Saltwater Anglers toward Marine Environmental Threats Using Multilayer Perceptron Neural Network Framework

Yeong Nain Chi[1], and Jennifer Chi[2]
[1]University of Maryland Eastern Shore, [2]University of Texas at Dallas
[1]ychi@umes.edu, [2]jxc126831@utdallas.edu

Corresponding author email: ychhi@umes.edu

***Abstract*-The data for this study extracted from the 2013 National Saltwater Angler Survey is to understand saltwater recreational anglers' concerns to the threats of marine environment and to identify groups exhibiting common patterns of response. Concerns of marine environmental threats from these participants were examined through factor analysis which identified three reliable factors. Cluster analysis was employed to identify three prominent groups. The multilayer perceptron neural network was utilized as a predictive model in deciding saltwater recreational anglers' perceptions toward marine environmental threats. In the architectural point of view, it showed a 13-6-3 neural network construction. The results revealed that "overfishing in commercial fisheries" contributed most in the neural network model construction, followed by "industrial pollution", "oil and gas extraction", "marine habitats loss or degradation", "climate change", and so forth. Results of this study may provide insight regarding the concerns of marine environmental threats from saltwater recreational anglers as an indicator of potential participation and behavior of saltwater recreational fisheries management.**

***Keywords*-Saltwater, Recreational Anglers, Marine Environmental Threats, Factor Analysis, Cluster Analysis, Discriminant Analysis, Multilayer Perceptron Neural Network.**

1. Introduction

The marine environment provides a range of important ecological goods and services for our society. To ensure the sustainability of this marine ecosystem, we need to require an understanding of the beneficial implications it has to human visitors but also the risks our actions may have on marine life. With these perspectives in mind, there is a need for policy that helps promote long-term resilience and sustainability.

In the United States, the National Ocean Policy was created by Executive Order No. 13547 on July 19, 2010. Out of the National Ocean Policy, the interagency National Ocean Council, which consists of twenty-seven federal agencies, departments, and offices, was made to work on the nation's ocean management and research efforts [28]. As environmental conditions worsen through the effects of global climate change, mixed with an ever-growing human population and carbon footprint, the National Ocean Policy is a progressive step towards the right direction for ocean policy.

The National Ocean Policy focuses on nine primary goals that seek to address the most pressing issues regarding the ocean, coastal, and Great Lakes ecosystems and their resources. Among the nine goals, they included how to shift regulators to a more holistic ecosystem-based management perspective, how to better integrate scientific information into policy decisions, and how to create a spatial-planning process for determining what kinds of activities should take place in different parts of the U.S. waters [28]. Torres et al. [36] also puts a heavy emphasis on strategies and agency-specific tasks that may benefit long-term sustainability.

However, there are concerns for the future state of the marine environment. Certain themes that are highlighted by the National Ocean Policy include pressing issues such as the ocean economy, safety and security, coastal and ocean resilience, local choices, and scientific information. Emerging areas like illegal, unregulated and unreported fishing and seafood fraud, harmful algal blooms/hypoxia, regional marine plans, ocean acidification, coastal resilience and sea level rise tools, and coastal mapping further highlight issues relating to human health, economic stability, aquatic health and protection [28].

Current threats toward marine ecosystem come in various forms, such as the dramatic loss of marine biodiversity and habitat [3] [29] [33]; overexploitation and harvesting [4]; the introduction of exotic species; waste pollution (i.e. plastic debris) [15]; developing offshore wind power [1]; and the potentially serious effects of global climate change.

Crain, et al. [12] highlighted human threats to coastal marine ecosystem including habitat loss, overexploitation, eutrophication and hypoxia, pollution, invasive species, altered salinity, altered sedimentation, climate change, ocean acidification, and disease.

The National Oceanic and Atmospheric Administration (NOAA) Fisheries Service's report about 88% of saltwater recreational anglers ranked overfishing in commercial fisheries, 86% ranked industrial pollution, and 79% ranked marine habitats loss or degradation as severe or moderate threats to

the marine environment. Conversely, 67% of respondents ranked alternative energy (e.g. wave or wind) development, and 51% ranked shipping as not a threat at all or not very severe threats to the marine environment [7].
Based on an online survey from people in ten European nations, Gelcich, et al. [19] found out ocean pollution, overfishing, and ocean acidification were highly concerned. Furthermore, Hughes [24] listed significant historical, current and developing threats to USA recreational fisheries including intensified land use, hydro-morphological modification, substrate and riparian modification, chemical contaminants, eutrophication and hypoxia, overfishing, non-native fish, endocrine disrupting chemicals, nanoparticles, and climate change.
Although human perceptions, understandings, and responses have been widely explored through some environmental problems, much less attention has been given to human impacts on marine environment. Not many systematic studies have been conducted on understanding how saltwater recreational anglers perceive marine environmental threat(s), specifically on profiling this interest group by using the marine environmental threat scale approach. If there is qualitative data conducted on these anglers, their insight could contribute to more efficient strategies for long-term fisheries management.
Neural networks have been used in fisheries-related research in forecasting, classification, distribution, and fisheries management since 1978 [34]. Very few detailed studies have been carried out on understanding how saltwater recreational anglers perceive marine environmental threats and specifically on the classification of this interest group of recreational anglers by using the statistical model identification based neural network approach. The results of this study may provide baseline information about saltwater recreational anglers' understanding towards marine environmental threats and which groups and issues should be targets for marine fisheries awareness and management campaign.

