Use of Cross-Sectional Variables to Complement the Prediction of Magnitude of Flood Along Foma River Areas

Chindo, A. A.
National Centre for Hydropower Research and Development, University of Ilorin
Correspondent Email - *chindo4september@gmail.com*
Phone Number: +2347030494872

Abstract
Flood hazards have been overwhelming in recent years. The hazard tends to impact on human lives and resulting to huge economic damage across the World. However, the prediction of the magnitude of flood has been faced with challenges such as inaccurate data, poor assessment of drainage basin, river pollution and encroachment, especially in Nigeria across the river areas. In this study, Geographical Information System (GIS) was used to derive cross-sectional variables in order to complement the prediction of magnitude of flood along Foma river areas. The Foma river was buffered 15 and 30 meters to expose 49 and 105 structures which were highly and fairly vulnerable to flood hazards respectively out of 530 structures captured. However, the remaining 376 structures were classified not vulnerable to flood hazard along Foma river areas. The accuracy of the ordered logit model was 81%, while the classification error due to the harmonization of the precision value (0.8026) and the recall value (0.6386) was put at 10%. In addition, the cross-sectional variables that were found to be significant at $\alpha = 0.005\%$ are the river watersheds, the vulnerability status classification of structure across the river areas, the vulnerable structures identified, inadequate bridges and culverts along the river areas, inappropriate size of bridges and culverts, and extreme pollution along the river areas. This study is recommending the use of significant cross-sectional variables to complement the prediction of the magnitude of flood along the river areas.

Keywords: Buffering, Cross-sectional, Georeferenced, Magnitude, Spatial

1. Introduction

The periodic occurrence of flood and its overwhelming hazard tends to impact on human lives and result in serious economic damage across the world. Its strength tends to threaten the entire world due to the underlining effect of climate change [6]. However, evaluating the possibility and magnitude of flood have seriously been hindered byclimate change, inaccurate data, poor assessment of drainage basin, pollution, and encroachment [1]. Many previous studies have reported some difficulties in the sampling technique of conventional rain and discharge measurement which have prevented the accurate evaluation of the magnitude of the flood, especially along the river areas.

Incidentally, the assessment of rivers tends to indicate that the level of flood quite differs from one river to the other even despite being in the same geographical location. This can be attributed to both natural and human factors such as watershed, drainage basin, drainage capacity, level of pollution, encroachment activities, and many others. Studies mostly focus on the relationship between the amount of rainfall and the magnitude of floods. This assessment might not be quite accurate due to disparity in distribution of rainfall along the same geographical location, and contributing streams, tributaries or rivers with their peculiar factors and determinants [4].

Similarly, studies have established that Geographic Information System (GIS) is a very powerful tool that that can be used to consider factors considered being peculiar in determining the magnitude of flood along river areas. The realization of data with the use of GIS techniques will give a complementary approach to determine cross-sectional variables which are significant to predicting the magnitude of flood along the river course. Cross-sectional variables can be observed at the local scale. The procedures involve numerical data about intrusion and runoff dynamics [10]. The variables have some peculiar characteristic that dictates the direction of flow of flood in each river or stream rather than just a prediction through generalization which may not be so accurate.

1.1 Flood Forecasting and Mitigation in Nigeria

Nigeria is likely to face the consequences of climate change due to its geographical location. The country is bounded by Atlantic ocean to the south and the Sahara Desert to the north. This, by implication, may lead to an increase in the temperature that influences the rainfall pattern and resulting in the rise of extreme drought and flood [1]. Due to its location, several cases of flooding in Nigeria have been reported in recent times, mostly in Sokoto, Lagos, Ibadan, Abeokuta, Gusau, and Makurdi. Not less than 39 people were killed due to flooding in central Nigeria, Plateau State, towards the end of July 2012. The Lamingo dam had an overflow and swept across several localities in Jos, and about 200 houses were inundated or devastated after protracted rain. At least 35 people were reported missing, prompting the head of the Red Cross organization to announce that relief efforts were being initiated [3]. The spatial distribution of areas extremely affected by the flooding in Nigeria is shown in Figure 1.1

