to the gelatin. GelMA hydrogels were fabricated with different degrees of methacrylation (DM), which were calculated based on Equation 3 (14.3%, 28.6%, 71.4%, and 86%) by employing 4, 8, 12, and 20 mL of MA, respectively (16,27).

$$Degree\ of\ methacrylation\ (DM)\ \% = \frac{Area\ under\ curve\ of\ GelMA\ signal\ at\ 2.9\ ppm}{Area\ under\ curve\ of\ Gelatin\ signal\ at\ 2.9\ ppm} * 100 \quad (3)$$

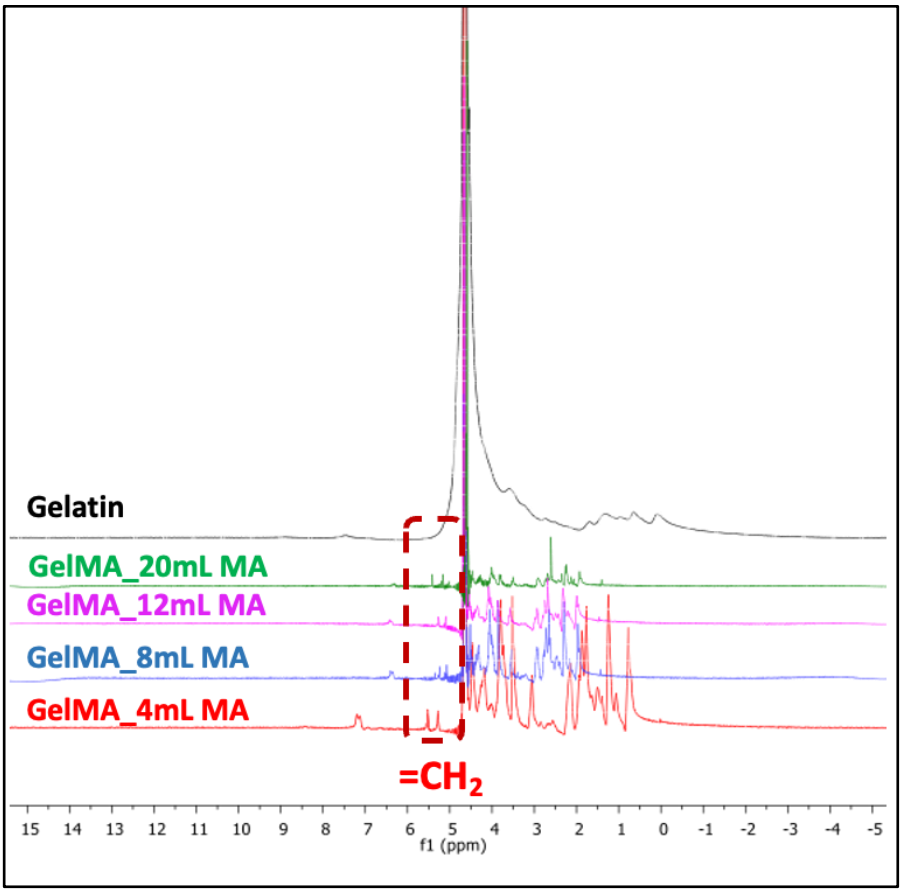


Figure 2: ^{1}H-NMR Analysis of pristine Gelatin and GelMA samples.

The morphological characterization, distribution, and variations in the porosities of the four hydrogels and samples are shown in Figure 3.

