



## Artificial Neural Network (ANN) and regression model for predicting the Albumin to Globulin (A/G) ratio in a serum protein electrophoresis test

Akshansh Mishra<sup>1\*</sup>

<sup>1</sup>Founder and Project Scientific Officer, Stir Research Technologies, Uttar Pradesh, India. E-mail: [akshansh.frictionwelding@gmail.com](mailto:akshansh.frictionwelding@gmail.com)

### Article Info

Volume 1, Issue 3, July 2019

Received : 12 May 2019

Accepted : 19 June 2019

Published : 8 July 2019

doi: [10.33472/AFJBS.1.3.2019.40-49](https://doi.org/10.33472/AFJBS.1.3.2019.40-49)

### Abstract

Multiple myeloma affects the several parts of bodies such as the spine, skull, pelvis and ribs. The cause of multiple myeloma is not known properly. The poor prognoses is associated with most cancers creates a sense of urgency for the brains behind healthcare Artificial Intelligence (AI) research. AI is able to detect cancer and other diseases earlier than possible through standard diagnostic methods, which could be lifesaving for future patients. The main objective of the research paper is to predict the Albumin to Globulin (A/G) ratio obtained by the electrophoresis test by developing regression model and Artificial Neural Network (ANN) model. The results obtained showed that the Mean Square Error (MSE) obtained by ANN model is less than the MSE obtained by the regression model.

**Keywords:** Artificial Neural Network (ANN), Regression model, Multiple myeloma, Electrophoresis

© 2019 African Journal of Biological Sciences. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

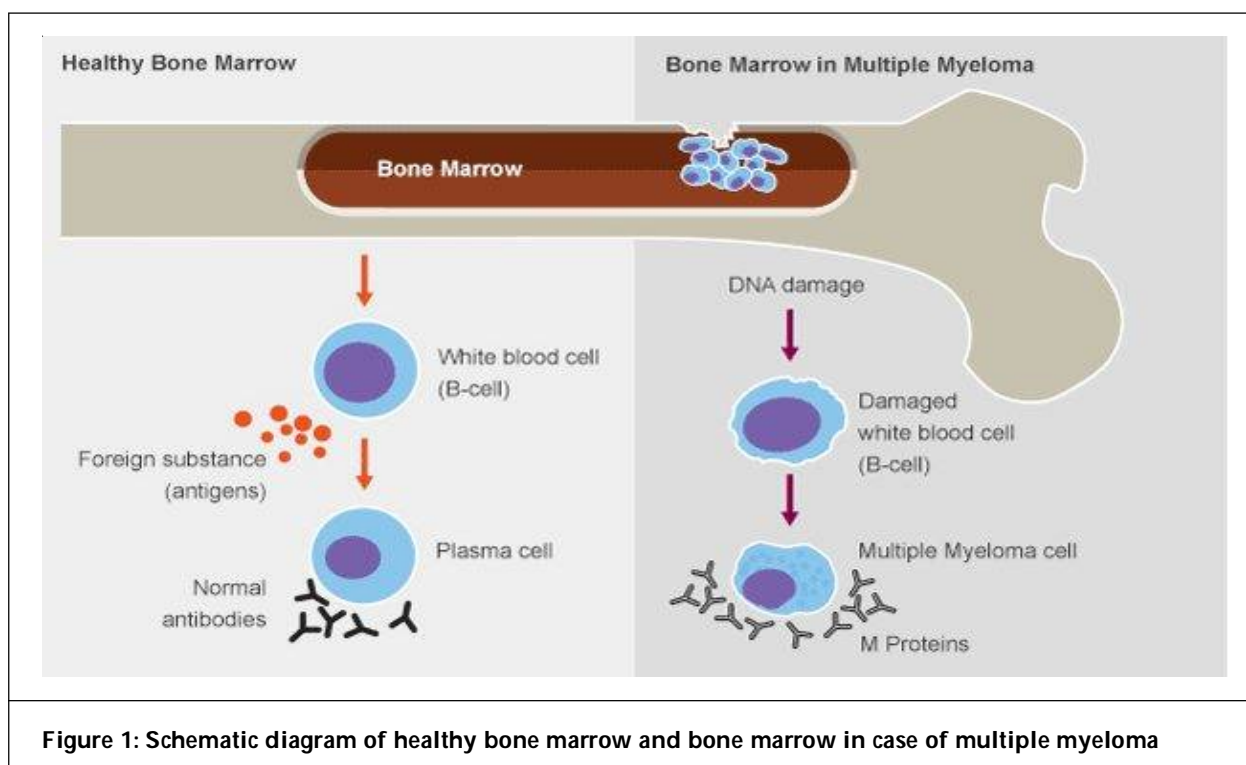
### 1. Introduction

Protein electrophoresis test is used to measure the amount of specific proteins in the blood. On the basis of their electrical charge, proteins are separated in the test. Generally, protein electrophoresis test is used to find an abnormal substance called Monoclonal proteins (M proteins). The presence of M proteins indicates the presence of a type of cancer called myeloma or multiple myeloma. Myeloma generally affects the WBC called plasma cells in the bone marrow.

In the case of multiple myeloma, cancer cells pile up in the bone marrow where they surround healthy blood cells. Instead of producing helpful antibodies these cancer cells produce abnormal proteins which further leads to various complications. The myeloma cells generally produce abnormal antibodies unlike healthy plasma cells which cannot be used by our body. The M proteins are the abnormal antibodies which build up in the body and cause problems such as damage to the kidneys. It also leads to bone marrow damage as shown in the Figure 1 and also increases the risk of broken bones.

Protein electrophoresis test can also be used to diagnose thyroid, diabetes, anaemia, liver diseases, poor nutrition or inability to absorb nutrients and certain autoimmune diseases. This test is needed in the condition if the particular person's healthcare provider suspects that he/she have a condition that affects plasma cells.

\* Corresponding author: Akshansh Mishra, Founder and Project Scientific Officer, Stir Research Technologies, Uttar Pradesh, India. E-mail: [akshansh.frictionwelding@gmail.com](mailto:akshansh.frictionwelding@gmail.com)



The symptoms of the latter condition are unexplained weight loss, bone pain, fatigue, weakness, nausea, constipation, unusual thirst, frequent urination, frequent illness or fevers, back pain, high levels of calcium in the blood and bones that fracture easily.