Figure 1.1: Distribution of Areas Affected by Extreme Floods in Nigeria (Chindo, *et al*. 2019)

The city of Ilorin which is Kwara state capital is located along the River Niger area, north-central part of Nigeria. The state is found between the latitude $8^0 24'$N and $8^0 36'$N and between longitude $4^0 10'$W and $4^0 36'$E. During the 2017 raining season, the city of Ilorin experienced a devastating flood hazard. Many residential buildings were reported to have submerged after a protracted rain that lasted for hours. The heavy rain, which was accompanied by flooding, washed away asphalt on some township roads. The ravaging flood also washed away bridges and destroyed valuable properties, as reported in the Nigeria Tribune newspaper [2]. Figure 1.2 captured the Alagbado bridge along Foma river which was washed away during the 2017 heavy raining season.

Figure 1.2: Alagbado bridge along Foma-river washed away by the flood.

The aim of this study is to develop a supervised model to complement the prediction of the magnitude of flood along the banks of Foma river. Other sub-objectives are to examine the river buffering in 15 meters and 30 meters across the Foma river floodplain areas, identify the cross-sectional variables in complementing the prediction of the magnitude of flood along the buffering areas of Foma river, determine the significance of the cross-sectional variables in complementing the prediction of the magnitude of flood along Foma river banks using Ordered Logistic Regression (OLR) model, and evaluate the performance of OLR in complementing the prediction of the magnitude of flood along Foma river using performance measurement metrics.

2. Data Generation Using GIS and Cross-sectional Approach

A cross-sectional study is an established approach to estimate the outcome of interest at a particular time, for a specified location and it is usually applied for health planning, hazard, or risk exposure. Cross-sectional study reflected a short period of exposition and has some characteristics associated with a specific period [5]. Among the reasons for carrying out the cross-sectional study is to describe the survey exercise, which usually does not have a hypothesis. The main aim is to describe some groups or sub-groups about the outcome of risk factors. Also, the goal is to elicit the prevalent outcome of interest for a descriptive population or group at a given time [12].

Similarly, the GIS approach to flood hazard evaluation and management has not been an often used method until the year 2000. The work of [8], initially used the GIS to estimate several risks in many areas of Colorado, to determine the suitability of land. The development of GIS modelling for excess rainfall was the approach adopted by [11]. In Nigeria, It was observed that the difficulty in the sampling technique of the conventional rain coupled with discharge measurement networks makes it challenging to observe and predict flood accurately [7]. While the work of [9] elicited some technical deficiencies that have been preventing Nigeria from having an accurate rainfall data. The research enumerated the present capacity of Nigeria's rain gauge network and the need according to the World Meteorological Organization's (WMO) guidelines. Nigeria presently has 87 rain gauges, instead of 1057. In essence, the country needs extra gauges of 970 to achieve a gauge density of 874 km^2 per gauge for the appropriate measurement of rainfall. The inaccurate records of rain data led Nigeria to be hugely affected by the devastating flood of September 2012. This event had a negative effect on the economy, roads, ports, rail lines, and most especially the water infrastructures.

3. Research Designs and Methodology

This study focused on the assessment of a complementary approach to flood prediction using GIS software. The software was initiated through Global Positioning System (GPS) to obtain the coordinates of the river channels, while the images of the earth are referenced in eastern (X) and northern (Y) coordinates. The processes elicited some cross-sectional variables from the river areas which are significant in determining the magnitude of flood along the Foma river channel. Arc GIS 9.3® software was used to analyze high-resolution imagery from Google earth.

3.1 Research Designs

The problem focused upon and addressed in this study is to develop a supervised model of cross-sectional variables to complement the prediction of the magnitude of flood along the Foma river. The flow chart in Figure 3.1 demonstrated the processes in which the cross-section variables are generated and evaluated.

