ACTA SCIENTIFIC MEDICAL SCIENCES

Volume 3 Issue 7 July 2019

Research Article

Classification based Computation: Nanofibrous Scaffold Architecture using Support Vector Regression Analysis

Sikiru Afolabi Adebileje, Hammed Tanimowo Aiyelabegan, Tajudeen Ayodele Adebileje* and Taoreed Olakunle Owolabi

Tehran University of Medical Sciences-International Campus (TUMS-IC), Iran

***Corresponding Author:** Tajudeen Ayodele Adebileje, Tehran University of Medical Sciences-International Campus (TUMS-IC), Iran.

Received: May 17, 2019; **Published:** June 12, 2019

Abstract

The aim of tissue engineering is to produce tissues/organs which looks like their natural counterparts with the help of good scaffolding architecture. The scaffold is a 3D artificial substrate for cells that serves as a template on which new tissue is regenerated. Nanofibrous scaffolds mimic the structural features of the extracellular matrix which provide cues to a regenerate tissue with the use of collagen fiber. The collagen fiber structure which is noted for cell attachment, migration, proliferation and differentiation in tissue culture serves well in estimating the band intensity of nanofibrous scaffolding architecture. In this paper, we have presented a machine learning approach through the use of support vector regression analysis in estimating the band intensity of nanofibrous scaffolding. The SVR was built, trained and validated using some experimental values of band intensity from the literature and the accuracy of 76.63% and 99.51% were obtained during the training and testing phase respectively.

Keywords: Tissue Engineering; Scaffolds; Nanofibrous; Band Intensity; Support Vector Regression

Introduction

Langer and Vacanti [1], the pioneers of Tissue Engineering (TE), defines it as a cross-disciplinary field that employs life sciences and engineering principles toward the development of biological replacements that maintain, restore, or improve biological tissue function or a whole organ. TE is a promising method of resolving transplantation challenges which include xenograft rejection and shortage of donor tissues (organs) [1-3].

In the field of TE, extracellular matrix (ECM) scaffold has an important role in aiding cell adhesion, migration, proliferation, differentiation, neo tissue generation, and three-dimensional (3D) organization. This scaffold is a 3D artificial substrate for cells, it serves as a template on which new tissue is regenerated, it should be biodegradable (decompose at a controlled manner without leaving anything foreign in the body) and should finally be replaced by the cell-produced by ECM.

Collagen; the main constituent of mammalian connective tissues, is found in every major tissue that needs flexibility and strength such as skin and bone. Collagen proteins are characterized by a unique triple-helix formation which extends over a large portion of the molecules. So far, about 25 different collagen alpha chains had been described, where each were encoded with a separate gene [4] and the most common is type I. The characteristics of type I collagen molecule contains a long, stiff, and triple-stranded helical structure, where three of the collagen polypeptide chains are inter-wound on one another in a ropelike superhelix surface. After the secretion of collagen into extracellular space, these molecules then assemble into higher-order polymers know as collagen fibrils, and the fibrils are then amassed into collagen fibers possessing 50 to 500 nm diameter range.

Collagen is the main ECM component, type I collagen is widely used in the fabrication of biomaterials, particularly for soft tissue repair [5]. Collagen is relatively bioinert [6] due to its well-conserved primary sequence and helical structure, The use of collagen as an ECM material is yet to be fully explored to overcome the difficulties in handling the problem of pathogen transmission, and less control over the biodegradability, mechanical properties, and batch-to-batch constancy of natural materials from biological sources [7].

One of the essential goals of TE is to produce tissues/organs which looks like their natural counterparts. One best technique towards "ideal" scaffold fabrication is the biomimetic methodology. Collagen fiber structure has long been noted for cell attachment, migration, proliferation, and differentiation in tissue culture [8]. Cell attachment, growth, and migration, on the polymer surfaces, are believed to be assisted by proteins, either secreted by the cells or adsorbed from serum proteins. Woo., *et al*. 2003 [9] were the first to report that nanofibrous architecture built on 3-D scaffolds to improve the characteristics of protein adsorption, promoting cell interactions with scaffolds. From their work, it is known that scaffolds possessing nanofibrous pore walls adsorbed serum proteins four times more when compared to solid pore walls scaffolds. In addition, the nanofibrous architecture selectively mediated protein adsorption such as vitronectin and fibronectin, even though both scaffolds were fabricated from the same PLLA material. According to the result of their study, nanofibrous band intensity scaffolds showed 1.7 times osteoblastic cell attachment as compared to solid pore walls scaffolds. Their results also demonstrated that the biomimetic nanofibrous with estimated band intensity architecture serves as superior scaffolding materials for TE [9].

