Feasibility of sexing of Bluegill *Lepomis macrochirus* from images of scales

Baker Herrin

**Abstract:**

This study was a proof of concept for whether Bluegills could be sexed from images of their scales. Total samples collected were 120 scales from four different adult Bluegill, which were used to train a machine learning model using PCA to select meaningful features and SVM as the algorithm. Results suggest that scales could be highly accurate in sex determination, and that scales from the caudal region are more effective than those in the region surrounding the pectoral fins. This has the potential to supply a new tool in on-site fish sexing and end the need for dissection to determine the sex of young of year Bluegill.

**Introduction:**

Sexing of fish is not only of immense importance to the scientific community, but also in the commercial fisheries and hobbyist trade. For the scientific community it is needed for studies comparing sex, as well as in the creation of fishing and other environmental regulations. Sexing is necessary for commercial fisheries and hobbyists in identifying brood stock or breeding pairs, or for the separation of sexes to avoid reproduction in grow out ponds. Sex determination is a difficult problem to address as the ability and method of determination depends on multiple factors and varies widely across species. A particular species may present clear differences in fin structure or color, while another may present no external differences of note. This makes the process of sex determination very species specific, with the only method that ensures certainty being dissection. Bluegill specifically are commonly sexed via the shape and size of the urogenital cavity; however, this alone cannot be used with certainty as it is prone to difficulty in the separation of the sex of subadult and non-spawning adults. Greater accuracy can be achieved by examining four other secondary sexual characteristics (James L. Brauhn 1972), but this too does not work well for smaller Bluegill (this method was 80% accurate with Bluegill of 5 cm total length).

Using image data to reveal information about fish is a growing field and is commonly employed in areas such as fish recognition systems for determining abundance (E A Awalludin *et al* 2020). However, while there has been work done in using sex as a feature for determining other parameters such as more accurate estimates of age (Sergio Bermejo 2007) there is no recent literature on enhancing the way sex is determined, which still as previously mentioned relies on expert knowledge. Bluegill were chosen as the subject of this study since young of year Bluegill require dissection for sex determination, and since they act as a representative of a commercial freshwater fishery, particularly in Florida where the study took place. This is important as Bluegill along with other sunfish species in Florida are highly valued as both a recreational and commercial fishery. The use of the images of Bluegill scales has the potential to be more accurate as well as faster in sex determination, since it would not be prone to human error in identification and does not rely on expert knowledge in recognition of sexual traits.

**Methods:**

The study was conducted on the University of Florida campus at Lake Alice (Figure 1). The lake serves as an outdoor classroom and stormwater pond for campus runoff. Lake levels are stabilized via two drainage wells.

## For the study, an electrofishing boat was used (equip with a 5000-Watt Honda Generator) to harvest four adult Bluegill, with the aim of getting two Bluegill of each sex (part of the motivation behind choosing adult Bluegill was the increased ability to differentiate between male and female from physical appearance). Once returned to shore, the data collector recorded weight and length (Table 1) of each specimen. All Bluegill were harvested on March 25th, 2021. Later that day dissection of each specimen proved the original assumptions of each specimen’s sex to be correct. Immediately afterwards, thirty scales were collected via tweezers from each specimen: fifteen from the right caudal region, and fifteen from the right side (Figure 2) of the fish. This yielded a total of one hundred and twenty scale samples in all, sixty of each sex. To produce high resolution images of each scale, a microscope was used. This was done to ensure the model would be able to find the differences in texture along the surface of the scale, which would otherwise not be visible (Figure 3). The scale images were then processed into two separate python numpy files, one for the data and one for the labels of male and female, designated with a zero or one for female or male respectively. The images were further divided by gender and body region to evaluate if one region was more informative than the other, referred to as the combined, R1, and R2 datasets and models. Principle Component Analysis (PCA) was used to decide which features of each image explained the most variance, and the Support Vector Machine (SVM) algorithm was used to train a machine learning model on the data. To determine which number of components (features) was best for PCA, and which combination of parameters together produced the most accurate classification results, a program was written to evaluate several potential SVM parameters for PCA components in a range of two to fourteen. The combination of PCA components and SVM parameters that yielded a model with the highest accuracy score when evaluated on a test set of scales was selected as the best model for this experiment. This is the model that was used by the region one, region two, and combined (regions one and two) datasets for classification. Precision, recall, F1-scores, and accuracy (Table 5) along with a ROC curve and confusion matrix were calculated for each dataset to assess the efficacy of classification *(for access to the python scripts written for this study, see* [*https://github.com/abubake/Bluegill\_Sex\_Program*](https://github.com/abubake/Bluegill_Sex_Program) *).*