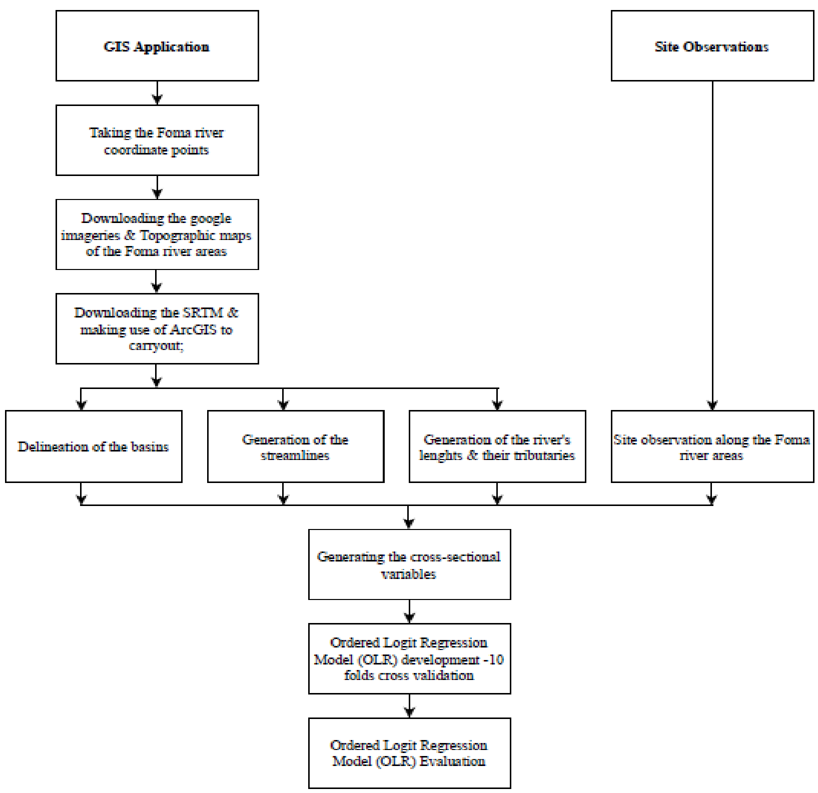

Figure 3.1: Flow Chart Showing the Research Design.

The use of GIS tools and methods ensure the generation and observation of some cross-sectional variables that are suspected to be significant in predicting the magnitude of flood along the Foma river. Table 3.1 presents the cross-section variables that were derived through the application of GIS and site observations.

Table 3.1: Cross-sectional Variables from Foma River Areas

Variable Name	Task	Values	Data Type
River Watershed	Input	Shed1, Shed2, Shed3, Shed4	Nominal
Drainage Density	Input	0.0001, 0.0002, 0.0005, 0.0007	Ordinal
Vulnerable Status	Input	Not Vulnerable, Fairly Vulnerable, Highly Vulnerable	Ordinal
Vulnerable Structures	Input	Hospital, Police post, Fishery ponds, Abattoir, Educational, Commercials, Slum, Agriculture, Residential	Nominal
Bridges and Culverts	Input	CAIS, Apalara, Oke-foma, Foma-bridge, Ajetunmabi, Oloje-bridge, Abata Baba-oyo, Alagbado Bridge, Sobi-bridge	Nominal
Size of Bridges and Culvert (m)	Input	2.1, 4.5, 7.2, 11.2, 14.9, 15, 19.5, 60.8	Ordinal
River Point	Input	Source, Middle, Extreme, Terminal	Nominal
River Pollution	Input	Fair, High, Severe, Extreme	Ordinal
Magnitude of Flood	Target	Mild, Moderate, Severe, Extreme	Ordinal

Source: Field Work (2019)

3.2 The Research Methodology

The study captured the vulnerability status of structures induced by flood activities along the course of the Foma river using remote sensing techniques. This was carried out on flood-prone areas and the buffering was examined using Arc-GIS. Structures located within 15 meters of the river bank were considered highly vulnerable to flood hazards, while those structures within 30 meters of the river were considered fairly vulnerable (The map of Ilorin west was acquired to create a database for the buffering). Also, Foma river map was extracted, georeferenced, and digitalized into 1:50,000 from the topographical map of Kwara state. The digitalization of the map involves the process of electronic scanning in order to convert it to points and lines using on-screen digitization. Specifications were then made to identify the objects on the map so that the Arc-GIS was linked using the spatial data with attributes of identified structures.

