

# Data Analysis for Hyaluronic Acid Production Using Principal Component Analysis

Medyan Riza<sup>1\*</sup>, Mukhlisien<sup>1</sup>, Andriansyah<sup>2</sup>

<sup>1</sup>Chemical Engineering Department Universitas Syiah Kuala, Banda Aceh, Indonesia

<sup>2</sup>Laboratory of Industrial Computation and optimization Industrial Engineering Department, Universitas Syiah Kuala, Banda Aceh Indonesia.

\*Corresponding author. Email: [medyan.riza@unsyiah.ac.id](mailto:medyan.riza@unsyiah.ac.id)

## ABSTRACT

Hyaluronic Acid is widely used in proprietary pharmaceutical products such as viscoelastic fluids in ophthalmic surgery, vitreous substitutes, and additional products in Visco for orthopedic disorders: osteoarthritis therapy and sports medicine. To produce HA can be through the growth of microorganisms through experiments. However, experiments have limitations that require a lot of resources. This study tries to find information on experimental data using statistical correlation measurement techniques and principal component analysis. The variables observed from the experimental data were time, pH, and total HA production with changes in the initiation of glucose concentration. The results obtained that the glucose concentration of 20% resulted in the amount of HA with a correlation of 99%.

**Keywords:** Hyaluronic Acid, production, glucose, component analysis.

## 1. INTRODUCTION

Biopolymer is an interesting product that can be produced naturally through the growth of microorganisms. Biopolymer applications are currently very widely used in the health and pharmaceutical fields. Most medical devices and pharmaceuticals use the application of biopolymers. Even in the business world, these biopolymers are used, such as in cosmetic equipment. Getting the desired biopolymer characteristics is not an easy thing. Lots of experiments to be done. However, the production of this biopolymer can be done naturally is a viscoelastic biopolymer belonging to by utilizing microorganisms. Hyaluronic Acid (HA) is a viscoelastic glycosaminoglycan family. HA is commonly used in pharmaceutical products and is approved by the FDA for application to health products. One method to produce HA is by fermentation method. Fermentation involves the growth of microorganisms over time. The experiments carried out have limitations and use a large per of resources, and use a large number

Hyaluronic Acid (HA) and is used in proprietary pharmaceutical products such as viscoelastic fluids in ophthalmological surgery and sports medicine. HA and its derivatives have been widely used in it a variety of

applications including, time-release drug delivery systems, in cosmetics as anti-aging and moisture-retaining agents, and neuronal and plastic surgery. HA was isolated for the first time from the bovine vitreous body by aqueous acetone extraction and from the human umbilical cord by chloroform extraction. HA can be produced commercially by two methods: chemical extraction of animal tissue and microbial fermentation of [4]. However, HA was produced by the previous route, mainly from the cock's combs [5].

Experiments are needed to produce data and behavior from the variables that will be seen following the research objectives. In research, HA has an important role in the use of various products today, so studies on increasing the amount of HA significantly are needed. One of these studies can be done by using the latest methods today, such as machine learning and deep learning. This study has been carried out in the [1] research, but there are still shortcomings in this research. Before using methods, Therefore, in research, the approach to prediction would be very good data that has been collected, given treatment, testing, or extracting information first. Another experiment on HA has been carried out by (2) Acidic on degradable HA. Experimental design can be used to find the optimality of this degradation. [3] is also

looking for optimization in HA production. The research was conducted using statistical optimization to find the relationship between prediction and experiment. Modeling accuracy is done by calculating the resulting standard error. The resulting error will be greater if there is no determination of the variables that will be used for modeling. Machine learning algorithms predict based on training, and each algorithm has a basic set of parameters to fit a multidimensional function that can be modified to increase the accuracy of 16 and (7). The deep learning method can predict multiple output properties simultaneously and 7. In this study, information was extracted about the data from the experimental results. Data information is needed as an early stage in decision making. Information was retrieved using statistical measures such as the correlation between variables and principal component analysis (PCA) methods. The results of extracting information affect the determination of the data ready to be predicted or re-experimentation is needed. PCA aims to determine variables that have a strong relationship so that in forecasting, predictor variables can be determined

## **2. METHODS**

The philosophy for the experiment has been arranged cautiously and expressed as just as could be expected, to address the issue for a minimized, economical, effective work base on broad references and logical articles (van Dam, 1992 and Case, and Johnson. 1984). The philosophy exemplifies the systems and methods, which are fundamental and broadly pertinent in research center examination of their inclination and employments. In this way, the substance remembers segments for cleansing, culture conservation, sets up the fermenter framework, aging and division Interaction, and methodology for examination test and estimation information. The overall schematic portrayal of the exploration philosophy for the creation of HA.

