MLP kernel-based to predict the optimal conditions of transglutaminase on protein polymerization

Zengyan Peng
School of Computer Science and Intelligence Education, Lingnan Normal University
Zhanjiang, Guangdong, 524048, China
pengzy@lingnan.edu.cn

Shiang-Liang Chen
Department of Animal Science, National Chiayi University
Chiayi City 600, Taiwan
shiangliangchen@mail.ncyu.edu.tw

Miao-Hsin Hsu
Department of Animal Science, National Chiayi University
Chiayi City 600, Taiwan
miao.king533@gmail.com

Dong-Meau Chang
School of Computer Science and Intelligence Education, Lingnan Normal University
Zhanjiang, Guangdong, 524048, China
morganch@me.com

Chun-Chi Chen *
Department of Animal Science, National Chiayi University
Chiayi City 600, Taiwan
kennath@mail.ncyu.edu.tw

Abstract

The research used the MLP kernel-based numerical simulation and estimation method. To analyze the effect of transglutaminase (TGase) on goat milk protein. Find the optimum temperature, the optimum amount of TGase to add, and the optimum reaction time. The experimental method was to remove the fat by high-speed centrifugation at 10,000 xg, 20 min, and 4 °C in goat milk and add different contents of TGase (0, 0.25, 0.1, 0.2, and 0.3 g). Then, a sample with a content of TGase was reacted at different temperatures (30, 40, and 50 °C) and different reaction times (1, 2, and 3hr), and then SDS gel electrophoresis was performed. The cross-linking conditions at different concentrations were observed. After the reaction, SDS gel electrophoresis was carried out, and at the end, it was fixed with a fixative solution and stained with a staining solution, and finally, the results were obtained by destaining. To get the best translation repair, an artificial neural network based on the MLP kernel was used as the estimation engine. Using the experimental results as simulation data, parameters such as optimal experimental temperature, TGase dosage, and experimental duration were calculated. The experimental results showed that the optimal conditions were the strongest at 47 °C, 0.1 g/mL, and 1 hr, with TG, and had the greatest effect on casein in goat milk. TGase acts most deeply on $\kappa - casein$, making casein polymerization more pronounced. This was consistent with the estimation conclusion of the numerical simulation.

1. Introduction

Many machine learning models have used a multilayer perceptron classifier (MLP) as the kernel. An MLP has multiple neurons, each of which was activated using a set of activation functions such as the ReLU, sigmoid, and tanh

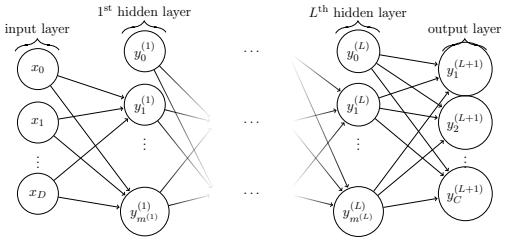

Figure 1. Network graph of a (L + 1)-layer perceptron with D input units and C output units. The lth hidden layer contains m(l) hidden units.