The buffering of the river revealed the number of structures that were highly vulnerable, fairly vulnerable and those that cannot be affected by flood hazards. Plate 2 exhibits the status of vulnerable structures along the river areas, while Table 3.2 reflects the delineation of the vulnerable status and number of structures within each drainage area along the Foma river.

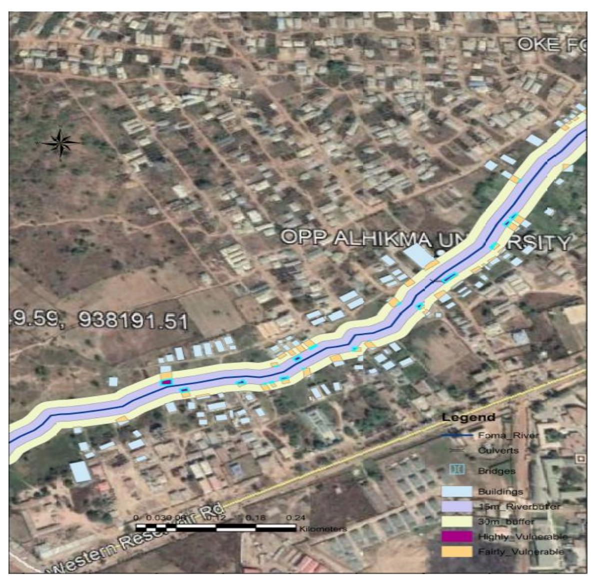

Plate 2. Showing Vulnerable status of structures along Foma river areas

Table 3.2: Vulnerable Status Classification along the River

ID	Description	Frequency
0	Not Vulnerable	377
1	Fairy Vulnerable	105
2	Highly Vulnerable	49

To carry out the pre-classification exercises, the original sample was split into 90/10 % repeated seed training/ testing sets. A non-exhaustive cross-validation *k*-fold was used with *k=10* so that the original sample be randomly divided into *k* equal-sized subsamples. Thus, taking out the subsample to be known as validation variables to test the model, where outstanding *k-1* subsamples were considered as training data. The process is repeated until every *k*-fold serves as the test set, such that the average record scores *(E)* of the 10 folds become the performance metric of the model. Where *E* as defined in equation (1) is the addition of performance scores in the iteration. The cross-validation technique in the study is demonstrated in Figure 3.2.

$$\text{Where } E = \frac{1}{10}\sum_{i=1}^{10} E_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

