

## RESEARCH ARTICLE

### A binary logistic regression model for prediction of feed conversion ratio of *Clarias gariepinus* from feed composition data

Familusi Oluwatosin Adekunle <sup>1\*</sup> 

<sup>1</sup> Ekiti-State University, Faculty of Agricultural Sciences, Forestry Wildlife and Fisheries Department, PMB. 5363, Ado-Ekiti, Nigeria

#### ARTICLE INFO

Article History:  
Received: 29.05.2020  
Received in revised form: 08.12.2020  
Accepted: 09.12.2020  
Available online: 03.01.2021

Keywords:  
*Binary logistic regression*  
*Clarias gariepinus*  
*Feed conversion ratio*  
*Jaccard similarity score*  
*Numerical optimizer*  
*Proximate analysis*

#### ABSTRACT

Aquaculture in developing countries faces a lot of challenges that are barely being addressed. With feed taking nearly 70% of the total production cost, it becomes imperative to develop means of optimizing how research is conducted into feed development. Feed conversion ratio as a measure of feed quality can be used to quantify in retrospect the appropriateness of feed fed to livestock, particularly, *Clarias gariepinus*. From the study, binary logistic regression can in simple terms, determine if prospective feed will perform below or above the acceptable level of 1.5, based on its composition and proximate analysis values. Data from similar experiments are normalized and split into train and testing data to fit a logistic regression model, three numerical optimizers were used including liblinear, Newton-CG, SAG and accuracy of the models were compared using the confusion matrix, and Jaccard similarity score. An accuracy value of 0.8 was observed in the model regardless of the numerical optimizer, this indicates the appropriateness of the model in predicting either high or low FCR for feed types. The probability of prediction showed disparity among liblinear and SAG/Newton-CG solvers. Liblinear solver showed close probabilities in predicting if values will be 1 or 0. While a similar prediction was made by all solvers, this indicates a possible affinity for error when the solver is used. This is also indicated with a logloss of 0.65 as compared to 0.51 in both SAG and Newton-CG solvers.

#### Please cite this paper as follows:

Adekunle, F. O. (2021). A binary logistic regression model for prediction of feed conversion ratio of *Clarias gariepinus* from feed composition data. *Marine Science and Technology Bulletin*, 10(2): 131-141.

#### Introduction

The variety of factors that could be responsible for an observed outcome in any biological system requires the need to

conduct several experiments and trials to get the desired result. In aquaculture, factors including the composition of a single feedstuff contribute alongside all other factors in the performance of the entire nutrient performance of the

\* Corresponding author  
E-mail address: [familusiadekunle@gmail.com](mailto:familusiadekunle@gmail.com) (F. O. Adekunle)



compounded feed. This accounts for the use of a large amount of human and capital resources.

Aquaculture plays a very important role in food systems, especially in middle-income regions where the industry employs labor and serves as a major source of animal protein. As stated by FAO (2016), global fish production reached a peak of 171 million tonnes, 47% of which was produced by aquaculture. However, if aquaculture is to remain an alternative to dwindling capture fishery stocks, there's the need to reduce the cost of production which is mostly accounted for by the cost of feeding, taking between 60% and 80% of the total cost of production (Ng et al., 2013).

Fishmeal is an indispensable component of fish feed due to its amino-acid profile, fatty acids, flavor, and other essential nutrients. The ecological cost of fishmeal and high demand from other livestock species necessitates the need for a similar substitute. (Farahiyah et al., 2016). Hence optimization of the protein content of feed relies heavily on successfully substituting fishmeal with other more affordable feedstuff for optimal growth performance (Degani et al., 1989). The FCR is simply the amount of feed it takes to grow a kilogram of fish. For example, if it requires two kilograms of feed to grow one kilogram of fish, the FCR would be two, this means that when a feed has a low FCR, it takes less feed to produce one kilogram of fish than it would if the FCR were higher. A low FCR is a good indication of a high-quality feed. FCR is a valuable and powerful tool for the fish farmer. It allows for an estimate of the feed that will be required in the growing cycle. Knowing how much feed will be needed then allows a farmer to determine the profitability of an aquaculture enterprise. This means that FCR allows the farmer to make wise choices in selecting and using the feed to maximize profitability. (USAID-HARVEST, 2011)

Several factors can influence the way fish respond to feed. Stage of culture, size, water quality, genetics, pond management, and the composition of other feedstuff.

