2013

The Integration of Artificial Neural Networks and Geometric Morphometrics to Classify Teeth from Carcharhinus Sp

K. James Soda

FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

THE INTEGRATION OF ARTIFICIAL NEURAL NETWORKS AND GEOMETRIC MORPHOMETRICS TO CLASSIFY TEETH FROM *CARCHARHINUS* SP.

By

K. JAMES SODA

A Thesis submitted to the
Department of Scientific Computing
in partial fulfillment of the
requirements for the degree of
Master of Science

Degree Awarded:
Fall Semester, 2013

K. James Soda defended this thesis on November 8, 2013.

The members of the supervisory committee were:

Dennis E. Slice
Professor Directing Thesis

R. Dean Grubbs
Committee Member

Anke Meyer-Baese
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the thesis has been approved in accordance with university requirements.

To my family for their constant encouragement, love, and patience.

ACKNOWLEDGMENTS

I am extremely grateful for the advice, assistance, and support provided to me by my advisor, Dennis Slice, throughout the entirety of this project. In addition, I owe a great debt to Gavin Naylor and Les Marcus, without whose data none of this work would have been possible. The work presented here also greatly benefited from comments and advice from R. Dean Grubbs, Lisa Whitenack, Matthew Kolmann, Anke Meier-Baese, Michelle Perry, and the members of Florida State University's Morphometrics Laboratory. Finally, I thank the students, faculty, and staff of Florida State University's Departments of Scientific Computing and Biology for additional comments and support.

TABLE OF CONTENTS

List of Tables vi
List of Figures viii
Abstract ix
1. INTRODUCTION 1
2. MATERIALS AND METHODS 9
2.1 Data Collection and Processing 9
2.2 Model Construction and Validation 12
2.3 Effect of Position on Classification 15
2.4 Classification accuracy under less ideal conditions 15
3. RESULTS 18
3.1 Ideal Dataset 18
3.2 Effects of Position on Classification 20
3.3 Reduced Dataset and Challenge Dataset 21
4. DISCUSSION 23
4.1 Model Performance for the Ideal Dataset 23
4.2 Effects of Position on Classification 24
4.3 Model Performance under Less Ideal Scenarios 25
4.4 Further Implications 26
REFERENCES 29
BIOGRAPHICAL SKETCH 31

LIST OF TABLES

1 The specimens present in the ideal training and test sets according to the specimen numbers provided in Naylor and Marcus (1994) 11

2 The specimens present in the reduced training and test sets according to the specimen numbers provided in Naylor and Marcus (1994) 16

3 The specimens present in the challenge training and test sets according to the specimen numbers provided in Naylor and Marcus (1994) 17

4 Linear Discriminate Analysis-Ideal Dataset. A confusion matrix reporting the classification accuracy of *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* for a linear discriminate analysis with fifteen individuals in the training set and using every canonical variate. Each row represents the true species affiliation, and each column represents the species to which the linear discriminate analysis classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column .. 18

5 MLP-Ideal Dataset. A confusion matrix reporting the classification accuracy of *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* for a multilayer perceptron with one hidden layer containing 15 units and with fifteen individuals in the training set. Each row represents the true species affiliation, and each column represents the species to which the multilayer perceptron classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column 19

6 RBFNN-Ideal Dataset. A confusion matrix reporting the classification accuracy of *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* for a radial basis function neural network with four units in its hidden layer and with fifteen individuals in the training set. Each row represents the true species affiliation, and each column represents the species to which the radial basis function neural network classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column 19

7 MLP-Reduced Dataset. A confusion matrix reporting the classification accuracy of *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* for a multilayer perceptron with one hidden layer containing 15 units when only five individuals are included in the training set. Each row represents the true species affiliation, and each column represents the species to which the multilayer perceptron classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the training set for each species appears in its respective box of the first columMLP-Reduced Dataset. A confusion matrix reporting the classification accuracy of *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* for a multilayer perceptron with one hidden layer containing 15 units when only five individuals are included in the training set. Each row represents the true species affiliation, and each column represents the species to which the multilayer perceptron classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column 21

8 MLP-Challenge Dataset. A confusion matrix reporting the classification accuracy of *C. altimus*, *C. obscurus*, and *C. plumbeus* for a multilayer perceptron with one hidden layer containing 15 units and with five individuals in the training set. Each row represents the true species affiliation, and each column represents the species to which the multilayer perceptron classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column.....22

LIST OF FIGURES

1 A schematic, following Figure 5.1 in Ripley (1996), illustrating the structure of a simple multilayer perceptron. Each circle represents one unit and, therefore, one activation function. Arrows represent the direction that values are passed, and each **w** value represents a weight by which the passed value will be scaled. At the end of the network, the output layer provides the probability that the specimen in question belongs to a given species 2

2 A schematic, loosely following Figure 5.5 in Haykin (1999), illustrating the structure of a radial basis function neural network. Circles represent the units in the output layer and input layer, which have linear activation functions (Moody and Darken 1988). Units in the hidden layer are represented as Gaussian-like curves to illustrate three key features of these units: First, each unit has a nonlinear activation function that quickly decays, like, for example, Gaussians. Second, each unit has a distinctive width, like the width of the Gaussian. And, third, each unit has a mean, that is to say a location, in input space, just as the Gaussian has a mean. Finally, the X represents the location of the vector of input variables in input space. The output of the hidden units is a function of how far the input vector is from the unit in question relative to the unit's associated width (Moody and Darken 1988). Like in Figure 1, each **w** represents the weight by which the hidden unit's output is scaled before the output units receive them, and the output units provide the probability that the specimen belongs to the given species 3

3 The average upper tooth series for each species in the ideal training set. The shape of each tooth in the series is the average shape for a tooth in that position, assuming that position relative to the symphysis indicates homology. This assumption is used only for graphical purposes and does not factor into subsequent analyses 7

4 The average upper tooth series for each species in the challenge training set. The shape of each tooth in the series is the average shape for a tooth in that position, assuming that position relative to the symphysis indicates homology. This assumption is used only for graphical purposes and does not factor into subsequent analyses 8

5 The percentages of teeth from the ideal test set that were misclassified according to their relative position in the jaw from the symphysis (0.0) to the jaw angle (1.0) (see text). Each bin represents a 0.1 long range of relative position values, and the value below each bar represents the upper bound for the bin. Results for a) linear discriminate analysis, b) a multilayer perceptron, and c) a radial basis function neural network 20

ABSTRACT

The advent of geometric morphometrics and the revitalization of artificial neural networks have created powerful new tools to classify morphological structures to groups. Although these two approaches have already been combined, there has been less attention on how such combinations perform relative to more traditional methods. Here we use geometric morphometric data and neural networks to identify from which species upper-jaw teeth from carcharhiniform sharks in the genus *Carcharhinus* originated, and these results are compared to more traditional classification methods. In addition to the methodological applications of this comparison, an ability to identify shark teeth would facilitate the incorporation of shark teeth's vast fossil record into evolutionary studies. Using geometric morphometric data originating from Naylor and Marcus (1994), we built two types of neural networks, multilayer perceptrons and radial basis function neural networks to classify teeth from *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus*, as well as classifying the teeth using linear discriminate analysis. All classification schemes were trained using the right upper-jaw teeth of 15 individuals. Between these three methods, the multilayer perceptron performed the best, followed by linear discriminate analysis, and then the radial basis function neural network. All three classification systems appear to be more accurate than previous efforts to classify *Carcharhinus* teeth using linear distances between landmarks and linear discriminate analysis. In all three classification systems, misclassified teeth tended to originate either near the symphysis or near the jaw angle, though an additional peak occurred between these two structures. To assess whether smaller training sets would lead to comparable accuracies, we used a multilayer perceptron to classify teeth from the same species but now based on a training set of right upper-jaw teeth from only five individuals. Although not as accurate as the network based on 15 individuals, the network performed favorably. As a final

test, we built a multilayer perceptron to classify teeth from *C. altimus*, *C. obscurus*, and *C. plumbeus*, which have more similar upper-jaw teeth than the original four species, based on training sets of five individuals. Again, the classification system performed better than a system that combines linear measurements and discriminate function analysis. Given the high accuracies for all three systems, it appears that the use of geometric morphometric data has a great impact on the accuracy of the classification system, whereas the exact method of classification tends to make less of a difference. These results may be applicable to other systems and other morphological structures.