2. Multilayer Perceptron Neural Network

Neural network is a computing technique designed to simulate the human brain's method in problem-solving. It is one of the most popular machine learning methods which is able to perform clustering and prediction tasks in a more reliable manner. According to Haykin [21], neural networks form a directed graph by connecting the artificial neurons, the basic processing components of the network.
The neuron is the basic information processing unit of a neural network. Mathematically, the output on the neuron can be expressed as follows:

$$f(x) = \varphi(\Sigma_{i=1} x_i w_i + b) \quad i = 1, \ldots, n$$

where the x_i are the input features, the w_i are the weights of respective inputs, b is the bias, which is summed with the weighted inputs to form the net inputs, and φ is the non-linear activation function. Bias and weights are both adjustable parameters of the neuron. The output of a neuron can range from -∞ to +∞. The neuron doesn't know the boundary. So we need a mapping mechanism between the input and output of the neuron. This mechanism of mapping inputs to output is known as activation function [21].
A simple perceptron is a linear classifier that produces a single output based on several real-valued inputs by forming a linear combination using its input weights. Input is typically a feature vector x multiplied by weights w and added to a bias b: $f(x) = wx + b$. Gardner and Dorling [18] define multilayer perceptron (MLP) as: "a system of simple interconnected neurons, or nodes, which is a model representing a nonlinear mapping between an input vector and an output vector". MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. Figure 1 shows a single hidden layer MLP with scalar output.

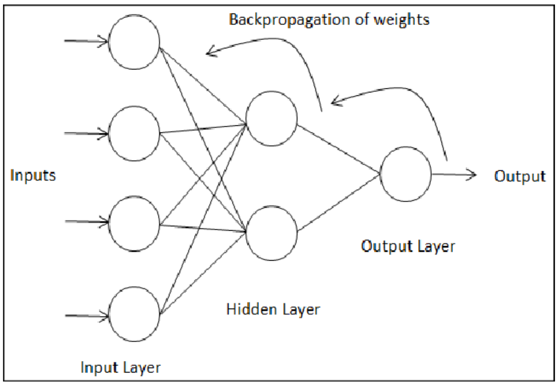

Figure 1: Single Hidden Layer MLP (adopted from Hassim and Ghazali [21])

MLP utilizes a supervised learning technique called back propagation to classify instances [37]. It analyzes the data set in three stages. The first stage is the "training process" that tries to perceive the association between the variables in the dataset. Based on what was learned during the first stage, it will attempt to discern a model which is done in a hidden layer. In this process, the optimal functions in the model are produced and dependent variables are assigned weights. Lastly, in the third stage, a new model is estimated and is called the output process [27].
MLP neural network has been frequently employed for modeling, prediction, classification, clustering, and optimization purposes [2] [13] [6] [38] [30] [14], which is one of the most widely used neural network techniques in data analysis [31] [8]. In a theoretical manner, MLP neural network is universal

approximator, and with respect to its inherent nature, it has a tremendous capacity of constructing any nonlinear mapping to any extent of accuracy [23]. It does not need a priori model to be assumed or a priori assumptions to be made on the properties of data [5].

3. Materials and Methods

For this study, the data was extracted from the 2013 National Recreational Angler Survey [7], which was developed by NOAA Fisheries Service and collected by the CIC Research, targeted saltwater recreational anglers, sixteen years of age and older, who had been saltwater fishing at least once in their life. The purpose of the survey was to elicit their perceptions, preferences, and attitudes about saltwater recreational fishing and its fisheries management. This survey was implemented in six regions in the United States, including North Atlantic, Mid-Atlantic, South Atlantic, Gulf of Mexico, West Coast, and Alaska.

In the survey, respondents were asked, "In your opinion, how much of a threat, if any, does each of the following factors pose to the marine environment?", to indicate thirteen statements regarding the threats of marine environment, using a Likert-type scale that ranged from 1 (not a threat at all) through 4 (severe threat), and 5 (I am unsure). This study examined the psychometric properties of marine environmental threat scale from the 7,763 saltwater recreational anglers who provided complete information for all 13-item marine environmental threats (Table 1).

The objectives of this study were to understand saltwater recreational anglers' perceptions toward marine environmental threats and to identify saltwater recreational angler groups exhibiting common response patterns. First, the dimensionality of the 13-item marine environmental threat scale was assessed by examining its factor solution [20]. A principal component analysis was used to determine the factors identified in this sample size. Second, a K-means cluster analysis was conducted to identify respondent groups exhibiting common response patters. Third, a multilayer perceptron neural network model was employed as a predictive model in deciding the saltwater recreational anglers' perceptions toward the threats of marine environment.

Table 1: Descriptive Statistics of Marine Environmental Threat Scale

In your opinion, how much of a threat, if any, does each of the following factors pose to the marine environment?	Mean	S.D.	Communalities
Industrial pollution	3.47	0.758	0.560
Oil and gas extraction	3.10	0.995	0.678
Climate change	2.72	1.106	0.520
Ocean acidification	3.38	1.134	0.458
Shipping	2.60	1.101	0.495
Overfishing in commercial fisheries	3.59	0.708	0.640
Overfishing in recreational fisheries	2.59	1.090	0.388
Non-native species	3.33	1.006	0.454
Aquaculture	3.04	1.392	0.552
Alternative energy (e.g. wave or wind) development	2.28	1.358	0.464
Algal blooms	3.46	1.048	0.515
Marine habitats loss or degradation	3.46	0.838	0.521
Dams/barriers	3.10	1.135	0.428

(Not a threat at all = 1, Not a very severe threat = 2, Moderate threat = 3, Severe threat = 4, I am unsure = 5)

4. Results

4.1 Factor Analysis

Factor analysis uses mathematical procedures for the simplification of interrelated measures to discover patterns in a set of variables [9]. In this study, the 13-item marine environmental threat scale was analyzed with varimax rotation, providing a clearer separation of the factors. Specifically, the amount of variance explained by the extracted factors (i.e., their eigenvalues) was noted. In addition, item-factor correlations (i.e., factor loadings) and other indices of model adequacy were examined. The factor loading of the three resulting factors were shown in Table 2. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.880, which met the fundamental requirements for factor analysis. The Bartlett's test of Sphericity showed that nonzero correlations existed at the significance level of 0.001.