Binary logistic regression studies the association between a category of the dependent variable and a set of independent variables. Logistic regression is used when the outcome has only two possible values (0 and 1), and is opposed to multinomial regression where the outcome could be three or more possible outcomes or prediction. Logistic regression as opposed to linear regression is used for the prediction of categorical response variables. It is assumed to be more suited for modeling because it does not assume a normal distribution for the independent variables (NCSS, 2020).

## Materials and Methods

### Data Generation and Preprocessing

FCR reported, based on specific feed composition as reported by Chor et al. (2013), Oyekanmi et al. (2013), Dudusola and Akinlade (2014), Falaye et al. (2015), and Aniebo et al. (2009) were used as historical data. Experimental results from feeding trials on *Clarias gariepinus* comprising of feed components used in each of the trials. Some components are present in nearly all the trials, i.e. fishmeal and lipids. Other components include maggot-meal, feather meal, blood-meal, etc. Feed proximate analysis data, with similar experimental design and analytical procedure as outlined by AOAC (1990) were collected as relevant to the feed composition data. The initial entry was done on excel spreadsheets, feed component and proximate data are loaded into rows, columns are based on feed trial indicator, and source.

Feed component data included in the model comprises the most utilized feedstuff for the formulation of feed for African catfish (Table 1). This is expected to facilitate the ease of using the model by a third party in the prediction of Feed conversion ratio.

Five code indicators are used in the columns to indicate the source of the data, they include FTM, MGT, CMGT, MAIZE, and FSHML representing Feather-meal, Maggot-meal, Maggot-meal, and Fishmeal respectively, each referring to the theme of feeding trial from which the corresponding data was obtained.

### Binary Classification

A feed conversion ratio of 1.8 to 1 was observed by Li et al. (2014) to be typical in experimental set-up and that was used to categorize the FCR values in the historical data. FCR values between 0 and 1.5 were categorized as 1 while FCR values greater than 1.5 were classified as 0 as shown in Table 2.

### Regression

Logistic regression uses the independent variable from historical data (feed composition and proximate analysis as data shown above) to produce a formula that predicts the probability of the class label (FCR churn). Logistic regression fits a special s-shaped curve by transforming the numeric estimate into a probability using the sigmoid function. Hence the model predicts the particular class for which a hypothetical feed composition belongs (1 meaning good FCR and 0 meaning bad FCR), and also gives the probability of having that class.