Support vector machine (SVM) is a machine learning techniques based on a statistical theory with some good beneficial features. SVM is good enough to model nonlinear relationships between variables using different types of kernels and they produce globally optimum results by solving a convex optimization problem [10]. The application of SVM cut across major areas like classification [11] and regression problems [12]. Support vector regression (SVR), a subfield of machine learning, is a computational method that tackles most real-life problems using artificial intelligence principles. Its great predictive ability is employed in tackling various problems in the medical field [28], material science [13-17], oil and gas industries [15,16] and response prediction in literature. This study aims to acquire a pattern that exists between the proteins band intensity of nanofibrous PLLA scaffolds and adopts the acquired pattern for future estimation of the unknown protein adsorption and cells attachment. The good predictive and generalization ability of SVR to solve numerous problems and in addition to having an accurate, direct, and effective way of predicting protein adsorption that contributes to cell attachment with the use of nanofibrous poly (l-lactic acid) (PLLA) scaffolding architecture serves as motivation for carrying out this research work.

Proposed Method

This research work uses SVR derived from the learning theory of support vector machine which was proposed by Vapnik for the sole aim of classification [21]. The SVR uses this theory to develop a machine learning model through which the band intensity of nanofibrous scaffold was estimated. The ϵ-insensitive loss function employed in SVR does not only controls the flatness of generated pattern but also maximize the tolerable deviations of the targets from the estimated values for all training dataset under the consideration with the number of samples. Equation (1) represents a linear function in which ‹w, x› denotes the dot product in the space of R'.

$$f(x,a) = \langle w, x \rangle + b \qquad (1)$$

where w ∈ R' and b ∈ R.

To ensure from equation (1) that the goal of flatness in SVR is reached, a small value of w is desired through the minimization of the Euclidean norm $\|w\|^2$ which makes the optimization problem of the regression looks like the one described in equation (2)

$$\text{minimize } \frac{1}{2}\|w\|^2 \text{ subject to } \left\{ \begin{array}{l} y_i - \langle w, x_i \rangle - b \le \varepsilon \\ \langle w, x_i \rangle + b - y_i \le \varepsilon \end{array} \right\} \qquad (2)$$

Existence of a function that is capable of providing error which is less than ε for all training pairs of the dataset is the condition under which equation (2) holds. The slack variables (ξ_i and ξ^*_i) are introduced in order to create room for another kind of error that may arise while dealing with real-life problems. Therefore, equation (3) is modified and presented in equation (4)

$$\text{minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k}(\xi_i + \xi_i^*) \text{ subject to } \left\{ \begin{array}{l} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \forall i = 1,2,...,k \end{array} \right\} \quad (3)$$

The problem of optimization in equation (3) is solved in dual space representation. However, Lagrangian ($\eta_i, \eta_i^*, \lambda_i$ and λ_i^*) are invoked to transform the problem into a well simplified dual space representation. Therefore, the Lagrangian from equation (3) is presented in equation (4) below

$$L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k}(\xi_i + \xi_i^*) - \sum_{i=1}^{k}\lambda_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \quad (4)$$

$$-\sum_{i=1}^{k}\lambda_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) - \sum_{i=1}^{k}(\eta_i\xi_i + \eta_i^*\xi^*_i)$$

The Lagrangian function with saddle point defined in equation (4) is easily located by equating the partial derivatives of the Lagrangian (with respect to *w,b*, ξ_i and ξ^*_i) to zero. These mathematical transformations give rise to the expression presented in equation (5), (6), and (7).

$$w = \sum_{i=1}^{k} (\lambda_i^* - \lambda_i) x_i \qquad (5)$$

$$\eta_i = C - \lambda_i \qquad (6)$$

$$\eta_i^* = C - \lambda_i^* \qquad (7)$$

The optimization equation is maximized by simply substituting equations (5-7) into (4) to give equation (8)

$$\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}(\lambda_i^* - \lambda_i)(\lambda_j^* - \lambda_j)\left(x_j \middle| x_i\right) - \varepsilon\sum_{i=1}^{k}(\lambda_i^* + \lambda_i) + \sum_{i=1}^{k} y_i(\lambda_i^* - \lambda_i) = 0$$

$$\text{subject to } \sum_{i=1}^{k}(\lambda_i^* - \lambda_i) = 0, 0 \le \lambda_i^* \text{ and } \lambda_i \le C \quad (8)$$