functions. MLP was a state-of-the-art multiple regression model with kernels. The structure of MLP is shown in Figure 1. Many researchers have used the regression MLP kernel as an estimation model in machine learning. Krithik et al. [1] reported a novel MLP model that improved a mixture model of a random forest classifier with a multilayer perceptron regressor. The model was used for real-time rainfall prediction and the prediction accuracy was improved to 0.92537. In addition, Maheshwari and Lamba [2] compared the differences in estimates of honey production among different regression algorithms. Mahajan and Kotecha [3] demonstrated a hybrid model for network traffic prediction and wireless mesh networks. They also compared the prediction results of several different algorithms. Maheshwari and Lamba [2] used the MLP model to predict air quality. Their experimental results indicated the accuracy of the model. Aboubakar et al. [4] used different algorithms to compare power consumption estimation models and verified the predictive performance and accuracy of different models from experimental values. In common, ones include fresh goat's milk, yogurt, cheese, and goat's milk powder. The different processing methods affected the product characteristics of goat milk dairy products. Transglutaminase (TGase) improved or modified the characteristics of proteins. For example, the TGase enzyme catalyzed the hydrolysis of amine residues in proteins [5]. Further, there was improved flavor and texture of dairy products and an increase in the nutritional value of goat milk. Often used to improve the processing characteristics of sheep dairy products. Glutamine aminotransferase was extracted from animals, plants, and microorganisms [6]. Glutamine transaminase, also known as transglutaminase (TGase), is a protein composed of 331 amino acids. The molecular weight was about 38,000 which improved or modified the molecules of protein polypeptides and used covalent bonding reactions to make molecules and molecules combinations or connections. There was modified the structure and function of protein, such as: improving the properties of protein foaming, emulsification, emulsion stability, heat stability, water retention, and gelling ability, thereby improving the sensory characteristics, texture, and appearance of food [7]. The main goal of this study was to use numerical simulation to find the optimal experimental conditions for mTGase experiments on goat milk conversion rate under different conditions. The experimental results showed that the optimal conditions were the strongest at 47 °C, 0.1 g/mL, and 1 hr, with TG, and had the greatest effect on casein in goat milk. TGase acts most deeply on $\kappa - casein$, making casein polymerization more pronounced. This was consistent with the estimation conclusion of the numerical simulation.

2. Materials and Methods

2.1. Preparation of samples (goat milk pretreatment)

The raw goat milk used was obtained from the farm of the Department of Animal Science, Chiayi University (Chiayi, Taiwan). Centrifuge raw goat milk at 5,000xg for 15 minutes to remove fat and store in a refrigerator at 4°C.

2.2. Sampling of samples

To complete the sampling under various experimental conditions, the experiment was divided into two groups according to different temperatures and different mTGase doses. First, samples were taken to examine the effects of different times (0, 1, 2, and 3 hours) and temperatures (30, 40, and 50°C) on goat milk protein polymerization. The experimental sampling steps were the following: 1. Add 0.3 g of TGase to 1 mL of goat milk, then let stand at 30°C for 0, 1, 2, and 3 hours respectively. Let stand at 40°C for 0, 1, 2, and 3 hours respectively. Let stand at 50°C for 0, 1, 2, and 3 hours respectively. 2. When preparing the goat milk sample, take out 50μL goat milk and mix it with 150μL sample Buffer to make a goat milk sample. 3. Goat milk samples from different groups were heated at 95°C for 3 minutes to inactivate TGase and then analyzed by electrophoresis. Secondly, different doses of mTGase 01. Added different doses of 02. When preparing the goat milk sample, take out 50μL goat milk and mix it with 150μL sample buffer to make a goat milk sample. 3. Heat goat milk samples from different groups at 95°C for 3 minutes to inactivate TGase, and then conduct electrophoresis analysis.

2.3. Experimental data simulation

The experimental data was shown in Table 1, which were the density volumes measured by HPLC under different conditions. Different experimental conditions obtained different measurement values. Table 1. The different components estimated by HPLC results under 50°C, 0.3g of TGase

However, to use the MLP kernel model, must have enough data points to make effective estimates. Therefore, the study used the mathematical interpolation method of cu-

Time-(hr)	αs-casein	α2-casein	β-casein	κ-casein	β-lactoglobulin
0	60.4215	0	368.749	39.7965	50.2616
1	26.5257	38.2083	67.0589	0.497384	77.7979
2	28.5757	28.1941	64.5111	0	51.4865
3	27.2134	27.7111	52.4333	0.834663	43.9347

Table 1. The different components estimated by HPLC results under 50°C, 0.3g of TGase.

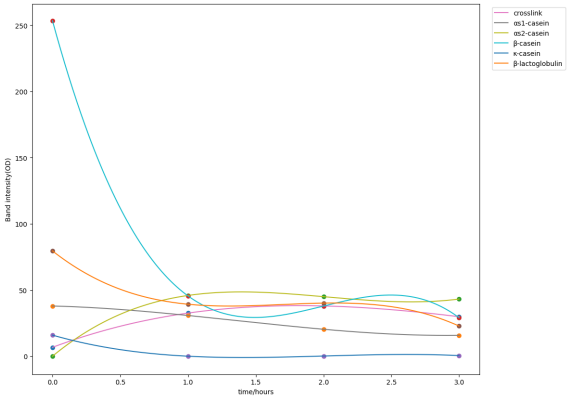

Figure 2. Band intensity(OD) Variation with time at 30 °C.

bic spline to simulate possible experimental values, such as the experimental points in Table 1. The time was subdivided into 400 points, and then the value of each parameter was calculated. The results are shown in Figure 2.