CHAPTER ONE

INTRODUCTION

Over the past twenty years, classification and morphometrics have undergone intellectual renaissances that have yielded new and powerful tools. In morphometrics, the so-called geometric morphometric "revolution" spurred the renaissance (Rohlf and Marcus 1993). Unlike traditional methods, geometric morphometric methods do no0t focus on simple measurements of length or angle, but rather on the location of landmarks on the specimen after position, orientation, and size have been standardized (Rohlf and Marcus 1993; Adams et al 2004; Adams et al 2013). The main advantage of these approaches is that geometric morphometric methods are able to describe the shape of an object without losing any information on the shape's geometry (Adams et al 2004; Adams et al 2013). As a result, more information content is available for comparison or classification.

Concurrent with the rise of geometric morphometrics, computational scientists developed renewed interest in artificial neural networks for classification and generalization applications. Although the term "artificial neural network" (ANN) refers to a cornucopia of methods, two of most common types of ANN are multilayer perceptrons (MLPs; see Figure 1) and radial basis function neural networks (RBFNNs; see Figure 2). Both MLPs and RBFNNs consist of individual elements, referred to as either units or neurons, which receive numeric inputs and provide numeric outputs (Rosenblatt 1962; Moody and Darken 1988; Ripley 1996; Haykin 1999). Each unit within the network has a function, often referred to as the activation function, that controls the exact value that the unit outputs (Rosenblatt 1962; Haykin 1999). In some

cases, one unit's output is the input for a second unit, in which case the two units are said to be connected (Rosenblatt 1962; Ripley 1996; Haykin 1999). In addition, units form groups called layers that differ in where their inputs originate, to which units they are connected, and how their

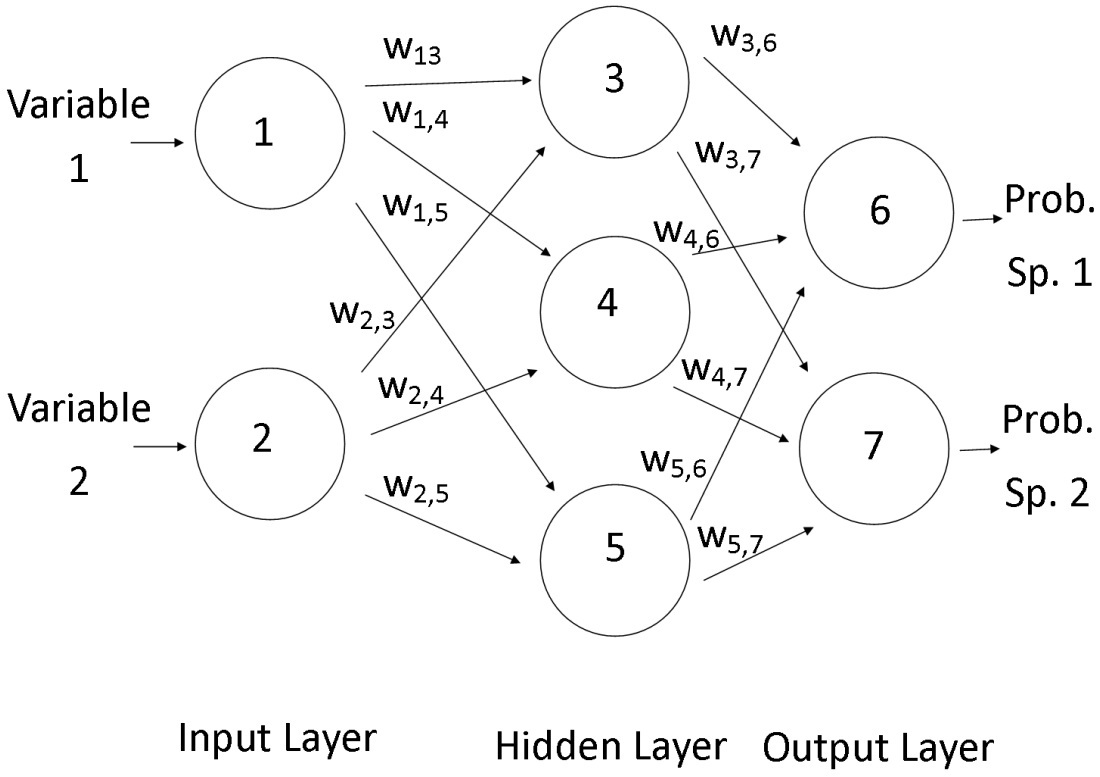

Figure 1. A schematic, following Figure 5.1 in Ripley (1996), illustrating the structure of a simple multilayer perceptron. Each circle represents one unit and, therefore, one activation function. Arrows represent the direction that values are passed, and each **w** value represents a weight by which the passed value will be scaled. At the end of the network, the output layer provides the probability that the specimen in question belongs to a given species.

outputs are utilized (Moody and Darken 1988; Ripley 1996; Haykin 1999). As a general rule, layers may be categorized into three types: input, hidden, and output (Haykin 1999). The units