The Cronbach's alpha, developed by Lee J. Cronbach [11] in 1951, is widely used to measure how closely related a set of items are as a group. The internal consistency coefficient score of the 13-item marine environmental threat scale showed the Cronbach's alpha of 0.824 was acceptable. Each of these three factors had a satisfactory Cronbach's alpha of 0.736, 0.722, and 0.521, respectively, which explained a cumulative 51.338 percent of the variance in the statement response (Table 2).

Table 2: Factor Analysis of Marine Environmental Threat Scale

In your opinion, how much of a threat, if any, does each of the following factors pose to the marine environment?	*Environmental Change*	*Industrial Development*	*Fisheries Activities*
Aquaculture	0.732		
Algal blooms	0.673		
Alternative energy development	0.604		
Dams/barriers	0.569		
Non-native species	0.559		
Ocean acidification	0.544		
Oil and gas extraction		0.807	
Climate change		0.693	
Industrial pollution		0.673	
Shipping		0.606	
Overfishing in commercial fisheries			0.784
Overfishing in recreational fisheries			0.532
Marine habitats loss or degradation			0.512
Eigenvalue	2.698	2.371	1.605
% of variance	20.757	18.237	12.344
Cumulative %	20.757	38.994	51.338
Reliability Alpha Coefficient	0.736	0.722	0.521
Reliability Alpha Coefficient of All 13 Statements = 0.824			
Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy = 0.880			
Bartlett's Test of Sphericity: Approx. Chi-Square = 23703.761, $df = 78$, $p < 0.001$			

As a result of exploratory factor analysis, three factors were identified. Each factor was named after a defined variable that made the greatest contribution in each dimension. An initial interpretation of these factors suggested that Factor 1 named *Environmental Change* comprised of six items (structure coefficients ranging from 0.732 to 0.544) and explained 20.757 percent of the variance with an eigenvalue of 2.698. Factor 2 had an emphasis in *Industrial Development* which comprised of four items (structure coefficients ranging from 0.807 to 0.606) and explained 18.237 percent of the variance with an eigenvalue of 2.371. Lastly, Factor 3 focused on *Fisheries Activities* which comprised of three items (structure coefficients ranging from 0.784 to 0.512) and explained 12.344 percent of the variance with an eigenvalue of 1.605 (Table 2).

4.2 Cluster Analysis

Cluster analysis technique assigns objects to groups so that there is as much similarity within groups, and difference between groups, as possible [10]. In this study, the K-means clustering analysis was applied to identify a solution with a specified number of clusters to the saved factor scores. The factor scores of marine environmental threat dimensions were used to cluster saltwater recreational anglers. Consequently, a three-cluster solution was agreed upon, which was labeled as *Utilized Concern*, *Environmental Concern,* and *Developmental Concern* clusters (Table 3).

The *Utilized Concern* cluster: this cluster was the largest group, comprising of approximately 45.0 percent of respondents, named because of the strongly positive factor score associated with *Industrial Development* and *Fisheries Activities* factors, negatively identified with *Environmental Change* factor among these respondents. The *Environmental Concern* cluster: this was the smallest group comprising of approximately 27.0 percent of the respondents. These respondents were positively connected with *Environmental Change* and *Fisheries Activities* factors, particularly negative and strongly identified with *Industrial Development* factor. The *Fisheries Concern* cluster: with 28.0 percent of the respondents, this group was named after the negatively strong association with *Fisheries Activities* and *Environmental Change* factors, but positively identified with *Industrial Development* factor (Table 3).

Table 3: Cluster Analysis of Saltwater Recreational Anglers

	Utilized Concern	*Environmental Concern*	*Fisheries Concern*
Environmental Change	-0.1352	0.3559	-0.1257
Industrial Development	0.6243	-1.1209	0.0778
Fisheries Activities	0.5132	0.3693	-1.1775
n = 7763	3490	2095	2178
Percentage	45.0	27.0	28.0

4.3 Discriminant Analysis

Discriminant analysis is a statistical technique to classify the target population into the specific categories or groups based on the certain attributes (predictor variables or independent variables) [16] [35]. Results of the cluster analysis were tested for accuracy using the linear discriminant analysis employed as a useful complement to cluster analysis, which is used primarily to predict membership in two or more mutually exclusive groups. In this case, the Wilk's Lambda scores were 0.199 (χ^2 = 12517.402, df = 6, p < 0.001) and 0.455 (χ^2 = 6112.717, df = 2, p < 0.001) for both discriminant functions, respectively, indicating that group means were significantly different. The canonical correlation results were both above 0.7, supporting that there were strong relationships between the discriminant score and the cluster membership (Table 4).

Table 4: Canonical Correlation of Discriminant Functions

Function	Eigenvalue	% of Variance	Canonical Correlation
1	1.283	51.7	0.750
2	1.199	48.3	0.738

* First two canonical discriminant functions were used in the analysis.

Two discriminant functions were formulated (Table 5). The first function is a function for discriminating between *Utilized Concern*, *Environmental Concern* and *Fisheries Concern* clusters combined, and the second function for discriminating between *Environmental Concern* and *Fisheries Concern* clusters, respectively. The first function is the most powerful differentiating dimension, but the second function may also represent additional significant dimensions of differentiation. Though mathematically different, each discriminant function is a dimension which differentiates a case into categories of the dependent variable, the three identified angler groups, based on its values on the independent variables. Furthermore, the territorial map is a tool for assessing discriminant analysis results by plotting the group membership of each case on a graph (Figure 2).