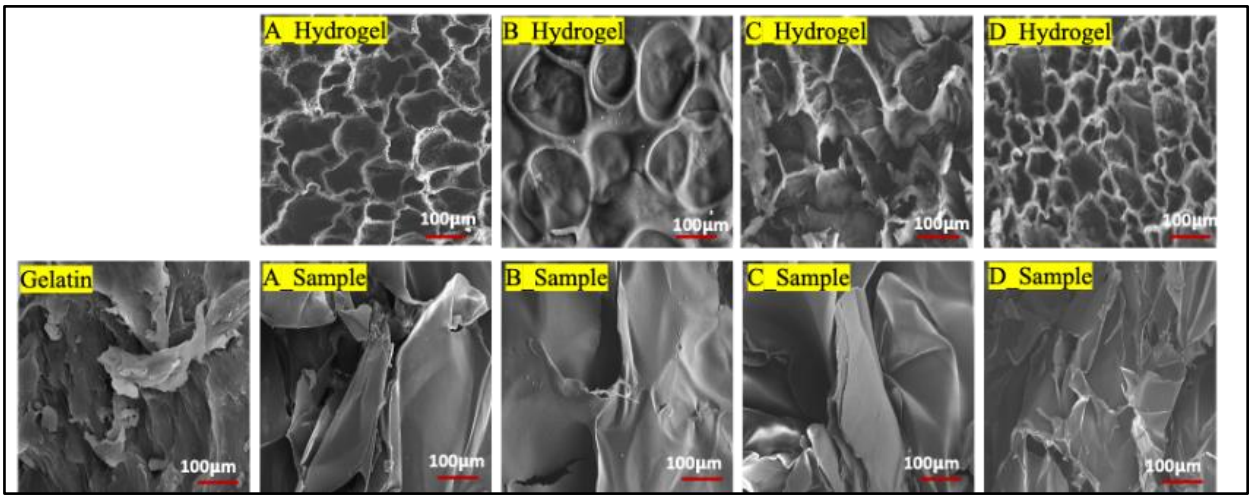


Figure 3: The SEM images of different amounts of methacrylated modified gelatin hydrogels (GelMA) (A-4 mL, B-8 mL, C-12 mL, D-20 mL) and samples (Gelatin, A-4 mL, B-8 mL, C-12 mL, D-20 mL).

The microstructure and porosity of the synthesized GelMA hydrogels and samples at four different MA concentrations (4 mL, 8 mL, 12 mL, and 20 mL) were investigated using SEM, as shown in Figure 3. The porosity of GelMA hydrogels is inversely correlated with GelMA concentration and cooling rate (23), and the hydrogel porosity can be controlled by adjusting the substitution amount of MA (24,28). According to the images in Figure 3, an increase in DM% does not result in any increase in porosity and uniformity.

The data collection is also explained in detail. Two datasets are obtained. Data were obtained at different methacrylation degrees and at different times for the physical properties, which were separated by the swelling and degradation ratios of the GelMA biomaterials. For each specific time point, the average of six measurements was obtained and processed as one data point.

First, data processing was applied to model the degradation properties of the GelMA hydrogels. A panda library was imported to create a data frame (Figure 4). Subsequently, the degradation results were entered according to the time of the experiment and transferred to Python. Data processing was started to load the data into the Python environment.

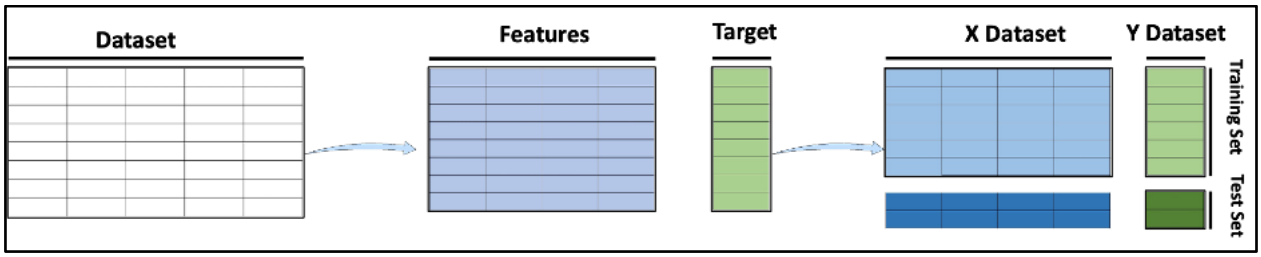


Figure 4: Train test data split.

The dataset was divided into variable data and columns to be predicted. Independent variables and allocated columns were divided into test and training data.

Actual and Predicted values are indicated. The data obtained from the application of the machine learning model are compared with the real values in Figure 5.

The graph for the degradation properties of the GelMA hydrogels was written in Python and was the result of the machine learning model. This helped us interpret the results of the model. Residuals were defined as the difference between observed and predicted values. A symmetric bell-shaped histogram, which was evenly distributed around zero, indicates that the normality assumption was likely to be true. Positive values on the y-axis indicate that the estimates were too low. Negative values indicated that the estimate was too high. A value of zero indicated that the prediction was correct.

