The Journal of Animal & Plant Sciences, 28(4): 2018, Page: 996-1001
ISSN: 1018-7081

PREDICTION OF 305 DAYS MILK YIELD FROM EARLY RECORDS IN DAIRY CATTLE USING ON FUZZY INFERENCE SYSTEM

O. Gorgulu[1]

[1]Department of Biostatistics & Medical Informatics, Faculty of Medicine, Ahi Evran University, Kırşehir, Turkey

Corresponding author's email:ozkangorgulu@gmail.com; ogorgulu@ahievran.edu.tr

ABSTRACT

In the present investigation, Adaptive Neuro Fuzzy Inference System (ANFIS) was implemented to predict 305 d milk yield using partial lactation records of Jersey dairy cattle. The input variables for the system in the study were age, lactation number and milk yields for the first three test-days. The output variable from the system was 305 d milk yield. ANFIS results related to the milk yields were compared with observed values. Three criteria considered in order to control the reliability of system predictions were the ratio of mean, determination coefficient, and root mean square error. In addition to, the accuracies of ANFIS were compared using the absolute difference between the observed and predicted 305 d milk yield. R^2, RMSE, and RoM values are in the acceptable range. As a conclusion, ANFIS predictions at the beginning of the lactation are related closely to the observed 305 d-lactation yield. The results indicated that ANFIS can be successfully applied for 305 d milk yield early prediction.

Key words: ANFIS, dairy cow, Fuzzy logic, Inference System, 305d lactation yield

INTRODUCTION

Predictions of lactation yield are useful for producers making management and breeding decisions essential for genetic evaluation. Early estimates of 305 d lactation yield are crucial to providing shorter generation interval as well as early culling of bulls and cows, exhibiting low breeding values for milk yield. They also; contribute to farm economics as a result of increasing selection intensity. The ability to predict the complete lactation period of a cow from its part yields would determine the success of dairy herd culling programs (Ranjan *et al*., 2005; Acikgoz *et al*., 2006; Gantner *et al*., 2009).

Different statistical methods such as multiple regression, empirical Bayes method, artificial neural networks, nonlinear auto regressive model and regression tree method based on CHAID and Exhaustive CHAID data mining algorithms were specified to predict 305 d lactation yield from test-day milk yield in dairy cattle (Atil, 1999; Grzesiak *et al*., 2003; Pereira *et al*., 2004; Ranjan *et al*., 2005; Sharma *et al*., 2007; Topal *et al*., 2010; Gorgulu, 2012; Eyduran *et al*., 2013).

Adaptive Neuro Fuzzy Inference System (ANFIS), a hybrid system of fuzzy inference system (FIS) and artificial neural networks (ANN), was used in this study. However, more detailed documentation regarding ANFIS is still needed. Nowadays, FIS have been accepted as a very useful modelling and forecasting technique for complex non-linear structures (Bakhtyar *et al*., 2008). FIS is one of the most widely used applications of fuzzy logic (FL) and fuzzy set theory. Fuzzy inference is the process of developing the mapping from a given input to an output using fuzzy logic which then offers a base from which decisions can be made or patterns perceived. The classical logic has only two truth values, true or false, and so the process of inference is simplified as compared to fuzzy logic, where we have to be concerned not only with propositions but also with their truth values (Kunhimangalam *et al*., 2013). Conventional methods require more time and specific algorithms to solve complex problems, but ANFIS algorithms solve complex problems in a very simple and fast manner. The most important advantage of this system is modelling of linguistic information in deciding what the experts use (Bakhtyar *et al*., 2008).

The purpose of this study was to reveal the usefulness of ANFIS in predicting 305 d milk yield of Jersey cows at the beginning of lactation.

MATERIALS AND METHODS

In this study data on age, lactation number and the first three months test-day milk-yield records of 595 Jersey cows from the KOCAS State Farm (Aksaray, Turkey) were utilized. Modeling of the records was accomplished with the ANFIS editor in MATLAB (Matrix Laboratory). The independent t-test was employed used to compare observed and predicted milk yield values. The Pearson correlation analysis was performed to calculate the relationship between observed and predicted values.

Different fuzzy inference methods have been used in modeling FL. The most widely used methods are Mamdani, Larsen, Tsukamato and Sugeno or TSK (Takagi-Sugeno-Kang). The principles of these methods are generally similar to one another except for the creations of IF-THEN rules (Görgülü, 2007). Sugeno fuzzy inference system (Sugeno FIS), which uses as a hybrid of artificial neural networks (ANN) and FL, was used in the present study.