Table 5: Standardized Canonical Discriminant Function Coefficient

	Function 1	Function 2
Environmental Change	-0.421	-0.061
Industrial Development	0.908	0.435
Fisheries Activities	-0.423	0.907

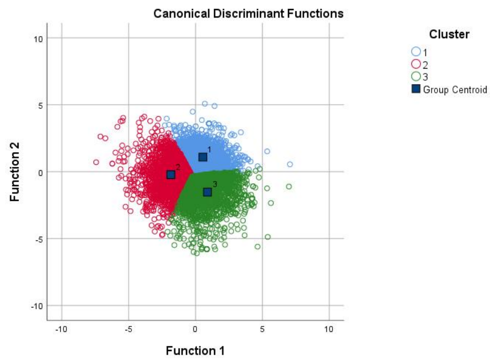

Figure 2: Territorial Map (1 = *Utilized Concern* cluster; 2 = *Environmental Concern* cluster; 3 = *Fisheries Concern* cluster)

The classification results based on discriminant analysis, 3490 cases fell into the *Utilized Concern* cluster, 2095 fell into the *Environmental Concern* cluster, and 2178 fell into the *Fisheries Concern* cluster in the original row total, which is the frequencies of groups found in the data. Across each row, the case amount in the group can be classified by this analysis into each group. For example, of the 3490 cases that were in the *Utilized Concern* cluster, 3459 were predicted correctly and 31 were predicted incorrectly (1 was predicted to be in the *Environmental Concern* cluster and 30 was predicted to be in the *Fisheries Concern* cluster) (Table 6).

Predicted group membership indicates the predicted frequencies of groups from the analysis. The numbers going down each column indicate how many were correctly and incorrectly classified. For example, of the 2113 cases that were predicted to be in the "*Environmental Concern*" cluster, 2084 were correctly predicted, and 29 were incorrectly predicted (1 cases were in the "*Utilized Concern*" cluster and 28 cases were in the "*Fisheries Concern*" cluster) (Table 6).

Table 6: Classification Results[a] Based on Discriminant Analysis

		Cluster	Predicted Group Membership			
			Utilized Concern	Environmental Concern	Fisheries Concern	Total
Original	Count	Utilized Concern	3459	1	30	3490
		Environmental Concern	8	2084	3	2095
		Fisheries Concern	9	28	2141	2178
	%	Utilized Concern	99.1	0.0	0.98	100

		Environmental Concern	0.4	99.5	0.1	100
		Fisheries Concern	0.4	1.3	98.3	100

a. 99.0% of original grouped cases correctly classified

4.4 Neural Network Analysis

After the formation of the three identified angler groups, a multilayer perceptron (MLP) neural network was employed as a predictive model in deciding the saltwater recreational anglers' perceptions toward marine environmental threats. The Multilayer Perceptron Module of IBM SPSS Statistics 26 was used to build the neural network model and test its accuracy. The MLP neural networks trained with a back-propagation learning algorithm uses the gradient descent to update the weights towards minimizing the error function.

The aim of this analysis was to examine whether a MLP neural network can help marine recreational fisheries managers correctly predict the threats of marine environment, by analyzing data obtained from the saltwater recreational anglers. The data was randomly assigned to training (70%) and testing (30%) subsets. The training dataset is used to find the weights and to build the model, while the testing data is used to find errors and prevent overtraining during the training mode (Table 7).

Table 7: Case Processing Summary

		N	Percent
Sample	Training	5430	69.9%
	Testing	2333	30.1%
Valid		7763	100.0%
Excluded		0	
Total		7763	

The neural network model is constructed with the multilayer perceptron algorithm. In order to find the best neural network, disparate possible networks were tested, and it was concluded that the neural network model with a single input layer, a single hidden layer, and a single output layer were the best option for this study. Previous studies have found that using this neural network layout is advantageous.

Sheela and Deepa [32] pointed out that as the number of neurons or the number of layers of a neural network increase, the training error also increases due to the overfitting. It is clear that using a single input layer, a single hidden layer, and a single output layer in the neural network will help decrease the probability of overfitting and require relatively lower computational time.

One of the most salient considerations in the construction of neural network is choosing activation functions for hidden and output layers that are differentiable. The results showed that a hyperbolic tangent activation function should be used for the single hidden layer of the model and linear activation function should be used for the output layer.

The Multilayer Perception Module of IBM SPSS Statistics 26 was used as the tool to choose the best architecture model automatically and it built the network with one hidden layer. From the thirteen independent variables, automatic architecture selection chose six nodes for the hidden layer, while the output layer had three nodes to code the depended variable *Cluster*. For the hidden layer, the activation function was the hyperbolic tangent. The output layer used the softmax function while a cross entropy was used as an error function (Table 8).

Table 8: Network Information

Input Layer	Covariates	1	Pollution
		2	Oil Extraction
		3	Climate Change
		4	Acidification
		5	Shipping
		6	Commercial Fishing
		7	Recreational Fishing
		8	Non-Native Species
		9	Aquaculture
		10	Energy
		11	Algal Blooms
		12	Habitats Loss
		13	Dams/Barriers
	Number of Units[a]		13
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		6
	Activation Function		Hyperbolic

			tangent
Output Layer	Dependent Variables	1	Cluster
	Number of Units		3
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit

The network diagram showed the thirteen input nodes, the six hidden nodes and the three output nodes representing the three identified angler categories. In the architectural point of view, it was a 13-6-3 neural network, means that there were total thirteen independent (input) variables, six neurons in the hidden layer and three dependent (output) variables (Figure 3).