The data obtained from the laboratory are presented in a table according to hours in Figure 5 (A). These values are written in Python. All Python codes used in this study are provided in detail in the Supplementary Information (EIS).

```
import pandas as pd
data = {
   'Hour': [1, 2, 3, 4, 5, 6, 7, 24, 168],
   '4ml': [19, 23, 22, 20, 19, 25, 20, 20, 23],
   '8ml': [23, 24, 24, 19, 26, 23, 23, 22, 23],
   '12ml': [20, 22, 20, 20, 16, 24, 22, 20, 22],
   '20ml': [15, 21, 17, 19, 15, 20, 22, 18, 15]
}
df = pd.DataFrame(data)
```

The train_test_split operation was used to split the data from the scikit-learn (sklearn) library. By assigning “x” and “y” variables, train and test sets were separated for each variable. Our “X” variable must be an independent variables. Our “y” variable was the set of variables from which wanted to obtain results and from which predictions to be used in machine learning were created. When the dependent variable was separated from the independent variables while separating the data, it was called the "y" column because the prediction was made on target 20 (mL) data.

X_train: This was part of the learning of the values predicted by the model. This was used to understand the relationship between the data during the learning process of the model.

X_test: This allows data trained with independent variables to be tested to evaluate the performance of the model.

Y_train was used to measure the performance of the model. In the x_train section, the learned relationship was compared with y_train, and the success of the model was determined.

Y_test: Used together with "X_test" data to measure how well the model works and the success of its predictions. The model's success was calculated by comparing its predictions with "y_test" data (such as the mean square error (MSE)).

Test _size=0.4 means 60% train and 40% reserved for test.

When Random_state is not defined in the code for each run, the results may vary. When the random_state = "constant integer" was entered, the training data were constant for each run.

```
from sklearn.model_selection import train_test_split
X = df.drop(['Hour', '20ml'], axis=1)
```

```
y = df['20ml']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=42)
```

Scaling was used to scale the features in the dataset and to improve the performance of the model. The main purpose was to transfer data to a specific range, such as between 0 and 1 or between -1 and 1 (29).

```
from sklearn.pre-processing import MinMaxScaler
scaler = MinMaxScaler()
X_normalized = scaler.fit_transform(X)
X_normalized = pd.DataFrame(X_normalized, columns= X columns).
```

In decision tree modeling, specifically the chosen empirical tree model, data are partitioned using a repeated division method (30). A regression model was selected because of its ability to explore the relationships between the dependent variable and a series of independent variables. The code sequence used to implement the model is outlined below:

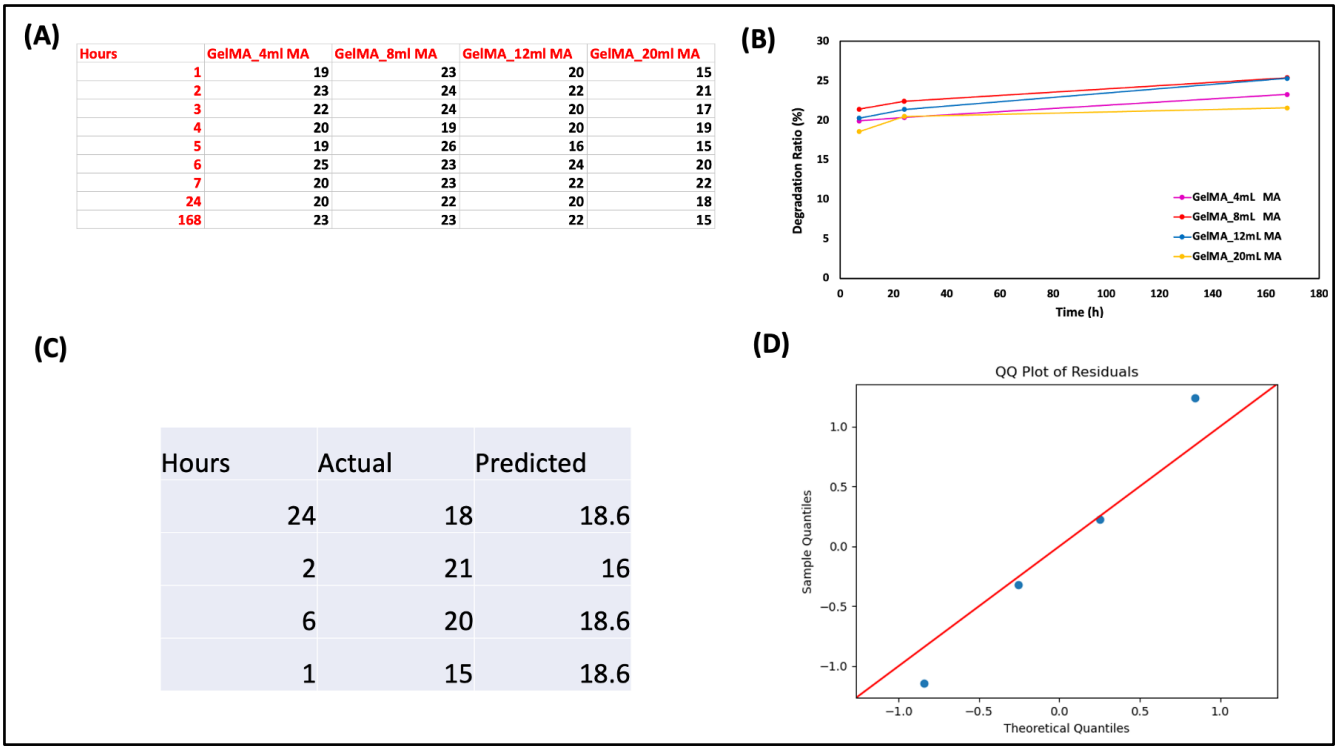
```
from sklearn.tree import DecisionTreeRegressor
model = DecisionTreeRegressor()
```

One of the factors that enhanced the model's performance was the hyperparameter setting (1). A key reason for using this method is its efficiency in handling large numbers of hyperparameters, particularly in complex datasets, thereby reducing time loss. Hyperparameter tuning has been employed in tree-based machine learning models and deep neural network modeling. The grid search method, a common approach in hyperparameter tuning, involves creating hyperparameter values on a preset grid and then using and validating the model for each combination. Consequently, combinations of hyperparameters that yielded the highest model performance were identified (31). This process was automated using Scikit-learn's GridSearchCV, which facilitated selection and implementation of the best model.

```
from sklearn.model_selection import GridSearchCV
param_grid = {'max_depth': [None, 5, 10, 15, 20, 25],
'min_samples_split': [2, 5, 10],
'min_samples_leaf': [1, 2, 4]}
grid_search = GridSearchCV(model, param_grid, cv=5, scoring='neg_mean_squared_error')
grid_search.fit(X_train_normalized, y_train)
best_model = grid_search.best_estimator_
```

Values were printed on the screen in this way to observe the actual and estimated values at 20 mL according to the hours in:

```
Output:
Hours: 24, Actual 20ml: 18, Predicted 20ml: 18.666666666666668
Hours: 2, Actual 20ml: 21, Predicted 20ml: 16.0
Hours: 6, Actual 20ml: 20, Predicted 20ml: 18.666666666666668
Hours: 1, Actual 20ml: 15, Predicted 20ml: 18.666666666666668
```



Hours	GelMA_4ml MA	GelMA_8ml MA	GelMA_12ml MA	GelMA_20ml MA
1	19	23	20	15
2	23	24	22	21
3	22	24	20	17
4	20	19	20	19
5	19	26	16	15
6	25	23	24	20
7	20	23	22	22
24	20	22	20	18
168	23	23	22	15

Hours	Actual	Predicted
24	18	18.6
2	21	16
6	20	18.6
1	15	18.6

Figure 5: Degradation data **(A)** and ratios (%) **(B)** were obtained according to time. Actual and Predicted Values **(C)**. QQ Plot of Residuals **(D)**.

A Q-Q Plot of residuals was used to assess whether the residuals were normally distributed. Ideally, these points should be close to the line. In the QQ Plot of Residuals, the percentiles of the theoretical normal distribution are on the x-axis, and the percentiles of the model's errors are on the y-axis. The histogram (Figure 5 (D)) shows the spread and evaluation of its success.