The application of neural network in medical field has widely been reported. Recent development in computer-aided diagnosis, medical image segmentation and edge detection towards visual content analysis are being done with the help of Artificial Neural Network (ANN) (Jiang *et al.*, 2010). Zhou *et al.* (2002) proposed an automatic pathological diagnosis procedure called Neural Ensemble-based Detection (NED) which worked on an ANN in order to identify the lung cancer cells in the images of the specimens of needle biopsies which were obtained from the bodies of the subjects to be diagnosed. Khan *et al.* (2001) demonstrated the potential application of ANN model for developing gene expression signatures in order to classify cancers. Abbas *et al.* (2002) developed an Evolutionary Artificial Neural Network (EANN) based on the Pareto Differential Evolution (PDE) algorithm to predict breast cancer. Djavan *et al.* (2002) used two ANNs for the early detection of prostate cancer in men with total Prostate-Specific Antigen (PSA) levels from 2.5 to 4 ng/mL and from 4 to 10 ng/mL.

No previous work has been done in the application of neural network in multiple myeloma domain. In this case study ANN model which trained on quasi-Newton algorithm and regression model is developed for predicting the Albumin to Globulin (A/G) ratio obtained from protein electrophoresis test of an early stage multiple myeloma patient.

## 2. Experimental Procedure

Firstly, bone marrow biopsy was performed on the early stage multiple myeloma patient. Bone marrow biopsy shows mostly fibrocartilaginous tissue. Only two to three normocellular marrow spaces seen showing interstitial prominence of plasma cells forming multiple clusters. MRI report showed altered marrow signal and partial collapse of L3 body with involvement of right pedicle.

Protein electrophoresis test is done with a blood sample. A needle is used to draw blood from a vein in patient's arm or hand. The patient's diet or lifestyle habits are not likely to affect the results of this test. Blood serum contains two major protein groups: albumin and globulin. Both albumin and globulin carry substances through the bloodstream. Using protein electrophoresis, these two groups can be separated into five smaller groups (fractions):

### 2.1. Albumin

Albumin proteins keep the blood from leaking out of blood vessels. Albumin also helps carry some medicines and other substances through the blood and is important for tissue growth and healing. More than half of the protein in blood serum is albumin.

### 2.2. Alpha-1 Globulin

High-Density Lipoprotein (HDL), the “good” type of cholesterol, is included in this fraction.

### 2.3. Alpha-2 Globulin

A protein called haptoglobin, which binds with hemoglobin, is included in the alpha-2 globulin fraction.

### 2.4. Beta Globulin

Beta globulin proteins help carry substances, such as iron, through the bloodstream and help fight infection.

### 2.5. Gamma Globulin

These proteins are also called antibodies. They help prevent and fight infection. Gamma globulins bind to foreign substances, such as bacteria or viruses, causing them to be destroyed by the immune system.

Each of these five protein groups moves at a different rate in an electrical field and together forms a specific pattern.

Now let’s talk about how this test is done. The health professional drawing blood will:

- Wrap an elastic band around the patient’s upper arm to stop the flow of blood. This makes the veins below the band larger so it is easier to put a needle into the vein.
- Clean the needle site with alcohol.
- Put the needle into the vein. More than one needle stick may be needed.
- Attach a tube to the needle to fill it with blood.
- Remove the band from the patient’s arm when enough blood is collected.
- Apply a gauze pad or cotton ball over the needle site as the needle is removed.
- Apply pressure to the site and then a bandage.

## 3. Results and Discussion

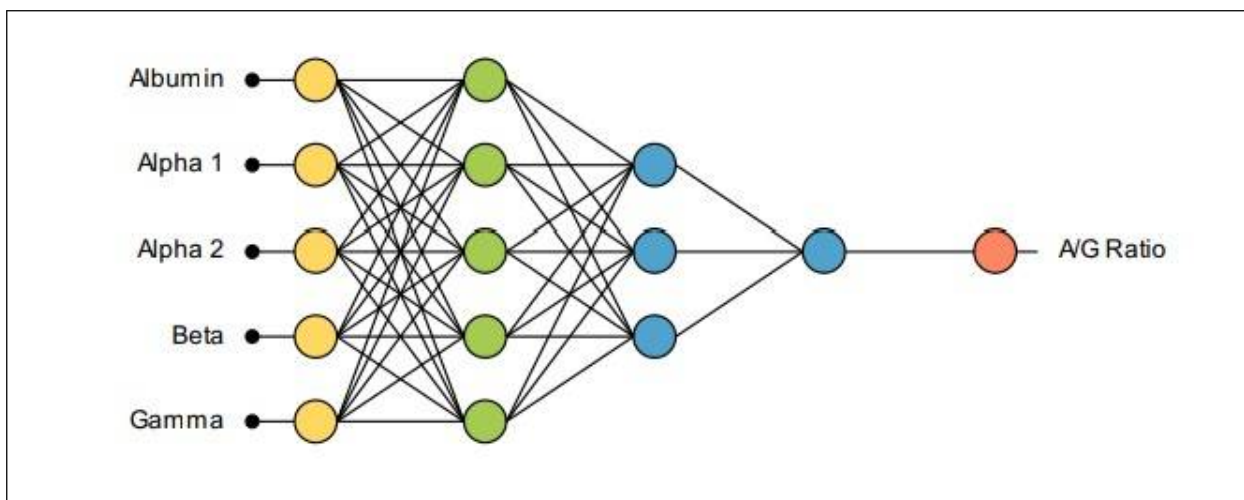
The experimental data obtained from the Serum Protein Electrophoresis (SPE) test is tabulated in the Table 1.

	A	B	C	D	E	F
1	Albumin	Alpha 1	Alpha 2	Beta	Gamma	A/G Ratio
2	54	2.2	9.8	10.9	23.1	1.17
3	41.5	2.7	13.6	9.9	32.3	0.71
4	63.8	2	14.7	14	5.5	1.77
5	56.7	3	17.4	16.2	6.7	1.31
6	52.3	1.4	10.4	10.2	25.7	1.1
7	58	1.8	9	9.9	21.3	1.77
8	39.9	5.5	12.9	8.8	30.4	0.66
9	52.31	2.65	12.54	11.41	20.71	1.21
10	41.3	2.5	13.3	9.8	32.4	0.74
11	63.7	1.9	14.6	13.8	5.3	1.73
12	40.1	3.4	14.5	15.2	6.3	0.85

