Belarusian State University of Informatics and Radioelectronics

# PRIP'2019

# Pattern Recognition and Information Processing

Proceedings of the 14th International Conference
(21-23 May 2019, Minsk, Belarus)

Minsk
«Bestprint»
2019

UDC 004.931(082)



The book collects the papers accepted for publication and presentation at the 14th International Conference on Pattern Recognition and Information Processing (PRIP '2019) that will be held at May 21-23 in Minsk, Belarus. The proceedings are prepared for publication by PRIP'2019 Program Committee and Belarussian Association for Image Analysis and Recognition.

Proceedings will be useful for students and researchers working in the following areas: Signal Processing, Image Analysis, Pattern Classification and Recognition, Information Processing, Distributed, Embedded Systems and Parallel Architectures for Signal and Image Processing, Knowledge – Based Expert and Decision Support System, Application of Pattern Recognition and Image Analysis.

Papers are reviewed by at least two reviewers and approved by Program Committee.

**UDC 004.931(082)**

**ISBN 978-985-90509-3-0**

# Determination of the Developmental Stage of Erythrocytes In the Common Nase (*Chodrostoma nasus*) Using Different Classification Methods

**Sladjana Spasić[1),4)], Srdjan Subotić[2)], Željka Višnjić-Jeftić[1)], Mirjana Lenhardt[3)]**

1) Institute for Multidisciplinary Research, University of Belgrade, Belgrade, Serbia,
e-mail: sladjana@imsi.bg.ac.rs,
http://www.imsi.bg.ac.rs/cv_template/?cv_title=z_sspasic&id_zap=091

2) Faculty of Biology, University of Belgrade, Belgrade, Serbia

3) Institute for Biological Research "Siniša Stanković", University of Belgrade, Belgrade, Serbia,
e-mail: lenhardt@ibiss.bg.ac.rs

4) University Singidunum, Faculty of Informatics and Computing, Belgrade, Serbia

*Abstract: Fish erythrocytes released from erythropoietic sites are immature. During circulation, they increase their own surface, while the cell body and nuclei begin to become more elliptical and less spherical during maturation. The relative abundance of different developmental stages represents an erythron profile which could be a more sensitive indicator of contamination than classic hematological indices. In this study, we analyzed blood smears of the Common Nase (Chodrostoma nasus) with the main goal to identify the determinants of the developmental stage of erythrocytes. Based on parameters developmental stages, erythrocytes are categorized into immature, intermediate or mature. In this investigation we have used four classification methods: the Two Step Cluster analysis, the K-Means Cluster analysis, and Neural Networks – Multilayer Perceptron and an Ordinal Regression Model. Our findings clearly justify that Multilayers Perceptron and OLR models are appropriate to classify the developmental stage of fish erythrocytes.*

*Keywords*: Ordinal Logistic Regression Analysis, Multilayer Perceptron, Two Step Cluster Analysis, K-Means Cluster Analysis, Developmental stage of erythrocytes, Shape factor of erythrocytes

## 1. INTRODUCTION

Fish erythrocytes released from erythropoietic sites are immature. During circulation, they increase their own surface, while the cell body and nuclei begin to become more elliptical and less spherical during maturation. The relative abundance of different developmental stages represents an erythron profile which could be a more sensitive indicator of contamination than classic hematological indices. Changes in the overall presence of certain developmental stages of erythrocytes, as well as changes in morphological characteristics, could point out to some environmental stressor and physiological problem. Traditional manual techniques can still be used in the diagnosis of diseases, analysis of cell morphology, histopathology of cells, classification of blood cells, etc. However, they are very tedious and long-lasting, and the accuracy of the results depends on human experience and expertise. Therefore, there is a constant need for an economical and robust automated system for the analysis of morphological and other characteristics of the cell [1]. The aim of this paper was to find the appropriate method for automatically determining the developmental stage of erythrocytes based on characteristics of their shape: the length, area, and perimeter of the erythrocytes and their nucleus were measured. Initially, the maturity of erythrocytes in the common nase (*Chodrostoma nasus*) was determined by experts.