ANFIS's general architecture is shown on a simple Sugeno FIS with two input and one output (Figure 1). In Sugeno FIS, the conclusion of IF-THEN rules – namely the next section of THEN can be a linear equation of input variables or a fixed value (Fahimifard *et al*. 2009; Wang *et al*. 2009). Rule structure used in this study was a linear equation of input variables. The generated IF-THEN rules structure was as follows:

Rule 1: IF x is A_1 and y is B_1 THEN $f_1 = p_1x+ q_1y+r_1$

Rule 2: IF x is A_2 and y is B_2 THEN $f_2 = p_2x+ q_2y+r_2$

Where x and y are input variables; p, q and r are the results parameters; A and B are membership functions and f is the output function (Vural *et al*., 2009; Esen, 2010). ANFIS basically consists of five layers (Figure 2) each performing different calculations.

Layer 1 (Input Layer): This stage is called fuzzification Fahimifard *et al*. (2009), and allows for this conversion. This step is provided to be converted to a linguistic expression (small, large etc.) of input values with a certain membership degree using the appropriate membership functions. A membership function (MF) allows for the calculation of a fuzzy set membership value, indicating its degree of belonging (Salehi *et al*., 2000; Görgülü, 2007). It can be shown as follow:

$$O_i^1 = \mu_{A_i}(x) \text{, i=1, 2} \qquad (1)$$

Where $\mu_{A_i}(x)$, is the MF of the fuzzy set. The value of this function is called the membership value of the fuzzy set of input-x and it can be shown as $\mu_{A_i}(x) \to [0, 1]$ Nascimento and Ortegan (2002); Pereira *et al*. (2004); Fahimifard *et al*. (2009); Boyacioglu (2010). Various MF can be used in this layer such as triangular, trapezoidal, Gaussian, S, Pi and sigmoidal MF. Gaussian MF was used in this study, and can be shown as follow:

$$\mu_A(x; m, \sigma) = \exp\left\{\frac{-(x-m)^2}{2\sigma^2}\right\} \qquad (2)$$

In Equation 2, m shows the center of the distribution of functions and σ shows the shape of the distribution (Baykal and Beyan, 2004; Görgülü, 2007).

Layer 2: It is called the rule of the stage. IF-THEN rules to be used in FIS are created at this stage. In this stage, the rules are created as ^{in} where p represents the number of membership functions; n represents the number of the input variables.

Two or more logical functions such as AND, OR can be combined using logical connections processors in IF-THEN rules. A threshold value (w_i) is obtained for each rule in this layer. If the rule was created with AND, a threshold value is calculated using Equation 3; otherwise, (with the OR) a threshold value is calculated using Equation 4 (Görgülü, 2007).

$$O_i^2 = w_i = \min[\mu_A(x), \mu_B(y)] \qquad (3)$$

$$O_i^2 = w_i = \max[\mu_A(x), \mu_B(y)] \qquad (4)$$

The number of IF-THEN statements was 243 (3^5) based on 5 input variables (age of the animal, number of lactation, test day 1, test day 2 and test day 3 milk yield) consisting of 3 membership functions (low, medium and high) for each, in the present study.

Layer 3: Layer 3 is defined as the mean or normalization layer. The power of each rule of the threshold value is determined in this layer. For this, a threshold value of the i^{th} rule divided by the sum of the threshold values in other rules (Equation 5) (Subasi, 2007; Rezaeeian *et al*., 2008; Boyacioglu, 2010).

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \text{, i=1, 2} \qquad (5)$$

Layer 4: This layer is called layer results $\bar{w}_i$ value calculated in a previous layer is multiplied by the result equation (Equation 6).

$$O_i^4 = \bar{w}_i\, f = \bar{w}_i(p_i x + q_i y + r_i) \text{, i=1, 2} \qquad (6)$$

Where p, q, and r are the parameter results. These parameters have different values for each rule. The initial value of the parameters is not important. These values will receive different values during continuous training of the network Fahimifard *et al*. (2009); Vural *et al*. (2009).