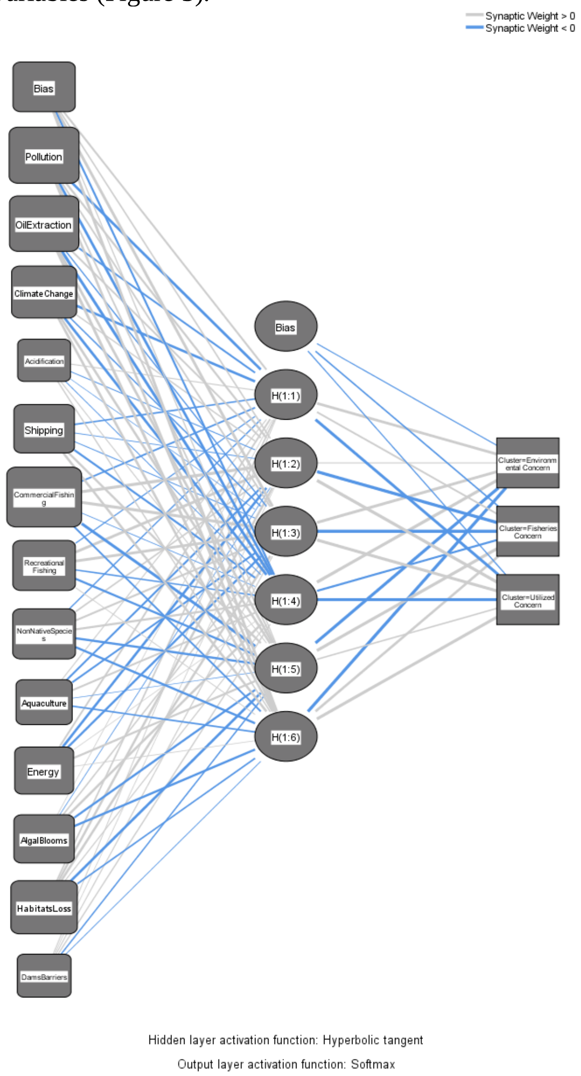

Figure 3: Network Diagram

The model summary provided information related to the results of training and testing sample (Table 9). Cross entropy error is displayed because the analysis is based on softmax activation function, and is given for both training and testing sample since is the error function that minimizes the network during training phase [25]. The value of cross entropy error (= 144.165) indicated the power of the model to predict the three identified angler groups. The cross entropy error was less for the testing sample compared with the training data set, meaning that the network model had not been over-fitted to the training data and has learned to generalize from trend. The result justified the role of testing sample which was to prevent overtraining.

In this study the percentage of incorrect prediction was equal to 0.4% in the training sample. So the percentage of correct prediction was 99.6% which is an excellent prediction in a qualitative study for determining the results of marine environmental threats for the three identified angler groups. The learning procedure was performed until one consecutive step with no decrease in error function was attained from the training sample.

Table 9: Model Summary

Training	Cross Entropy Error	144.165
	Percent Incorrect Predictions	0.4%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.89
Testing	Cross Entropy Error	74.155
	Percent Incorrect Predictions	0.9%

a. Error computations are based on the testing sample. (Dependent Variable: Cluster)

Using the training sample only, MLP neural network utilized synaptic weights to display the parameter estimates that showed the relationship between the units in a given layer to the units in the following layer (Table 10). Note that the number of synaptic weights can become rather large, and that these weights are generally not used for interpreting network results [25].

Table 10: Parameter Estimates

Predictor		Predicted								
		Hidden Layer 1						Output Layer		
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	Environmental Concern	Fisheries Concern	Utilized Concern
Input Layer	(Bias)	0.628	0.383	1.386	-0.738	0.249	1.356			
	Pollution	-1.254	0.773	0.477	-1.212	0.492	1.029			
	Oil Extraction	-0.604	0.557	-0.270	-1.724	1.212	1.432			
	Climate Change	-1.105	0.300	-0.317	-1.230	1.037	1.053			
	Acidification	0.220	-0.202	-0.024	-0.155	0.209	0.270			
	Shipping	-0.446	-0.151	-0.174	-0.731	0.848	1.020			
	Commercial Fishing	-0.497	2.186	1.969	-0.110	-1.643	-0.590			
	Recreational Fishing	-0.244	1.374	1.342	-0.357	-0.815	0.133			
	Non-Native Species	-0.275	0.416	0.712	0.808	-0.966	-0.921			
	Aquaculture	0.216	-0.654	-0.621	0.609	-0.010	-0.477			
	Energy	0.498	-1.088	-0.876	0.051	0.783	0.092			
	Algal Blooms	0.077	-0.095	0.350	1.011	-0.859	-0.923			
	Habitats Loss	0.443	0.995	1.009	0.227	-1.181	-0.459			
	Dams/Barriers	0.170	0.092	0.225	0.258	-0.328	-0.120			
Hidden Layer 1	(Bias)							-0.230	-0.411	-0.250
	H(1:1)							1.250	0.505	-1.744
	H(1:2)							0.238	-2.873	2.504
	H(1:3)							1.539	-2.990	2.072
	H(1:4)							2.404	-0.762	-1.963
	H(1:5)							-2.497	2.368	0.369
	H(1:6)							-2.831	1.145	2.165

Based on the MLP neural network, a predictive model was developed and displayed a classification table (i.e. confusion matrix) for categorical dependent variable the three identified angler groups, by partition and overall (Table 11). As can be seen, the MLP neural network correctly classified 5406 recreational anglers out of 5430 in the training sample and 2312 out of 2333 in the testing sample. Overall, 99.6% of the training cases were correctly classified. The predictive model developed had excellent classification accuracy.

Using the training sample only, it was able to classify 1482 *Environmental Concern* recreational anglers as *Environmental Concern* group, out of 1491. It held 99.4% classification accuracy for the *Environmental Concern* group. Similarly, the same model was able to classify 1494 *Fisheries Concern* recreational anglers as *Fisheries Concern* group out of 1500, and 2430 *Utilized Concern* recreational anglers as *Utilized Concern* group out of 2439. It was able to generate 99.6% classification accuracy for both the *Fisheries Concern* and *Utilized Concern* groups (Table 11).