**Note:** Depending on these dataset ANN model and fegression model are constructed.

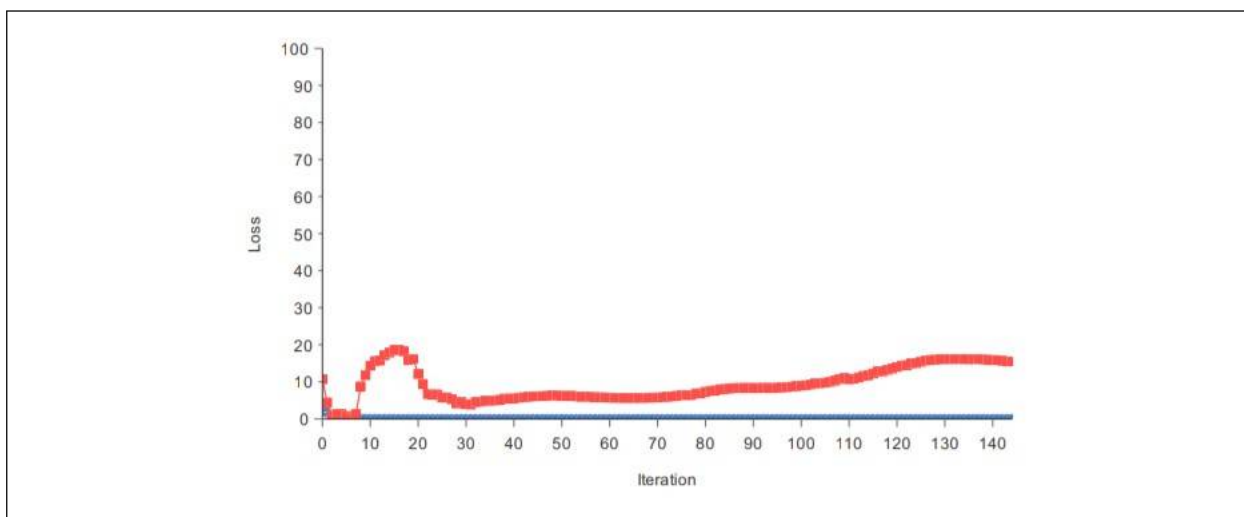
### 3.1 Artificial Neural Network (ANN) Model

A graphical representation of the network architecture is depicted in the Figure 2. It contains a scaling layer, a neural network and an unscaling layer. The yellow circles represent scaling neurons, the green circles the principal components, the blue circles perceptron neurons and the red circles unscaling neurons. The number of inputs is five, the number of principal components is five, and the number of outputs is one. The complexity, represented by the numbers of hidden neurons, is three.



**Figure 2: Artificial Neural Network (ANN) architecture**

The loss index plays an important role in the use of a neural network. It defines the task the neural network is required to do, and provides a measure of the quality of the representation that it is required to learn. The choice of a suitable loss index depends on the particular application. The normalized squared error is used here as the error method. It divides the squared error between the outputs from the neural network and the targets in the data set by a normalization coefficient. If the normalized squared error has a value of unity then the neural network is predicting the data ‘in the mean’, while a value of zero means perfect prediction of the data. The neural parameters norm is used as the regularization method. It is applied to control the complexity of the neural network by reducing the value of the parameters. The value of neural parameters norm weight is 0.001. The quasi-Newton method is used here as training algorithm. It is based on Newton’s method, but does not require calculation of second derivatives. Instead, the quasi-Newton method computes an approximation of the inverse Hessian at each iteration of the algorithm, by only using gradient information. Figure 3 shows the losses in each iteration. The initial value of the training loss is 3.36237, and the final value after 144 iterations is 0.00506113. The initial value of the selection loss is 10.6446, and the final value after 144 iterations is 15.4535.



**Figure 3: Quasi-Newton method losses history**

A graphical representation of the resulted deep architecture is depicted in Figure 4. It contains a scaling layer, a neural network and an unscaling layer. The yellow circles represent scaling neurons, the green circles the principal components, the blue circles perceptron neurons and the red circles unscaling neurons. The number of inputs is five, the number of principal components is five, and the number of outputs is one. The complexity, represented by the numbers of hidden neurons, is four.

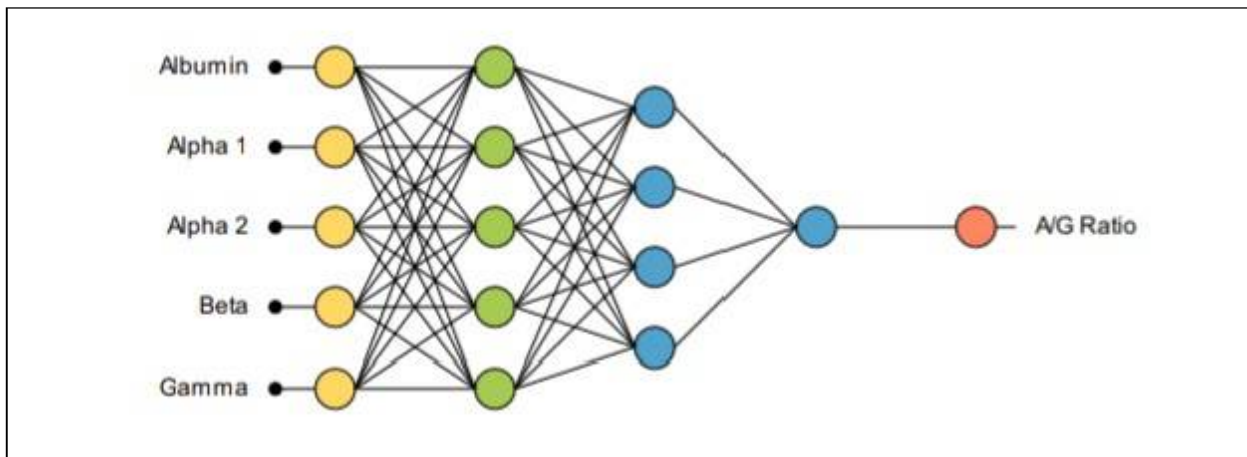


Figure 4: Final architecture

### 3.2 Regression Model

The hypothesis function for linear regression is:

$$y = a_0 + a_1x_1 + a_2x_2$$

The linear regression model is used from scikit-learn library. It uses the Ordinary Least Squares solver from scipy, to converge to the global minimum. The code is shown below:

```
from sklearn import linear_model
regr = linear_model.LinearRegression( )
regr.fit(x_train, y_train)
y_pred = regr.predict(X_test)
```

The scatter chart plots of A/G ratio with respect to other input parameters are shown from Figures 5 to 9.