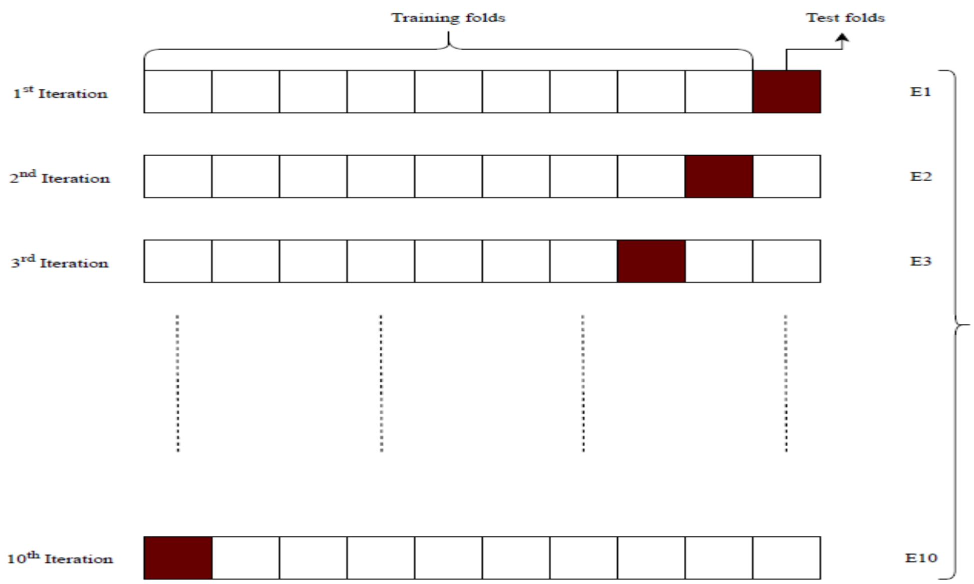

Figure 3.2: Cross-validation technique in the study

However, the dependent variable (magnitude of the flood) is taking more than two categories. Thus, we employed the use of the ordered logit approach due to its capability to predict the presence or absence of a dependent variable, and its uniqueness in predicting the probability of each character in the model because the chance is a ratio. The interpretation of results in the odd ratio, parameter estimate, and probability is quite an added advantage in the area of the results' analyses. Since the dependent variable has more than two categories, and the interval between the categories was in the relative sequential order in a way that the value is indeed higher than the previous one, then the ordered logit approach would be deemed applicable.

Ordered response models are usually applied when the dependent variable is discrete and when there is an ordered measurement. In general, consider an ordered response variable Y, which can take the value y + 1, 0, 1, 2……j. Such that the general linear function

$$\hat{Y} = X\beta + \varepsilon \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

The latent variable $\hat{Y}$ is not directly observed, thus, the threshold set by which the observed value change as predicted, otherwise known as 'CUT POINT'. Cut points establish the relationship between $\hat{Y}$ and Y, let α_i be the threshold. Then.

$$Y = \begin{cases} Y_0 & \text{iff} \quad \hat{Y} < \propto_0 \\ Y_1 & \text{iff} \quad \propto_0 \leq \hat{Y} \leq \propto_1 \\ Y_2 & \text{iff} \quad \propto_1 \leq \hat{Y} \leq \propto_2 \\ Y_3 & \text{iff} \quad \hat{Y} \geq \propto_2 \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

The response variable Y takes four value categories: 0= mild flood, 1= moderate flood, 2= severe flood, and 3= extreme flood. Therefore, the unknown parameters α_i are estimated jointly with β_s via maximum likelihood. The $\hat{\propto}_i$ estimates are reported on Gretl as cut_1, cut_2 , and cut_3 in this case. In other to apply the models in Gretl, the dependent variable must either take only non-negative integer values or be explicitly marked.

3.3 Determination of Classification Error

In the multi-class measurement, errors in classification have different implications. Errors in classifying *Y* as *X* may likely to have different weighted implications than classifying *C* as *D,* and many more of such errors. The accuracy measure does not take any of such problem into account. The pre-determined assumption was that the sample distribution among classes is balanced.

Thus, in the case of imbalanced distribution, the most commonly used classification approach repeatedly produces a disappointing estimate. In this case, the conventional approaches need to be re-examined to address the problem of imbalanced data classification. However, the confusion matrix will create an error table to derive the measurement metrics.

In order to determine the level of accuracy of the significant classifications, the study developed 4 by 4 confusion matrices for each of the 10 folds. The matrices enabled the derivation of the measurement metrics (accuracy, F1-Score, precision, and recall). Previous studies have established that accuracy works well in describing balanced data and misleading the performance in imbalanced data. Additionally, F1-score has proven to be a useful metric when the data is imbalanced.

4. Classification Performance

The 10-folds cross-validation classification accuracy is demonstrated in Table 4.1.