Layer 5: This layer is called the output layer. Here, a single value is calculated with the aid of Equation 7 Fahimifard *et al*. (2009; Vural *et al*. (2009).

$$O_i^5 = \sum_{i=1}^{2} \bar{w}_i\, f_i = \frac{\sum_{i=1}^{2} w_i f_i}{\sum_{1}^{2} w_i} \text{, i=1, 2} \qquad (7)$$

The predicted values calculated using ANFIS compared with observation values. No matter how close the predicted values observed value means it will be a much smaller error.

During the network training phase, changing the result parameters (p, q, and r) result in recalculating the error value when it is not in the desired limit. They are the values that keep within the limits required parameters of the final error value.

Observed error squares values from the trained network of were used check whether the predictions were close to the RMSE, the RoM and R^2 values. Calculation formulas are as follows: RMSE- Equation 8, RoM- Equation 9 and R^2- Equation 10.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i - p_i)^2} \quad (8)$$

$$RoM = \frac{average\ (p)}{average(o)} \quad (9)$$

$$R^2 = \frac{\sum_{i=1}^{n}(o-p)^2}{\sum_{i=1}^{n}o^2} \quad (10)$$

Where o is: the observed value, p: is the value predicted, and n is predicted the number of observations (Friedrich *et al*., 2008; Fahimifard *et al*., 2009; Vural *et al*., 2009).

Two different methods are used in the training phase of the generated models, including hybrid and back propagation in Sugeno FIS. In this study, the hybrid method was used: Least squares method and the back propagation method (Wang *et al*., 2009).

Before training, network data were for training and testing in ANFIS. The total 595 head of cattle data were randomized such as 60% of the data (357 cows) training data and 40% (238 cows) testing data.

In the study, 5 input variables (age, lactation number, control 1, control 2 and control 3 milk yield) and 1 output variable (305 d-milk yields) were included in the model. A total of 3^5=243 pieces of IF-THEN rules has been established. Created ANFIS's general structure shown in Figure 3.

RESULTS AND DISCUSSION

The cause of variation in milk yield can be breed, season, herd management, health status, lactation number, and parity as well as several genetic and environmental factors (Jovanovac *et al*., 2008; Gantner *et al*., 2009). There was a very significant positive relationship between the number of lactation and 305 d milk yield. Similarly, there was the highest relationship between 305 d milk yield and the age of animal (Gorgulu, 2011). 305-day milk yields were predicted by using partial lactation records (the first three test-day milk yield), age and lactation number of Jersey dairy cattle (Table 1). The selection of the right input variables facilitates accurate predictions with the ANFIS approach.

In the present study, various membership functions and iteration counts were tested. The shape of the MF, x position on the axis and iteration number was found by trial and error path in the model to predict the value closest to the observation value. The closest predictions to the observed values were obtained from the model using the Gaussian membership functions. Sugeno-type fuzzy inference system is specified in the model that has three Gaussian membership functions and 150 iterations (Figure 3).

The results of the statistics for observed and predicted 305 days milk yield from the ANFIS are presented in Table 1. The average predicted milk yield was very close to that the average observed milk yield for training and testing data sets. There was no significant differences between observed and predicted milk yield (p>0.05). The correlations between predicted and observed values were highly compatible for ANFIS.

Table 1. RMSE, RoM and R^2 values of train and test data.

	N	Observed Yield Mean (kg)	Predicted Yield Mean (kg)	RMSE	RoM	R^2
Train	357 (%60)	6590^a	6586^a	0.393	1.00	0.940
Test	238 (%40)	6537^b	6521^b	0.576	0.98	0.865

[a,b] Row means with different superscripts differ significantly at *P* <0.05

RMSE, RoM and R^2 values were used as a goodness of criteria with the aim of evaluating the prediction accuracy of the studied algorithm. The R^2, RMSE and RoM values for training and testing data reflected that the predicted values by ANFIS were in agreement with the observation values (Table 1). The coefficient of determination (R^2) values for test and train data sets were 0.865 and 0.940, respectively. Olori *et al*. (1999) defined that $R^2 \geq 0.70$ implies a very good fit for a model. Another criterion, RMSE values were found to be in an acceptable range. RMSE values of testing and training data were 0.576 and 0.393, respectively. These values are better than several other studies (Grzesiak *et al*., 2003; Ranjan *et al*., 2005; Njubi *et al*., 2010; Gorgulu, 2012. Atil (1999) found the value of R^2 between 0.040 (for the first month) and 0.959 (for the first nine months) in the study which used ratio and regression in order to predict 305 d milk yield. Similarly, Ashmawy *et al*. (1985) using another herd of Friesian cattle in Egypt found that the seven months are the best of predictions 305 day milk yield (R^2= 0.92). In this study, we used the first 3 test days yield (the first 3 month yield) and R^2 value was found 0.94 for train data. Accuracies coefficients such as RMSE, R^2 and ROM showed that ANFIS was more successful than classical prediction methods.