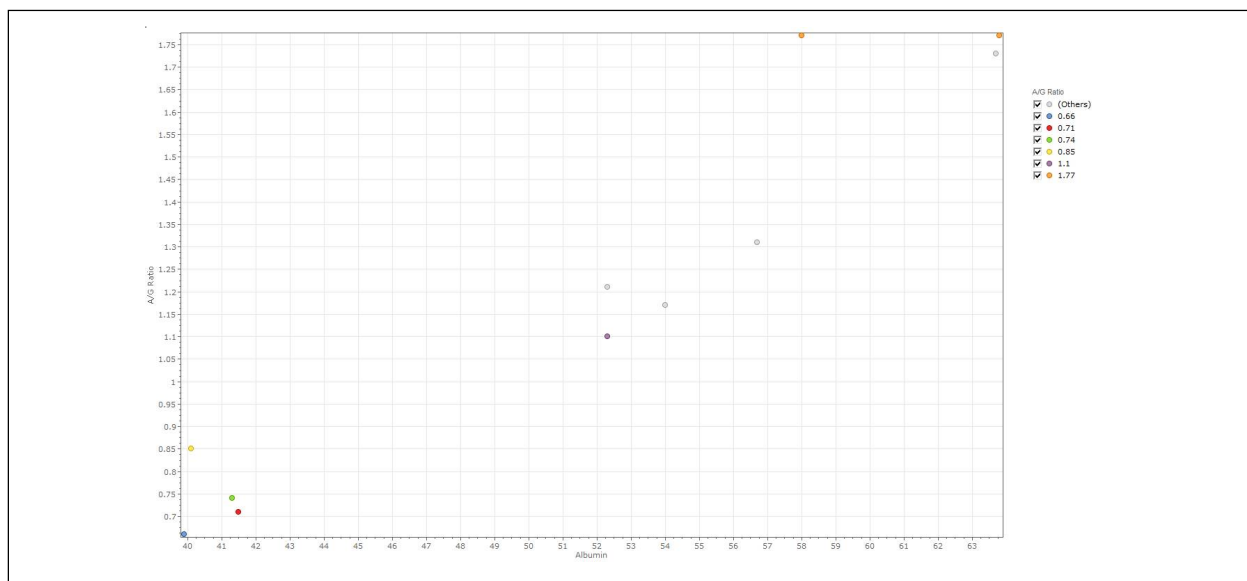


Figure 5: Scatter plot of A/G ratio with respect to albumin

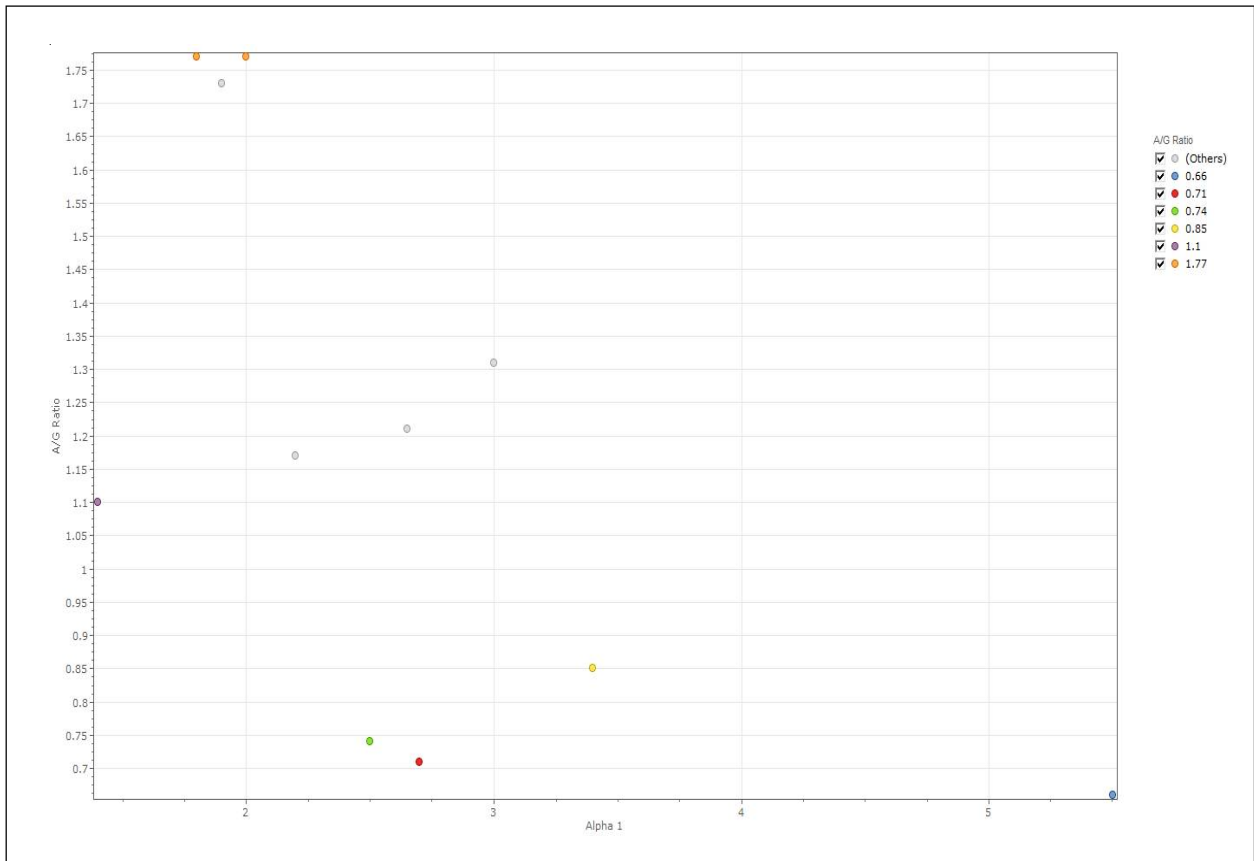


Figure 6: Scatter plot of A/G ratio with respect to alpha 1