## 2. MATERIAL AND METHODS

### 2.1 Sampling – Data and Variables

Thirty common nase (Chodrostoma nasus) individuals were caught at the Danube River, near Belgrade (Serbia). Blood smears were prepared on microscope slides and stained using a Bio-Diff kit (Bio Optica, Milano, Italy). Stained slides were observed under DM RB photomicroscope (Leica, Wetzlar, Germany) and random fields were photographed (Fig. 1). Pictures were analyzed in Digimizer image analysis software which automatically calculated cellular (AreaC, PerimeterC and LengthC) and nuclear parameters (AreaN, PerimeterN and LengthN). Based on morphological characteristics of erythrocyte cell and its nucleus, developmental stages of erythrocyte are categorized into the three groups - immature, intermediate or mature as it used in the study [2]. These groups were visually classified by three experts.

There is heterogeneity of fish erythrocytes in circulation which represents different developmental stages of red blood cells which is expressed as the extensive morphological changes during the process of maturation. The differences are mainly in length of the major axis, one-sided surface area, form factor and axis ratio [3] with form factor > 0.94 for immature erythrocytes [4]. Here, fifty erythrocytes were photographed for each individual and, based on [5], a subsample of 60 erythrocytes (20 for each type) was used for further analysis. The Shape Factor (*SF*) was calculated for each cell based on the following formula:

$$SF = \frac{4\pi \ \text{AreaC}}{\text{LengthC}} \qquad (1)$$

Shape factor (*SF*) is a normalized measure of roundness with values ranging in the real interval [0, 1] so that a perfect circle has a shape factor value equal to 1.

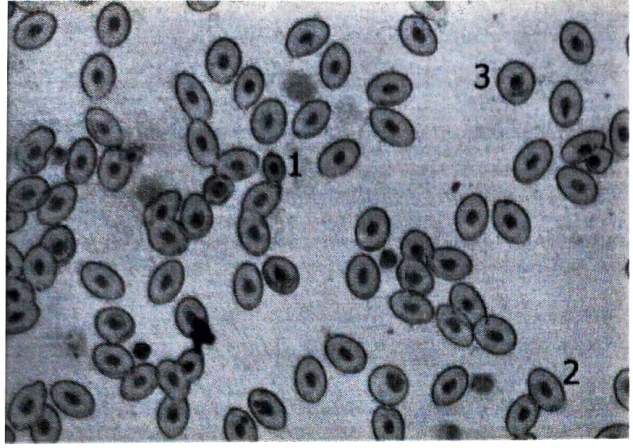


**Fig. 1. Photograph of the stained slide observed under DM RB photomicroscope: immature (1), intermediate (2), and mature (3) erythrocytes of common nase (Chrondrostoma nasus).**

### 2.2 Classification analysis

Generally, we could say that the objective of classification is grouping similar data points together and discover underlying patterns. In contrast to supervised classification methods, the unsupervised methods bring conclusions from data sets using only input variables without referring to known outcomes. Supervised and unsupervised classification methods have been used to predict the maturity of erythrocytes based on morphological characteristics - their size and shape. To solve the main task in this investigation we have used four classification methods: the Two Step Cluster (TSC) analysis and the K-Means Cluster (KMC) analysis as unsupervised methods and Neural Networks – Multilayer Perceptron (MLP) and an Ordinal Regression Model as supervised methods. Since the developmental stage is ordinal variable, an Ordinal Regression procedure, or PLUM Polytomous Universal Model (PLUM) has been developed to find predictors of developmental stage of erythrocytes and to check the applicability of the OLR model.

#### *2.2.1. Two Step Cluster Analysis*

The Two Step Cluster (TSC) analysis is specially developed and incorporated into the SPSS software as an unsupervised method designed to handle very large data sets. This method uses a likelihood distance measure which assumes that variables in the cluster model are independent in order to classify categorical and continuous variables. The TSC algorithm allows the user to set the maximum number of clusters or let the algorithm automatically determine the number of clusters.