Conclusion: The milk yield in dairy cows tend to rise day by day after calving and reaches a peak in 3 to 8 weeks; then gradually decreases until the dry period. The monthly test-day milk yields were used to predict 305 day lactation yield by using ANFIS. Results of this study showed that the ANFIS was successful in attaining high accuracy in the prediction of 305 d milk yield from an early stage of lactation. It has a much more flexible structure than other prediction methods in the literature and does not require any assumptions, it can be said that ANFIS can be widely used in 305 d milk yield estimations and in forecasting models in very different areas related to animal husbandry. ANFIS is a new method that can be successfully used in the prediction of 305 d milk yield in early lactation. It can be used for obtaining higher accuracy for predicting 305 days milk yields.

Acknowledgements: The authors are grateful to the Kocas State Farm, which was the subject of General Directorate of Agricultural Enterprises.

REFERENCES

Acikgoz, A., A. Kaygisiz, M. Sahin (2006). Estimation of 305 days milk yield using partial milk yield in Holstein cattle. J. Agric. Sci. 12(4): 307–312.

Atil, H. (1999). Ratio and regression factors for predicting 305 day productions from part lactation malik records in a herd of Holstein Friesian Cattle. Pakistan. J. Biol. Sci. 2(1): 31-37.

Ashmawy, A. A., A. S. Khattab and M.K. Hamed (1985). Ratio and regression factors for predicting 305-day production from part-lactation milk records in Friesian cattle in Egypt. Bulletin of Faculty of Agriculture, Cairo University, Egypt.

Bakhtyar, R., A.Y. Bakhtiary and A. Ghaheri (2008). Application of neuro-fuzzy approach in prediction of runup in swash zone. Appl. Ocean. Res. 30(1): 17-27.

Baykal, N., and T. Beyan (2004). Fuzzy logic principles and foundations. Bıcaklar Bookstore, Ankara, Turkey (in Turkish).

Boyacioglu, M.A., and D. Avci (2010). An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange. Expert. Syst. Appl. 37(12): 7908-7912.

Esen, H., and M. Inalli (2010). ANN and ANFIS models for performance evaluation of a vertical ground source heat pump system. Expert. Syst. Appl. 37(12): 8134-8147.

Eyduran, E., I. Yilmaz, M.M. Tariq, and A. Kaygisiz (2013). Estimation of 305-d milk yield using regression tree method in brown Swiss Cattle. The J. Anim. Plant. Sci. 23(3): 731-735.

Fahimifard, S. M., M. Homayounifar, M. Sabouhi, and A.R. Moghaddamnia (2009). Comparison of ANFIS, ANN, GARCH and ARIMA techniques to exchange rate forecasting. J. Appl. Sci. 9(20): 3641-3651.

Friedrich, J.O., N.K.J. Adhikari, and J. Beyene, (2008). The ratio of means as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: A simulation study. BMC. Med. Res. Method. 8(32): 1-15.

Gantner , A., S. Jovanovac, M. Klopcic, M. Cassandro, N. Raguz, and K. Kuterovac (2009). Methods for estimation of daily and lactation milk yields from alternative milk recording scheme in Holstein and Simmental cattle breeds. Ital. J. Anim. Sci. 8(4): 519-530.

Görgülü, Ö. (2007). Study on the fuzzy logic theory and its applicability to agriculture. M.Sc. thesis (unpublished), Department of Zootechnic, University of Mustafa Kemal, Hatay, Turkey.

Gorgulu, O. (2011). Path analysis on effective factors affecting 305-d milk yield in Simmental cattle. Int. J. Agric. Biol. 13(3): 381–385.

Gorgulu, O. (2012). Prediction of 305-Day milk yield in Brown Swiss cattle using Artificial Neural Networks, S. Afr. J. Anim. Sci. 42(3): 280-287.

Grzesiak, W., R. Lacroix, J. Wójcik, and P. Blaszczyk, (2003). A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. Can. J. Anim. Sci. 83(1): 307-310.