The solutions (λ_i^*, λ_i) obtained from equation (8) were also substituted into equation (1) and presented in equation (9)

$$f(x,\alpha) = \sum_{i=1}^{k}(\lambda_i^* - \lambda_i)\langle x_i, x\rangle + b \qquad (9)$$

The idea of Kernels function is useful in SVR algorithm for solving non-linear problems in which data is mapped into a higher dimensional feature space. The regression function in this feature space can be written as shown in equation (10) which includes the kernel function $K\langle x_i, x\rangle$

$$f(x,\alpha) = \sum_{i=1}^{k}(\lambda_i^* - \lambda_i)K\langle x_i, x\rangle + b \qquad (10)$$

The variables in kernel function control the structures of most high dimensional feature space that measures the complexity of the final solution. Equations (11-14) describes most kernel functions obtainable in literature [17]. These kernels are Polynomial, Linear, Gaussian, and Sigmoid functions respectively which are described by Equation (11-14).

$$K(\vec{x_i}, \vec{x_j}) = (\vec{x_i}.\vec{x_j} + 1)^d \qquad (11)$$

$$K(x_i, x_j) = x_i^T x_j \qquad (12)$$

$$K(\vec{x_i}, \vec{x_j}) = \exp\left(-\gamma \left\|\vec{x_i} - \vec{x_j}\right\|^d\right) \qquad (13)$$

$$K(x_i, x_j) = \tanh\left(\gamma x_i^T x_j + r\right) \qquad (14)$$

Where $\gamma, r,$ and d represent kernel parameters.

Working Principle of SVR

The SVR adopts the principles of artificial intelligence as applied to SVM in its operation. It aims to learn a generalized pattern from the descriptors and target which helps in predicting an unknown target. The measures of the approximate size of protein in kDa are the property of protein adopted as a descriptor for developing the SVR model through which the band intensity of nanofibrous can be estimated.

The algorithm of SVR is made up of variables which are to be varied and tuned by the user until the desired performance is achieved from the model. The regularization factor is one of the variables that control a trade-off between the amount to which deviations larger than E is allowed and the flatness of the acquired pattern [20]. It can also be referred to as a penalty factor that has a wide limit of variation and controls the model's fitness. The Epsilon, kernel option and hyper-parameter are among the variables of SVR that affect the performance of the model. The maximum tolerable deviations of the all estimated values from the target values are well represented by the epsilon. Hyper-parameter minimizes the error of the model by selecting a good hyper-plane while the structure of a high dimensional feature space that controls the complexity of the developed model is determined by kernel options.

For a developed SVRM, the training period of the model entails learning and acquiring patterns that need to be generalized for future estimation of an unknown target. The generalization of each of the acquired pattern during the training period of the model is achieved by the model in the course of computing each generated target with the actual value so as to ensure the generalized pattern. The accuracy, the fitness of the generalized pattern and efficiency can be validated through testing the unknown values of the target values in which the trained model employs the acquired pattern during the training period to evaluate the unknown target using input descriptors.

Performance evaluation of the developed model

The generalized performance of the developed model was evaluated using correlation coefficient (CC), root means square error (RMSE) and absolute error (Ea). These parameters were respectively obtained through equation (15), (16), and (17).

$$cc = 1 - \left[\sum_{i=1}^{n} \frac{e_i}{E_{\exp}^2}\right] \qquad (15)$$

$$rmse = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2\right)} \quad (16)$$

$$Ea = \sum_{i=1}^{n} |e_i| \quad (17)$$

Where e_i, E_{exp} and n represent error (the difference between the experimental and estimated data), experimental and the number of data point respectively.

Empirical study

Description of the dataset

The development of the SVRM through which the band intensity of nanofibrous was estimated employs fifteen experimental values of the band intensity of nanofibrous given the corresponding approximate size of the protein. The descriptors and target were drawn from the literature [9] and presented in Table 1. Statistical analysis was carried out on the dataset and the results are showed in Table 2. The correlation coefficients presented in Table 2 shows a negative correlation which shows that relationship between the chosen descriptors and the target is best learned by support vector regression.

Approximate Size of Proteins (kDa)	Band Intensity (nanoporous)
180	0.107
150	0.118
120	0.146
105	0.053
90	0.086
81	0.196
75	0.104
70	0.983
68	0.259
60	0.25
55	0.257
45	0.111
40	0.073
26	0.123
25	0.285

Table 1: Dataset used for modeling SVRM.

	Approximate Size of Proteins (kDa)	Band Intensity (nanoporous)
Mean	0.21	79.33
Median	0.12	70
Standard Deviation	0.23	44.14
Maximum	0.98	180
Minimum	0.05	25
The correlation coefficient is -0.18		

Table 2: Statistical Analysis.