Table 11: Predictive Ability and Classification Results

Classification					
Sample	Observed	Predicted			
		Environmental Concern	Fisheries Concern	Utilized Concern	Percent Correct
Training	Environmental Concern	1482	4	5	99.4%
	Fisheries Concern	2	1494	4	99.6%
	Utilized Concern	1	8	2430	99.6%
	Overall Percent	27.3%	27.7%	44.9%	99.6%
Testing	Environmental Concern	595	2	7	98.5%
	Fisheries Concern	4	672	2	99.1%
	Utilized Concern	0	6	1045	99.4%
	Overall Percent	25.7%	29.1%	45.2%	99.1%

Dependent Variable: Cluster

For the dependent variable *Cluster*, the following chart displays boxplots that classify the predicted pseudo-probabilities based on the whole dataset [25]. For each boxplot, the values above 0.5 show correct predictions. The first, from the left, boxplot showed the predicted probability of the observed *Environmental Concern* recreational anglers to be in the *Environmental Concern* category. The second and third boxplots showed that the probability for a recreational angler to be classified in *Environmental Concern* category although he/she really was in *Fisheries Concern* and *Utilized Concern* categories, respectively. The fourth boxplot showed, for outcomes that have observed category "*Fisheries Concern*", the predicted probability of category *Environmental Concern*. The right boxplot showed, the probability a recreational angler who really *Utilized Concern* category to be classified in the *Utilized Concern* category (Figure 4).

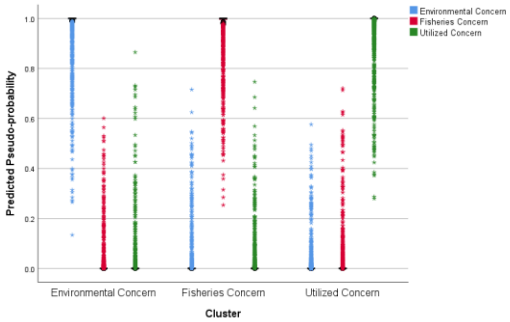

Figure 4: Predicted-by-Observed Chart

The ROC curve is a diagram of sensitivity versus specificity that shows the classification performance for all possible cutoffs [25]. It gives the sensitivity and specificity (= 1 – false positive rate) chart, based on the combined training and testing samples. The 45-degree line from the upper right corner of the chart to the lower left represents the scenario of randomly guessing the class. The more the curve moves away the 45-degree baseline, the more accurate is the classification (Figure 5).

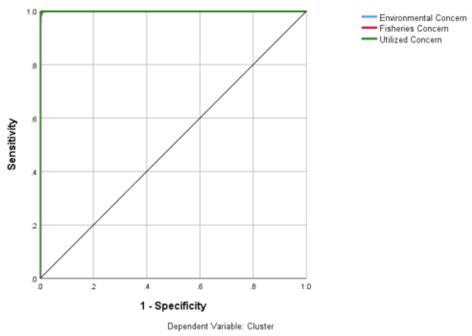

Figure 5: ROC Curve

The area under the ROC curve [25] showed that, if a recreational angler from the *Environmental Concern* category and a recreational angler from the *Fisheries Concern* category were randomly selected, there was 100% (1.000) probability that the model-predicted pseudo-probability for the first recreational angler of being in the *Environmental Concern* category, was higher than the model-predicted pseudo-probability for the second recreational angler of being in the *Environmental Concern* category (Table 12).

Table 12: Area under the Curve

		Area
Cluster	Environmental Concern	1.000
	Fisheries Concern	1.000
	Utilized Concern	1.000

The Cumulative Gains chart (Figure 6) is the presentence of correct classifications obtained by the MLP neural network model against the correct classifications that could result by chance (i.e. without using the model) [25]. Gains is a measure of the effectiveness of a classification model calculated as the percentage of correct predictions with the model versus the percentage of correct predictions obtained without a model (baseline). The farther above the baseline a curve lies, the greater the gain. A higher overall gain indicates better performance.
For example, the second point on the curve for the *Fisheries Concern* category was at (20%, 70%), meaning that if the network scores a dataset and sort all of the cases by predicted pseudo-probability of *Fisheries Concern*, it would be expected the top 20% to contain approximately 70% of all of the cases that actually take the category *Fisheries Concern*. The selection of 100% of the scored dataset, obtained all of the observed *Fisheries Concern* cases in the dataset (Figure 6).

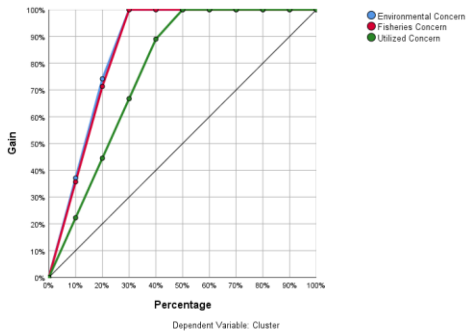

Figure 6: Cumulative Gains Chart

Lift chart, as well as cumulative gains charts, is visual aid for evaluating performance of classification models [25]. However, in contrast to the confusion matrix that evaluates models on the whole population, gains or lift chart evaluates model performance in a portion of the population. A lift chart uses a part of the dataset to give a clear view of the benefit to use a model in contrast to not using a model. The values from the gains diagram were used to calculate the lift factor (i.e. the benefit): the lift at 70% for the category *Fisheries Concern* was 70%/20% = 3.5 (Figure 7).

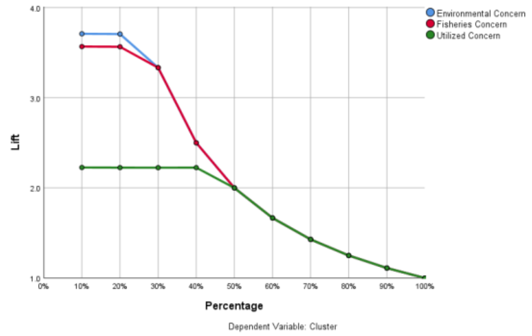

Figure 7: Lift Chart

The importance of the individual independent variables (factor influencing the threats of marine environment) is a measure of how much the network model predicted value changes for different independent variables [25]. The input parameters -- marine environmental threats which influenced the three identified angler groups have been ranked by the neural network model were given in the following Table 13.