The first step begins with the pre-clustering the cases into a Cluster Features (CF) Tree. The CF tree consists of the sub-clusters and a node that contains multiple cases contains a summary of variable information about those cases. The pre-cluster step uses sequential clustering by scanning the data cases one at a time. The next case should be added to one of the existing clusters or a new cluster is formed based on the distance measure as the similarity criterion. Thereafter, the second step clusters the sub-clusters of the CF tree into the final number of clusters using an agglomerative clustering algorithm. As a result of the agglomeration grouping, more solutions are obtained and the "best" cluster number is determined using Schwarz Bayesian Criteria (BIC) or Akaike Information Criterion (AIC) as the criteria for grouping.

We used the average Silhouette coefficient to evaluate clusters. An average Silhouette combines the concepts of cluster cohesion and cluster separation and belongs to real interval [-1, 1]. The average Silhouette coefficient is simply the average over all cases of the following calculation for each individual case:

$(B-A) / \max(A,B)$

where A is the distance from the case to the centroid of the cluster which the case belongs to and B is the minimal distance from the case to the centroid of every other cluster. The Silhouette coefficient and its average range between −1 (indicating a very poor model) and 1 (indicating an excellent model). The average silhouette greater than 0.5 indicates a reasonable partitioning of data, whereas when it is less than 0.2 means that the data do not show the cluster structure.

#### *2.2.2. K-Means Cluster Analysis*

K-Means Cluster (KMC) Analysis is an unsupervised procedure for identification relatively homogeneous groups of cases (n) based on selected variables using predefined the number of clusters (k). The K-Means cluster is a classic classification method that uses a partitioning algorithm and it is restricted to continuous variables. The K-means refers to averaging of the data finding the centroid as the imaginary or real location representing the center of the cluster. Distances are computed using Euclidean distance. We selected option for classifying cases with updating cluster centers iteratively and classifying. This means that the algorithm K-means starting with the first group of the randomly selected centroid, which is used as the initial solution, and then perform iterative calculations to optimize the centroid positions (the cluster center) in order to reach the final solutions. The clustering has been successful when there is no change in the centroids values or the defined number of iterations has been achieved. Then, the creating and optimizing clusters are finished.

#### *2.2.3. Ordinal Regression Analysis - Polytomous Universal Model (PLUM)*

The Ordinal regression analysis also belongs to the supervised methods and involves the application of the proportional odds model which is a commonly used model for the analysis of ordinal categorical data [6]. The proportional odds model is used to estimate the odds of being at or beyond a particular level when the response variable is polytomous as is in our case.

In this paper, an ordinal logistic regression model e.g. Polytomous Universal Model (PLUM) was developed using SPSS. The outcome variable is ordered with three levels of the developmental stage of red blood cells: 1 - immature, 2 - intermediate or 3 - mature. Cellular and nuclear measures and shape factor were used as independent variables. The ordinal logit model is based on the following equation:

$$\log\left(\frac{P(Y \le j|x_1,\ldots,x_p)}{P(Y > j|x_1,\ldots,x_p)}\right) = \alpha_j - \beta_1 X_1 - \ldots - \beta_p X_p \qquad (2)$$

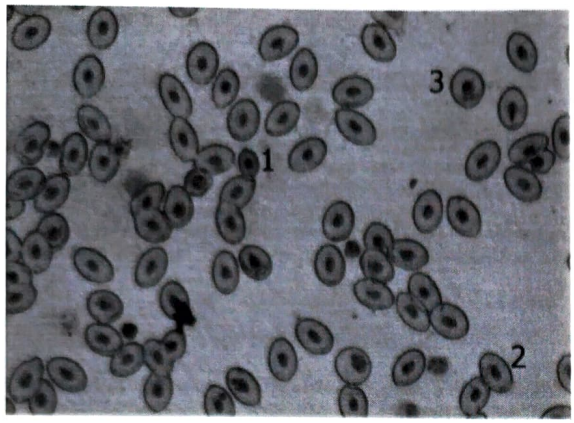


**Fig. 1. Photograph of the stained slide observed under DM RB photomicroscope: immature (1), intermediate (2), and mature (3) erythrocytes of common nase (Chrondrostoma nasus).**

## 2.2 Classification analysis

Generally, we could say that the objective of classification is grouping similar data points together and discover underlying patterns. In contrast to supervised classification methods, the unsupervised methods bring conclusions from data sets using only input variables without referring to known outcomes. Supervised and unsupervised classification methods have been used to predict the maturity of erythrocytes based on morphological characteristics - their size and shape. To solve the main task in this investigation we have used four classification methods: the Two Step Cluster (TSC) analysis and the K-Means Cluster (KMC) analysis as unsupervised methods and Neural Networks – Multilayer Perceptron (MLP) and an Ordinal Regression Model as supervised methods. Since the developmental stage is ordinal variable, an Ordinal Regression procedure, or PLUM Polytomous Universal Model (PLUM) has been developed to find predictors of developmental stage of erythrocytes and to check the applicability of the OLR model.