Second, data processing was applied to model the swelling properties of the GelMA hydrogels. The results of the swelling physical measurements were modeled by making inferences from the machine learning performed in the degradation physical measurement. Physical measurement of swelling was based on the results of the two methacrylation degrees and hours.

Actual and Predicted values are indicated. After machine learning modeling was performed to observe the results, the y_test set was estimated. The results were compared with real values, as shown in Figure 6.

Data Pre-processing began by importing the pandas library. The data obtained from the experiment were written in the form of a data frame according to the given time (Figure 6 (A)).

```
import pandas as pd
data = {
   'Hour': [1, 2, 3, 4, 5, 6, 7, 24, 168, 336],
   '4ml': [225, 230, 234, 319, 250, 288, 274, 242, 276, 274],
   '8ml': [199, 219, 229, 253, 202, 243, 193, 224, 228, 238],
   '12ml': [199, 190, 227, 180, 220, 182, 202, 279, 299, 325],
   '20ml': [255, 304, 293, 307, 261, 381, 276, 208, 250, 225]
}
df = pd.DataFrame(data)
```

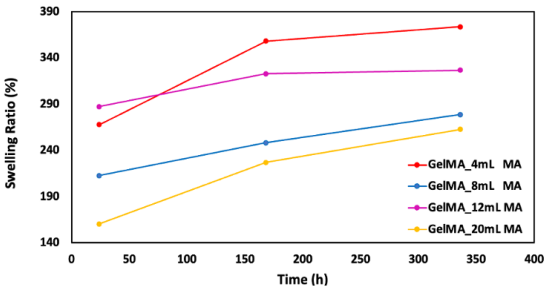
After applying the model, the results were printed on the screen to compare the predicted values for 20 mL with the actual values according to the given hours. In addition, these values are listed in the table for comparison and observation.

Output: Actual vs Predicted 20 mL Values:
Hours: 168, Actual 20ml: 250, Predicted 20ml: 208.0
Hours: 2, Actual 20ml: 304, Predicted 20ml: 307.0
Hours: 6, Actual 20ml: 381, Predicted 20ml: 307.0
Hours: 1, Actual 20ml: 255, Predicted 20ml: 261.0

(A)

Hours	GelMA_4ml MA	GelMA_8ml MA	GelMA_12ml MA	GelMA_20ml MA
1	225	199	199	255
2	230	219	190	304
3	234	229	227	293
4	319	253	180	307
5	250	202	220	261
6	288	243	182	381
7	274	193	202	276
24	242	224	279	208
168	276	228	299	250
336	274	238	325	225

(B)



(C)

Hours	Actual	Predicted
168	250	208
2	304	307
6	381	307
1	255	261

(D)

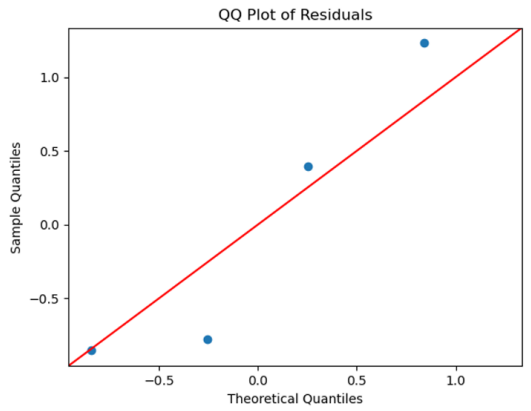


Figure 6: Swelling data **(A)** and ratios (%) **(B)** were obtained according to time. Actual and Predicted Values **(C)**. QQ Plot of Residuals **(D)**.

While designing this study, we asked the following questions.

- How does the variation in methacrylation amount during GelMA synthesis affect the physical changes in the GelMA molecule structure?
- Can hydrogel-based materials be designed using machine learning?
- Can a machine learning algorithm be created by creating a GelMA synthesis dataset?

We hypothesized that increasing and/or decreasing the amount of methacrylamide in the GelMA hydrogel would alter the physical properties of the prepared hydrogel. A model developed using machine learning methods can predict the physical properties of a hydrogel by taking the amount of methacrylic anhydride in the GelMA hydrogel as input.