According to the above steps, the simulation values obtained under various conditions were organized into a data table. The purpose of this study was to estimate the optimal TGase concentration, reaction time, and reaction temperature. Therefore, the obtained simulation data, concentration, reaction time, and reaction temperature were taken out as input features. In addition, using the percentage of crosslinks as the label, the MLP model was used for data training and optimization of experimental repair estimation.

3. Results and discussion

3.1. Experimental data at different reaction temperatures and reaction times

As shown in Figure 1, the reaction time, each component, and cross-linking degree were measured at different temperatures and different TGase dosages. Then, the simulation values were calculated from the experimental data. Figures 3 and 4 show the relationship between reaction time and ingredients at different temperatures.

Secondly, temperature was an important influencing factor. Therefore, the reaction time was fixed to simulate the changes in ingredients at different temperatures. The simulation results as shown in Figures 5 to 7.

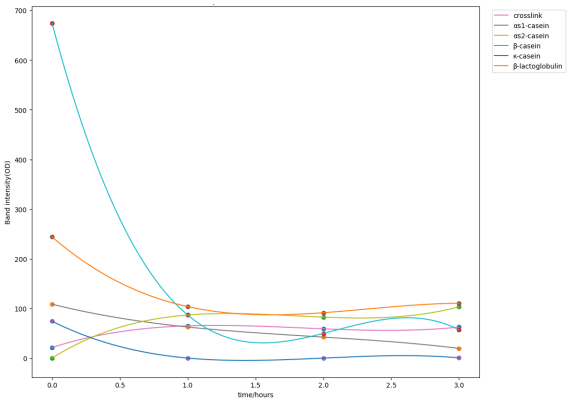

Figure 3. Band Intensity(OD) Variation with time at 40 °C.

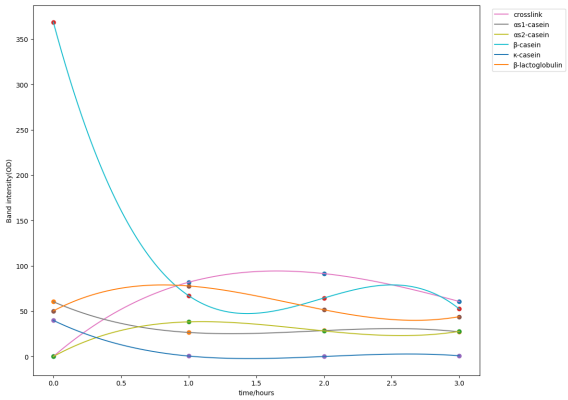

Figure 4. Band Intensity(OD) Variation with time at 50 °C.

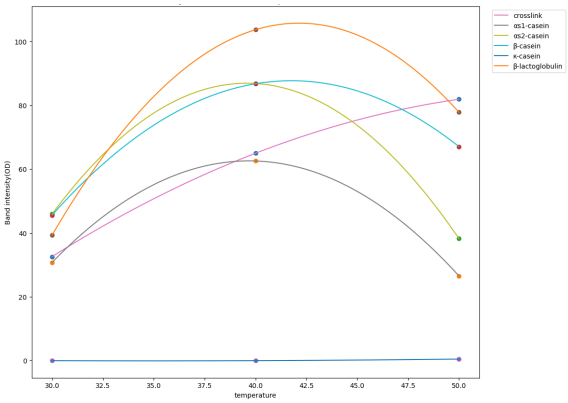

Figure 5. Band Intensity(OD) Variation with temperature after 1hr curation.