### *2.2.1. Two Step Cluster Analysis*

The Two Step Cluster (TSC) analysis is specially developed and incorporated into the SPSS software as an unsupervised method designed to handle very large data sets. This method uses a likelihood distance measure which assumes that variables in the cluster model are independent in order to classify categorical and continuous variables. The TSC algorithm allows the user to set the maximum number of clusters or let the algorithm automatically determine the number of clusters.

The first step begins with the pre-clustering the cases into a Cluster Features (CF) Tree. The CF tree consists of the sub-clusters and a node that contains multiple cases contains a summary of variable information about those cases. The pre-cluster step uses sequential clustering by scanning the data cases one at a time. The next case should be added to one of the existing clusters or a new cluster is formed based on the distance measure as the similarity criterion. Thereafter, the second step clusters the sub-clusters of the CF tree into the final number of clusters using an agglomerative clustering algorithm. As a result of the agglomeration grouping, more solutions are obtained and the "best" cluster number is determined using Schwarz Bayesian Criteria (BIC) or Akaike Information Criterion (AIC) as the criteria for grouping.

We used the average Silhouette coefficient to evaluate clusters. An average Silhouette combines the concepts of cluster cohesion and cluster separation and belongs to real interval [-1, 1]. The average Silhouette coefficient is simply the average over all cases of the following calculation for each individual case:

(B−A) / max(A,B)

where A is the distance from the case to the centroid of the cluster which the case belongs to and B is the minimal distance from the case to the centroid of every other cluster. The Silhouette coefficient and its average range between −1 (indicating a very poor model) and 1 (indicating an excellent model). The average silhouette greater than 0.5 indicates a reasonable partitioning of data, whereas when it is less than 0.2 means that the data do not show the cluster structure.

### *2.2.2. K-Means Cluster Analysis*

K-Means Cluster (KMC) Analysis is an unsupervised procedure for identification relatively homogeneous groups of cases (n) based on selected variables using predefined the number of clusters (k). The K-Means cluster is a classic classification method that uses a partitioning algorithm and it is restricted to continuous variables. The K-means refers to averaging of the data finding the centroid as the imaginary or real location representing the center of the cluster. Distances are computed using Euclidean distance. We selected option for classifying cases with updating cluster centers iteratively and classifying. This means that the algorithm K-means starting with the first group of the randomly selected centroid, which is used as the initial solution, and then perform iterative calculations to optimize the centroid positions (the cluster center) in order to reach the final solutions. The clustering has been successful when there is no change in the centroids values or the defined number of iterations has been achieved. Then, the creating and optimizing clusters are finished.

### *2.2.3. Ordinal Regression Analysis - Polytomous Universal Model (PLUM)*

The Ordinal regression analysis also belongs to the supervised methods and involves the application of the proportional odds model which is a commonly used model for the analysis of ordinal categorical data [6]. The proportional odds model is used to estimate the odds of being at or beyond a particular level when the response variable is polytomous as is in our case.

In this paper, an ordinal logistic regression model e.g. Polytomous Universal Model (PLUM) was developed using SPSS. The outcome variable is ordered with three levels of the developmental stage of red blood cells: 1 - immature, 2 - intermediate or 3 - mature. Cellular and nuclear measures and shape factor were used as independent variables. The ordinal logit model is based on the following equation:

$$\log\left(\frac{P(Y \le j|x_1,\ldots,x_p)}{P(Y > j|x_1,\ldots,x_p)}\right) = \alpha_j - \beta_1 X_1 - \ldots - \beta_p X_p \qquad (2)$$

where $\alpha_j$ ($j$ =1, ...,$J$-1) are the thresholds, $\beta_i$ ($i$ =1, ...,$p$) are logit coefficients.