### **2.1. Fermentation Process**

The accompanying segments incorporate planning complex media, set up the fermenter with its adornments and fringe research center gear, the readiness of inoculum for the fermenter, disinfection of media and fermenter. The test considerations were acted in a 500-ml Shake carafe and a 2 liters fermenter with the method of fermenter tasks group and took care of a bunch utilizing Micro DCU 200 framework (B.Braun Biotech International GmbH).

### **2.2. Shake Flask Fermentation**

Fermentation was conveyed with 150 ml of the Complex medium contained (in g/l): Glucose (Ajax), 20,

Yeast remove, 10; MgSO<sub>4</sub>. 7H<sub>2</sub>O, 0.5; Na<sub>2</sub>HPO<sub>4</sub>. 12H<sub>2</sub>O, 25) in a 500 ml erlenmeyer carafe. The media were sanitized at (121 °C for 20 min). Glucose arrangements were autoclave independently and permitted to cool and added aseptically to the disinfected medium not long before immunization. In the wake of autoclaving the unadulterated mucoid settlements as *Streptococcus Equi subs Zoopidemicus* (ATCC 39920) were aseptically immunize with sterile circle from SBA inclination or SBA Petri plate into 250 ml complex medium in each a-500 ml Erlenmeyer jar and close each the Erlenmeyer flagon with a freely fitting cap or cotton plug, (11 carafes incorporate 1 jar clear for control). All cups were hatched at 37°C on turning shaker with 150 rpm for 24 hours and the pH of the maturation stock was not kept up with. The points of shake jar maturation to examine the impact of glucose focus and age culture on the development of *Streptococcus Equi subs Zoopidemicus* and the creation of HA, change of pH, and to decide dry cell weight standard bend.

### **2.3. Software and Analysis**

To obtain information, statistical correlation calculations and PCA were performed using the R software. R is currently one of the most popular software in modeling and data mining. Besides being open source, there are many modules provided by this software. The analysis was carried out by looking at the variables of pH and total HA with time changes. The initiation variables used were glucose concentrations of 20 g/l, 40 g/l, and 60 g/l.

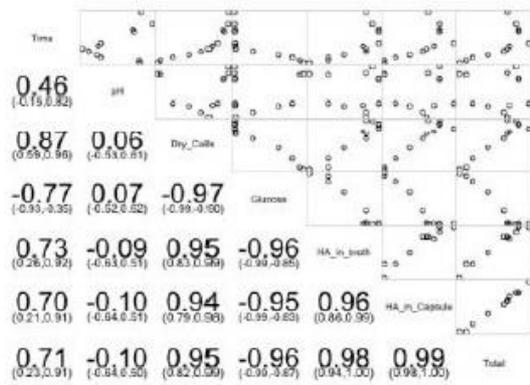
## **3. RESULTS AND DISCUSSION**

The data used in this study are the same data from the [1] study, but in this study, there are additional data with a glucose concentration of 60 g/l. The Table 1 is the data from the shake flask experiment with a glucose concentration of 60 g/l. This data is additional data from the experimental results that have been carried out in the study.

In the Figure 1, you can see the correlation between time, pH, and total HA. A strong correlation can be shown by the pH and time variables. This means that the time variable affects the amount of HA at a glucose concentration of 20 g/l. Figure 4 is a graph of PCA at a glucose concentration of 20 g/l. From this figure, it can be explained that the close relationship between all variables at this concentration can ensure that the time variable can be used as a predictor to determine the amount of pH and the amount of HA in the experiment.