## 

## 

**Results:**

When running the three SVM models trained from the combined, R1, and R2 datasets on the test set made of images of scales the model had not seen before, the accuracy scores (Table 5) were all significantly higher than random guessing. The datasets had balanced proportions of male to female scales to ensure the high accuracy scores were not due to majority class prediction (L. Lusa and R. Blagus 2012). Cross validation ensured consistency of our accuracy scores through generalization of the model. Surprisingly, R2 outperformed both the R1 and combined datasets. The effectiveness of classification is seen in each AUC found (Figures 5, 8, 11) with a perfect AUC for R2 (Figure 11). This is due to not having enough data to test on, as it is unlikely R2 would serve as a perfect classifier.

The quality of each classifier model was assessed with precision, recall, and F1-scores. For precision, the general trend is more precision in predicting males than females. In fact, all the models had perfect precision in males. So, when a male was identified, the model was certain of it. This was not so for females, which for the models varied from 75% to 100% precision (the 100% being the model that produced perfect classification for R2). Recall was a vastly different story. The relationship is inverse of precision, with perfect recall of females for all models, but resulting in male recall worse than the precision percentages for females. The worst of which was R1, with a recall equivalent to that of random guessing. To sum it up, on average the models had trouble predicting a sample as male, favoring a female prediction. However, when a male happened to be predicted the model was always correct. This qualitative description is summarized via the f1-score, which averages the results of precision and recall into an overall score. In f1-score, R2 outperformed the other two models. Precision, Recall, and f1-scores are all summarized in Tables 2, 3 and 4. Lastly, the confusion matrixes (Figures 6, 9, 12) support the claim that on average, the models had a harder time identifying males than females. It is also of note that in testing, older iterations of the model would classify as all females. This issue stopped once the number of PCA components increased.

**Discussion:**

This study shows that scales can be an effective tool for sex prediction in Bluegills, with the potential to be highly accurate. Furthermore, it shows the possibility of specific regions of the Bluegill having more accurate predictive capabilities, in particular the caudal scale region. The effectiveness of classification is most clearly seen in Figures 4, 7 and 10, where the predictions on scales in the test set are displayed.

There are broader implications of this study as this technique can be used for cases of Bluegill sex determination where it is desirable to avoid mortality of the fish, such as within the aquarium trade or with breeding stock. The success of this study also opens two other distinct advances in sex determination. One is the extension of this study to other fish species, and the other is improvements upon this method so that it can be used in the field for on-site sex determination with high accuracy. To extend this study to other species, all that would need to be done is to collect scale images from the same regions of other species, and test the models using the program written for this study. For the extension to field use, the principal change that would need to be made is achieving a suitably high accuracy with images taken in the field on smartphone cameras, rather than the extremely clear microscope images of the scales that were taken for this study. This may not be possible and would require another study.

This study has two main limitations. The lack of algorithms tested, and the lack of data collected. Only PCA combined with SVM was used for this study, which may not have been a pareto-optimal solution (Charles J. Petrie, Teresa A. Webster, Mark R. Cutkosky, 1995). Another solution which would have the potential to yield even higher accuracy would have been a convolutional neural network, which excels in binary classification problems such as this, additionally working well with PCA as implemented in this study. For the lack of data collected, this issue is two-fold and directly affected the results of the study. The first of which is its effect on the test set. The test sets were small for the R1 and R2 datasets, which increased the variability in the accuracy scores of those two models, which decreases their trustworthiness, as a difference of one incorrect vs. correctly identified image could drastically change the resulting metrics of the model. The other issue that arises is due to only having scales from mature fish, which means it may not be able to generalize well to smaller Bluegill, where sexual characteristics are much less developed, making it impossible to differentiate males and females without explicitly finding the reproductive organs. This means the results may not hold for that case, where the information would be much more useful than the results proven in this study.