Jang, J.S.R. (1993). ANFIS: Adaptive-network-based fuzzy inference systems. IEEE Trans. Syst. Man. Cybern. 23(3): 665–685.

Jovanovac, S., V. Gantner, K. Kuterovac, and M. Klopčič and M. Cassandro (2008). Estimation of daily and lactation milk yield from alternative milk recording scheme. Stočarstvo. 62(2):91-102.

Kunhimangalam, R., S. Ovallath and P.K. Joseph (2013). A novel fuzzy expert system for the identification of severity of carpal tunnel syndrome. Biomed. Res. Int. vol. 2013, Article ID 846780, doi:10.1155/2013/846780

Nascimento, L.F.C., and N.R.S., Ortega (2002). Fuzzy linguistic model for evaluating the risk of neonatal death. Rev. Saude. Publica. 36(6): 686–692.

Njubi, D.M., J.W. Wakhungu, and M.S. Badamana (2010). Use of test day records to predict first lactation 305 day milk yield using artificial neural network in Kenyan Holstein-Friesian dairy cows. Trop. Anim. Health. Prod. 42(4): 639-644.

Olori, V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk (1999).Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. Livest. Prod. Sci. 58(1): 55–63.

Pereira, J.C.R., P.A. Tonelli, L.C. Barros, and N.R.S. Ortega (2004). Clinical signs of pneumonia in children: Association with and prediction of diagnosis by fuzzy sets theory. Braz. J. Med. Biol. Res. 37(5): 701-709.

Ranjan, S.K., M.L. Lakhera, M. Singh (2005). Development of prediction equations for 300 days milk yield from part yields, cumulative part and sequential monthly records in Sahiwal cows. Int. J. Agri. Biol. 7(4): 574-578.

Rezaeeian, A., A.Y. Koma, B. Shasti, and A. Doosthoseini (2008). ANFIS Modeling and feedforward control of shape memory alloy actuators. Int. J. Math. Models. Methods. Appl. Sci. 2(2): 228-235.

Salehi, F., K.M. Lacroix, K.M. Wade (2000). Development of neuro-fuzzifiers for qualitative analyses of milk yield. Comput. Electron. Agric. 28(3): 171–186.

Sharma, A.K., R.K. Sharma, H.S. Kasana (2007). Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modelling. Appl. Soft. Comput. 7(3): 1112-1120.

Subasi, A., (2007). Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction. Comput. Biol. Med. 37(2): 227-244.

Topal, M., V. Aksakal, B. Bayram, and A.M. Yağanoğlu (2010). An analysis of the factors affecting birth weight and actual milk yield in Swedish Red cattle using Regression Tree analysis. The J. Anim. Plant. Sci. 20(2): 63-69.

Vural, Y., D.B. Ingham, M. Pourkashanian (2009). Performance prediction of a proton exchange membrane fuel cell using the ANFIS model. Int. J. Hydrogen. Energy. 34(22):9181–9187.

Wang, W.C., K.W. Chau, C.T. Cheng, and L. Qiu (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J. Hydrol. 374 (3-4): 294–306.

Appendices: This study was presented at the 7. National Animal Science Congress 14–16 September, 2011, published in Turkish proceedings book.

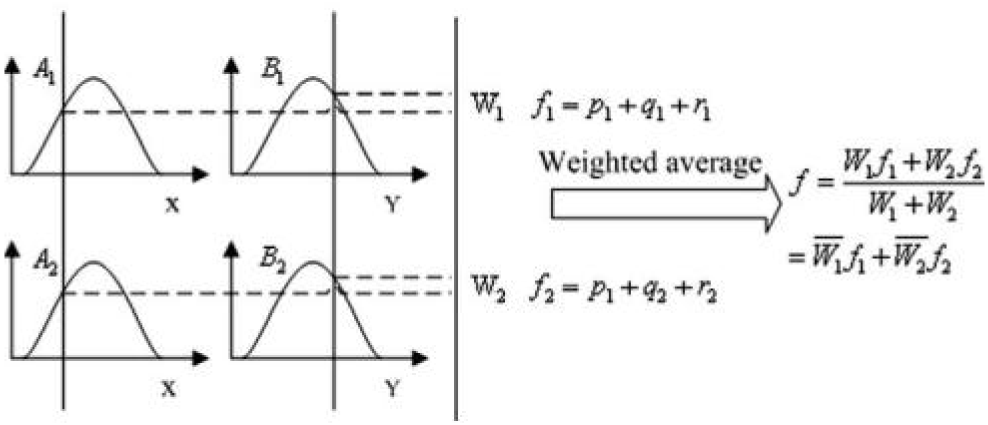

Figure 1. The graphical representation of simple Sugeno FIS with two input variable Jang (1993)