3.2. Experimental data of different TGase doses

In addition to reaction time and temperature, TGase dosage was a very important factor. Therefore, it was also necessary to fix the reaction time for different doses and

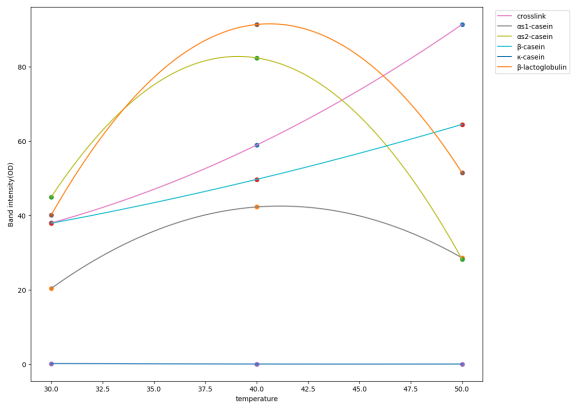

Figure 6. Band Intensity (OD) Variation with temperature after 2hr curation.

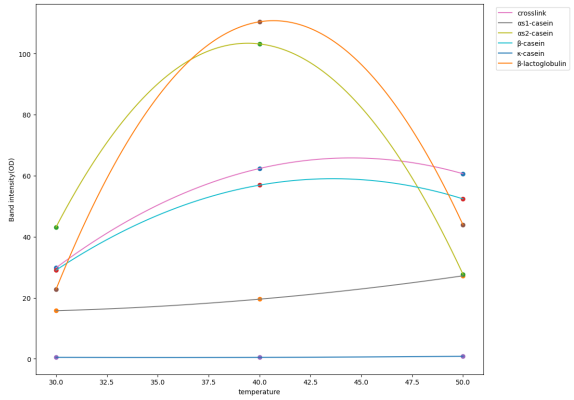

Figure 7. Band Intensity(OD) Variation with temperature after 3hr curation.

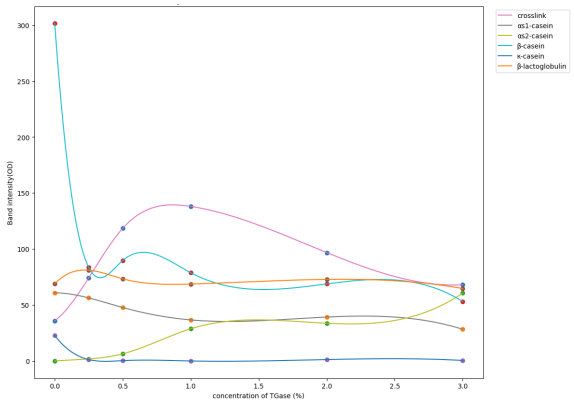

Figure 8. Band Intensity(OD) Variation with concentration of TGase after 1hr curation time.

observe the changes in each component. Then, there was simulated the distribution of different doses based on experimental data. Figures 8 and 9 show the simulation results at different times.

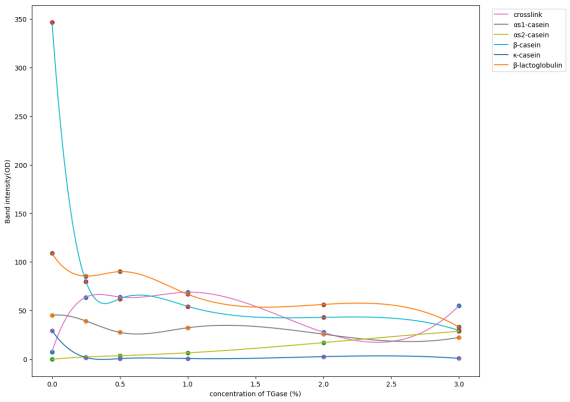

Figure 9. Band Intensity(OD) Variation with concentration of TGase after 2hr curation time.

3.3. Estimation of optimal reaction time, temperature, and TGase dose

To find out the best parameter conditions, first do MLP model training. The model was designed to use the MLP Regressor model, which used 50 hidden layers and the max iterations were 1000. From the calculated value, took out the required features and put them into the input data set, used crosslink % as the label, and split the 3200 data into a training set and a test set, with a ratio of 8:2. The trained model has R2=0.77 and RMSE = 0.08, which was a very good model. Next, there was the need to estimate the optimal parameters. From the experimental data and simulation result graphics, it was found that the reaction time, reaction temperature, and TGase dosage caused the ingredients to reach maximum values in the ranges of 1 to 2 hours, 37 to 47 °C, and 0.5-1.0. Therefore, these three parameter intervals were subdivided, each case was combined, brought into the model to calculate the predicted value, and find the maximum cross-linking degree. The program code is shown in Table 2.