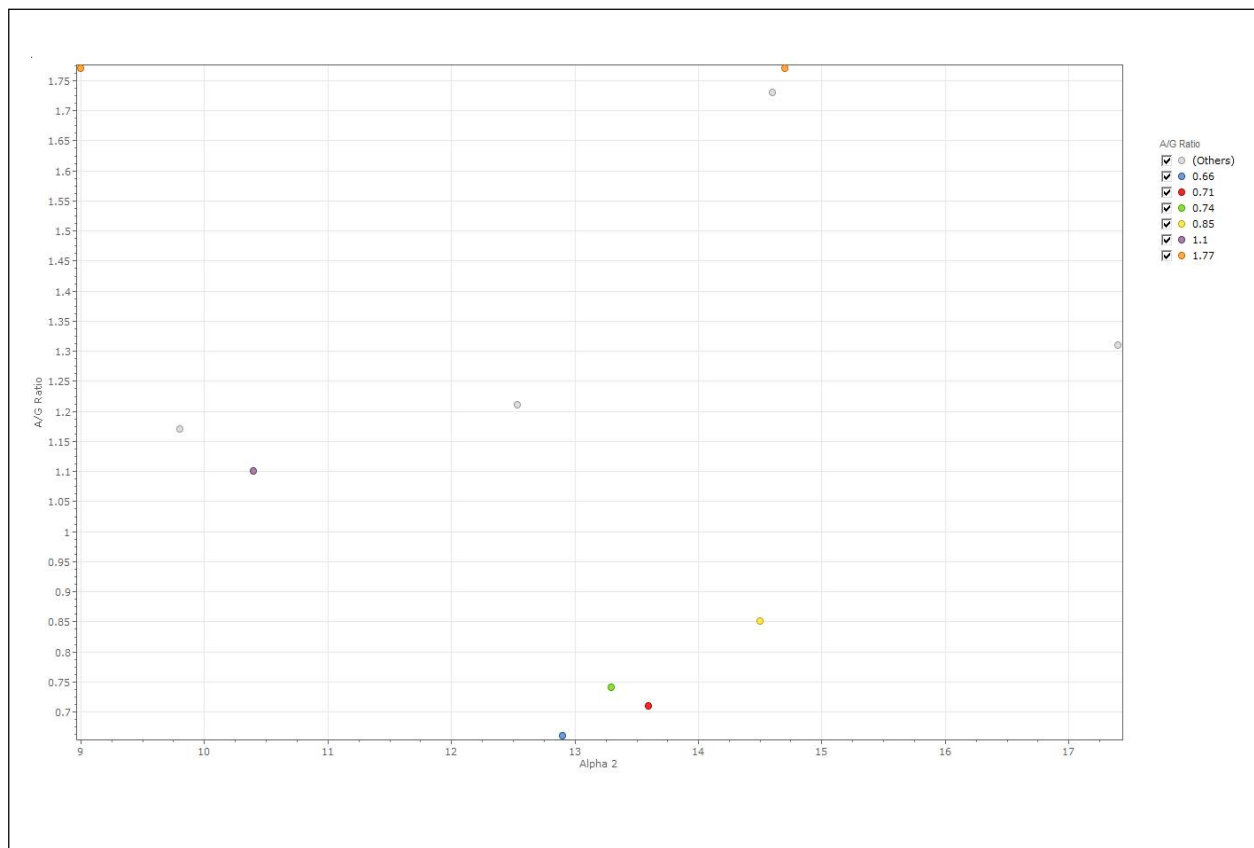


Figure 7: Scatter plot of A/G ratio with respect to alpha 2

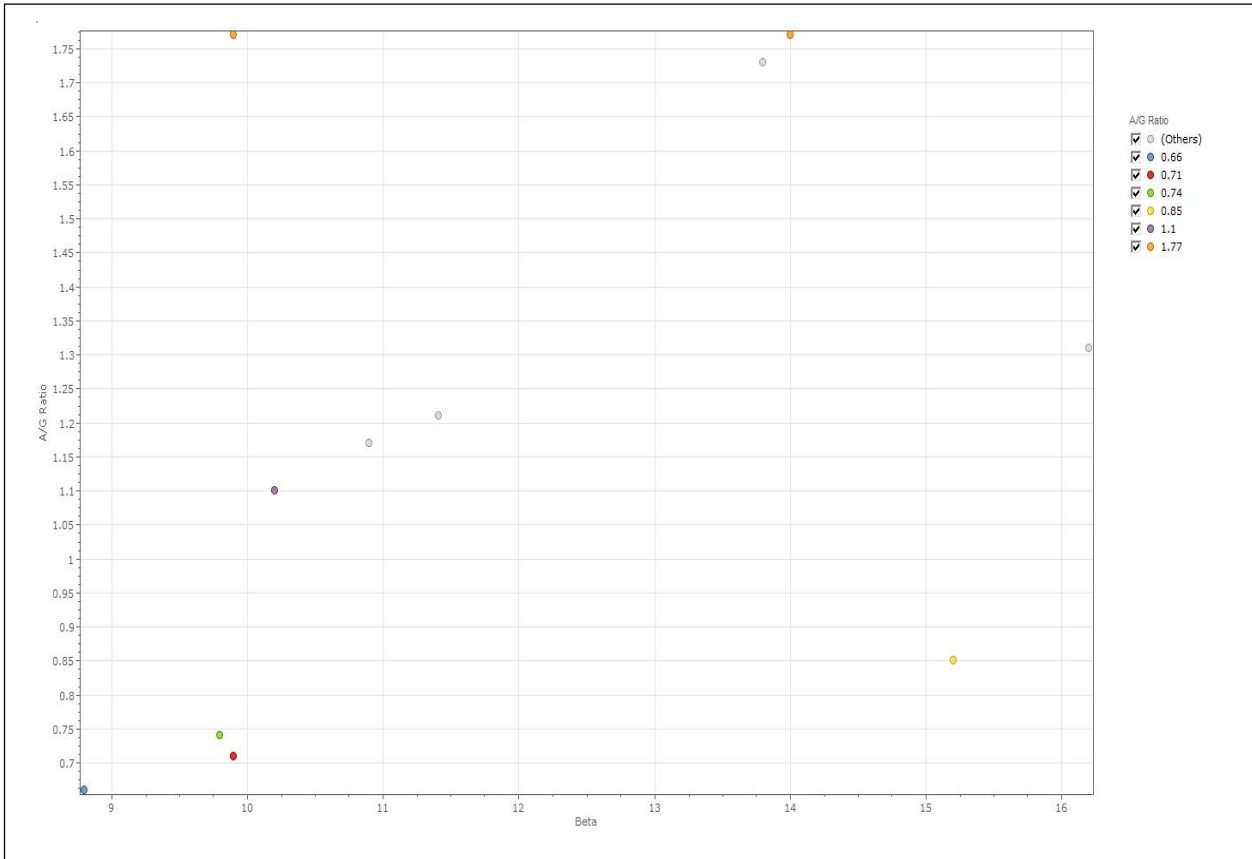


Figure 8: Scatter plot of A/G ratio with respect to beta

