Prediction of Immunoglobulin G in Lambs with Artificial Intelligence Methods

Pınar CİHAN 1,a (*) Erhan GÖKÇE 2,b Onur ATAKIŞI 3,c Ali Haydar KIRMIZIGÜL 2,d Hidayet Metin ERDOĞAN 4,e

1 Department of Computer Engineering, Faculty of Çorlu Engineering, Tekirdağ Namık Kemal University, TR-59860 Tekirdağ - TURKEY
2 Department of Internal Medicine, Faculty of Veterinary Medicine, University of Kafkas, TR-36100 Kars - TURKEY
3 Department of Chemistry, Faculty of Arts and Science, University of Kafkas, TR-36300 Kars - TURKEY
4 Department of Internal Medicine, Faculty of Veterinary Medicine, University of Aksaray, TR-68100 Aksaray - TURKEY

ORCIDs: a 0000-0001-7958-7251; b 0000-0003-2674-1010; c 0000-0003-1183-6076; d 0000-0002-6660-2149; e 0000-0003-1261-4352

Article ID: KVFD-2020-24642    Received: 25.06.2020    Accepted: 06.01.2021    Published Online: 07.01.2021

Abstract

The health, mortality and morbidity rates of neonatal ruminants depend on colostrum quality and the amount of Immunoglobulin G (IgG) absorbed. Computer-aided estimates are important as measuring IgG concentration with conventional methods is costly. In this study, artificial neural network (ANN), multivariate adaptive regression splines (MARS), support vector regression (SVR) and fuzzy neural network (FNN) models were used to predict the serum IgG concentration from gamma-glutamyl transferase (GGT) enzyme activity, total protein (TP) concentration and albumin (ALB). The correlation between parameters was examined. IgG positively correlated with GGT and TP and negatively correlated with ALB (R = 0.75, P<0.001; R = 0.67, P<0.001; R = -0.17, P<0.01, respectively). IgG, GGT, and TP cut-off values were determined for mortality, healthy, and morbidity in neonatal lambs by decision tree method. IgG ≤113 mg/dL (P<0.001), GGT ≤191 mg/dL (P=0.001), and TP ≤45 g/L (P<0.001) were determined for mortality. IgG >575 mg/dL (P=0.02), GGT >191 mg/dL (P<0.001), and TP >55 g/L (P<0.001) were determined for healthy. It has been observed that the FNN is the most successful method for the prediction of IgG value with a correlation coefficient (R) of 0.98, root mean square error (RMSE) of 234.4, and mean absolute error (MAE) of 175.8.

Keywords: Artificial neural network, Decision tree, Fuzzy neural network, Immunoglobulin G, Multivariate adaptive regression splines, Support vector regression

Yapay Zeka Yöntemleri İle Kuzularda İmmüunoglobulin G Tahmini

Öz

Yenidoğan ruminantların sağlığı, ölüm ve hastalık oranları, kolostrum kalitesine ve emilen Immunoglobulin G (IgG) miktarına bağlıdır. Konvansiyonel yöntemler IgG koncentrasyonunun ölçümesi maliyetli olduğundan, bilgisayar destekli tahminler önemlidir. Bu çalışmada, gamma-glutamil transferaz (GGT) enzim aktivitesi, toplam protein (TP) ve albümin (ALB) değerlerinden serum IgG konstantrasyonunu tahmin etmek için yapay sinir ağı (YSA), çok değişkenli uyarlanabilir regresyon eğrileri (MARS), destek vektör regresyonu (SVR) ve bulanık sinir ağı (FNN) modelleri kullanılmıştır. Parametreler arasındaki korelasyon incelemiş ve serum IgG konstantrasyonunun, GGT ve TP ile pozitif, ALB ile negatif korelasyonu olduğu görülmüştür (srasıyla R = 0.75, P<0.001; R = 0.67, P<0.001; R = -0.17, P<0.01). Yenidoğan kuzularda ölüm, sağlıklık ve hastalık için eşik değerler karar açısı yöntemine belirlenmiştir. Ölüm için IgG ≤113 mg/dL (P<0.001), GGT ≤191 mg/dL (P=0.001), ve TP ≤45 g/L (P<0.001) olarak belirlenmiştir. Sağlıklı için IgG >575 mg/dL (P=0.02), GGT >191 mg/dL (P<0.001) ve TP >55 g/L (P<0.001) olarak belirlenmiştir. 0.98 korelasyon katsayısı (R), 234.4 hata kareler ortalamasının karekökü (RMSE) ve 175.8 ortalama mutlak hata (MAE) ile IgG değerini tahmin etmede en başarılı yöntemi FNN olduğu görülmüştür.