After all comparisons, the optimal conditions were found to be a reaction temperature of 47°C, a TGase dosage of 0.1 g/mL, and a reaction time of 1 hr.

4. Conclusion

The study used MLP kernel-based numerical simulation and estimation methods to analyze the effect of transglutaminase (TGase) on goat milk protein and find the optimal temperature. The optimal amount of TGase to add and the optimal reaction time. The experiment was based on fresh goat milk that was centrifuged at 10,000 xg for 20 minutes at 4 °C to remove fat and then different amounts of transglutaminase (0.025, 0.05, 0.1, 0.2, and 0.3 g). After the basic data was obtained, the cubic spline interpolation method was used to simulate the data under different con-

```
def findOptimize(threshvalue = 0.7):
    optim = []
    value = 0.0
    concs = np.linspace(0.5,1.0, 200)
    times = np.linspace(1.0,2.0, 200)
    temperatures = np.linspace(37.0,47.0,
200)
    for conc in concs:
        for tm in times:
            for tmpt in temperatures:
                testdata = [tmpt,conc,tm]
                testv =
np.array(testdata).reshape(1, -1)
                testval =
extrem_v als.transform(testv)
                val = nn.predict(testval)
                value = val.tolist()[0]
                if value ¿= threshvalue:
                    optim = testdata
                    threshvalue = value
    return optim,threshvalue
```

Table 2. The Python code for estimation

ditions. The artificial neural network based on the MLP kernel was used as the estimation engine to calculate the optimal experimental temperature, TGase addition amount, and experimental duration parameter conditions. The experimental results showed that the optimal conditions were the strongest at 47°C, 0.1 g/mL, and 1 hr, with TG, and had the greatest effect on casein in goat milk. TGase acts most deeply on $\kappa - casein$, making casein polymerization more pronounced. This was consistent with the estimation conclusion of the numerical simulation.

5. Acknowledgement

The study was funded by the scientific research project of Lingnan Normal University(LY1814): "Research on Student Campus Behavior Analysis System Based on Hadoop Big Data Technology". The 2022 Guangdong Province Undergraduate College Teaching Quality and Teaching Reform Engineering Construction Project (Guangdong Jiao Gaohan [2023] No. 4). "Research and reform practice of "Big Data Core Technology" course teaching based on OBE concept in the context of engineering education professional certification" was hereby gratefully acknowledged.

References

[1] S. P. K. P. V. S. Krithik, M. T. A. Raghavan and K. Vasanth. Real time rainfall prediction for hyderabad region using machine learning approach. *2021 Innovations in Power and Advanced Computing Technologies (i-PACT), Kuala Lumpur, Malaysia*, pages 1–6, 2021. 2

[2] K. Maheshwari and S. Lamba. Air quality prediction using supervised regression model. *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India*, pages 1–7, 2019. 2

[3] Harikrishnan R. S. Mahajan and K. Kotecha. Prediction of network traffic in wireless mesh networks using hybrid deep learning model. *IEEE Access*, 10:7003–7015, 2022. 2

[4] I. Quenel M. Aboubakar and A. A. Abba Ari. A predictive model for power consumption estimation using machine learning. *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa*, pages 1–5, 2021. 2

[5] Li WenTai Chang KenLin Kuo MengI Chen ChaoJung Chen ChunChi, Chen LiangYu and Hsieh JungFeng. Influence of chymosin on physicochemical and hydrolysis characteristics of casein micelles and individual caseins. *Nanomaterials*, 11(10):1–10, 2021. 2

[6] Umeda K etal. Ando H, Adachi M. Purification and characteristics of a novel transglutaminase derived from microorganisms. *Journal of the Agricultural Chemical Society of Japan*, 53(10):1–8, 1989. 2

[7] Mc Cullough and F. S. W. Nutritional evaluation of goat's milk. *British Food Journal*, 105(4):239–251, 2003. 2