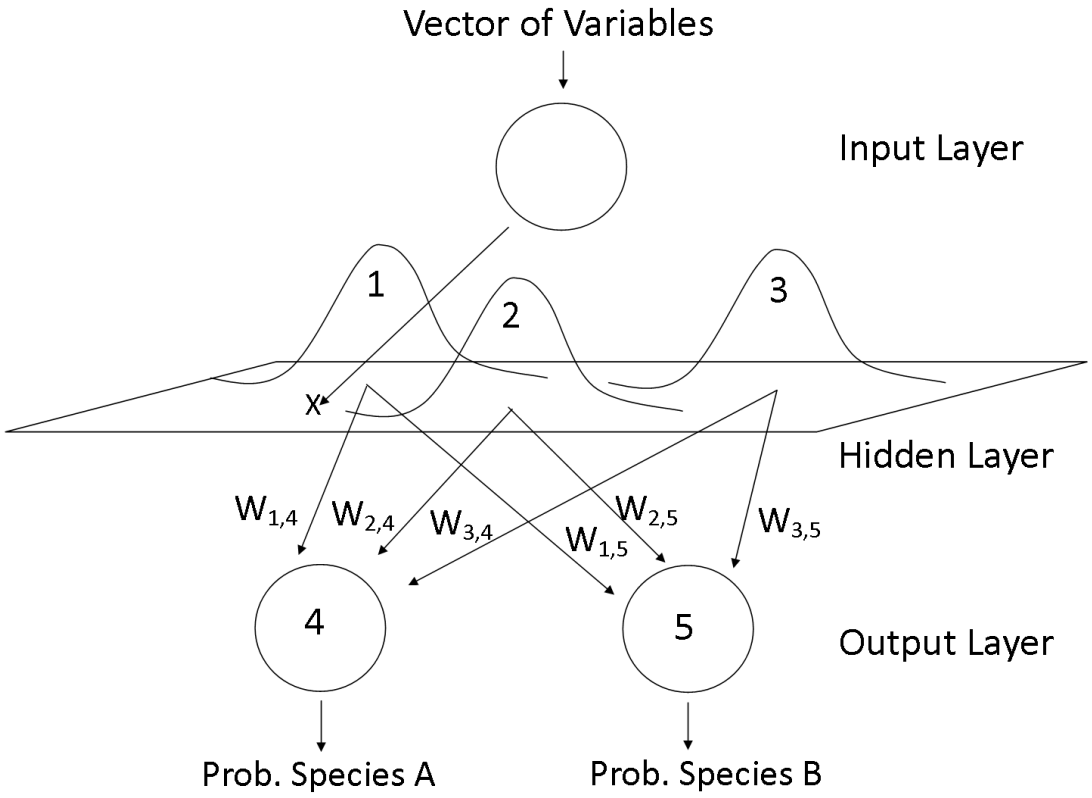

Figure 2. A schematic, loosely following Figure 5.5 in Haykin (1999), illustrating the structure of a radial basis function neural network. Circles represent the units in the output layer and input layer, which have linear activation functions (Moody and Darken 1988). Units in the hidden layer are represented as Gaussian-like curves to illustrate three key features of these units: First, each unit has a nonlinear activation function that quickly decays, like, for example, Gaussians. Second, each unit has a distinctive width, like the width of the Gaussian. And, third, each unit has a mean, that is to say a location, in input space, just as the Gaussian has a mean. Finally, the X represents the location of the vector of input variables in input space. The output of the hidden units is a function of how far the input vector is from the unit in question relative to the unit's associated width (Moody and Darken 1988). Like in Figure 1, each **w** represents the weight by which the hidden unit's output is scaled before the output units receive them, and the output units provide the probability that the specimen belongs to the given species.

in an input layer, also known as S-units, receive raw variable values (Rosenblatt 1962). Each S-unit subjects its variable value to its activation function (Rosenblatt 1962). In practice, though, an S-unit's function may be the identity function (Ripley 1996). Next, each S-unit passes the activation function's output to every unit to which it is connected at the input end, but, as the value is passed, it is weighted according to its recipient (Ripley 1996). Similarly to how an S-unit's activation function may be the identity function, though, the weight may be one for all units (Haykin 1999). The next units are referred to as A-units and form one of potentially many hidden layers (Rosenblatt 1962; Ripley 1996). Each A-unit sums its inputs, if more than one input is provided, and subjects this sum to its own activation function (Moody and Darken 1988; Ripley 1996). Just as with the S-units, the A-unit passes its output to every unit to which the A-unit is connected at the input end, weighting the value according to recipient (Ripley 1996). After potentially more hidden layers, the network will terminate with the R-units in an output layer (Rosenblatt 1962; Ripley 1996). Like the A-units, the R-units sum their inputs and submit the sum to their activation functions (Ripley 1996). However, in the case of classification problems, each R-unit corresponds to one candidate group, and each R-unit's output is the Bayesian *a posteriori* probability that the specimen belongs to that group (Richard and Lipman 1991; Ripley 1996; Haykin 1999). The specimen is assigned to whichever group has the greatest probability of inclusion (Richard and Lipman 1991; Ripley 1996; Haykin 1999).

Although they share their basic structure, MLPs and RBFNNs have crucial differences as well. First, in MLPs the inputs are scalar values whereas in RBFNNs the inputs are combined into a single vector (Moody and Darken 1988; Ripley 1996; Haykin 1999). This has computational consequences. In an MLP, each A-unit in the hidden layers must sum its inputs because it

receives multiple inputs whereas an A-unit in a RFNN receives a point in sample space that does not require summation (Haykin 1999). This reflects a second key difference between the two network types: MLPs and RBFNNs use different activation functions. Although other activation functions are possible, MLPs generally use sigmoidal nonlinear functions, such as logistic or hyperbolic functions (Ripley 1996; Haykin 1999). Radial basis function neural networks utilize more complicated activation functions. Just as the input for an A-unit in an RBFNN has a location in sample space, each A-unit itself has a location, usually assigned via a k-means algorithm (Moody and Darken 1988). The activation function is then some function of the distance between the input's location in sample space and the unit's location. Since the activation function generally must rapidly approach zero, Gaussian or logistic functions are popular choices (Moody and Darken 1988). There is a second key difference between the activation functions used in MLPs and those used in RBFNNs. In an MLP, the R-units used for a classification problem are generally nonlinear whereas an RFNN always uses a linear activation function in its S- and R-units (Haykin 1999). As a final difference, multilayer perceptrons and RBFNNs differ in how many layers may be present. In an MLP, there is no limit to how many layers are present because there may be multiple hidden layers. The only constraint is that, in instances where multiple hidden layers are present, a unit may only pass its output to units within a hidden layer closer to the output layer (Ripley 1996). In contrast, a RBFNN may only have a single hidden layer (Moody and Darken 1988; Haykin 1999). This distinction is logical as RBFNNs' activation functions rely on the distance between a sample point's location in sample space and the unit's location in sample space; since there is only one sample space, there can only be one hidden layer in an RBFNN.

Combined together, geometric morphometric methods and ANNs could be powerful classification tools for the biological sciences, especially in fields such as taxonomy where there is growing interest in automated methods for species identification (MacLeod et al. 2010). Although some studies have already integrated the two (e.g. Dobigny et al. 2002; Baylac et al. 2003; Bignon et al. 2005; van den Brink and Bokma 2011) or contrasted the two (e.g. MacLeod 2008), less attention has been placed on assessing how well methods that combine geometric morphometrics and ANNs compare to other combinations of methods, such as traditional morphometrics and discriminate function analysis. Teeth from carcharhiniform sharks in the genus *Carcharhinus* are an ideal application for combined geometric morphometric and ANN methods. Many species within *Carcharhinus* can be reliably identified based on tooth morphology and number (Naylor and Marcus 1994). Therefore, it is reasonable to believe that computational methods would be able to distinguish the species from which teeth originated. Doing so would also be of great practical value to the field of evolutionary biology. Shark teeth have an expansive fossil record, but much of the available specimens cannot be identified to a species without information on where in the jaw the tooth originated (Maisey 1984; Naylor and Marcus 1994). If computation methods were able to correctly and reliably identify teeth to species, a large volume of material would become available for research. Finally, *Carcharhinus* teeth are ideal for testing and demonstrating how well combined geometric morphometric data and ANNs may operate because there have already been attempts to classify *Carcharhinus* teeth using computation methods. Naylor and Marcus (1994) outlined a method to classify *Carcharhinus* teeth using linear measurements and linear discriminate analysis. Although the study achieved some success, the overall classification accuracy never exceed 80% and only certain teeth between the symphysis and the jaw angle could be accurately classified. Using a

subset of the data presented in Naylor and Marcus (1994), we attempted to establish methods involving both geometric morphometrics and ANNs to classify *Carcharhinus* teeth without incorporating any information on where in the jaw the tooth originated. The overall procedure was developed using upper-jaw teeth from *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus*. These four species represent an especially ideal situation as many specimens of this species were available for training, and these four species have easily distinguished teeth (Grubbs, *personal communication*; see Figure 3). Once the methodology was fine-tuned, we