Anahtar sözcükler: Bulanık sinir ağı, Çok değişkenli uyarlanabilir regresyon eğrileri, Destek vektör regresyonu, Immunoglobulin G, Karar açısı, Yapay sinir ağı

How to cite this article?
DOI: 10.9775/kvfd.2020.24642

(*) Corresponding Author
Tel: +90 282 250 2456
E-mail: pkaya@nku.edu.tr (P. Cihan)

This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)
**INTRODUCTION**

Ruminants are born agammaglobulinemic due to the placental structure. Because of this placental structure, it largely inhibits the transfer of immunoglobulins (Ig) to the fetus. Failure of passive transfer (FPT) develops when neonatal ruminants do not receive enough and sufficient quality Ig. Ruminants with FPT cannot be protected from infectious diseases, thus getting the disease more and the mortality rate increase [1,2].

The first 28 days of life, which is referred to as the neonatal period, are vital for ruminants. FPT is one of the most important factors affecting the health of ruminants in the neonatal period. FPT causes enteritis and septicemia, especially in the neonatal period, leading to death and mortality. Mortality of ruminants in the neonatal period leads to losses for both the enterprise and the economy [3-6].

Colostrum has high energy, protein and vitamin content and contains a high proportion of immunoglobulin G (IgG) [7]. Colostrum quality is directly proportional to the amount of IgG it contains [8]. Ruminants with high IgG concentrations are reported to have higher survival rates than ruminants with low IgG levels in colostrum [9,10]. IgG levels can be measured by direct radial immunodiffusion (SRID) [11] and Enzyme-linked immunosorbent assay (ELISA) [12] methods. These methods are difficult, expensive, time-consuming and incapable of testing of a large number of samples at once [2,8]. The fast, cheap and reliable prediction of the IgG concentration is required [1,12,13]. Therefore, in the present study, IgG value was predicted by different AI models. A regression is a type of supervised learning problem and it refers to the prediction of a continuous dependent variable from one or more independent variables [14,15].

Recently, regression methods have been used for a wide range of applications in civil engineering [16], environmental science [17], veterinary [18], agriculture [16], and medicine [16]. However, it is known that artificial intelligence (AI) regression models are not used sufficiently in the veterinary field [18,19].

The purposes of this study were (1) to examine the correlation between parameters: Immunoglobulin G (IgG), gamma-glutamyl transferase (GGT), total protein (TP), and albumin (ALB) in neonatal lamb, (2) to determine the most successful/appropriate AI model(s) (artificial neural network, ANN; multivariate adaptive regression splines, MARS; support vector regression, SVR; and fuzzy neural network, FNN) that predict(s) IgG (mg/dL) levels from GGT, TP and ALB values in blood, (3) to determine the cut-off point that is important for the risk of disease and death in lambs for GGT, TP, and ALB.

**MATERIAL AND METHODS**

**Animals**

In this study, using the hematological-immunological data (IgG, GGT, TP and ALB) obtained from the TOVAG 1080847 project [1], which was carried out with the approval numbered KAU-HADYEK-2008-23, serum IgG concentration predicted by artificial intelligence regression methods (ANN, MARS, SVR and FNN). In addition, cut-off values for IgG, GGT, TP and ALB were determined with the machine learning decision tree method. Briefly, 347 Akkaraman lambs on two neighboring farms with similar management practices were included in the study. The lambs were weighed before taking colostrum. After this procedure, the lambs were allowed to suck their mothers naturally for one week. During this time the lambs were not fed with supplemental colostrum. After this period, the lambs were transferred to a separate pen and their mothers were allowed to feed twice a day for three months.

**Blood Sample Collections**

The blood samples were taken from Akkaraman crossbred lambs (n=347). The blood samples were taken 24±1 h after birth, centrifuged at 3000 rpm for 5 min and the serum samples were stored at -20°C till analysis.

**IgG and Other Test Assay**

Serum IgG concentrations were measured using the ELISA kit (Bio-X Competitive ELISA Kit for Ovine blood serum IgG Assay-BIO K350, Bio-X Diagnostics, Belgium). Gammaglutamyltransferase (GGT), total protein (TP) and albumin (ALB) analyses were done using commercial spectrophotometric kit (TML, Tani Medical, Turkey). These tests were performed and the results were calculated according to the manufacturer’s instruction manual.