In this study, multiple independent variables (features) were obtained from different degrees of methacrylation in the dataset. Different datasets for the physical properties, swelling, and degradation were created and evaluated using different models. When many machine learning models were compared for degradation measurements, their performance was determined using the Mean Squared Error (MSE) (32). All details are included in the Supplementary Information (ESI). The MSE is the

difference between the actual and predicted values, and the square of the errors is calculated; thus, the performance of the regression model is calculated. Many methods can be used to measure model performance; examples of these methods include Mean Square Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). In this study, the calculations were performed using the MSE value. MSE is the mean of the squares of the errors, that is, the difference between the actual and predicted values (4). It is expected that the resulting amount of data will not be very high. It is understood that performance is good when the result is close to zero. For the RMSE and MAPE performance results, z' was the predicted value. z is the measured value, and N is the total number of observations. The lowest RMSE value indicated the highest performance (33,34). For the MAPE calculation, the performance result is obtained as a percentage, and a low percentage is a measure of good performance (35,36). MAE properties are commonly referred to as absolute errors (5).

When the data structure was observed, there were four different variables, and each variable had nine data points for the degradation results. In addition, there were four different variables, and each variable had 10 data points for the swelling results. The most suitable model for this dataset is the decision tree regressor (31,37). This modeling was performed to create predictions of 20 mL. The prediction output was obtained as a result of four data modeling from column 20 mL. After the modeling was applied, the MSE values were checked to evaluate the performance.

It should be noted that this dataset was small. The present research aims to demonstrate how the applications aimed in this study can be carried out. As this was a small dataset, it was anticipated that the margin of error would be high. This study aimed to demonstrate the applicability of modeling and obtain prediction results.

Scaling and hyperparameter adjustments were performed in Python to minimize the errors made by the model and avoid overfitting. The MSE, RMSE, MAE, and MAPE were evaluated based on a hyperparameter application to test the performance of the model.

4. CONCLUSION

This study aimed to measure the physical properties according to different degrees of methacrylation, create a dataset with the results, and interpret them by modeling them using machine learning. Swelling and degradation data were experimentally collected for each degree of methacrylation. For four different degrees of methacrylation (four distinct variables), 10 data points were used in the swelling studies, whereas nine data points were used in the degradation studies. Experimental results were initially obtained using a 20 mL methacrylation degree as a sample. Subsequently, the results were presented by comparing the experimental data with machine learning data for the 20 mL methacrylation degree. After determining the predicted values, a decision tree regressor was applied, which was selected as the appropriate model. Performance metrics were calculated to assess the accuracy of the results. To visualize the model's outcomes, the predicted values were observed, and graphs were generated. Following data processing and partitioning, a machine-learning model was applied. The performance metrics were computed after the implementation of the model. To visually interpret the results, a graph displaying the actual versus the predicted values was generated. Numerically, predicted and actual values were documented in code, facilitating result analysis.

As a result, the degradation and swelling physical properties of the GelMA biomaterial for different degrees of methacrylation were experimentally studied, and data were obtained. The obtained data were compared with experimental results using machine learning and decision tree regression for a 20 mL methacrylation degree. The Mean Squared Error (MSE) value for degradation was calculated as 10.16, with a Root Mean Squared Error (RMSE) of 3.1885, Mean Absolute Error (MAE) of 2.6667, and Mean Absolute Percentage Error (MAPE) of 14.66%. For swelling, the MSE value was calculated to be 1821.25, with an RMSE of 3.1885, MAE of 2.6667, and MAPE of 14.66%. In future studies, it is anticipated that the performance of the model will improve with the expansion of the experimental dataset for swelling measurements. When the results were compared, it was determined that the accuracy of the study increased as the experimental dataset increased. As the number of data points increases, the resulting performance increases. In addition, it is important to avoid overfitting in machine learning, that is, the situation in which machine learning imitates the data exactly.

5. CONFLICT OF INTEREST

The authors declare that they have no affiliations with or involvement in any organization or entity with any financial interest in the subject matter or materials discussed in this manuscript.

6. ACKNOWLEDGMENTS

The TUBITAK 2209-A University Student Research Project Support Program supported this study. Project number: 1919B012307651.

7. DATA ACCESS STATEMENT

Research data supporting this publication are available from the Visual Studio Code, Python, and Conda repositories located at https://code.visualstudio.com/download.