The first three significant dominant factors that have been found were "Overfishing in commercial fisheries" (100%), contributed the most in the neural network model construction, followed by "Industrial pollution" (80.9%), and "Oil and gas extraction" (78.4%), had the greatest effect on how the recreational anglers' perceptions, in terms of marine environmental threats. The next two important factors have been "Marine habitats loss or degradation" (66.4%) and "Climate change" (60.8%). The other factors were relatively not important such as "Alternative energy (e.g. wave or wind) development" (36.1%), "Aquaculture" (25.5%), "Dams/barriers" (15.7%), and the least important factor which was identified is "Ocean acidification" (12.2%).

Table 13: Independent Variable Importance Analysis

In your opinion, how much of a threat, if any, does each of the following factors pose to the marine environment?	Importance	Normalized Importance	Rank
Industrial pollution	0.122	80.9%	2
Oil and gas extraction	0.118	78.4%	3
Climate change	0.092	60.8%	5
Ocean acidification	0.018	12.2%	13
Shipping	0.065	42.8%	8
Overfishing in commercial fisheries	0.151	100.0%	1
Overfishing in recreational fisheries	0.077	51.0%	7
Non-native species	0.078	51.9%	6
Aquaculture	0.038	25.5%	11
Alternative energy (e.g. wave or wind) development	0.054	36.1%	10
Algal blooms	0.064	42.3%	9
Marine habitats loss or degradation	0.100	66.4%	4
Dams/barriers	0.024	15.7%	12

The independent variable importance chart showed the impact of each independent variable in the MLP neural network model in terms of relative and normalized importance [25]. The independent variable importance chart also depicted the importance of the independent variables, i.e. how sensitive is the model is the change of each input variable (Figure 8).

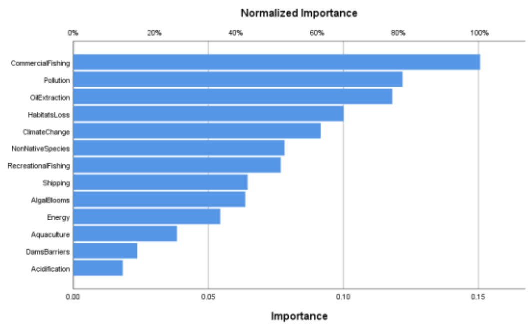

Figure 8: Independent Variable Importance Chart

5. Discussion and Conclusions

Understanding the concerns of saltwater recreational anglers about some of the major marine environmental threats could be one of many critical factors in implementing effective programs for ecosystem-based marine resource management throughout the United States. This study utilized cross-sectional data to understand and identify groups exhibiting common patterns of responses.

This empirical study seeks to provide an up-to-date assessment of using factor analysis for data preparation, cluster analysis for data examination, and discriminant analysis for classification. Three distinct angler groups -- *Utilized Concern*, *Environmental Concern*, and *Developmental Concern* groups -- were formulated, using K-means clustering analysis. These groups differed significantly in three dimensions through factor analysis from the 13-item marine environmental threat scale -- *Environmental Change*, *Industrial Development*, and *Fisheries Activities* – which were used to determine group placement. In conclusion, marine fisheries managers must understand there were three groups of saltwater recreational anglers identified in the study, each with different wants and needs for their specific concerns of marine environmental threats.

In the modeling of marine environmental threat scale, data shown by the neural network model offers excellent insight because it performs linear and nonlinear mapping without any preliminary information exists in the data. Therefore, neural network provides flexible, consistent and reliable appraisals compared to other statistical methods used for modeling the marine environmental threat scale data. The multilayer perceptron neural network model was trained to obtain better classification accuracy through the configuration of the neural network architecture. Overall, the multilayer perceptron neural network performed well on predicting the classification of the saltwater recreational anglers, and is easier to use than the conventional neural network software offered to behavioral researchers. It is also concluded that in neural network analysis, predictability is optimal in management decisions of dependent and independent variables repeatedly.

In an era that demands both protection and productivity of our nation's waters, the National Ocean Policy is a step towards long-term sustainability: a strong, coherent national policy based on science and local stakeholders. This study illustrated the diversity of saltwater recreational anglers' concerns and contradict the concept of an "average" angler. This study may also place a strong emphasis on the importance of understanding marine ecosystem structure, its function and processes, and how effects of human activities, including the socio-economic implications. Thus, all sectors of the community should take their individual steps. Thinking globally and acting locally is a fundamental intention to reduce such an environmental threat.

This study had both theoretical and practical implications. With updated testing of the well-developed conceptual framework of the marine environmental threat scale among saltwater recreational anglers, this research contributed to existing decision-making literature by either providing more evidence of the validity and robustness of this framework or by providing suggestions for adaptation in applying this framework to understand saltwater recreational angler groups across different socio-demographic backgrounds. Also, this research added more to the existing literature on the dynamically changing saltwater recreational anglers.

The results of this study would assist saltwater recreational fishery organizations in designing practical recreational management strategies to address concerns of anglers and to benefit saltwater fisheries populations. This research may also provide practical marketing implications for environmental education by proposing effective ways to understand and target these consumers.

References

[1] J. Acheson, "Attitudes toward offshore wind power in the midcoast region of Maine," *Marine Policy Review*, vol. 21, no. 2, pp. 42-55, 2012.

[2] F. E. Ahmed, "Artificial neural networks for diagnosis and survival prediction in colon cancer," *Molecular Cancer*, vol. 4:29, pp. 1-12, 2005.

[3] T. Beatley, "Protecting biodiversity in coastal environments: introduction and overview," *Coastal Management*, vol. 19, no. 1, pp. 1–19, 1991.

[4] J. R. Beddington, D. J. Agnew, and C. W. Clark, "Current problems in the management of marine fisheries," *Science*, vol. 316, pp. 1713-1716, 2007.