**Statistical Analysis**

The prediction methodology of IgG (mg/dL) values is as follows: After the pre-processing step was applied (e.g. sigmoid normalization) to the dataset, 80% of the dataset was used for model training and 20% was used to test the model performance. K-fold cross validation technique was used in the model training phase in which the dataset is randomly divided into k equal size of subsamples. One of the subsample is used as the validation data for testing the model and the remaining k-1 subsamples are used as training data. This cross-validation process repeats k-times (fold). ANN, MARS, SVR and FNN models were used for the prediction of IgG values. After the IgG value was predicted by regression models, the model prediction accuracy was measured using the test dataset. The correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE) statistical criteria were used to compare the prediction performance of the regression models.

*Artificial neural networks (ANN)* can be defined as a system designed to model the way the brain performs a function. ANN consists of several ways of connecting artificial nerve with one another and is usually arranged in layers. The input layer includes neurons that receive raw data from
outside. Only the input values are transmitted to the next layer without any processing. The output layer is the layer containing the neurons that transmit the outputs. The input and output layers include a single layer. The hidden layer contains hidden nodes. Hidden nodes have no direct connections to the outside world. They perform calculations and transfer information from input nodes to output nodes. The hidden layer may comprise one or more layers. Activation functions are important for ANN to learn [21]. The parameters of the developed ANN model in the study are as follows: 3 input layers (GGT, TP, ALB), 2 hidden layers and 1 output layer (IgG) was used. Logistic function was used for the activation function, back propagation algorithm was used for training of neural networks, number of epochs was 1000 and learning rate was 0.01.

Multivariate Adaptive Regression Splines (MARS) is a version of the iterative separation method and stepwise linear regression. MARS makes no assumptions about the functional relationship between dependent and independent variables. Instead, it creates relationships between different input variables based on the basic functions and coefficients. MARS establishes a flexible regression model using basic functions corresponding to different ranges of independent variables. In other words, the basic idea is based on the method of divide-and-conquer strategy [22].

Support Vector Machine (SVM) is a kernel-based method used for classification and regression problems. Regression with support vector machines is called support vector regression (SVR). In the case of linear separation of data, the aim is to find the best hyperplane that maximizes the margin. In the case of non-linear separation of data, mapping is performed to the higher space from which the data can be parsed linearly with the help of dataset [23]. The parameters of the developed SVR model in the study are as follows: Polynomial kernel function was used with gamma=0.001 and degree=3. Cost of constraints violation was 1.

Fuzzy Neural Network (FNN) or neuro-fuzzy systems based on combining the ability of artificial neural networks (ANNs) to learn and find the most appropriate with the advantages of fuzzy logic to make decisions like a human and provide expert knowledge [24]. In FNN, ANN combines with fuzzy rule based systems (FRBS). FRBS are based on the fuzzy set theory and proposed by Zadeh [25]. FRBS are also known as fuzzy inference systems (FIS) and fuzzy models. FRBS are used for classification and regression problems. Generally, fuzzy rule based system (FRBS) architecture consists of fuzzification, knowledge base, inference, and defuzzification steps [26,27]. At the fuzzification step, input variables (crispy values) are converted to linguistic term (fuzzy sets) using membership functions. Knowledge base is composed rule base and database components. Rule base contains fuzzy if-then rules and database includes the fuzzy sets. At the inference engine step, the fuzzy output is generated from the fuzzy inputs using by inference model (Mamdani, Takagi-Sugeno-Kang). At the defuzzification step, the fuzzy output of the inference engine is mapped into a numerical output [28,29]. In this study, method type was selected Adaptive-network-based fuzzy inference system. Takagi Sugeno Kang type fuzzy model used for linguistic rules. Max iteration was 5, step size was 0.01 and ZADEH type of implication function was used for a value representing.

In order to evaluate the prediction performance of the methods, it should be examined how much the actual value and the estimated value match. In the present study, R, RMSE and MAE statistical criteria were used to evaluate model prediction accuracy. As a result of testing the prediction methods, it is desirable that the R-value is high and the RMSE and MAE value is low. R (Eq. 1.), RMSE (Eq. 2.) and MAE (Eq. 3.) were determined as follows:

\[ R = \frac{\sum_{i=1}^{n} (a_i - \bar{a})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^{n} (a_i - \bar{a})^2 \sum_{i=1}^{n} (p_i - \bar{p})^2}} \]  
Eq. 1.

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (a_i - p_i)^2} \]  
Eq. 2.

\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |a_i - p_i| \]  
Eq. 3.