Computational methodology

This research work utilizes MATLAB computing environment for training and testing the SVR through which band intensity of nanofibrous was estimated. The MATLAB environment was also made used while validating the developed model for determining band intensity. The dataset for developing SVRM was normalized and reshuffled purposely to enhance efficient computations. The normalized dataset was further split into training and testing phase in the ratio of 8 to 2 (which means 80% of the fifteen data-point were used to train the SVR while the remaining 20% was used to test the model). The developed SVRM (well trained and tested using SVR) was used.

Strategy adopted in searching for optimum parameters

The efficiency, accuracy and the fitness of SVR depend greatly on the adopted strategies used in searching for optimum performance of the model. The developed SVRM performs optimally at certain values of regularization factor, hyper-parameter, kernel option and epsilon for a particular kernel function. We optimized each of these parameters using the test-set cross-validation technique where the effect one of the parameters on the performance of the model is determined while others are kept at constant values. The values of SVR parameters through which the developed model achieves its optimum performance are presented in Table 3 and Figure 1 shows the convergence of the GSA for model optimization.

SVR hyperparameters	Optimum Value
C	0.5
Lambda	e-7
Epsilon	0.0002
Kernel option	8
Kernel function	Gaussian

Table 3: Optimum Parameter.

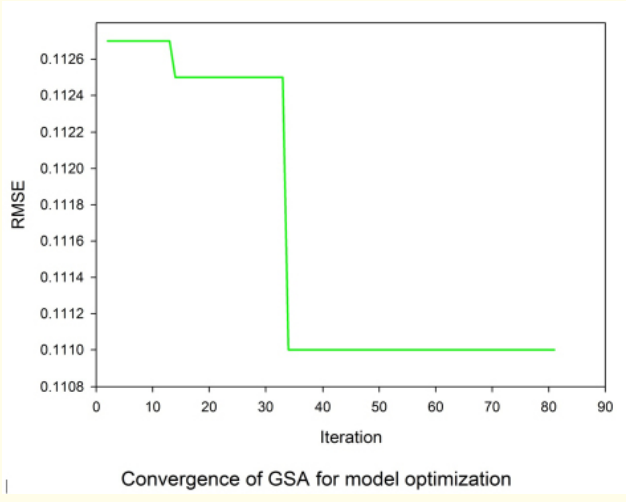

Figure 1: The convergence of GSA for model optimization.

Results

The development of SVRM that was employed in estimating the band intensity of nanofibrous involves training and testing SVR using fifteen values of experimental data of the band intensity. The correlation between the experimental and estimated band intensity of nanofibrous in the course of training and testing the model are presented in Figure 2 and Figure 3 with the correlation of 76.63% and 99.51% respectively as illustrated in Table 4. Table 4 of the developed SVRM is characterized by a high coefficient of correlation (cc), absolute error (Ea) and low root mean square error (rmse). Since correlation shows the degree of similarity or closeness between two variables, these results mean that the band intensity of nanofibrous obtained from SVRM are 99.51% accurate and are close to the experimental values.

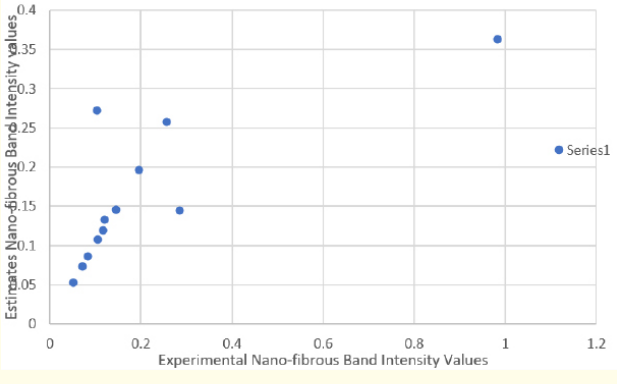

Figure 2: Correlation between experimental and estimated Nano-fibrous band intensity while training SVRM.

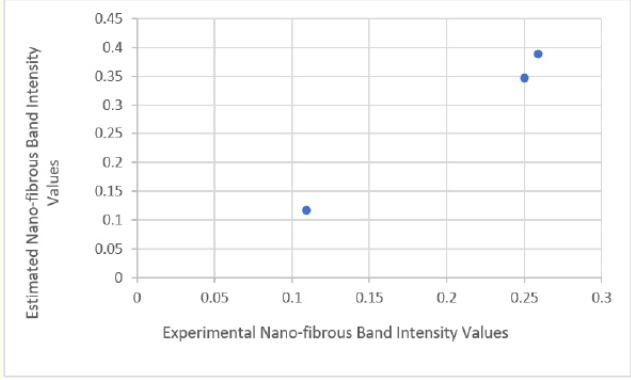

Figure 3: Correlation between experimental and estimated Nano-fibrous band intensity while testing SVRM.