Another extension of this study that could improve results would be using more pre-processing, since normalization of the pixels to be between zero and one was the only technique used. Using OpenCV for testing several pre-processing techniques such as thresholding and cropping of the images would carry out an extension of the pre-processing done in this study. Thresholding has the potential to drastically improve results, as it is likely that one of the PCA components used accounts for noise in the images caused by the background, which is undesirable as it reduces the ability of the model to generalize.

**References:‌**

References:‌

Klumb, Robert & Bozek, Michael & Frie, Richard. 2011. Validation of three back-calculation models by using multiple oxytetracycline marks formed in the otoliths and scales of bluegill × green sunfish hybrids. Canadian Journal of Fisheries and Aquatic Sciences. pg. 352-364.

L. Lusa and R. Blagus. 2012. The Class-Imbalance Problem for High-Dimensional Class Prediction. 11th International Conference on Machine Learning and Applications. pg. 123-126.

James L. Brauhn 1972. A Suggested Method for Sexing Bluegills, The Progressive Fish-Culturist, 34:1, 17-17.

Sergio Bermejo. 2007. Fish age classification based on length, weight, sex and otolith morphological features, Fisheries Research, Volume 84, Issue 2. pg. 270-274.

[Journal of Physics: Conference Series](https://iopscience.iop.org/journal/1742-6596), [Volume 1529](https://iopscience.iop.org/volume/1742-6596/1529). 2019. [The 2nd Joint International Conference on Emerging Computing Technology and Sports (JICETS). pg. 25-27.](https://iopscience.iop.org/issue/1742-6596/1529/5)

Marini, S., Fanelli, E., Sbragaglia, V. *et al.* 2018. Tracking Fish Abundance by Underwater Image Recognition. *Sci Rep* 8.

Charles J. Petrie, Teresa A. Webster, Mark R. Cutkosky. 1995. Using Pareto Optimality to Coordinate Distributed Agents.

**TABLES:**

TABLE 1: Weight and Length for each Bluegill specimen collected

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Female 1 | Female 2 | Male 1 | Male 2 |
| Weight | 192 | 167 | 248 | 340 |
| Length | 221 | 211 | 225 | 242 |

TABLE 2: Summary of results for the combined classifier, using 8 principal components for PCA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Precision | Recall | F1-score | Support |
| Female | 0.85 | 1.00 | 0.88 | 11 |
| Male | 1.00 | 0.78 | 0.80 | 9 |
| Weighted avg | 0.92 | 0.90 | 0.90 | 20 |

TABLE 3: Summary of results for the R1 classifier, using 8 principal components for PCA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Precision | Recall | F1-score | Support |
| Female | 0.75 | 1.00 | 0.86 | 6 |
| Male | 1.00 | 0.5 | 0.67 | 4 |
| Weighted avg | 0.85 | 0.80 | 0.78 | 10 |

TABLE 4: Summary of results for the R2 classifier, using 8 principal components for PCA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Precision | Recall | F1-score | Support |
| Female | 1.00 | 1.00 | 1.00 | 6 |
| Male | 1.00 | 1.00 | 1.00 | 4 |
| Weighted avg | 1.00 | 1.00 | 1.00 | 10 |