Where \( n \) is the number of data, \( a \) is the actual value, \( p \) is the predicted value, \( \bar{a} \) is mean of actual values and \( \bar{p} \) is the mean of the predicted values. In the study, the R programming language was used for statistical computations and AI models development.

**Results**

Correlation between the serum IgG concentration with GGT, TP and ALB are illustrated in Fig. 1. Parameters’ distribution is shown in the diagonal; correlation coefficient and significance levels (P<0.05 *, P<0.01 **, P<0.001 ***) are in the upper triangle; bivariate relationships are shown in the lower triangle of Fig. 1. IgG was significantly and positively associated with GGT (R=0.75, P<0.001) and TP (R=0.67, P<0.001). There was an inversely and low correlation between ALB and IgG (R=0.17, P<0.01). Also, when the distributions of the features in the dataset were examined, it was seen that variables were not normally distributed.

In this study, it was aimed to define a cut-off point for serum IgG concentration, serum TP concentration, and serum GGT activity associated with the risk of death in lambs. The decision tree method was used to determine cut-off levels. Decision tree is a machine learning method used for classification and regression. The decision tree splits the input variables (IgG, GGT, TP) from the cut points where the target variable can be grouped most homogeneously.
(healthy, ill, death). The results are presented in Fig. 2. It was also tried to determine the threshold value for ALB, but it was not possible to find an exact range for ALB.

The IgG threshold value for lambs in the neonatal period was determined as ≤113 by the decision tree method (Fig. 2-A). The mortality rate of lambs with IgG value ≤113 was 100% (11/11) (P<0.001). Lambs with IgG values were in the range of 113-575 with the morbidity rate of 89% (8/9). For the IgG value >575, the healthy, morbidity and mortality rates in lambs were 87.4% (286/327), 12% (39/327), and 0.6% (2/327) (P=0.02), respectively.
In neonatal lamb mortality, the threshold value for TP ≤ 45 (Fig. 2-B). For the TP value ≤ 45, the mortality rate was 71% (10/14) (P<0.001). The morbidity rate was 73% (11/15) when TP was in the range of 45-55 (P<0.001). In the case of TP > 55, the healthy, morbidity, and mortality rates in lambs were 89% (283/318), 10.4% (33/318) and 0.6% (2/318) (P<0.001), respectively.

The GGT threshold value ≤ 191 was determined for the lamb's mortality (Fig. 2-C). The rate of mortality was 85% (11/13), when the GGT value ≤ 191 (P<0.001). For the GGT value > 191, the healthy, morbidity and mortality rates in lambs were 86% (287/334), 13.4% (45/334), and the 0.6% (2/334) (P<0.001), respectively.

ANN, MARS, SVR and FNN models were used to estimate IgG output value from GGT, TP and ALB input values. After the completion of the training process, the predictive performances of the models were tested with test set. Randomly selected 277 samples were used for regression models training and remaining unseen 70 samples were used to predict IgG value. For these unseen 70 samples, IgG values were estimated and compared with the actual IgG values and R, RMSE and MAE metrics were calculated. Actual IgG values and predicted IgG values by regression models are given in Fig. 3. The prediction performance of AI models was presented in Fig. 4.

When comparing the prediction accuracy of the models, the highest R-value and the smallest RMSE and MAE values are desirable. Examining Fig. 4, the FNN model shows superior performance than other models (R=0.98). According to R-value, the FNN model is followed by MARS (R=0.85), SVR (R=0.82) and ANN (R=0.80) respectively. Likewise, according to the statistical results of R, RMSE and MAE, the FNN model produced more successful results than other models. These statistical results indicate that the FNN model is the best for predicts the IgG (mg/dL) concentrations.

**Discussion**

Because of the lambs' placental structure, they are born hypogammaglobulinemic. Hence, ingestion and absorption of maternal antibodies in colostrum is necessary for providing humoral immunity in the neonatal period. This
process is denominated passive transfer and is determined by measuring serum IgG concentrations.

Numerous studies in the last thirty years have associated neonatal diseases with insufficient serum IgG, in other words with FPT in animals, thus demonstrating the importance of IgG in preventing infections and increasing growth performance in neonatal animals [3,4,30-34]. It is known that if the lambs received sufficient volume and quality of colostrum in the first 12 h of life, the adequate passive immune transfer is provided [30,35-38]. Otherwise, a secondary immunodeficiency develops, called Passive Transfer Failure (FPT).