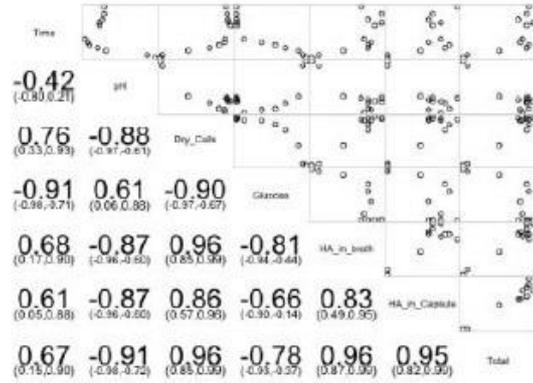
**Table 1.** Data for shake flask at an initial glucose concentration of 60 g/l.

| Time | pH   | Dry cells (g/l) | Glucose | HA in broth | HA in capsule | Total |
|------|------|-----------------|---------|-------------|---------------|-------|
| 0    | 6.84 | 0               | 63.35   | 0           | 0             | 0     |
| 2    | 6.82 | 0.02            | 62.18   | 0.02        | 0             | 0.02  |
| 5    | 5.74 | 0.84            | 58.54   | 0.03        | 0.04          | 0.07  |
| 7    | 5.51 | 1.5             | 55.15   | 0.09        | 0.06          | 0.15  |
| 9    | 5.12 | 2.26            | 45.92   | 0.23        | 0.09          | 0.32  |
| 11   | 5.01 | 2.69            | 37.46   | 0.23        | 0.07          | 0.3   |
| 13   | 4.89 | 2.69            | 24.33   | 0.27        | 0.07          | 0.35  |
| 22.5 | 4.76 | 2.53            | 4.81    | 0.34        | 0.11          | 0.46  |
| 25   | 4.79 | 2.59            | 1.95    | 0.32        | 0.12          | 0.44  |
| 28.5 | 4.98 | 3.07            | 1.69    | 0.27        | 0.15          | 0.42  |
| 31   | 5.15 | 2.92            | 1.3     | 0.31        | 0.09          | 0.41  |

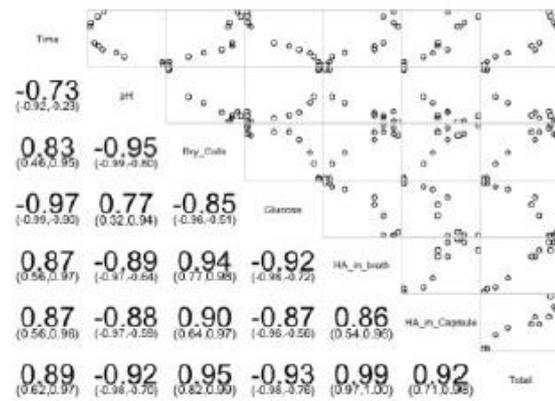


**Figure 1.** Correlation Matrix for Initial Glucose Concentration of 20 g/l

Figure 2, shows the correlation between time, pH, and total HA at a glucose concentration of 40 g/l Compared with the results of a glucose concentration of 20 g/l, this level showed worse results in several variables, especially the variables that were the focus of this study. Visually, this behavior can be seen in the Figure 2. The figure explains the relationship between time and pH away from each other so that the time variable is not good enough if it is used as a predictor variable for pH at this glucose concentration, but it is good enough for the total variable HA.



**Figure 2.** Correlation Matrix for Initial Glucose Concentration of 40 g/l



**Figure 3.** Correlation Matrix for Initial Glucose Concentration of 60 g/l.

Furthermore, Figure 3 shows the best concentration in this experiment. This is evidenced by the relationship between time and pH at a glucose concentration of 60 g/l but produces the lowest relationship in producing the amount of HA. This is acceptable because the correlation value is still above 90%. Visually all the relationships and behavior between variables at this concentration can be seen in Figure 6.

The experimental results also show that the higher the concentration of glucose that is initiated, does not guarantee an increase in the production of pH and the amount of HA. On the other hand, the lower the glucose concentration involved, the better the experiment. To find out the best glucose concentration that can significantly affect the pH value and also increase pH production, more experiments are needed. However, this exploratory data analysis can help in decision-making so that the number of experiments required is less.