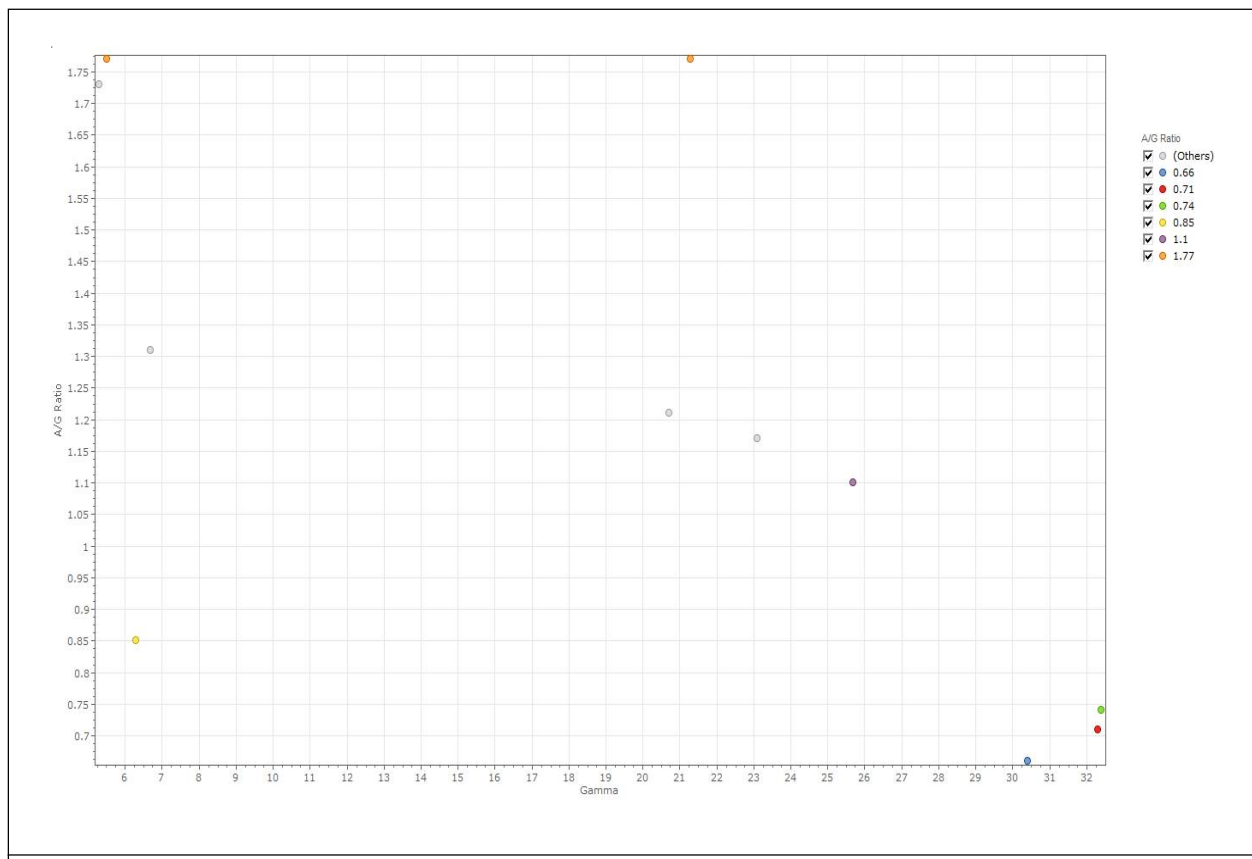


Figure 9: Scatter plot of A/G ratio with respect to gamma

The contour maps of the regression model are depicted from Figures 10 to 13.