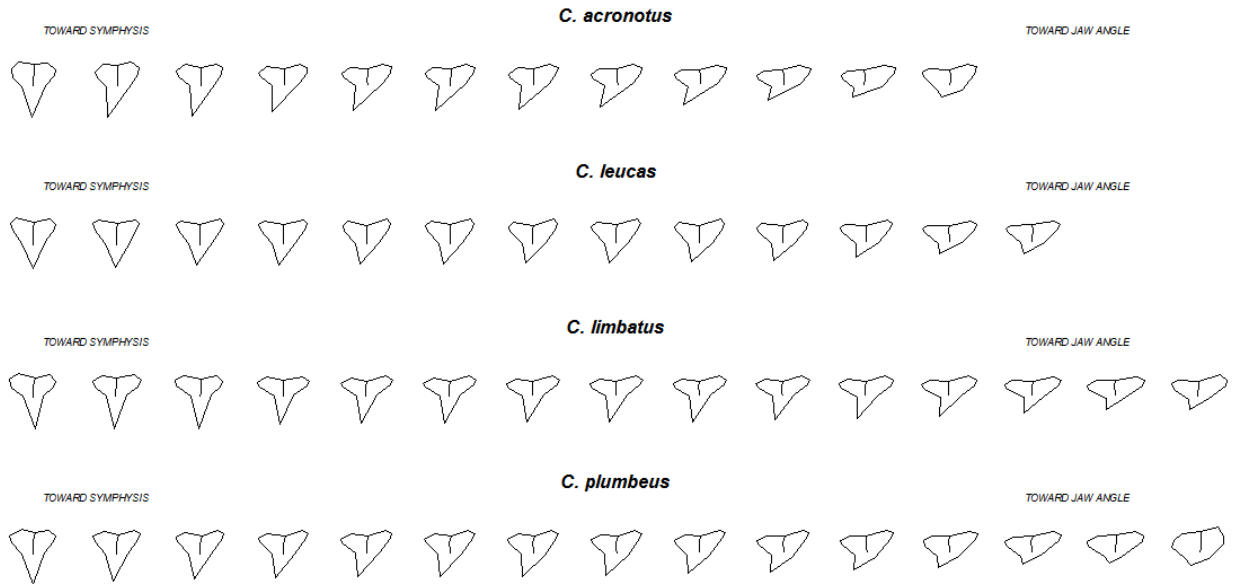

Figure 3. The average upper tooth series for each species in the ideal training set. The shape of each tooth in the series is the average shape for a tooth in that position, assuming that position relative to the symphysis indicates homology. This assumption is used only for graphical purposes and does not factor into subsequent analyses.

used the same framework to establish a procedure to classify upper-jaw teeth from *C. altimus*, C. *obscurus*, and *C. plumbeus*. These three species share a close phylogenetic relationship, and the

interspecies differences between their teeth are not as pronounced (Naylor 1992; Grubbs, *personal communication*; see Figure 4). Further, there were fewer individuals of these species available for training. As a result, this latter instance assessed how well the classification scheme operates in less ideal situations.

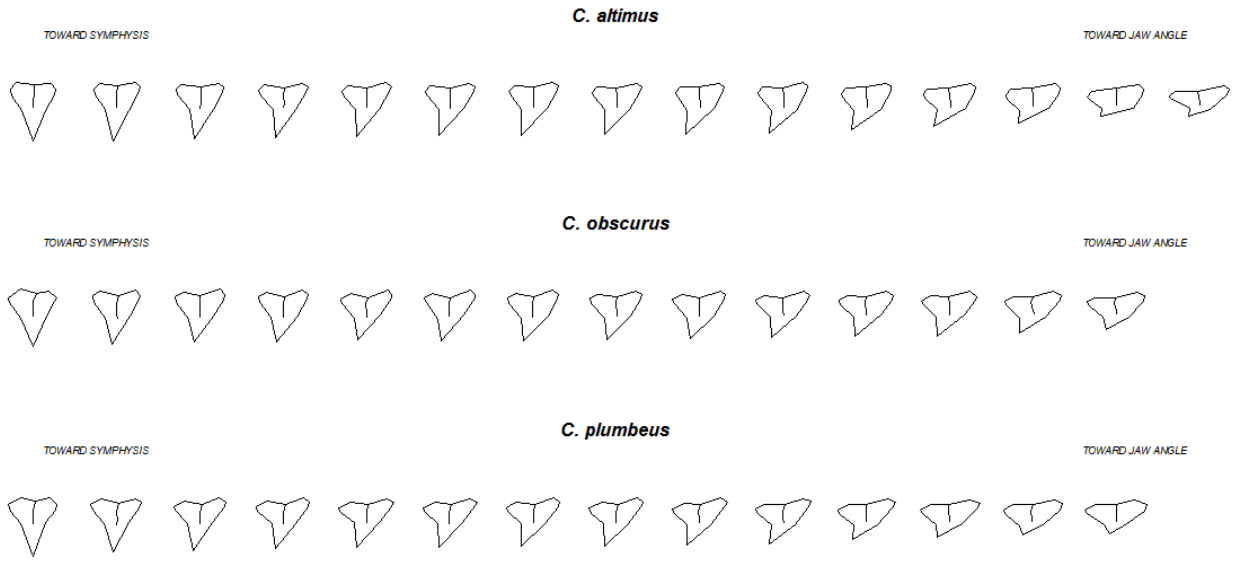

Figure 4. The average upper tooth series for each species in the challenge training set. The shape of each tooth in the series is the average shape for a tooth in that position, assuming that position relative to the symphysis indicates homology. This assumption is used only for graphical purposes and does not factor into subsequent analyses.

CHAPTER TWO

MATERIALS AND METHODS

2.1. Data Collection and Processing

All data originated from that used in Naylor and Marcus (1994). As such, the data were composed of teeth from both the lower and upper right jaw, each one described by 13 landmarks. All information on how specimens were acquired, how teeth were photographed, and how landmarks were digitized is contained within Naylor and Marcus (1994). The data from 20 of the original 22 data files were reformatted and Procrustes aligned using Morpheus et al ver. 1.7 (Slice 2005); data for *C. amblyrhynchos* and *C. sorrah* were missing.

To determine which classification method operated with the greatest accuracy, we composed a dataset with optimal features for classification (an obviously distinct morphology), a dataset that we term the "ideal dataset." The sample size was the first factor that was controlled. The size of the training set influences the performance of all three methodologies. Therefore, in order to prevent small sample sizes from confounding the results, we initially focused on the four most well represented species in the set, *C. acronotus*, *C. leucas*, C. *limbatus*, and *C. plumbeus*. These four species also have the additional fortuitous quality that they have relatively distinctive tooth shapes, which should further aid classification (Figure 3). To further emphasize the difference, we chose to focus on only upper jaw teeth, which tend to be more distinctive than lower jaw teeth (Naylor and Marcus 1994). Second, the data were refined according to completeness. Certain individuals within the original Naylor and Marcus (1994) dataset were missing teeth in the jaw. If left unchecked, individuals with complete jaws would provide more information

content to the classification schemes than individuals without complete tooth sets. In order to ensure that each individual provided the same amount of information, we removed any individual from these four species that were clearly missing one or more teeth. Finally, we controlled for inequality in sample sizes. Unequal sample sizes can bias each methodology's performance, as the species which has the largest sample size will have a larger effect on the pooled covariance matrix in linear discriminate analysis and will provide a neural network with more material on which to learn (Naylor and Marcus 1994). However, the hierarchical structure of the data complexified the generation of a balanced data set. Since each individual specimen may differ in tooth number, the data set cannot simultaneously be balanced in terms of number of teeth and number of specimens. Since we sought a system to identify teeth from any position in the upper jaw, it was necessary to allow every position of tooth to be present in the training set. Therefore, we chose to balance the training set according to the number of individuals rather than the number of teeth, thus generating a training set with every tooth from a set number of individuals. Fifteen specimens with complete jaw series, but varying numbers of teeth, were randomly selected to represent each species using routines in the random module of Python ver. 2.7 (van Rossum and Drake 2001); the individual teeth from these individuals composed the ideal training set. Any remaining specimens with complete jaw series were allocated to the ideal test set. Once the classification scheme was built using the teeth in the ideal training set, the scheme was used to classify each individual tooth in the ideal test set to species. Table 1 enumerates which specimens appear in each dataset according to the specimen numbers in Naylor and Marcus (1994).