Table 4.1: Ordered Logit Classification Performance Estimate

Ordered Logit Accuracy for the Folds										
Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Fold 6**	**Fold 7**	**Fold 8**	**Fold 9**	**Fold 10**	**Average**
80.7	80.5	80.5	80.5	80.3	81.1	81.3	80.1	81.6	80.3	80.7%

It was observed that the average number of cases correctly predicted is 80.7%. By this impression, the OLR model is said to be approximately 81% good to predict the magnitude of flood along the Foma river areas. With this classification accuracy, the variables are well fitted to complement the prediction of the magnitude along the Foma river flood. This correct percentage classification is quite high and explains how strongly significant the variables are. Similarly, this study presented eight (8) cross-sectional variables in predicting the magnitude of flood along the Foma river for classification. However, six (6) out of the eight (8) variables' average P-values were less than 0.05. The six variables were found significant and relevant to complement the prediction of the magnitude of flood along the Foma river flood channel. The 6 cross-sectional variables are the river watersheds, vulnerable status, vulnerable structures, bridges, and culverts (B & C), size of bridges and culverts and river pollution. Meanwhile, the 2 other cross-sectional variables were omitted due to exact collinearity, which indicated serial linearity between the two variables; they are the river drainage density and river points along the river channel.

There was an indication of a continuous increase in the probability of the magnitude of flood along the river which was demonstrated by the cut point estimates. The estimates of P-values were highly significant all through the folds, and their coefficients were equally positive. The significance of the P-value is an indication that there is a steady and continuous rise in the level of magnitude of flood across the Foma river areas. Meanwhile, due to the imbalanced data distribution, this study further evaluates the level of significance of the cross-sectional variables using the measurement metrics.

4.1 Discussion on Level of Classification Accuracy

The OLR model estimate was quite high which is at 81%. This suggested a high level of classification of the cross-sectional variables in complementing the prediction of the magnitude of flood along Foma river. This study further described the classification performance of OLR using the measurement metrics due to the high disparity in the sampling distribution. Figure 4.1 demonstrates the level of sampling disparity in the study.

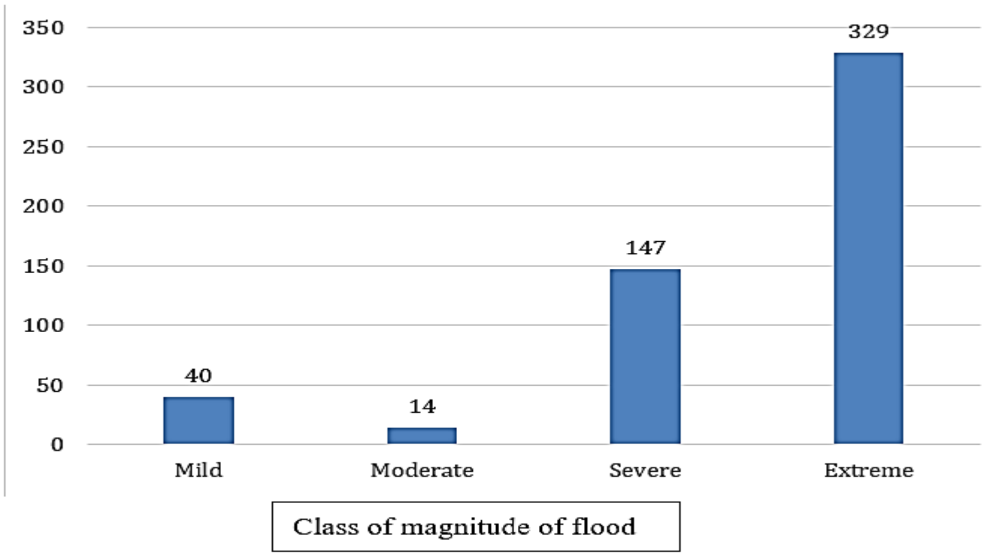

Figure 4.1: Level of magnitude of flood along Foma river areas.

There was an indication of high disparity in the magnitude of flood along the Foma river areas. Thus, the prediction of the magnitude of flood tends to favour the higher categories compared to the lower categories. In order to describe the performance of the OLR model, F1-score metric was used to measure the OLR performance and minimize the sampling disparities through the use of precision and recall.

The weighted average of precision and recall was used to measure how good the OLR classification is at predicting the magnitude of flood along Foma river. The four-measurement metrics employed in this study are the accuracy, precision, recall and F1-score to determine the strength of the prediction. The results of the four-measurement metrics for the models are presented in Table 4.2.