8. REFERENCES

1. Meyer TA, Ramirez C, Tamasi MJ, Gormley AJ. A user’s guide to machine learning for polymeric biomaterials. ACS Polym Au [Internet]. 2023 Apr 12;3(2):141–57. Available from: <URL>.

2. Tutar R, Koken SY, Tuncaboylu DC, Çelebi-Saltik B, Özeroğlu C. In situ formation of biocompatible and ductile protein-based hydrogels via Michael addition reaction and visible light crosslinking. New J Chem [Internet]. 2023;47(22):10759–69. Available from: <URL>.

3. Basu B, Gowtham NH, Xiao Y, Kalidindi SR, Leong KW. Biomaterialomics: Data science-driven pathways to develop fourth-generation biomaterials. Acta Biomater [Internet]. 2022 Apr;143:1–25. Available from: <URL>.

4. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol [Internet]. 2022 Jan 13;23(1):40–55. Available from: <URL>.

5. Inza I, Calvo B, Armañanzas R, Bengoetxea E, Larrañaga P, Lozano JA. Machine learning: An indispensable tool in bioinformatics. In 2010. p. 25–48. Available from: <URL>.

6. Peng GCY, Alber M, Buganza Tepole A, Cannon WR, De S, Dura-Bernal S, et al. Multiscale modeling meets machine learning: What can we learn? Arch Comput Methods Eng [Internet]. 2021 May 17;28(3):1017–37. Available from: <URL>.

7. Castelli V, Cover TM. On the exponential value of labeled samples. Pattern Recognit Lett [Internet]. 1995 Jan 1;16(1):105–11. Available from: <URL>.

8. Reddy YCAP, Viswanath P, Reddy BE. Semi-supervised learning: a brief review. Int J Eng &Technology [Internet]. 2018;7(1):81–5. Available from: <URL>.

9. Li Y. Deep Reinforcement Learning: An Overview [Internet]. 2017. 85 p. Available from: <URL>.

10. Ayodele TO. Types of Machine Learning Algorithms. In: New Advances in Machine Learning [Internet]. InTech; 2010. Available from: <URL>.

11. Mahesh B. Machine learning algorithms - A review. Int J Sci Res [Internet]. 2020 Jan 5;9(1):381–6. Available from: <URL>.

12. Aery MK, Ram C. A review on machine learning: Trends and future prospects. An Int J Eng Sci [Internet]. 2017;25:89–96. Available from: <URL>.

13. Pathak S, Mishra I, Swetapadma A. An assessment of decision tree based classification and regression algorithms. In: Proceedings of the International Conference on Inventive Computation Technologies (ICICT-2018) [Internet]. 2018. p. 92–5. Available from: <URL>.

14. Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. J Chemom [Internet]. 2004 Jun 4;18(6):275–85. Available from: <URL>.

15. Süren SM, Tutar R, Özeroğlu C, Karakuş S. Versatile multi-network hydrogel of acrylamide, sodium vinyl sulfonate, and N,N′-methylene bisacrylamide: A sustainable solution for paracetamol removal and swelling behavior. J Polym Environ [Internet]. 2024 Jan 20;32(1):164–81. Available from: <URL>.

16. Tavafoghi M, Sheikhi A, Tutar R, Jahangiry J, Baidya A, Haghniaz R, et al. Engineering tough, injectable, naturally derived, bioadhesive composite hydrogels. Adv Healthc Mater [Internet]. 2020 May 24;9(10):1901722. Available from: <URL>.

17. Van Den Bulcke AI, Bogdanov B, De Rooze N, Schacht EH, Cornelissen M, Berghmans H. Structural and rheological properties of methacrylamide modified gelatin hydrogels. Biomacromolecules [Internet]. 2000 Mar 14;1(1):31–8. Available from: <URL>.

18. Nichol JW, Koshy ST, Bae H, Hwang CM, Yamanlar S, Khademhosseini A. Cell-laden microengineered gelatin methacrylate hydrogels. Biomaterials [Internet]. 2010 Jul;31(21):5536–44. Available from: <URL>.

19. He J, Sun Y, Gao Q, He C, Yao K, Wang T, et al. Gelatin methacryloyl hydrogel, from standardization, performance, to biomedical application. Adv Healthc Mater [Internet]. 2023 Sep 15;12(23):2300395. Available from: <URL>.