Table 1. Feed components included in the model

Feed	FSHML	PRN	FTML	BRWWT	SBML	BLDML	Lipid	MGTML	WTBRN	YMZ	GNC	CMC	Vit	Chromic	Minerals	Calcium	Cellulose	Tapioca
FTM0	56.5	0	0	0	0	0	6.43	0	0	0	0	2.36	3	0.5	4	1	5.34	28.25
FTM20	45.32	0	9.86	0	0	0	7.54	0	0	0	0	2.36	3	0.5	4	1	6.1	28.39
FTM40	33.99	0	19.7	0	0	0	8.63	0	0	0	0	2.36	3	0.5	4	1	6.88	28.53
FTM60	22.66	0	29.6	0	0	0	9.73	0	0	0	0	2.36	3	0.5	4	1	7.66	28.67
FTM80	11.33	0	39.4	0	0	0	10.8	0	0	0	0	2.36	3	0.5	4	1	8.43	28.81
FTM100	0	0	49.3	0	0	0	11.9	0	0	0	0	2.36	3	0.5	4	1	9.21	28.94
MGT0	31.73	0	0	0	0	0	0.25	0	11.18	22.4	31.7	0	0.5	0	1.5	1	0	0
MGT50	22.77	0	0	0	0	0	0.25	22.77	9.57	19.1	22.8	0	0.5	0	1.5	1	0	0
MGT33	31.89	0	0	0	0	0	0.25	15.95	11.07	22.2	16	0	0.5	0	1.5	1	0	0
MGT66	17.76	0	0	0	0	0	0.25	35.57	8.66	17.3	17.8	0	0.5	0	1.5	1	0	0
MGT75	14.55	0	0	0	0	0	0.25	43.65	8.8	16.7	14.5	0	0.5	0	1.5	1	0	0
CMGT0	25	0	0	0	34	10.3	5.2	0	11	11	0	0	0.2	0.2	0.15	3	0	0
CMGT12	12.5	0	0	0	39	10	3.5	12.5	8	11	0	0	0.2	0.2	0.15	3	0	0
CMGT25	0	0	0	0	43.5	10	1	25	6	11	0	0	0.2	0.2	0.15	3	0	0
MAIZE0	13.63	0	0	0	27.26	0	1	0	0	28.9	27.3	0	0.5	0	0.5	2	0	0
MAIZE25	13.3	0	0	0	26.63	0	1	0	0	21.7	26.6	0	0.5	0	0.5	2	0	0
MAIZE50	13.3	0	0	0	26.63	0	1	0	0	14.7	26.6	0	0.5	0	0.5	2	0	0
MAIZE75	13	0	0	0	26	0	1	0	0	7.65	26	0	0.5	0	0.5	2	0	0
MAIZE100	13	0	0	0	26	0	1	0	0	0	26	0	0.5	0	0.5	2	0	0
FSHML100	62	0	0	4.25	0	9	2.5	0	0	15	3	0	0.3	0	0.3	2.9	0	0.5
FSHML75	47	15	0	4.25	0	9	2.5	0	0	15	3	0	0.3	0	0.3	2.9	0	0.5
FSHML50	31	31	0	4.25	0	9	2.5	0	0	15	3	0	0.3	0	0.3	2.9	0	0.5
FSHML25	15	47	0	4.25	0	9	2.5	0	0	15	3	0	0.3	0	0.3	2.9	0	0.5
FSHML0	0	62	0	4.25	0	9	2.5	0	0	15	3	0	0.3	0	0.3	2.9	0	0.5

**Note:** Abbreviations in header row and meanings: FSHML- Fishmeal; SBML- Soyabean Meal; PRN - Rocky-prawn; BRWWT- Brewers Waste; BLDML- Blood-meal; MGTML- Maggot-Meal; WTBRN - Wheat-bran; YMZ- Yellow maize; GNC- Groundnut Cake; CMC- Carboxymethyl Cellulose; Vit - Vitamin premix

**Table 2.** Proximate analytical data for compounded feed

Feed	Protein	Fat	Ash	Crude fiber	NFE	Moisture	Culture period	Fish weight	FCR	FCR churn
FTM0	39.82	11.45	86.55	4.811	25.85	4.78	4	2.85	1.24	1
FTM20	39.51	11.23	87.43	4.811	25.85	5.12	4	2.85	1.34	1
FTM40	41.25	11.87	89.1	4.811	25.85	5.03	4	2.85	2.9	0
FTM60	38.95	12.14	89.7	4.811	25.85	5.89	4	2.85	3.08	0
FTM80	40.34	11.5	90.7	4.811	25.85	4.43	4	2.85	2.85	0
FTM100	40.67	11.26	91.27	4.811	25.85	5.37	4	2.85	1.91	0
MGT0	44.2	4.76	13.37	3.74	23.98	9.96	10	21.73	5.07	0
MGT50	43.5	4.6	12.7	3.76	25.47	9.9	10	21.8	3.96	0
MGT33	44.23	4.5	13.11	3.74	24.52	9.8	10	21.67	3.57	0
MGT66	43.48	4.76	13.16	3.85	24.8	9.94	10	21.8	4.16	0
MGT75	44.03	4.85	13.6	4.1	28.54	4.9	10	21.77	3.13	0
CMGT0	40.76	9.2	10.59	4.1	25.85	9.65	10	10	1.15	1
CMGT12	40.59	8.98	9.1	4.87	25.85	9.22	10	10.02	1.17	1
CMGT25	40.74	8.51	8	5.22	25.85	9.87	10	10	1.16	1
MAIZE0	37.42	4.36	15.78	4.2	28.45	9.96	12	4.03	0.65	1
MAIZE25	35.96	4.98	15.88	5.53	28.2	9.78	12	4.06	0.68	1
MAIZE50	36.42	5.24	16.02	5.51	27.3	9.89	12	4.06	0.7	1
MAIZE75	38.25	5.46	15.95	5.6	24.92	9.93	12	4.05	0.68	1
MAIZE100	40.2	6.66	16.22	5.58	22.63	9.88	12	4.05	0.62	1
FSHML100	38.3	6.81	8.61	3.44	39.96	8.88	13	10.73	2.33	0
FSHML75	37.6	6.45	8.78	3.56	34.2	9.41	13	10.57	2.33	0
FSHML50	38.9	7.63	9.44	3.73	30.47	9.83	13	10.56	2.16	0
FSHML25	37.8	7.49	9.53	3.63	32.09	9.46	13	10.58	2.3	0
FSHML0	36.6	7.33	9.59	3.69	33.3	9.49	13	10.61	2.62	0