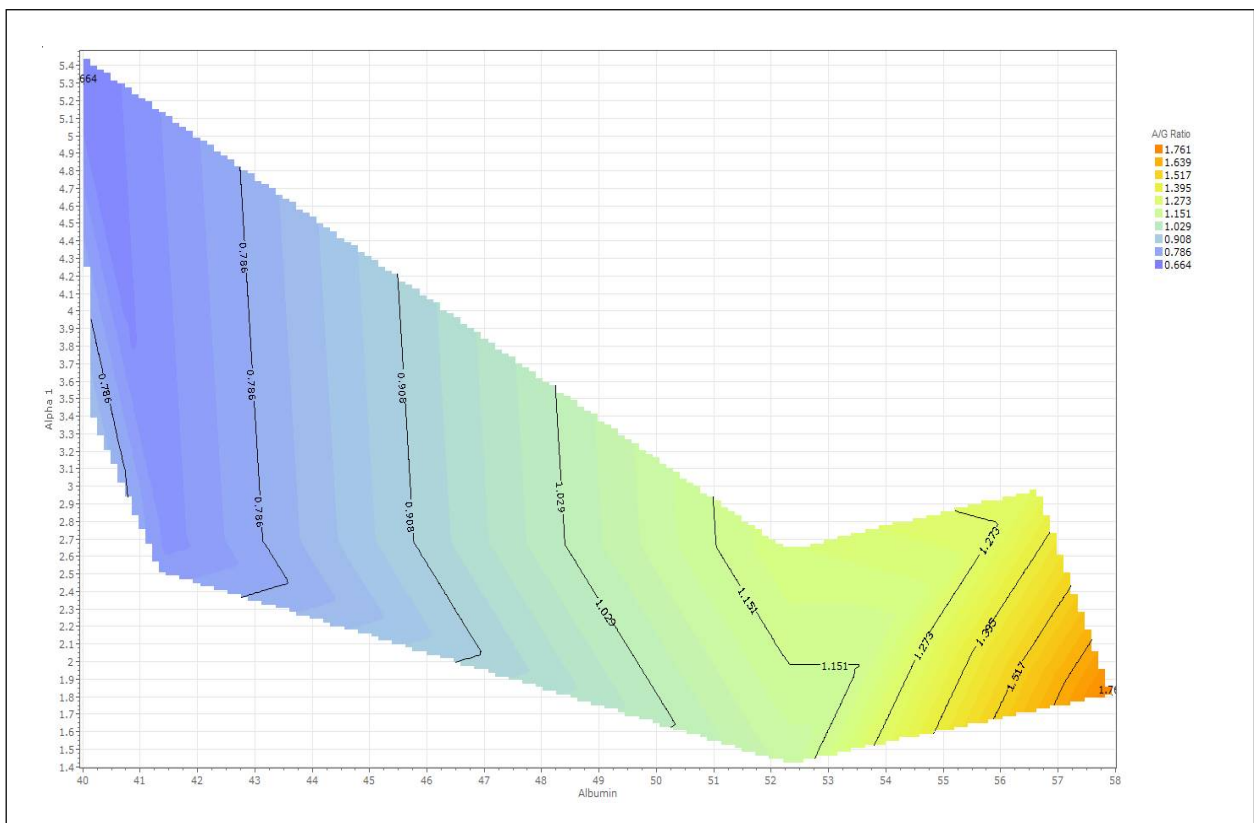


Figure 10: Variation of A/G with respect to albumin and alpha 1

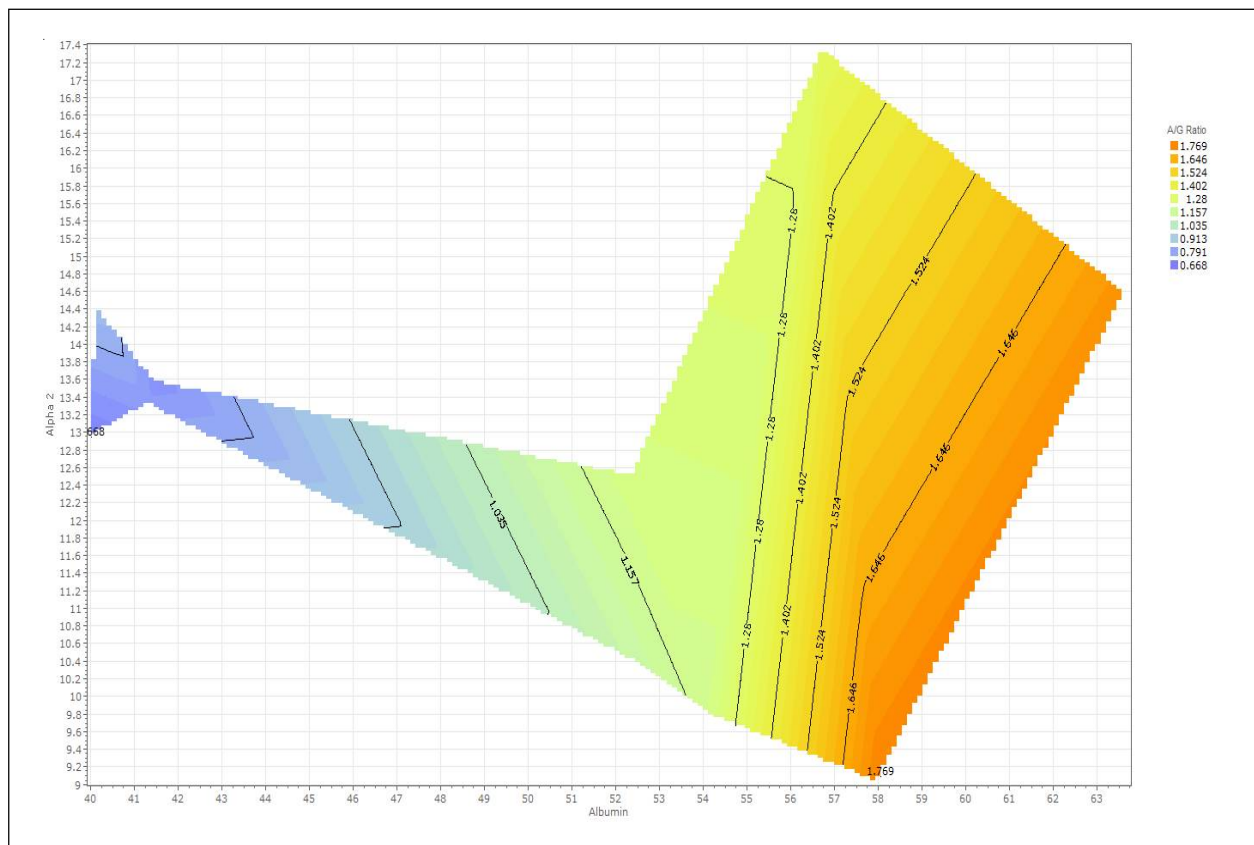


Figure 11: Variation of A/G ratio with respect to albumin and alpha 2



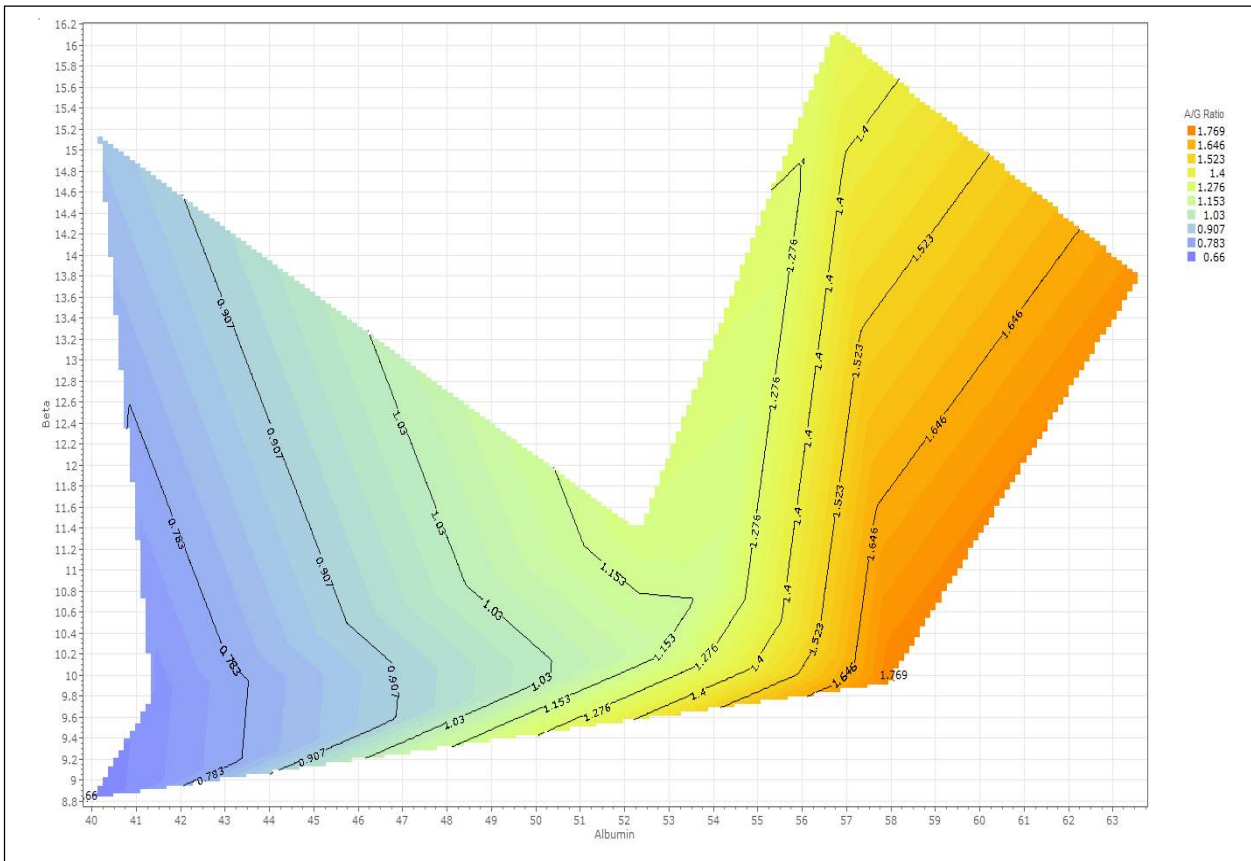


Figure 12: Variation of A/G ratio with respect to albumin and beta

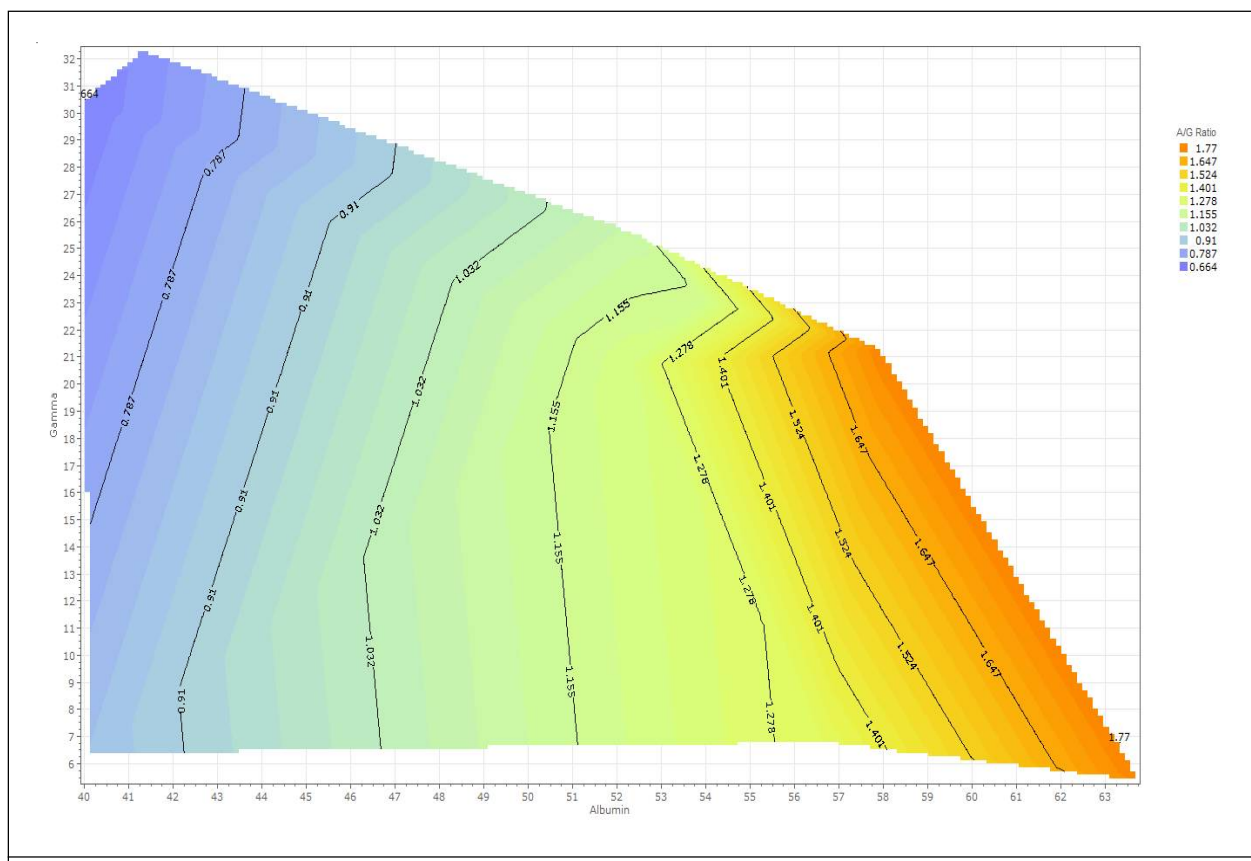


Figure 13: Variation of A/G ratio with respect to albumin and gamma

#### 4. Conclusion

The Mean Square Error (MSE) obtained during training, selection and testing of the ANN architecture are 0.164585, 0.0276843 and 0.26813. For regression model, MSE obtained is 2.664 which are higher than the ANN architecture model. It can be concluded that ANN models can be used for prediction purpose in multiple myeloma case with maximum accuracy.

#### References

- Abbass, H.A. (2002). [An evolutionary artificial neural networks approach for breast cancer diagnosis. \*Artificial Intelligence in Medicine\*. 25 \(3\), 265-281.](#)
- Djavan, B., Remzi, M., Zlotta, A., Seitz, C., Snow, P. and Marberger, M. (2002). [Novel artificial neural network for early detection of prostate cancer. \*Journal of Clinical Oncology\*. 20 \(4\), 921-929.](#)
- Jiang, J., Trundle, P. and Ren, J. (2010). [Medical image analysis with artificial neural networks. \*Computerized Medical Imaging and Graphics\*. 34 \(8\), 617-631.](#)
- Khan, J., Wei, J.S., Ringner, M., Saal, L.H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C.R., Peterson, C. and Meltzer, P.S. (2001). [Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. \*Nature Medicine\*. 7 \(6\), 673.](#)
- Zhou, Z.H., Jiang, Y., Yang, Y.B. and Chen, S.F. (2002). [Lung cancer cell identification based on artificial neural network ensembles. \*Artificial Intelligence in Medicine\*. 24 \(1\), 25-36.](#)

**Cite this article as:** Akshansh Mishra (2019). [Artificial Neural Network \(ANN\) and regression model for predicting the Albumin to Globulin \(A/G\) ratio in a serum protein electrophoresis test. \*African Journal of Biological Sciences\* 1 \(3\), 40-49.](#)