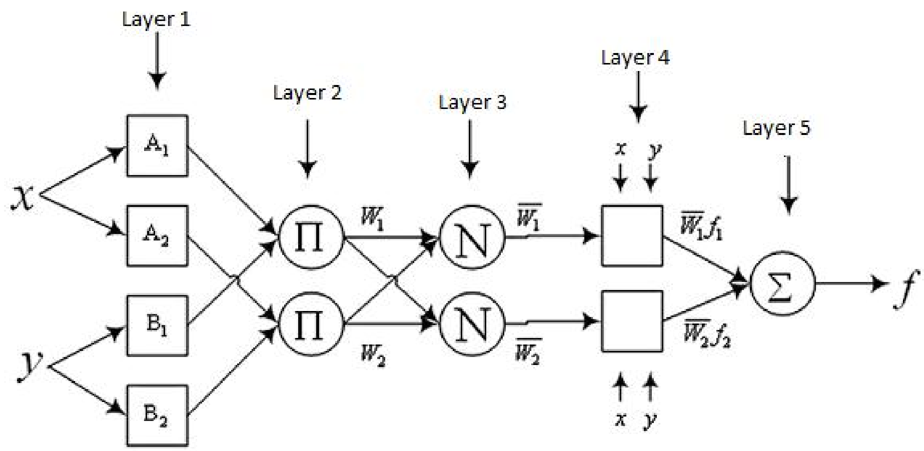

Figure 2. Two input variables ANFIS architecture Jang (1993); Rezaeeian *et al*. (2008

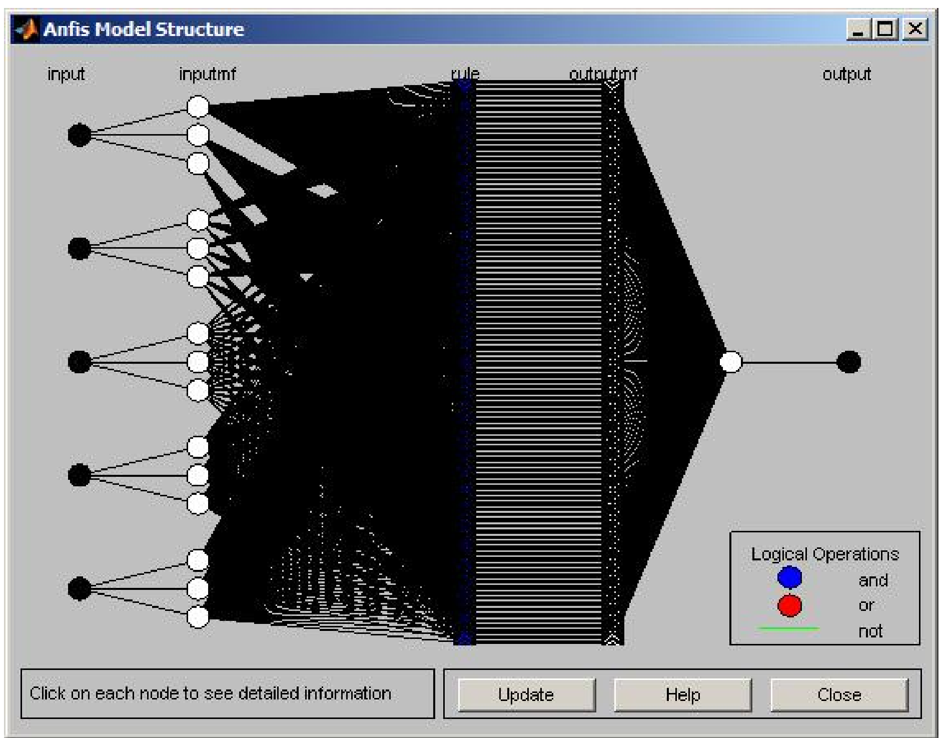

Figure 3. The general structure of ANFIS created for the predicted 305 daily milk yield.