**Note:** Abbreviations used in index column and meanings (values attached in the table indicates level of inclusion in the compounded feed): FTM-Feather meal; MGT-Maggot meal; CMGT-Maggot meal as reported by Aniebo et al. (2009); MAIZE-Maize inclusion meal; FSHML- Fish Meal.

Mathematically, the sigmoid function is represented as:

$$\sigma: h(x)=\sigma(\theta TX)=\frac{e^{(\theta_0+\theta_1x_1+\theta_2x_2+\dots)}}{1+e^{(\theta_0+\theta_1x_1+\theta_2x_2+\dots)}} \quad (1)$$

Where  $\theta TX$ = regression results (sum of variables weighted by the coefficients),  $(\theta TX)$  =sigmoid or logistic function. Probability of a class calculated by Equation (2).

$$P_i(Y=1|X)=\sigma(\theta TX_i)=\frac{e^{\theta TX}}{1+e^{\theta TX}} \quad (2)$$

Where  $P_1$  = Probability of having an acceptably high value for FCR,  $X_i$  is a vector of explanatory variables,  $\theta TX$  is unknown parameters to be estimated.

**Normalization, Train, and Test Splitting**

Normalization was done using Standard scaler from preprocessing Sci-kit learn library to have an equal representation of each feature within the groups. Using train\_test\_split library, data was split into the train and testing set. Test size is set at 20%, while 80% is used for training the model.

**Modelling and Fitting**

The inverse of the regularization strength also known as the ‘C’ parameter is set at 0.01, the numerical optimizer is set as liblinear, SAG, and Newton-CG solvers are also applied to know the optimal solver. A test set comprising of 5 data points was used to test the model.

**Measurement of Accuracy**

Jaccard index/ Jaccard similarity score: is estimated using metrics from sklearn. The index is a measure of size of the intersection divided by size of the union of two label sets (0 and

1), i.e. if all predicted labels for a particular set matches with the true labels, subset accuracy is 1.0, if none match, and it is 0.0.

Confusion matrix shows the number of correctly predicted points versus wrong predictions side by side. The first row is for FCR whose actual churn value in the test set is 1, the second row is for FCR with an actual churn value of 0. The first column holds total number of correct predictions and second column holds the number of wrong predictions. These values can then be interpreted as true positives, false positives, true negatives, and false negatives. Classification report comprises precision, recall, and F1 score. Precision measures accuracy provided class label has been predicted by Equation (3).

$$\text{Precision}=\frac{TP}{(TP+FP)} \quad (3)$$

Recall is the rate of true positives. Recall calculated by Equation (4).

$$\text{Recall}=\frac{TP}{(TP+FN)} \quad (4)$$

F1 score is the harmonic average of both precision and recall. Best value = 1 while worst = 0. Log loss measures the performance of the classifier where output to be predicted is binary.