#### 4. CONCLUSION

This research is an improvement from the previous research by [1]. In making predictions, it is very important to extract information from the data so that the results obtained are better results with high accuracy. In addition, finding information through data, whether experimental or modding, can help in making decisions that will be used in the future. In this study, the experiment did not stop at the initiation of glucose concentration of 60 only. Based on the analysis that has been done based on this research, the experiment can be carried out again with the initiation of glucose concentrations in the range of 40 gr/l-60 gr/l

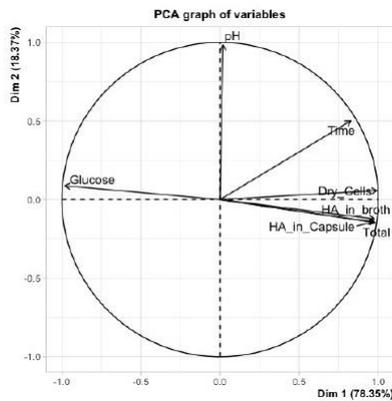


Figure 4. PCA Graph for initial glucose concentration of 20 gr/l

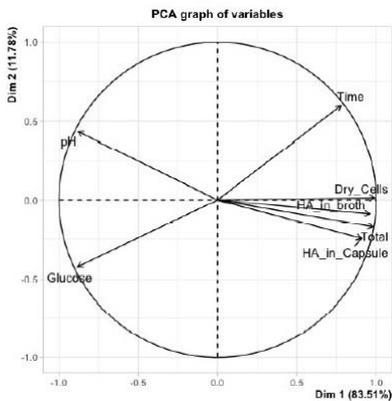


Figure 5. PCA Graph for initial glucose concentration of 40 gr/l

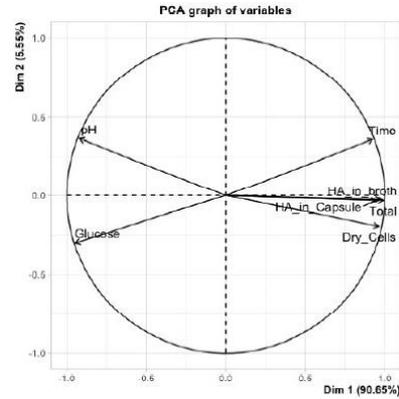


Figure 6. PCA Graph for initial glucose concentration of 60 gr/l

Based on the correlation matrix and the PCA graph shown. For further research, the use of artificial neural networks or other prediction algorithms can be tried again, but specifically for variables that have a high correlation based on research.

#### ACKNOWLEDGMENT

This research is supported by the Universitas Syiah Kuala, Ministry of Education and Culture of the Republic of Indonesia, in accordance with the Letter of Appointment Agreement of Research Professor of Fiscal Year 2021 Number 21/UNI12.1/PT 0103/PNBP/2021. Thanks and high appreciation to Rector and Head of LPPM Universitas Sviah Kuala.

#### REFERENCES

- [1] Riza M, Mukhlisyon, Sentia PD, Andriansyah, Dewi R. 2019 Modeling Bopolymer and Glucose ar Carbon Source Using Artificial Neural Network. IOP Conf Ser. Mater Sci Eng.
- [2] Sandqvist Wedin E. 2019 Optimization of Acidic Degradation of Hyaluronic Acid using Design of Experiments
- [3] Gedikli S. Gungör G. Toptas Y. Sezgin DE, Demirbilek M, Yazıhan N, et al 2018 Optimization of hyaluronic acid production and its cytotoxicity and degradability characteristics. Preparative Biochemistry & Biotechnology. 48(7):610-8
- [4] Crescenzi V. Imbriaco D. Velasquez CL., Dentini M. Ciferri A. 1995 Novel types of polysaccharidic assemblies Macromolecular Chemistry and Physics 196(9):2873-80
- [5] Kno JW, Swann DA, Prestwich GD. 1991 Chemical modification of hyaluronic acid by carbodumides. Bi-conjugate Chem 2(4):232-41.

- [6] Michie D. Spiegelhalter D, Taylor C 1999 Machine Learning, Neural and Statistical Classification Technometrics
- [7] Liu L, Sun J. Xu W. Du G, Chen J. 2009 Modeling and optimization of microbial hyaluronic acid production by
- [8] Streptococcus toorpidemicus using radial basis function neural network coupling quantum-behaved particle swarm optimization algorithm Biotechnology Progress.
- [9] Abdi H, Williams L1. 2010 Principal component analysis WIRE Computational Statistics 2(4):453-59