*2.2.4 Neural Networks – Multilayer Perceptron*

Neural Networks – Multi-layer Perceptron (MLP) can be used in predicting or in this case in the problem of classifying. This method belongs to the supervised methods that mean the results of prediction can be compared with values of the target variables. MLP used a similar idea as ORL, but it differs from logistic regression in that MLP has one or more non-linear hidden layers between the input and output layers. MLP learns a function

$$f: R^m \rightarrow R^n$$

by training on a dataset, where $m$ is the dimension of the input vector and $n$ is the dimension for output vector. For a given set of $m$-dim input vectors and $n$-dim target variables, a nonlinear function for classification can be found. MLP learning of the classification function takes place in two phases: training and testing of the neural network. The dataset is partitioned into the training data and testing data sample. When learning a neural network is completed, an MLP model is created that will be used to classify data. In the hidden layer, a hyperbolic tangent as the activation function is used, however, in the output layer the activation function was "Softmax".

## 3. RESULTS

All methods for classification analysis have been repeated with three groups input continuous variables denoted with superscript letters a, b, c on Tables 1, 2, 3, 4.

### 3.1. Two Step Cluster Analysis

Two Step Cluster analysis was performed with the following common parameters: the number of clusters was fixed to 3, maximum of branch number was 8 and maximum level of tree depth was 3. Quality of clustering is poor in case of seven input variables and became greater in the case of four or two input variables. However, the percentage of properly classified erythrocytes does not exceed 78.33%.

**Table 1. Erythrocytes Classification using Two-Step Cluster analysis**

| Sample | Observed | Predicted | | | | Average Silhouette [-1,1] / Cluster Quality |
|---|---|---|---|---|---|---|
| | | Im-mature | Inter-mediate | Mature | Percent Correct | |
| 7 input variables[a] | Immature | 19 | 0 | 1 | 95.00% | 0.4 Poor |
| | Intermediate | 0 | 16 | 4 | 80.00% | |
| | Mature | 0 | 13 | 7 | 35.00% | |
| | Overall % | 31.67% | 48.33% | 20.00% | 70.00% | |
| 4 input variables[b] | Immature | 19 | 1 | 0 | 95.00% | 0.5 Good |
| | Intermediate | 2 | 17 | 1 | 85.00% | |
| | Mature | 0 | 9 | 11 | 55.00% | |
| | Overall % | 35.00% | 27.00% | 12.00% | **78.33%** | |
| 2 input variables[c] | Immature | 19 | 1 | 0 | 95.00% | 0.6 Good |
| | Intermediate | 3 | 17 | 0 | 85.00% | |
| | Mature | 0 | 11 | 9 | 45.00% | |
| | Overall % | 36.67% | 48.33% | 15.00% | 75.00% | |

a. AreaC, PerimeterC, LengthC, ShapeFactorC, AreaN, PerimeterN, LengthN;
b. AreaC, PerimeterC, LengthC, ShapeFactorC;
c. PerimeterC, ShapeFactorC.

### 3.2. K-Means Cluster Analysis

K-Means Cluster Analysis has been used for classifying cases with iterative updating of cluster centers, while the number of clusters was set to 3. Convergence achieved with convergence criterion 0.000 that means the maximum absolute coordinate change for any center was 0.000 while a number of iteration varied from 3 to 5 (see Table 2). The best result of classification was achieved with two input variables (PerimeterC and ShapeFactorC) and 61.67% correctly classified cases.

### 3.3. Ordinal Regression Analysis - PLUM

The classification results obtained by the PLUM method are shown in the matrix of confusion in Table 3. The best result of classification with application PLUM was achieved with all seven input variables and 96.67% cases are correctly classified.