**Results and Discussion**

**Results**

Accuracy metrics results for test set data are given in following tables (Table 3, Table 4, Table 5). Matrix for logistic regression using different solver are presented in following figures (Figure 1, Figure 2, Figure 3). Classification reports for different solvers are tabulated in Table 6, Table 7, and Table 8.

**Table 3.** Accuracy metrics using liblinear solver

0	1	Prediction	Jaccard Similarity Score	Log Loss
0.520076	0.479924	0	0.8	0.65
0.518691	0.481309	0		
0.539945	0.460055	0		
0.544009	0.455991	0		
0.438961	0.561039	1		

**Table 4.** Accuracy metrics using Newton-CG solver

0	1	Prediction	Jaccard Similarity Score	Log Loss
0.72107	0.27893	0	0.8	0.51
0.704345	0.295655	0		
0.652111	0.347889	0		
0.68603	0.31397	0		
0.341602	0.658398	1		

**Table 5.** Accuracy metrics using SAG solver

0	1	Prediction	Jaccard Similarity Score	Log Loss
0.72107	0.27893	0	0.8	0.51
0.704345	0.295655	0		
0.652111	0.347889	0		
0.68603	0.31397	0		
0.341602	0.658398	1		

**Table 6.** Classification report for liblinear solver

	Precision	Recall	F1-Score	Support
0	0.75	1	0.86	3
1	1	0.5	0.67	2
Accuracy			0.8	5
Macro avg	0.88	0.75	0.76	5
Weighted avg	0.85	0.8	0.78	5

**Table 7.** Classification report for Newton-CG solver

	Precision	Recall	F1-Score	Support
0	0.75	1	0.86	3
1	1	0.5	0.67	2
Accuracy			0.8	5
Macro avg	0.88	0.75	0.76	5
Weighted avg	0.85	0.8	0.78	5

**Table 8.** Classification report for SAG solver

	Precision	Recall	F1-Score	Support
0	0.75	1	0.86	3
1	1	0.5	0.67	2
Accuracy			0.8	5
Macro avg	0.88	0.75	0.76	5
Weighted avg	0.85	0.8	0.78	5

**Discussion**

Similar Jaccard similarity score from tables 3.1.1, 3.1.2 and 3.1.3 shows all solvers predicted the output of the 5 test data with an 80% accuracy, this also corresponds to the similar values in the prediction columns. However, probability of prediction observed using a liblinear solver has very low margins. The Probability of the first prediction made using a liblinear solver, are 57% for 0 and 42% for 1 were recorded. This

may make the solver more prone to error. Logarithmic loss is also highest under the liblinear solver when compared to the other solvers.

Newton-CG and SAG solvers show very high probabilities in the accurate predictions made. But when compared to liblinear solver, SAG and Newton-CG solver places very high probability on predicting the last data point which was wrong. An accurate value for point 5 was supposed to be 0 but was predicted as 1, while liblinear solver apportioned a probability