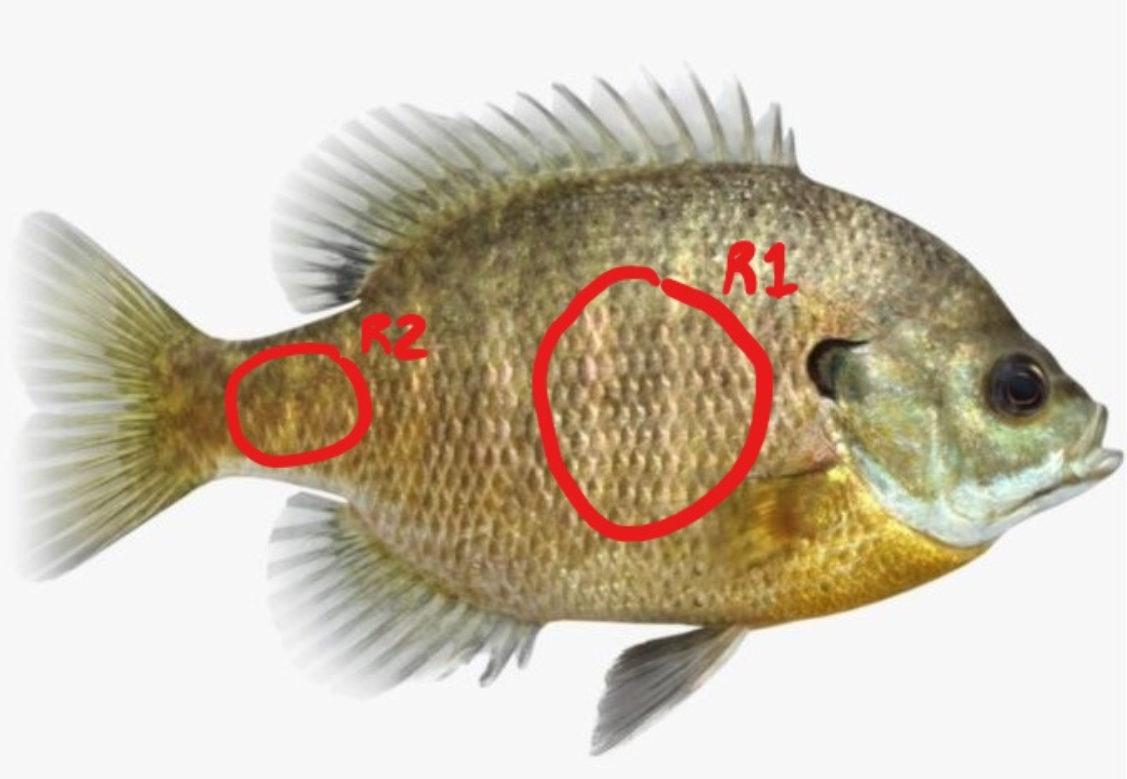
TABLE 5: Accuracy scores for different testing varieties

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grouping | Accuracy | Principle-Components | SVC C | SVC Gamma |
| R1 | 0.80 | 8 | 5 | 0.005 |
| R2 | 1.00 | 8 | 0.5 | 0.05 |
| Combined | .90 | 8 | .01 | 0.01 |

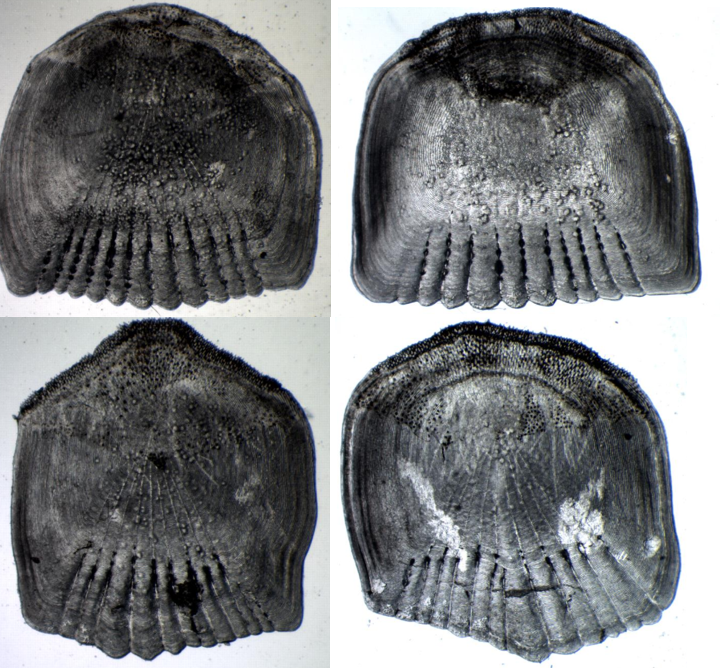
**Figures:**



**Fig. 1** *Lake Alice aerial view, approximately 9 Hectares*

**

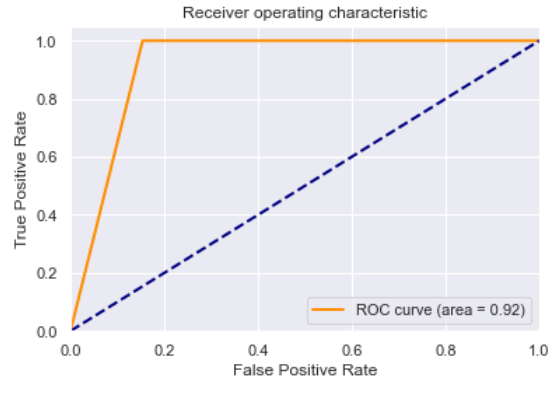
**Fig. 2** *Regions of scale removal on the Bluegill*