**Table 2. Erythrocytes Classification using K-Means Cluster analysis**

| Sample | Observed | Predicted | | | | Number of iteration / Cluster Quality |
|---|---|---|---|---|---|---|
| | | Immature | Intermediate | Mature | Percent Correct | |
| 7 input variables[a] | Immature | 11 | 0 | 9 | 55.00% | 3 / Poor |
| | Intermediate | 0 | 6 | 14 | 30.00% | |
| | Mature | 0 | 5 | 15 | 75.00% | |
| | Overall % | 18.33% | 18.33% | 63.33% | 53.33% | |
| 4 input variables[b] | Immature | 11 | 0 | 9 | 55.00% | 4 / Poor |
| | Intermediate | 0 | 6 | 14 | 30.00% | |
| | Mature | 0 | 5 | 15 | 75.00% | |
| | Overall % | 18.33% | 18.33% | 63.33% | 53.33% | |
| 2 input variables[c] | Immature | 11 | 9 | 0 | 55.00% | 5 / Poor |
| | Intermediate | 0 | 14 | 6 | 70.00% | |
| | Mature | 0 | 8 | 12 | 60.00% | |
| | Overall % | 18.33% | 51.67% | 30.00% | **61.67%** | |

a. AreaC, PerimeterC, LengthC, ShapeFactorC, AreaN, PerimeterN, LengthN;
b. AreaC, PerimeterC, LengthC, ShapeFactorC;
c. PerimeterC, ShapeFactorC.

**Table 3. Erythrocytes Classification using Ordinal Regression Analysis - PLUM**

| Sample | Observed | Predicted | | | | Cluster Quality |
|---|---|---|---|---|---|---|
| | | Immature | Intermediate | Mature | Percent Correct | |
| 7 input variables[a] | Immature | 20 | 0 | 0 | 100.00% | Excellent |
| | Intermediate | 0 | 19 | 1 | 95.00% | |
| | Mature | 0 | 1 | 19 | 95.00% | |
| | Overall % | 33.33% | 33.33% | 33.33% | **96.67%** | |
| 4 input variables[b] | Immature | 19 | 1 | 0 | 95.00% | Excellent |
| | Intermediate | 0 | 19 | 1 | 95.00% | |
| | Mature | 0 | 2 | 18 | 90.00% | |
| | Overall % | 31.67% | 36.67% | 31.67% | 93.33% | |
| 3 input variables[c] | Immature | 19 | 1 | 0 | 95.00% | Excellent |
| | Intermediate | 0 | 19 | 1 | 95.00% | |
| | Mature | 0 | 1 | 19 | 95.00% | |
| | Overall % | 31.67% | 35.00% | 33.33% | 95.00% | |
| 2 input variables[c] | Immature | 17 | 3 | 0 | 85.00% | Good |
| | Intermediate | 2 | 15 | 3 | 75.00% | |
| | Mature | 0 | 2 | 18 | 90.00% | |
| | Overall % | 31.67% | 33.33% | 35.00% | 83.33% | |

a. AreaC, PerimeterC, LengthC, ShapeFactorC, AreaN, PerimeterN, LengthN;
b. ShapeFactorC, AreaC, PerimeterC, LengthC;
c. ShapeFactorC, AreaC, PerimeterC;
d. ShapeFactorC, AreaC.

### 3.4. Multi-layer Perceptron

In all repetitions of MLP analysis based on 7 or 4 variables, only one case was misclassified. Interestingly, it was the same case. Parameters in the analysis: Number of hidden layers was equal to 1 and activation function in the hidden layer was hyperbolic tangent. Dependent Variable was the developmental stage of erythrocytes with three units in the output layer and activation function Softmax. Percent of correct predictions was 98.30%. Fig. 2 shows a one hidden layer MLP with four variables in the input layer and the type of erythrocytes in the output layer.

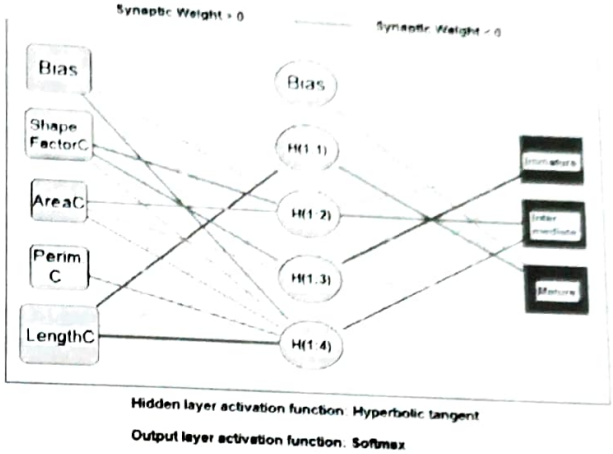


**Fig. 2. A one hidden layer MLP with four variables in the input layer and the developmental stage of erythrocytes in the output layer.**