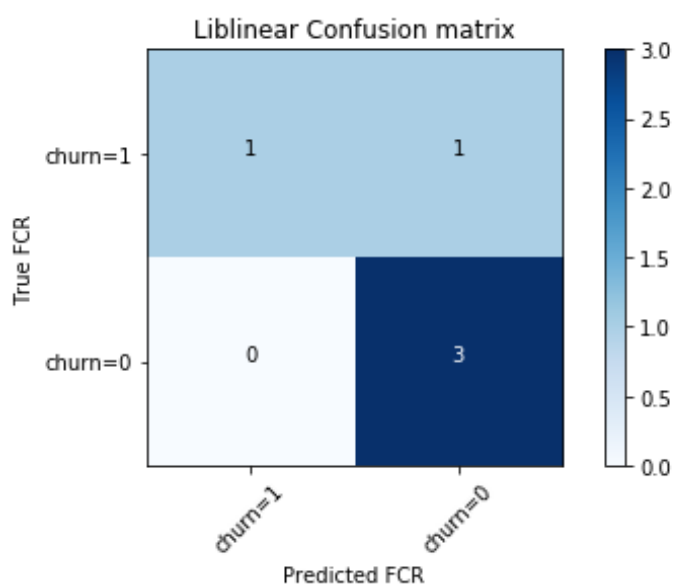


Figure 1. Matrix for logistic regression using liblinear solver

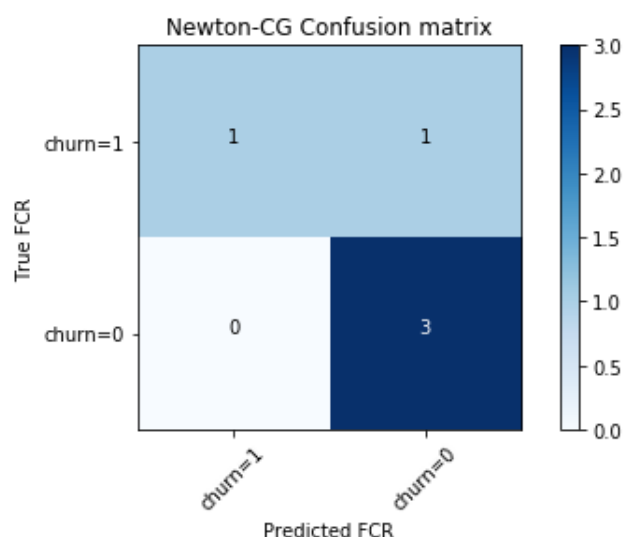


Figure 2. Matrix for logistic regression using Newton-CG solver

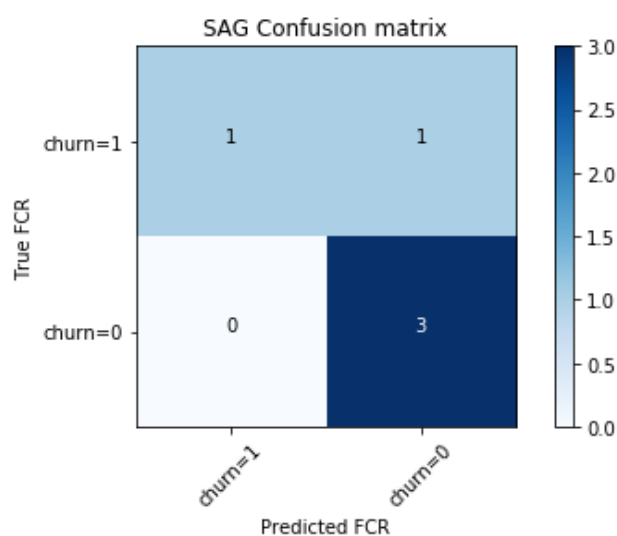


Figure 3. Matrix for logistic regression using SAG solver

of 43% on 0, both Newton-CG and SAG apportioned a probability of 56%.

Confusion matrix indicates similar numbers of true positive (1), false positive (1) and true negatives (3), no false negatives were observed. This translates to having 4 right and one wrong prediction regardless of the solver used.

### Conclusion and Recommendation

Binary logistic regression simply underlies the use of data to answer questions with two possible outcomes. This can be used in simply predicting either result will be high or low. This machine learning method can also be used to predict multiple outcomes (multinomial regression). Application to aquaculture especially when extensive laboratory experiments are unavailable will help rural farmers, feed manufacturers, and researchers have an idea of the expected Feed conversion ratio of feed being compounded.

From the probability results obtained, different solvers provide closely similar results but may differ in the probabilities of prediction made. While Newton-CG and SAG solvers perform better than liblinear solver, the results indicate the need to run the same prediction with multiple solvers and compare the resulting probabilities.

The study indicates logarithmic regression can be used to successfully predict the FCR of feed compounded for *Clarias gariepinus* as either high (1) or low (0). As long as feed composition contains any of the following set of feedstuff: Fish-meal, rocky-prawn, Feather-meal, brewers-waste, soybean-meal, blood-meal, Lipid, maggot-meal, wheat-bran, yellow-maize, groundnut-cake, Carboxymethylcellulose (CMC), Vitamin, Chromic, Minerals, Calcium, Cellulose and Tapioca.

Also, proximate analysis data needed for the model include: Feed Protein content, fat content, ash content, crude-fiber, Nitrogen Free Extract, moisture, culture period, and fish weight at the onset of the experiment.