20. Noshadi I, Hong S, Sullivan KE, Shirzaei Sani E, Portillo-Lara R, Tamayol A, et al. In vitro and in vivo analysis of visible light crosslinkable gelatin methacryloyl (GelMA) hydrogels. Biomater Sci [Internet]. 2017;5(10):2093–105. Available from: <URL>.

21. O'Connell CD, Zhang B, Onofrillo C, Duchi S, Blanchard R, Quigley A, et al. Tailoring the mechanical properties of gelatin methacryloyl hydrogels through manipulation of the photocrosslinking conditions. Soft Matter [Internet]. 2018;14(11):2142–51. Available from: <URL>.

22. Karaoglu IC, Kebabci AO, Kizilel S. Optimization of gelatin methacryloyl hydrogel properties through an artificial neural network model. ACS Appl Mater Interfaces [Internet]. 2023 Sep 27;15(38):44796–808. Available from: <URL>.

23. Van Vlierberghe S, Dubruel P, Schacht E. Effect of cryogenic treatment on the rheological properties of gelatin hydrogels. J Bioact Compat Polym [Internet]. 2010 Sep 4;25(5):498–512. Available from: <URL>.

24. Chen Y, Lin R, Qi H, Yang Y, Bae H, Melero-Martin JM, et al. Functional human vascular network generated in photocrosslinkable gelatin methacrylate hydrogels. Adv Funct Mater [Internet]. 2012 May 23;22(10):2027–39. Available from: <URL>.

25. Tutar R, Yüce-Erarslan E, İzbudak B, Bal-Öztürk A. Photocurable silk fibroin-based tissue sealants with enhanced adhesive properties for the treatment of corneal perforations. J Mater Chem B [Internet]. 2022;10(15):2912–25. Available from: <URL>.

26. Rahali K, Ben Messaoud G, Kahn C, Sanchez-Gonzalez L, Kaci M, Cleymand F, et al. Synthesis and

characterization of nanofunctionalized gelatin methacrylate hydrogels. Int J Mol Sci [Internet]. 2017 Dec 10;18(12):2675. Available from: <URL>.

27. Claaßen C, Claaßen MH, Truffault V, Sewald L, Tovar GEM, Borchers K, et al. Quantification of substitution of gelatin methacryloyl: Best practice and current pitfalls. Biomacromolecules [Internet]. 2018 Jan 8;19(1):42–52. Available from: <URL>.

28. Lee Y, Lee JM, Bae P, Chung IY, Chung BH, Chung BG. Photo-crosslinkable hydrogel-based 3D microfluidic culture device. Electrophoresis [Internet]. 2015 Apr 24;36(7–8):994–1001. Available from: <URL>.

29. Jamal P, Ali M, Faraj RH, Ali PJM, Faraj RH. Data normalization and standardization: A technical report. Mach Learn Tech Reports [Internet]. 2014;1(1):1–6. Available from: <URL>.

30. Tso GKF, Yau KKW. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy [Internet]. 2007 Sep;32(9):1761–8. Available from: <URL>.

31. Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing [Internet]. 2020 Nov;415:295–316. Available from: <URL>.

32. Hodson TO. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci Model Dev [Internet]. 2022 Jul 19;15(14):5481–7. Available from: <URL>.

33. Polat K, Güneş S. Automatic determination of diseases related to lymph system from lymphography data using principles component analysis (PCA), fuzzy weighting pre-processing and ANFIS. Expert Syst Appl [Internet]. 2007 Oct;33(3):636–41. Available from: <URL>.

34. İnal M. Determination of dielectric properties of insulator materials by means of ANFIS: A comparative study. J Mater Process Technol [Internet]. 2008 Jan;195(1–3):34–43. Available from: <URL>.

35. Amid S, Mesri Gundoshmian T. Prediction of output energies for broiler production using linear regression, ANN (MLP, RBF), and ANFIS models. Environ Prog Sustain Energy [Internet]. 2017 Mar 7;36(2):577–85. Available from: <URL>.

36. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res [Internet]. 2005 Dec 19;30(1):79–82. Available from: <URL>.

37. Loh W. Classification and regression trees. WIREs Data Min Knowl Discov [Internet]. 2011 Jan 6;1(1):14–23. Available from: <URL>.