	Training	Testing
cc	76.63	99.51
rmse	0.1902	0.0915
Ea	0.0785	0.0755

Table 4: Performance Evaluation.

Conclusion

We have established a platform for the estimation of the band intensity of a nanofibrous scaffolding architecture using SVRM developed through training and testing SVR with fifteen experimental values of band intensity using the test set cross validation optimization techniques. The SVRM approach which is precise, easy computing, fast and saves time is therefore recommended for the estimation of nanofibrous scaffolding.

Bibliography

1. Langer R and Vacanti JP. "Tissue Engineering". *Science* 260.5110 (1993): 920-926.

2. Nerem RM and Schutte SC. "The Challenge of Imitating Nature". In: Principles of Tissue Engineering: Fourth Edition (2013): 9-24.

3. Adabi M., *et al*. "Biocompatibility and nanostructured materials: applications in nanomedicine". *Artificial Cells, Nanomedicine, and Biotechnology* 45.4 (2017): 833-842.

4. Alberts B., *et al*. "Molecular Biology of the Cell". (2002): 1616.

5. Pachence JM. "Collagen-based devices for soft tissue repair". *Journal of Biomedical Materials Research* 33.1 (1996): 35-40.

6. Anselme K., *et al*. "Tissue reaction to subcutaneous implantation of a collagen sponge. A histological, ultrastructural, and immunological study". *Journal of Biomedical Materials Research* 24.6 (1990): 689-703.

7. Ma PX and Zhang R. "Synthetic nano-scale fibrous extracellular matrix". *Journal of Biomedical Materials Research* 46.1 (1999): 60-72.

8. Turner NJ., *et al*. "Canfield AE.?? 2 (VIII) collagen substrata enhance endothelial cell retention under acute shear stress flow via an?? 2?? 1 integrin-dependent mechanism: An in vitro and in vivo study". *Circulation* 114.8 (2006): 820-829.

9. Woo KM., *et al*. "Nano-fibrous scaffolding architecture selectively enhances protein adsorption contributing to cell attachment". *Journal of biomedical materials research. Part A* 67.2 (2003): 531-537.

10. Boser BE., *et al*. "A Training Algorithm for Optimal Margin Classifiers". Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (1992): 144-152.

11. Smola AJ and Schölkopf B. "A tutorial on support vector regression". *Statistics and Computing* 14.3 (2004): 199-222.

12. Kazem A., *et al*. "Support vector regression with chaos-based firefly algorithm for stock market price forecasting". *Applied Soft Computing* 13.2 (2013): 947-958.

13. Owolabi TO and Gondal MA. "Estimation of surface tension of methyl esters biodiesels using computational intelligence technique". *Applied Soft Computing* 37 (2015): 227-233.

14. Owolabi TO., *et al*. "Estimation of surface energies of hexagonal closed packed metals using computational intelligence technique". *Applied Soft Computing* 31 (2015): 360-368.

15. Owolabi TO., *et al*. "Estimation of surface energies of hexagonal close packed metals using computational intelligence technique". *Applied Soft Computing Journal* 31 (2015): 360-368.

16. Owolabi TO., *et al*. "Modeling of average surface energy estimator using computational intelligence technique". *Multidiscipline Modeling in Materials and Structures* 11.2 (2015): 284-296.

17. Owolabi TO., *et al*. "Development and validation of surface energies estimator (SEE) using computational intelligence technique". *Computational Materials Science* 101 (2015):143-151.

18. Akande KO., *et al*. "Investigating the effect of correlation-based feature selection on the performance of support vector machines in reservoir characterization". *Journal of Natural Gas Science and Engineering* 22 (2015): 515-522.

19. Olatunji SO., *et al*. "Predicting correlations properties of crude oil systems using type-2 fuzzy logic systems". *Expert Systems with Applications* 38.9 (2011): 10911-10922.

20. Pal M and Deswal S. "Support vector regressionbased shear strength modelling of deep beams". *Computers and Structures* 89.13-14 (2011): 1430-1439.

21. Cawley G C. "MATLAB support vector machine toolbox" (2000).

Volume 3 Issue 7 July 2019