Table 1. The specimens present in the ideal training and test sets according to the specimen numbers provided in Naylor and Marcus (1994).

Species	**Training Set**		**Test Set**	
C. acronotus	151	386	384	
	372	387	389	
	374	388	391	
	379	401	400	
	381	403	402	
	382	404	435	
	383	470	459	
	385		460	
C. leucas	250	751	465	
	255	752	562	
	370	753	760	
	458	754	756	
	472	755	745	
	740	758	738	
	742	761		
	744			
C. limbatus	405	444	153	211
	155	226	154	222
	468	473	158	227
	195	408	161	283
	353	453	162	302
	159	407	187	377
	156	170	196	464
	225		204	
C. plumbeus	242	449	236	
	378	455	371	
	414	463	424	
	415	555	447	
	418	557	556	
	419	564		
	421	570		
	425			

2.2. Model Construction and Validation under Ideal Conditions

Based on the ideal training set, a linear discriminate analysis, an MLP, and an RBFNN were built to classify individual teeth to species. R ver. 2.15 was used for the linear discriminate analysis, Weka ver. 3.7 was used to build the MLP, and the RBFNetwork ver. 1.0 package of Weka ver. 3.7 was used to build the RBFNN (Hall et al. 2009; Frank 2012; R core team 2012). Once the classification scheme was built, the scheme classified each tooth in the ideal test set to species to assess how well the methodology performed. The overall accuracy of the scheme was calculated as the total number of teeth in the ideal test set that was correctly classified over the total number of teeth in the test set. In addition, we calculated the species-specific accuracy in like manner.

In order to construct MLPs, an *a priori* designation of the network's architecture must be selected. The classification task pre-determines some aspects of the architecture. In this case, each tooth is described by thirteen landmarks digitized from two-dimensional photographs. Therefore, there are 26 (= 13 landmarks x 2 coordinate dimensions) Procrustes coordinates and 26 units in the input layer, as each coordinate is passed to one unit in the input layer. Further since the ideal set has four potential species to which to classify teeth, the output layer has four units, as each group receives one unit. Therefore, there are three aspects of the MLP architecture that may vary: the number of units in each hidden layer, the number of hidden layers, and which units in adjacent hidden layers are connected (Ripley 1996; Haykin 1999). The number of units in the hidden layer was determined via trial-and-error. Nineteen multilayer perceptrons with one hidden layer were trained, each with a varying number of units in the hidden layer. In each network, the learning rate was set to 0.3, the momentum was set to 0.2, and training was allowed for either 500 iterations or until 20 consecutive iterations failed to improve the results. The

effectiveness of a network's architecture was assessed according to the scheme's overall classification accuracy using the ideal test set. Initially, Weka's four "wildcard" numbers of hidden layers were tried (four units, 15 units, 26 units, and 30 units). Out of these four options, the 15 unit MLP classified teeth with the greatest overall accuracy. In order to ensure that the overall classification accuracy for 15 units was at least a local maximum, we next tested MLPs with between zero to 16 units. The overall classification accuracy ranged from 74.3929% (for one hidden unit) to 95.8057% (for 15 hidden units), so an MLP with 15 units had a local maximum in overall classification accuracy and had the global maximum for this range of units. To determine the optimal number of hidden layers, we constructed a multilayer perceptron with two hidden layers, both containing 15 units and with all units in adjacent layers connected. When the second hidden layer was added, the overall classification accuracy dropped to 93.5982%. Based on these results, we chose a multilayer perceptron with one hidden layer containing 15 units. Since only one hidden layer was present, there was no need to determine which units where connected.

Like an MLP, the task defines many aspects of an RBFNN's architecture. For this problem, the output layer again had four units, one for each group. Since RBFNNs always have a single hidden layer, the number of units in this lone hidden layer is the only aspect of the architecture that needs to be determined. We used the RBFNetwork package of Weka to construct six networks with differing numbers of units. For each network, the maximum number of iterations was 100,000, and the minimum standard deviation was 0.1. RBFNewtork also requires that the user provide a value for a ridge parameter, which was set to 10^{-8}. Again, the effectiveness of each network's architecture was assessed using the overall classification accuracy for the ideal

test set. The number of units in three of the networks was determined based on the number of species. These networks had two units (half the number of species), four units (the number of species), and eight units (twice the number of species). The number of units in two of the networks was determined based on the literature. The first of these had 796 units, the same number as points in the training set, because, as Tao (1993) points out, when the number of units equals the number of points in the training set, every point in the training set will be classified with 100% accuracy. Following another line of logic, Moody and Darken (1988) were relatively successful at predicting function outputs using an RBFNN with one tenth the number of points in their training set. Therefore, the second network had 80 units, which is roughly one tenth the number of training points in the ideal training set. Finally, since RBFNetwork's radial basis function neural network uses a K-means algorithm to choose the location of units, we tried one architecture that had as many hidden layers as the optimal number of units for a K-means clustering of the data (Frank 2012). Although many systems have been proposed to find the optimal number of clusters in a K-means clustering algorithm, a basic suggestion is to plot some criteria of fit, such as the total within cluster sums of squares, against the number of clusters and to determine where the addition of clusters ceases to substantially increase fit (Everitt et al. 2011). Using this criterion, we found six clusters to be the best candidate for optimal cluster number, and so we created a network with six hidden layers. The network with four units had the highest overall classification accuracy, so this network was used in the rest of the analysis.

To contrast the combination of geometric morphometric data and ANNs with geometric morphometric data and more traditional classification schemes, we ran a linear discriminate analysis on the Procrustes coordinates of each tooth in the ideal training set. Then, using all

three canonical variates that the previous linear discriminate analysis created, we attempted to classify each tooth in the ideal test set to species based on the tooth's Procrustes coordinates.

2.3. Effects of Position on Classification

In addition to classification rates, we assessed whether misclassified teeth tend to originate from certain regions in the jaw. After each method had classified the test set, we identified from how many tooth positions from the symphysis each misidentified tooth originated. However, since individuals may vary in the total number of teeth present in the jaw, the same tooth position in two different individuals may represent different locations along the jaw. In order to allow for biologically-relevant comparisons across individuals, each tooth position was divided by the total number of teeth on the upper right side of the jaw to come to what we term the relative position of the tooth. Finally, due to the nature of integer division, certain ranges of relative positions are better represented in the test set than other ranges. To control for these differences, we binned the misidentified teeth into 0.1 wide relative distance categories and then divided by the number of teeth within the ideal test set that fit into that bin so that each bin represents the proportion of misidentified teeth relative to the total number of teeth in the bin.

2.4. Classification Accuracy under Less Ideal Conditions

The classification schemes for sorting teeth to *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* represent an ideal case; each species has relatively distinctive teeth from each other, and a large sample of each species was available to build the schemes. Conditions may not be as ideal in practical applications, and it would be useful to assess how results from an ideal situation carryover to the less ideal. To gauge how well the scheme with the highest accuracy performs

when training set size is minimal, we developed a second dataset we call the reduced dataset. The reduced dataset has the same basic structure as the ideal dataset (right upper-jaw teeth from *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* and incorporating only individuals that are not missing teeth), but only five individuals from each species were randomly allocated to the training set; all other complete individuals were placed in a test set. Table 2 enumerates which specimen was placed into which dataset according to its identification number in Naylor and Marcus (1994). We trained a new classification scheme based on the reduced training set using the methodology that had the highest accuracy for the ideal dataset and assessed its accuracy

Table 2. The specimens present in the reduced training and test sets according to the specimen numbers provided in Naylor and Marcus (1994).