Table 4. Erythrocytes Classification using Neuronal Networks - Multilayer Perceptron

| Sample | Observed | Predicted | | | | | No of units in hidden layer/ Cluster Quality |
|---|---|---|---|---|---|---|---|
| | | Immature | Intermediate | Mature | Percent Correct | AUC | |
| 7 variables[a] | Immature | 20 | 0 | 0 | 100.0% | **1.000** | 5 / Excellent |
| | Intermediate | 0 | 19 | 1 | 95.0% | **0.999** | |
| | Mature | 0 | 0 | 20 | 100.0% | **0.999** | |
| | Overall % | 33.3% | 31.7% | 35.0% | **98.3%** | | |
| 4 variables[b] | Immature | 20 | 0 | 0 | 100.0% | 1.000 | 4 / Excellent |
| | Intermediate | 0 | 19 | 1 | 95.0% | 0.984 | |
| | Mature | 0 | 0 | 20 | 100.0% | 0.991 | |
| | Overall % | 33.3% | 31.7% | 35.0% | **98.3%** | | |
| 3 variables[c] | Immature | 18 | 2 | 0 | 90.0% | 1.000 | 4 / Excellent |
| | Intermediate | 0 | 19 | 1 | 95.0% | 0.984 | |
| | Mature | 0 | 0 | 20 | 100.0% | 0.991 | |
| | Overall % | 30.0% | 35.0% | 35.0% | 95.0% | | |
| 2 variables[d] | Immature | 19 | 1 | 0 | 95.0% | 0.994 | 3 / Excellent |
| | Intermediate | 0 | 19 | 1 | 95.0% | 0.962 | |
| | Mature | 0 | 1 | 19 | 95.0% | 0.989 | |
| | Overall % | 31.7% | 35.0% | 33.3% | 95.0% | | |
| 2 variables[e] | Immature | 19 | 1 | 0 | 95.0% | 1.000 | 2 / Excellent |
| | Intermediate | 0 | 18 | 2 | 90.0% | 0.986 | |
| | Mature | 0 | 1 | 19 | 95.0% | 0.988 | |
| | Overall % | 31.7% | 33.3% | 35.0% | 93.3% | | |

a. AreaC, PerimeterC, LengthC, ShapeFactorC, AreaN, PerimeterN, LengthN;
b. ShapeFactorC, AreaC, PerimeterC, LengthC;
c. ShapeFactorC, AreaC, PerimeterC;
d. ShapeFactorC, AreaC;
e. ShapeFactorC, PerimeterC.

## 4. DISCUSSION

Data on the application of models in studies on fish erythrocytes morphometry is lacking. Rowan [5] did erythrocytes classification, by a proportional odds model,

with the cellular and nuclear area and shape factor as independent variables. Since the aim of this study was not to replicate existing research, we tested different models with the different number of variables, with the idea to find the model best suited for complementing expert's visual analysis.

From that point of view, in this study, we have used four classification methods to classify the developmental stage of fish erythrocytes: the TSC analysis, the KMC analysis, MLP, and an Ordinal Regression Model. Our findings clearly justify that PLUM and MLP models are more successful than the other two. The best results in the classification of the maturity of erythrocytes were achieved using neural networks. MLP model with 7 or 4 input variables correctly classified 98.3% cases. The result received by application an Ordinal Regression Analysis - PLUM is slightly less accurate (96.67%) but with all 7 variables as the input. The best result in TSC analysis was obtained with four input variables: 78.33% accurately classified cases. The much lower percent of correct classification (61.67%) was achieved by using KMC analysis and only the two input variables.