Table 4.2: Values of the Measurement Metrics

Folds	Measurement metrics			
	Accuracy	Precision	Recall	F1-score
Fold – 1	0.8092	0.8778	0.6462	0.7444
Fold -2	0.8050	0.8764	0.6390	0.7390
Fold – 3	0.8050	0.8761	0.6397	0.7394
Fold – 4	0.8050	0.6262	0.6209	0.6236
Fold – 5	0.8029	0.8739	0.6381	0.7376
Fold – 6	0.8113	0.6321	0.6259	0.629
Fold – 7	0.8134	0.8822	0.6681	0.7601
Fold – 8	0.8008	0.6224	0.6221	0.6222
Fold – 9	0.8155	0.8833	0.6463	0.7466
Fold – 10	0.8029	0.8750	0.6396	0.7390
Average	**0.8071**	**0.8026**	**0.6386**	**0.7081**

The average values of each of the multi-class metrics derived in Table 4.2 were directed towards determining the performance of the OLR model in predicting the magnitude of flood along Foma river areas. Figure 4.2 illustrated the supervised model for complementing the prediction of the magnitude of flood using cross-sectional variables.

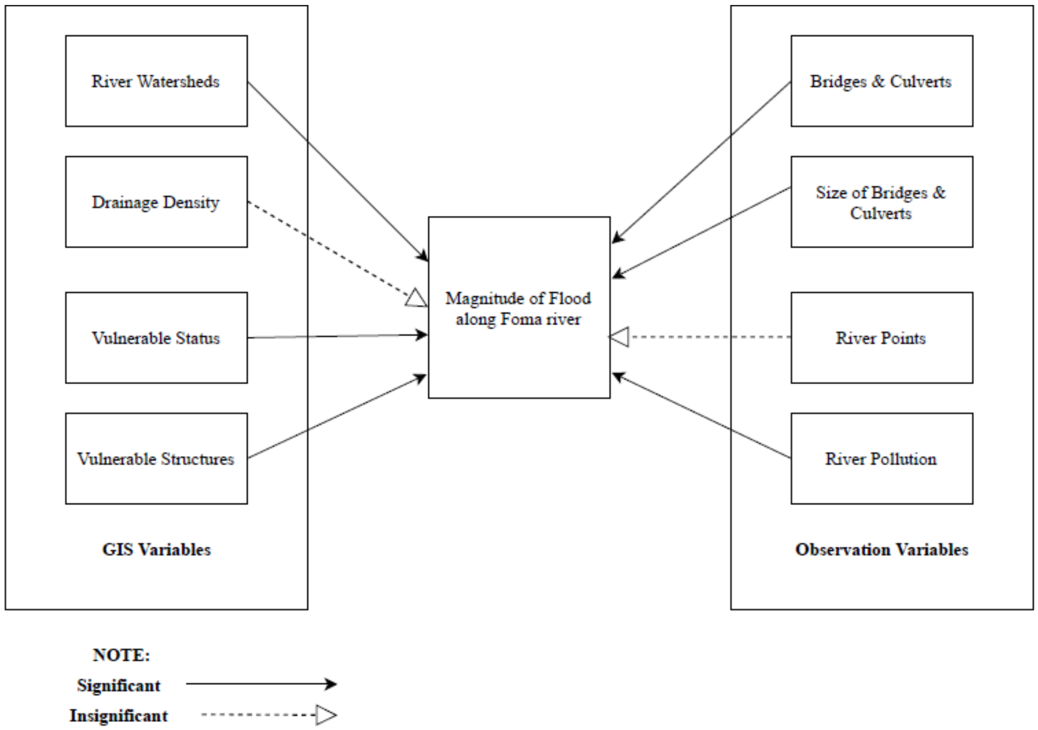

Figure 4.2: Supervised Model for Complementing the Prediction of Magnitude of Flood along Foma river using Cross-sectional Variables.