SRID and ELISA tests are used for a measure of serum IgG concentration by directly. But these kits are costly, time-consuming and incapable of testing of a large number of samples at once. Due to these disadvantages, it is urgent and important to predict accurately the IgG value by indirectly with alternative methods. In addition to this method, the SRID or ELISA can be used as a confirmatory diagnosis. GGT and TP are highly correlated with serum IgG concentration and are used to estimate IgG concentration. These indirect methods are fast, practical, inexpensive and easier to apply in the field. Therefore, indirect tests such as GGT and TP can be used for estimation of IgG concentration and direct tests are used only as confirmatory methods. In this study, serum GGT, TP and ALB activities were used as input variables to predict the IgG output variable using ANN, MARS, SVR and FNN model. To the best of our knowledge, there is no such study in the literature that examines and compare regression models for predicting the IgG (mg/dL) value.

There is no global accepted optimal IgG threshold by the veterinary community for FPT in lambs. And it’s known the passive immunity develops when the threshold value of IgG is <500 when ELISA used [35]. In the present study, a significant linear association was detected between TP and IgG (R=0.67, P<0.001). The cut-off level with TP≥45 g/L (P<0.001) were determined for morbidity (rate is 89%). TP>55 g/L (P<0.001) were determined for healthy (rate is 87.4%). IgG values 113-575 mg/dL range (P<0.001) were determined for morbidity (rate is 73%).

Albumin, which is synthesized in the liver and constitutes 50% of plasma proteins, is reported to be the highest stored and largest amino acid carrier. A negative and low correlation was observed between ALB and IgG concentration (R=-0.17, P<0.01). In this study, the specific cut-off point for ALB could not be determined.

This study presents the following two novel contributions. First, the prediction of serum IgG concentration from blood samples (GGT, TP, and ALB) is possibly using artificial intelligence methods. IgG concentration is measured by commercial kits to give an idea about the health status of ruminants. Measurement of IgG concentration with these kits is difficult, expensive, time-consuming and incapable of testing of a large number of samples at once. Because of these disadvantages of the kits, it is important to predict the IgG concentration indirectly using alternative methods. Second, it is also possible to determine the cut-off level for ALB could not be determined.

It is known the passive immunity develops when the threshold value of IgG <1000 when SRID used, whereas the threshold value of IgG is <500 when ELISA used [35]. In the present study, the cut-off level for serum IgG concentration was determined as ≤113 mg/dL (P<0.001) for mortality in neonatal lambs (rate is 100%). IgG >575 mg/dL (P=0.02) for healthy (rate is 87.4%). IgG values 113-575 mg/dL range (P=0.02) were determined for morbidity (rate is 89%).

GGT enzyme is produced from ductile cells in mammary glands and is present in high concentration in colostrum. It has been reported that serum and plasma GGT activity may be useful in the evaluation of passive transfer status in ruminants. GGT enzyme activities significantly correlated with colostrum IgG concentrations and it is known that it can be used to determine colostrum quality [39,40]. The results of this study indicated a significant linear correlation between GGT and IgG (R=0.75, P<0.001). GGT≤191 mg/dL (P<0.001) were determined for mortality (rate is 85%). GGT>191 mg/dL (P<0.001) were determined for healthy (rate is 86%).

The measurement of TP is the appropriate method for indirect assessment of immune status because of significant correlation [41]. Although the relationship between TP and neonatal diseases in calves is frequently studied, the relationship between TP and neonatal diseases in lambs is not known and there is no STPC threshold used for disease risk [30,38,41,42]. In the present study, a significant linear association was detected between TP and IgG (R=0.67, P<0.001). The cut-off level with TP≥45 g/L (P<0.001) were determined for mortality (rate is 71%). TP>55 g/L (P<0.001) were determined for healthy (rate is 89%). TP values 45-55 g/L range (P<0.001) were determined for morbidity (rate is 73%).

ACKNOWLEDGEMENTS

The dataset used in this study was collected within the scope of the TÜBİTAK project (Project Code: TOVAG 108 O 847).

AUTHOR CONTRIBUTIONS

P. CİHAN, performed the proposed models, processed and analyzed the data, P. CİHAN and E. GÖKÇE, writing-original draft preparation, E. GÖKÇE, O. ATAKIŞİ, A.H. KIRMIZİGÜL and H.M. ERDOĞAN, collected hematological-immunological data. All authors discussed the results and contributed to the final manuscript.

REFERENCES

1. Gokce E, Atakisi O, Kirmizigul AH, Unver A, Erdogan HM: Passive immunity in lambs: Serum lactoferrin concentrations as a predictor of IgG concentration and its relation to health status from birth to 12