The study utilized historical data in making predictive analysis, the quantity and quality of data used in training models determined the accuracy and robustness of such models. This can be made easier with the use of cloud relational databases that hold experimental data and make them easily accessible. This would enhance aquaculture development especially in areas where experimental funding is quite a challenge.

### Compliance with Ethical Standards

### Conflict of Interest

The author declares that there is no conflict of interest. Historical data utilized in the research is appropriately cited.

### **Ethical Approval**

For this type of study, formal consent is not required.

### **References**

- Aniebo, A. O., Erondue, E. S. & Owen, O. J. (2009). Replacement of fish meal with maggot meal in African catfish (*Clarias gariepinus*) diets. *Revista Científica UDO Agrícola*, **9**(3): 666–671.
- AOAC. (1990). *Official Methods of Analysis, 13th edition*. Association of Official Analytical Chemists.
- Chor, W. -K., Lim, L. -S. & Shapawi, R. (2013). Evaluation of feather meal as a dietary protein source for African catfish fry, *Clarias gariepinus*. *Journal of Fisheries and Aquatic Science*, **8**(6): 697–705. <https://doi.org/10.3923/jfas.2013.697.705>
- Degani, G., Ben-Zvi, Y. & Levanon, D. (1989). The effect of different protein levels and temperatures on feed utilization, growth and body composition of *Clarias gariepinus* (Burchell 1822). *Aquaculture*, **76**(3-4): 293-301. [https://doi.org/10.1016/0044-8486\(89\)90082-3](https://doi.org/10.1016/0044-8486(89)90082-3)
- Dudusola, A. & Akinlade, S. T. (2014). The use of maggot meal in African cat fish feeding. *Advances in Aquaculture and Fisheries Management*, **1**(5): 49–51. <https://doi.org/10.46882/AAFMM/1008>
- Falaye, A. E., Omoike, A. & Adesina, S. B. (2015). Growth performance and nutrient utilization of catfish *Clarias gariepinus* fed varying inclusion level of fermented unsieved yellow maize. *Continental Journal of Biological Sciences*, **8**(1): 14–23. <https://doi.org/10.5707/cjbiolsci.2015.8.1.14.23>
- FAO. (2016). *The State of World Fisheries and Aquaculture 2016. Contributing to Food Security and Nutrition for All*.
- Farahiyah, I. J., Zainal, A. A. R., Ahmad, A., Mardhati, M., Thayalini, K. & Yong, S. T. (2016). Evaluation of local feed ingredients based diets on growth performance of African catfish, *Clarias gariepinus*. *Malaysian Journal of Animal Science*, **19**(2): 39–46.
- Li, M. H., Robinson, E. H., Bosworth B. G. & Torrains, E. L. (2014) Growth and feed conversion ratio of pond-raised hybrid catfish harvested at different sizes, *North American Journal of Aquaculture*, **76**(3): 261-264. <https://doi.org/10.1080/15222055.2014.911225>
- NCSS. (2020). NCSS Statistical Software, Logistic Regression. NCSS, LLC. Kaysville, Utah, USA, <https://ncss.com/software/ncss>
- Ng, W. -K., Teh, S. -W, Chowdhury, K. M. A. & Bureau, D. P. (2013). On-farm feeding and feed management in tilapia aquaculture in Malaysia. In M. R. Hasan & M. B. New, (Eds.), *On-farm feeding and feed management in aquaculture*. FAO Fisheries and Aquaculture Technical Paper No. 583. Rome, FAO. pp. 407–431
- Oyekanmi, F., Omoniyi, I. & Akegbejo, S. (2013). Whole rocky fresh water prawns, *Caridina africana* as replacement for fish-meal in diets for African catfish (*Clarias gariepinus*). *Proceedings of 28<sup>th</sup> Fisheries Society of Nigeria Annual Conference*. pp. 366-370.
- USAID-HARVEST. (2011). *Feed Conversion Ratio* [Technical Bulletin]. [www.CambodiaHARVEST.org](http://www.CambodiaHARVEST.org)