Conclusion and Recommendations

It was observed that the ordered logit regression average prediction value 80.71% was vulnerable to error due to the high disparity in the sampling distribution. Consequently, the model was subjected to further evaluation using the F1-score which reduced the sampling error by approximately 10%. The supervised model's average capacity to predict the magnitude of flood along Foma river areas is 70.81%. Similarly, the model classification provided six (6) out of the eight (8) cross-sectional variables evaluated to be significant in complementing the prediction of the magnitude of flood along Foma river areas. While the other two variables were considered insignificant due to absolute collinearity.

The river buffer areas within 15 meters and 30 meters established the vulnerability status of structures along the Foma river floodplain. This exercise identified a total number of 154 structures to be vulnerable to flood hazards along the river bank areas. One hundred and five (105) of the structures were fairly vulnerable, while forty-nine (49) similar structures were at a very high risk of flood hazard along the river areas. In conclusion, this study is recommending the use of significant cross-sectional variables to complement the prediction of magnitude of flood along the river banks.

References

[1] Ayanshola, A., Olofintoye, O., &Obadofin, E. The Impact of Global Warming on Precipitation Patterns in Ilorin and the Hydrological Balance of the Awun Basin. Slovak *Journal of Civil Engineering*. 2018, 26(1), 40-46.

[2] Biola Azeez. Flood Damages Property in Ilorin-Latest News-Tribune Online. From https//tribuneonline.com. 2017, May 28, 2017

[3] Chindo, A. A., Mohd Shaharanee, I. N., & Mohd Jamil, J. A conceptual framework for predicting the effects of encroachment on magnitude of flood in Foma-riverarea, Kwara State, Nigeria using data mining. *TEST Engineering & Management*. 2019, *81*, 3913-3918.

[4] Du, J., Cheng, L., Zhang, Q., Yang, Y., & Xu, W. Different flooding behaviours due to varied urbanization levels within river Basin: A case study from the Xiang river Basin, China. *International Journal of Disaster Risk Science*. 2019, *10*(1), 89-102.

[5] Ezzatvar, Y., Calatayud, J., Andersen, L. L., Aiguadé, R., Benítez, J., &Casaña, J. Professional experience, work setting, work posture and workload influence the risk for musculoskeletal pain among physical therapists: A cross-sectional study. *International Archives of Occupational and Environmental Health*. 2020, 93(2), 189-196.

[6] Hasselaar, J. J. Hope in the Context of Climate Change: Jonathan Sacks' Interpretation of the Exodus and Radical Uncertainty. International Journal of Public Theology. 2020, 14(2), 224-240.

[7] Isma'il, M., & Saanyol, I. O. Application of remote sensing (RS) and geographic information systems (GIS) in flood vulnerability mapping: case study of River Kaduna. *International Journal of Geomatics and Geosciences*. 2013, *3*(3), 618.

[8] Mejía-Navarro, M., Wohl, E. E., & Oaks, S. D. Geological hazards, vulnerability, and risk assessment using GIS: model for Glenwood Springs, Colorado. In Geomorphology and Natural Hazards. 1994, 331-354. Elsevier.

[9] Ngene, B. U., Agunwamba, J. C., Nwachukwu, B. A., & Okoro, B. C. The Challenges to Nigerian Raingauge Network Improvement. *Research Journal of Environmental and Earth Sciences*. 2015, 7(4), 68-74.

[10] Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M.,&Holden, J. Land-use change impacts on floods at the catchment scale: Challenges and opportunities for future research. Water resources research. 2017, 53(7), 5209-5219.

[11] Schumann, A. H., Funke, R., & Schultz, G. A. Application of a geographic information system for conceptual rainfall-runoff modelling. *Journal of Hydrology*. 2000, 240(1-2), 45-61.

[12] Venkataramanan, V., Packman, A. I., Peters, D. R., Lopez, D., McCuskey, D. J., McDonald, R. I., & Young, S. L. A systematic review of the human health and social well-being outcomes of green infrastructure for stormwater and flood management. *Journal of environmental management*. 2019, *246*, 868-880.