Species	Training Set	Test Set				
C. acronotus	372	151	385	401	459	
	383	374	386	402	460	
	384	379	388	403	470	
	387	381	391	404		
	389	382	400	435		
C. leucas	755	250	472	745	761	
	758	255	738	751		
	562	370	740	753		
	754	458	742	756		
	752	465	744	760		
C. limbatus	377	153	161	204	227	407
	158	154	162	211	283	408
	187	155	170	222	302	444
	453	156	195	225	353	464
	473	159	196	226	405	468
C. plumbeus	236	242	421	463		
	378	371	424	556		
	418	414	447	557		
	555	415	449	564		
	425	419	455	570		

using the reduced test set.

Finally, to assess how well the classification scheme with the highest accuracy for the ideal dataset performs when the differences between teeth are more subtle, we assembled one other dataset, which we term the challenge dataset. The training set for the challenge dataset incorporated only upper-jaw teeth from five jaws that were not obviously missing a tooth from *C. altimus*, *C. obscurus*, and *C. plumbeus*. These three species have more similar upper-jaw teeth than *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* (*RD Grubbs, personal communication*; see Figure 4). Again, all remaining individuals' teeth were allocated to a test set. Once again, we used the challenge training set to develop a classification scheme based on the methodology that performed the most accurately in the ideal dataset, and assessed how well the scheme performed using the challenge test set. Table 3 lists which specimens were placed into each dataset using the identification numbers provided in Naylor and Marcus (1994).

Table 3. The specimens present in the challenge training and test sets according to the specimen numbers provided in Naylor and Marcus (1994).

Species	Training Set		Test Set	
C. altimus	720	735	717	721
	725	737	718	727
	730		719	
C. obscurus	94	586	95	584
	577	592	246	585
	579		583	
C. plumbeus	414	236	419	555
	425	242	424	556
	418	371	449	557
	447	378	455	564
	421	415	463	570

CHAPTER THREE

RESULTS

3.1 Ideal Dataset

Confusion matrices for the linear discriminate analysis, the MLP, and the RBFNN when these schemes are trained using the ideal training set and tested using the ideal test set are presented in Table 4, Table 5, and Table 6, respectively. Overall, the multilayer perceptron had the highest accuracy out the three methods (95.81% accuracy), followed by linear discriminate analysis (94.48%). Radial basis function neural networks had the lowest accuracy of the three (92.05%).

Table 4. Linear Discriminate Analysis-Ideal Dataset. A confusion matrix reporting the classification accuracy of *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* for a linear discriminate analysis with fifteen individuals in the training set and using every canonical variate. Each row represents the true species affiliation, and each column represents the species to which the linear discriminate analysis classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column.

Classified as ->	*C. acronotus*	*C. leucas*	*C. limbatus*	*C. plumbeus*
C. acronotus (N = 94)	**92.553%**	0.00%	2.13%	5.32%
C. leucas (N = 76)	1.32%	**94.74%**	0.00%	3.95%
C. limbatus (N = 215)	0.47%	1.40%	**95.81%**	2.33%
C. plumbeus (N = 68)	2.94%	2.94%	1.47%	**92.65%**

Table 5. MLP-Ideal Dataset. A confusion matrix reporting the classification accuracy of *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* for a multilayer perceptron with one hidden layer containing 15 units and with fifteen individuals in the training set. Each row represents the true species affiliation, and each column represents the species to which the multilayer perceptron classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column.

Classified as ->	*C. acronotus*	*C. leucas*	*C. limbatus*	*C. plumbeus*
C. acronotus (N = 94)	**93.62%**	1.06%	2.13%	3.20%
C. leucas (N = 76)	0.00%	**98.68%**	0.00%	1.32%
C. limbatus (N = 215)	0.47%	1.40%	**95.35%**	2.79%
C. plumbeus (N = 68)	1.47%	1.47%	0.00%	**97.06%**

Table 6. RBFNN-Ideal Dataset. A confusion matrix reporting the classification accuracy of *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* for a radial basis function neural network with four units in its hidden layer and with fifteen individuals in the training set. Each row represents the true species affiliation, and each column represents the species to which the radial basis function neural network classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column.

Classified as ->	*C. acronotus*	*C. leucas*	*C. limbatus*	*C. plumbeus*
C. acronotus (N = 94)	**93.62%**	1.06%	1.06%	4.26%
C. leucas (N = 76)	1.32%	**89.47%**	1.32%	7.89%
C. limbatus (N = 215)	4.19%	0.47%	**93.95%**	1.40%
C. plumbeus (N = 68)	5.88%	4.41%	2.94%	**86.76%**

3.2. Effect of Position on Classification

Figure 5 shows the percentage of teeth from the training set that were misclassified, separated

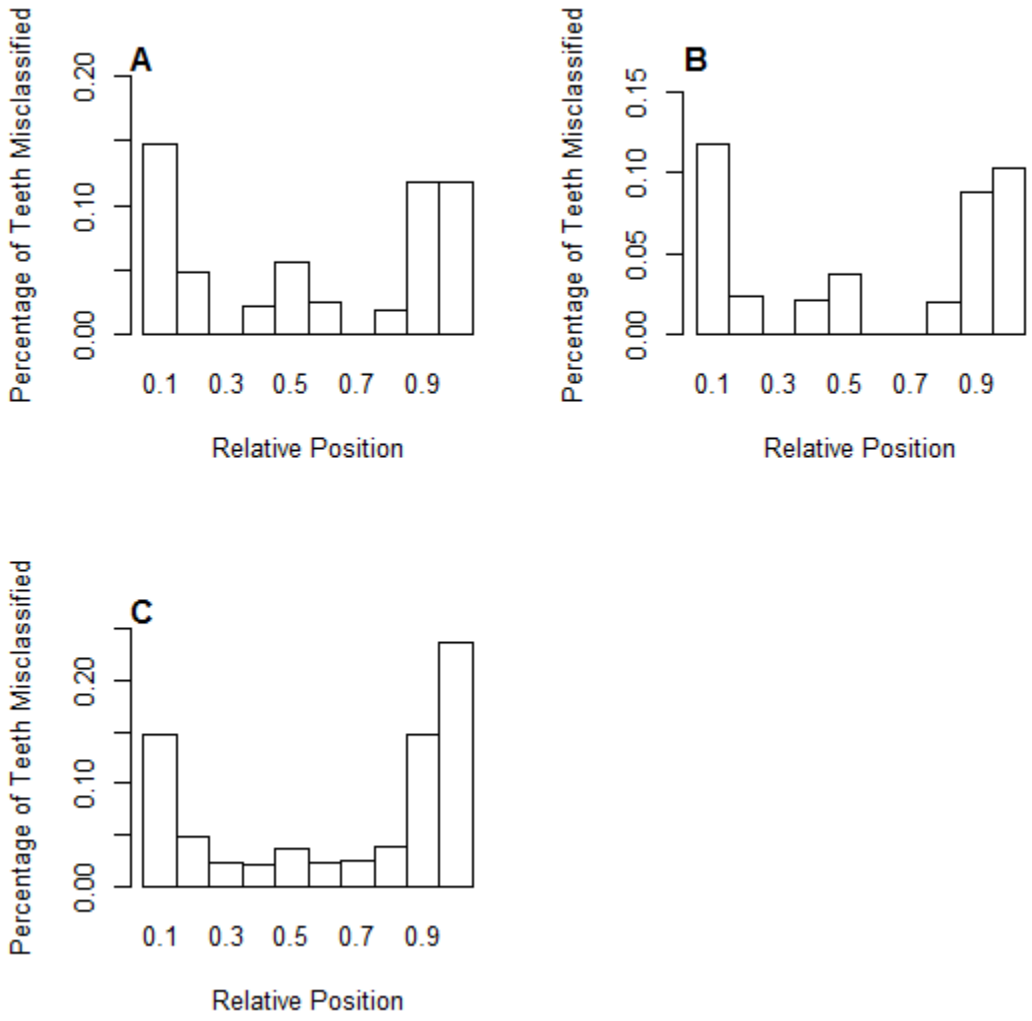

Figure 5. The percentages of teeth from the ideal test set that were misclassified according to their relative position in the jaw from the symphysis (0.0) to the jaw angle (1.0) (see text). Each bin represents a 0.1 long range of relative position values, and the value below each bar represents the upper bound for the bin. Results for a) linear discriminate analysis, b) a multilayer perceptron, and c) a radial basis function neural network.

according to their relative positions. Despite differences in accuracy, all three methods show relatively similar patterns in where misclassified teeth were located. Misclassification was most notable for teeth near the jaw angle (relative position 0.8-0.1) or near the symphysis (relative position 0.0-0.1). Further, although not as great in magnitude, there are local maxima in misclassification between relative positions 0.5 and 0.6. The exact location of local minima differs between methods, but there is always a local minimum between 0.2 and 0.4 and between 0.5 and 0.7. In the case of the MLP and of LDA, there are even bins where no tooth was misclassified.