The TSC method is used in the study [7] to disclose sperm subpopulations based on morphometric parameters of the sperm head and midpiece. The authors have demonstrated that a simple TSC procedure successfully revealed the existence of 6 subpopulations within a semen sample. They conclude that the great advantage of this system is that all the analyses can be done in one step more rapidly and easily than with the techniques used so far to disclose sperm subpopulations. Despite the advantages of this method and the aforementioned successful application of cell clustering, in our analysis, this method proved to be relatively unsuccessful. Generally, KMC performances are not as good as those of other more advanced clustering techniques, because small variations in data can lead to large variance in classification. However, in their study [8] successfully used KMC in the unsupervised morphological classification of ganglion cells in the mouse retina. The cells were imaged in three dimensions and the morphologies of a series of 219 cells were analyzed quantitatively. A total of 26 parameters were studied, however, the cells were effectively clustered only on the basis of three parameters. In our analysis, the best result was achieved by using two input variables, but it was not sufficient with respect to the quality of clustering.

In [1], the authors successfully used the model of a neural network model for the purpose of classification for diagnosing disease. The proposed system automatically counts white blood cells, accurately determines their size and classifies them into five types. The advantage of the MLP model is the ability to learn non-linear models even in real time. Although MLP has certain disadvantages such as the requirement for setting a specific number of hyperparameters (the number of hidden neurons, layers and iterations) or sensitivity to feature scaling; in this analysis they have not proved to be crucial because SF is standardized, and the variables area, perimeter, length are linearly increasing with erythrocyte maturation. Therefore, we had a relatively simple task that was solved using the MLP model. An interesting situation is observed for one specific erythrocyte. Experts classified it as an intermediary; however, both PLUM and MLP classified it as mature. In some cases, it is difficult to distinguish whether the erythrocyte is in the final intermediary phase or in the initial mature phase. This also highlights the usefulness of tested models as a tool to help or even correct the visual analysis of experts and it could help in better assessing contamination exposure by erythron profile [9].

## 5. CONCLUSION

Our findings have shown that Multilayers Perceptron and OLR models are appropriate for classification of the development phase of erythrocyte fish. These methods can be used as a tool in automatic classification or as a help to traditional manual techniques.

***Acknowledgment:*** This work was supported by the Ministry of Education, Science and Technological Development of Serbia, grant number OI173045.

## 6. REFERENCES

[1] S. Nazlibilek, D. Karacor, T. Ercan, M. H. Sazli, O. Kalender, Y. Ege Automatic segmentation, counting, size determination and classification of white blood cells, *Measurement*, 55 (2014), p.58-65.

[2] S. Nowaczewski, H. Kontecka, Haematological indices, size of erythrocytes and hemoglobin saturation in broiler chickens kept in commercial conditions, *Animal Science Papers and Reports*, 30(2) (2012), p.181-190.

[3] A.H. Houston, Review: Are the classical hematological variables acceptable indicators of fish health? Transactions of the American Fisheries Society, 126 (6) (1997), p. 879-894.

[4] A.H. Houston, A. Murad, Hematological characterization of goldfish Carassius auratus L. by image analysis: effect of thermal acclimation and heat shock. Canadian Journal of Zoology 69 (1991),p. 2041-2047.

[5] M.W. Rowan, Use of blood parameters as biomarkers in brown bullheads (Ameiurus nebulosus) from Lake Erie tributaries and Cape Cod ponds. Doctoral dissertation. The Ohio State University. 2007.

[6] L. Xing. Ordinal regression analysis: Fitting the proportional odds model using Stata, SAS, and SPSS, *Journal of Modern Applied Statistical Methods*, 8 (2) (2009) p.632-645, Article 30.

[7] I.N. Martinez, J.M. Moran, F.J. Pena, A Two-Step Cluster Procedure After Principal Component Analysis Identifies Sperm Subpopulations in Canine Ejaculates and Its Relation to Cryoresistance, *Journal of Andrology*, 27(4) (2006) p.596-603.

[8] J. Kong, D. R. Fish, R. L. Rockhill, R. H. Masland, Diversity of ganglion cells in the mouse retina: unsupervised morphological classification and its limits, *The Journal of comparative neurology*, 489(3) (2005),p.293-310.

[9] V. Aliko, M Qirjo, E. Sula, V. Morina, C. Faggio, Antioxidant defense system, immune response and erythron profile modulation in goldfish, *Carassius auratus*, after acute manganese treatment. *Fish & shellfish immunology* 76 (2018), p. 101-109.