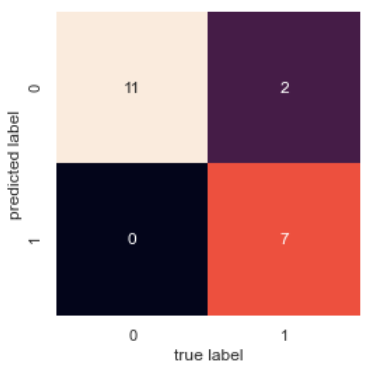
**Fig. 3** *Male and female scales; pictured top left is female region 1, top right is male region 1. The same ordering follows for the bottom two scales.*

**

**Fig. 4** *Combined accuracy results*

**

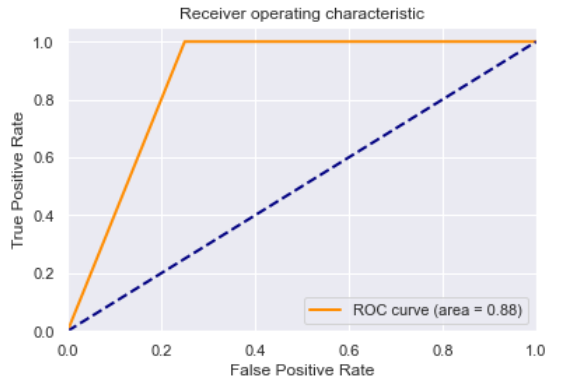
**Fig. 5** *Combined AUROC*

**

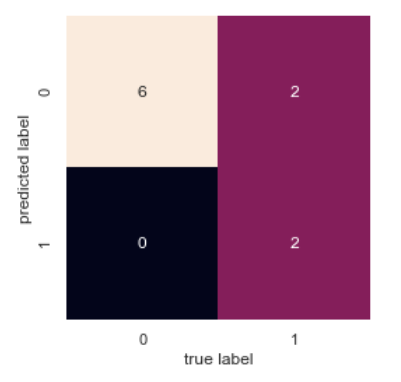
**Fig. 6** *Combined Confusion matrix*

**

**Fig. 7** *R1 classification results*

**

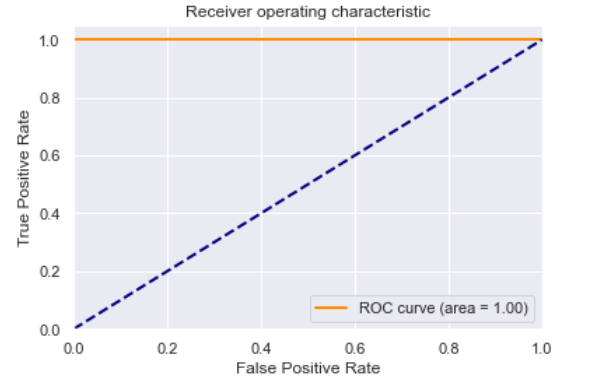
**Fig. 8** *R1 ROC curve*

**

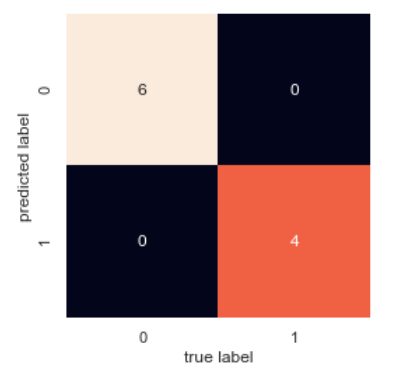
**Fig. 9** *R1 Confusion matrix*



**Fig. 10** *R2 classification results*



**Fig. 11** *R2 ROC curve*



**Fig. 12** *R2 Confusion matrix*