3.3. Reduced Dataset and Challenge Dataset

As the multilayer perceptron had the highest classification accuracy, we trained another

Table 7. MLP-Reduced Dataset. A confusion matrix reporting the classification accuracy of *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* for a multilayer perceptron with one hidden layer containing 15 units when only five individuals are included in the training set. Each row represents the true species affiliation, and each column represents the species to which the multilayer perceptron classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column.

Classified as ->	*C. acronotus*	*C. leucas*	*C. limbatus*	*C. plumbeus*
C. acronotus (N = 212)	**91.98%**	0.94%	0.94%	6.13%
C. leucas (N = 203)	0.99%	**89.66%**	5.91%	3.45%
C. limbatus (N = 359)	1.67%	0.28%	**91.36%**	6.69%
C. plumbeus (N = 209)	1.91%	6.70%	2.87%	**88.52%**

multilayer perceptron using the reduced training set to classify teeth from *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus*. Table 7 is a confusion matrix for this scenario. The overall classification accuracy was 90.54%. To draw further comparison, we trained a third multilayer perceptron using the challenge training set to classify teeth from *C. altimus*, *C. obscurus*, and *C. plumbeus*. Table 8 is a confusion matrix for the resulting scheme. The overall classification accuracy was 80.51%.

Table 8. MLP-Challenge Dataset. A confusion matrix reporting the classification accuracy of *C. altimus*, *C. obscurus*, and *C. plumbeus* for a multilayer perceptron with one hidden layer containing 15 units and with five individuals in the training set. Each row represents the true species affiliation, and each column represents the species to which the multilayer perceptron classified the specimen. Correct classifications appear along the diagonal in bold. The exact number of teeth present in the test set for each species appears in its respective box of the first column.

Classified as ->	*C. altimus*	*C. obscurus*	*C. plumbeus*
C. altimus (N = 70)	**97.14%**	1.43%	1.43%
C. obscurus (N = 74)	8.11%	**72.97%**	18.92%
C. plumbeus (N = 210)	6.19%	16.19%	**77.62%**

CHAPTER FOUR

DISCUSSION

4.1 Model Performance for the Ideal Dataset

The MLP more accurately classified teeth from *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus* than linear discriminate analysis or the RBFNN. Therefore, MLPs appear to be the best classification method for identifying *Carcharhinus* teeth under the ideal conditions of large differences and plentiful training specimens. However, it is important to note that all three systems achieved overall classification accuracies over 90%, and the classification accuracy for any particular species was always greater than 85%.

Direct comparisons with Naylor and Marcus (1994) are difficult, as this study and Naylor and Marcus (1994) have some crucial differences. Perhaps chief among these differences is how many species were incorporated into the classification scheme. Whereas our ideal dataset only incorporated four species, Naylor and Marcus (1994) incorporated 22 species. With additional species comes more ways to misclassify a tooth. Further, our ideal dataset incorporated every right upper-jaw tooth from 15 individuals (178-217 teeth) for every species. In contrast, Naylor and Marcus (1994) only included eight individuals, each individual represented by four pseudoteeth (see below), from each species. Since lower sample sizes are associated with a lower classification rate and because Naylor and Marcus (1994) had more species to which teeth could be classified, Naylor and Marcus (1994) should intuitively have a lower classification rate. However, in contrast, Naylor and Marcus (1994) also standardized the teeth into idealized forms called pseudoteeth, whereas this study trained its schemes based only on the Procrustes

coordinates of each landmark on the tooth; no further data processing was involved. If the pseudoteeth in Naylor and Marcus (1994) were successful in their intended purpose, that is in allowing for only biologically homogenous teeth to be compared, then all things being equal Naylor and Marcus (1994) would be expected to have a higher accuracy rate than this study. In any case, a basic idea of how well the methods in this study operated relative to methods that use linear measurements with linear discriminate analysis can be obtained by comparison with Naylor and Marcus (1994).

Naylor and Marcus (1994) achieved an overall classification accuracy of 70.86% for *C. acronotus*, *C. leucas*, *C. limbatus*, and *C. plumbeus*, and the classification accuracy for any of these four species was never greater than 85%. In this light, it appears that geometric morphometric data classified using MLPs, RBFNNs, or linear discriminate functions all perform relatively well. Further, the difference in accuracy between MLPs and RBFNNs, the classification system that performed the most poorly, is only 3.76 percentage points. In other words, out of every 100 teeth classified, we estimate that multilayer perceptrons will only correctly classify three to four more teeth than a radial basis function neural network. Based on this observation, the exact classification method may not be as crucial as the use of geometric morphometric data for training, except in scenarios where correct classification is absolutely crucial.

4.2. Effects of Position on Classification

Despite the differences in performance, both the data presented here and Naylor and Marcus (1994) demonstrate a somewhat similar phenomenon in regards to the effect of position on

accuracy. The classification accuracy tends to be lowest at the jaw angle and symphysis. However, in Naylor and Marcus (1994), the rise and fall in accuracy is uninterrupted whereas here we found an additional local maximum in misclassification at around 0.4-0.5. The exact cause of this discrepancy is not clear. However, all misclassified teeth that fell into the 0.4-0.5 relative position range originated from three individuals. Therefore, it is possible that these three individuals possess anomalous tooth shapes and that, for whatever reason, the classification methods that incorporated GM were more sensitive to these anomalies. In contrast, other patterns may have overcome these anomalies in Naylor and Marcus (1994), leading to the uninterrupted transition.

4.3. Model Performance under Less Ideal Scenarios

To separate the effects of decreased training set size from the effects of differences in tooth shape, we re-built the multilayer perceptron used to sort teeth to *C. acronotus*, *C. limbatus*, *C. leucas*, or *C. plumbeus* using only data from five separate individuals, a set we termed the reduced training set. Although the overall classification accuracy and species-specific classification accuracies are not as high when the reduced training set is used relative to the ideal training set, the results are still favorable overall. Once again, the overall accuracy (90.54%) was greater than the accuracy achieved in Naylor and Marcus (1994) (70.86%), and each species-specific accuracy was greater than 85%. Therefore, training sets that only contain five individuals may still yield competitive classification systems. In other words, even though the *C. altimus, C. obscurus*, and *C. plumbeus* training set only contains five individuals, it is possible that these species may be well-distinguished.

Although neither the overall classification accuracy, nor the species-specific classification accuracies for the challenge dataset, which contained teeth from *C. altimus, C. obscurus*, and *C. plumbeus*, are as high as for the original four species, the results still reflect favorably on the use of MLPs and geometric morphometric data. Although the overall accuracy was only 80.51%, the combined accuracy in Naylor and Marcus (1994) for these three species was 73.33%, and each species-specific accuracy was higher than in Naylor and Marcus (1994). So even when species have similar teeth, a combination of MLPs and geometric morphometric data still appears to classify teeth with better accuracy than previous methods and at relatively high rates overall.