[5] C. M. Bishop, Pattern Recognition and Machine Learning, New York, NY: Springer Science + Business Media, 2006.

[6] B. K. Bose, "Neural network applications in power electronics and motor drives - an introduction and perspective," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 1, pp. 14-33, 2007.

[7] A. A. Brinson, and K. Wallmo, "Attitudes and preferences of salter recreational anglers: Report from the 2013 National Saltwater Angler Survey, Volume I," U.S. Department of Commerce, NOAA Technical Memorandum NMFS-F/SPO-135, 45p. 2013.

[8] Y. Chauvin, and D. E. Rumelhart, (Eds.). Developments in Connectionist Theory. Back-propagation: Theory, Architectures, and Applications. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, 1995.

[9] D. Child, The Essentials of Factor Analysis, 3rd ed., New York, NY: Continuum International Publishing Group, 2006.

[10] G. A. Churchill, Jr. and D. Iacobucci, Marketing Research: Methodological Foundations, 9th ed., Mason, OH: Thomson/South-Western, 2005.

[11] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297-334, 1951.

[12] C. M. Crain, B. S. Halpern, M. W. Beck, and C. V. Kappel, "Understanding and managing human treats to the coastal marine environment," *Annuals of the New York Academy of Sciences*, vol. 1162, pp. 39-62, 2009.

[13] J. G. De Gooijer, and R. J. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, no. 3, pp. 443-473, 2006.

[14] L. N. N. Do, N. Taherifar, and H. L. Vu, "Survey of neural network-based models for short-term traffic state prediction," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 1, pp. 1-24, 2019.

[15] J. G. B. Derraik, "The pollution of the marine environment by plastic debris: a review," *Marine Pollution Bulletin*, vol. 44, pp. 842–852, 2002.

[16] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179-188, 1936.

[17] E. W. Forgy, "Cluster analysis of multivariate data: efficiency versus interpretability of classifications," *Biometrics*, vol. 21, pp. 768–769, 1965.

[18] M. W. Gardner, and S. R. Dorling, "Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences," *Atmospheric Environment*, vol. 32, no. 14, pp. 2627-2636, 1998.

[19] S. Gelcich, P. Buckley, J. K. Pinnegar, J. Chilvers, I. Lorenzoni, G. Terry, M. Guerrero, J. C. Castilla, A. Valdebenito, and C. M. Duarte, "Public awareness, concerns, and priorities about anthropogenic impacts on marine environments," *PNAS*, vol. 111, no. 42, pp. 15042-15047, 2014.

[20] D. W. Gerbing, and J. C. Anderson, "An updated paradigm for scale development incorporating unidimensionality and its assessment," *Journal of Marketing Research*, vol. 25, pp. 186-192, 1988.

[21] Y. M. M. Hassim,and R. Ghazali, "Training a functional link neural network using an artificial bee colony for solving a classification problems," Journal of Computing, vol. 4, no. 9, pp. 110-115, 2012.

[21] S. S. Haykin, Neural Networks and Learning Machines, 3rd ed., Upper Saddle River, New Jersey: Pearson Education, Inc, 2009.

[23] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.

[24] R. M. Hughes, "Recreational fisheries in the USA: economics, management strategies, and ecological threats," *Fisheries Science*, vol. 81, no. 1, pp. 1-9, 2015.

[25] IBM, IBM SPSS neural networks 26. Armonk, NY: IBM Corporation, 2019.

[26] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281-297, 1967.

[27] S. Manel, J. M. Dias, S. T. Buckton, and S. J. Ormerod, "Alternative methods for predicting species distribution: an illustration with Himalayan river birds," *Journal of Applied Ecology*, vol. 36, pp. 734–747, 1999.

[28] National Ocean Council. National Ocean Policy Implementation Plan. The White House, Washington, D.C., 2013.

[29] E. A. Norse, "A river that flows to the sea: marine biological diversity movement," *Oceanography*, vol. 9, no. 1, pp. 5-9, 1996.

[30] H. Ramchoun, M. A. Janati Idrissi, Y. Ghanou, and M. Ettaouil, "New modeling of multilayer perceptron architecture optimization with regularization: an application to pattern classification," *IAENG International Journal of Computer Science*, vol. 44, no. 3, pp. 261-269, 2017.

[31] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," In D. E. Rumelhart, J. L. McClelland and the PDP research group, (Eds). Parallel distributed processing: explorations in the microstructure of cognition, Volume 1. Cambridge, MA: MIT Press, 1986.

[32] K. G. Sheela, and S. N. Deepa, Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Mathematical Problems in Engineering*, Volume 2013, Article ID 425740, 11p. 2013.

[33] P. V. R. Snelgrove, "Getting to the bottom of marine biodiversity: sedimentary habitats," *BioScience*, vol. 49, pp. 129–138, 1999.

[34] I. Suryanarayana, A. Braibanti, R. S. Rao, V. A. Ramamc, D. Sudarsan, and G. N. Rao, "Neural networks in fisheries research," *Fisheries Research*, vol. 92, pp. 115–139, 2008.

[35] B. G. Tabatchnick, and L. S. Fidell, Using multivariate statistics, 6th ed., Boston: Pearson Education, Inc., 2013.

[36] H. Torres, F. Muller-Karger, D. Keys, H. Thornton, M. Luther, and K. Alsharif, "Whither the U.S. National Ocean Policy Implementation Plan?" *Marine Policy*, vol. 53, pp. 198-212, 2015.

[37] Wikipedia, Multilayer perceptron. [online] https://en.wikipedia.org/wiki/Multilayer_perceptron (Accessed 12 December 2019)

[38] N. Z. Zacharis, "Predicting student academic performance in blended learning using artificial neural networks," *International Journal of Artificial Intelligence and Applications*, vol. 7, no. 5, pp. 17-29, 2016.