4.4. Further Implications

Although focusing on shark teeth, these findings have implications for any structures that contain a phyletic signal. Geometric morphometric methods appear to capture phyletic signals more accurately than traditional linear measurements such that incorporating these methods into classification schemes may contribute more strongly to the success of the method than the choice of classification scheme itself. However, the type of classification scheme still influences the accuracy rate; certain artificial neural networks, such as MLPs, may improve classification rates relative to more common statistic methods, such as linear discriminate analysis. However, it is important to remember that "artificial neural network" is an umbrella term for numerous methods. Certain neural networks, RBFNNs in this case, may not perform better than more typical methods; the exact type of neural network must be assessed on a case by case basis.

In practice, though, the choice between artificial neural network and statistical methodologies may be a trade-off between convenience and accuracy. Methods like linear discriminate analysis have become common place in biology and are readily available, "pre-packaged" in many statistical programs. Artificial neural networks are still uncommon in biology and pre-existing software may not be as readily available for their construction, although software such as Weka has significantly ameliorated the difficulty in generating them. Further, linear discriminate analysis involves a relatively straightforward algorithm and requires few technical decisions on the part of the user. In contrast, MLPs and RBFNNs are iterative algorithms, requiring guess-and-check like strategies to build (Moody and Darken 1988; Ripley 1996; Gardner and Dorling 1998). The additional computational cost that these algorithms incur could become impractical, particularly for multilayer perceptrons (Moody and Darken 1988). In fact, the original inspiration for RBFNNs was to find a more rapid system for artificial neural network construction; since its original formulation, it has been known that RBFNNs require more information than MLPs to reach comparable accuracies (Moody and Darken 1988). In our experience, though, MLP construction did not take an unreasonable amount of time even on a personal computer (Windows 7, AMD Athelon II Duel Core M3000 , 2.0 GHz, 3.00 Gb), so the efficiency boost for radial basis function neural networks may not offset the additional data requirements. In addition, most artificial neural networks require the user to set certain parameters, most notably the number of units in hidden layers and, in the case of MLPs, how many hidden layers to use. Although there are "rules of thumb" and various algorithms to guide the researcher (see Ripley 1996; Haykin 1999), these methods seem to be mostly rough guidelines. The parameter values ultimately used in a study very well may not be the global optimum. In this study, many parameters were allowed to remain at their default value, and the

number of hidden units was based on guess-and-check methods and *a priori* ideas on what number may be optimal. Given these additional complications, a good default workflow may be to initially attempt classification via discriminate function analysis, and, if more accurate predictions are required, then attempt to build a more satisfying neural network.

REFERENCES

Adams, D. C., F. J. Rohlf, & D. E. Slice. (2004). Geometric morphometrics: ten years of progress following the "revolution." *Italian Journal of Zoology* 71: 5-16.

Adams, D.C., F.J. Rohlf, & D. E. Sice. (2013). A field comes of age: geometric morphometrics in the 21st century. *Hysterix, the Italian Journal of Mammalogy* 24: 7-14.

Baylac, M., C. Villemant, and G. Simbolotti. (2003). Combining geometric morphometrics with pattern recognition for the investigation of species complexes. *Biological Journal of the Linnean Society* 80: 89-98.

Bignon, O., M. Baylac, J.D. Vigne, and V. Eisenmann. (2005). Geometric morphometrics and the population diversity of late glacial horses in western Europe (*Equus caballus arcelini*): phylogeographic and anthropological implications. *Journal of Archaeological Science* 32: 375-391.

van der Brink, V. and F. Bokma. (2011). Morphometric shape analysis using learning vector quantization neural networks-an example distinguishing two microtine vole species. *Annales Zoologici Fennici* 48: 359-364.

Dobigny, G., M. Baylac, and C. Denys. (2002). Geometric morphometrics, neural networks and diagnosis of sibling *Taterillus* species (Rodentia, Gerbillinae). *Biological Journal of the Linnean Society* 77: 319-327.

Everitt, B.S., S. Landau, M. Leese, and D. Stahl. (2011). Cluster Analysis, *5th ed*. Chichester, West Sussex: Wiley

Frank, E. (2012). RBFNetwork: Classes that implement radial basis function networks. Available at http://prdownloads.sourceforge.net/weka/RBFNetwork1.0.4.zip?download.

Gardner, M.W. and S.R. Dorling. (1998). Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences. *Atmospheric Environment* 32: 2627-2636.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I.H. (2009) The WEKA data mining software: an update. *SIGKDD Explorations* 11: 10-18.

Haykin, S. (1999). Neural networks: a comprehensive foundation, *2nd ed.* Upper Saddle River, New Jersey: Prentice Hall.

Maisey, J.G. (1984). Higher elasmobranch phylogeny and biostratigraphy. *Zoological Journal of the Linnean Society* 82: 33-54.

MacLeod, N. (2008). Understanding morphology in systematic contexts: 3D specimen ordination and 3D specimen recognition. In Q. Wheeler (Ed.). *The New Taxonomy*. London: CRC Press, Taylor & Francis Group.

MacLeod, N., M. Benfield, and P. Culverhouse. (2010). Time to automate identification. *Nature* 467: 154-155.

Moody, J. and C. Darken. (1988). Learning with localized receptive fields. In D. Touretzky, G. Hinton, and T. Sejnowski (Eds.). *Proceedings of the 1988 Connectionist Models Summer School*. San Matco, CA: Morgan Kaufmann Publishers.

Naylor, G.J.P. (1992). The phylogenetic relationships among requiem and hammerhead sharks: inferring phylogeny when thousands of equally most parsimonious trees result. *Cladistics* 8: 295-318.

Naylor, G. J. P. & L. F. Marcus. (1994). Identifying isolated shark teeth of the genus *Carcharhinus* to species: relevance for tracking phyletic change through the fossil record. *American Museum Novitates* 3109, 1-53.

Slice, D. E. (2005). Morpheus et al.: platform independent software for morphometric analysis. Available at http://morphlab.sc.fsu.edu/software.html.

R Core Team. (2012). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-project.org/.

Richard, M.D. and R.P. Lippmann. (1991). Neural network classifiers estimate Bayesian a posteriori probabilities. *Neural Computation* 3: 461-483.

Ripley, B. D. (1996). Pattern Recogniton and Neural Networks. Cambridge: Cambridge University Press.

Rohlf, F.J. and L. F. Marcus. (1993). A revolution in morphometrics. *Trends in Ecology and Evolution* 8: 129-132.

Rosenblatt, F. (1962). Principles of Neurodynamics: perceptrons and the theory of brain mechanisms. Washington, DC: Spartan Books.

van Rossum, G. and F. L. Drake (eds). (2001). Python reference manual. Virginia, USA: PythonLabs. Available at http://www.python.org.

Tao, K. M. (1993). A closer look at the radial basis function (RBF) networks. *Conference record- Asilomar Conference on Signals, Systems, & Computer:* 401-405.

BIOGRAPHICAL SKETCH

K. James Soda was born September 2, 1987 in Colorado Springs, Colorado. He attended Western Washington University where he received a Bachelors of Science in Biology with an emphasis on marine biology. Upon completion of his Bachelors degree, he obtained a Masters of Arts from Stony Brook University's Department of Ecology and Evolution. In addition to these degrees, Mr. Soda has also held internships, apprenticeships, and volunteer positions with Shannon Point Marine Center, Friday Harbor Laboratories, and the National Museum of Natural History.