

Article

Integrating Deep Learning into Genotoxicity Biomarker Detection for Avian Erythrocytes: A Case Study in a Hemispheric Seabird

Martín G. Frixione ^{1,2}, Facundo Roffet ³, Miguel A. Adami ¹, Marcelo Bertellotti ^{1,4}, Verónica L. D'Amico ¹, Claudio Delrieux ³ and Débora Pollicelli ^{1,3,5,*}

- ¹ Centro para el Estudio de Sistemas Marinos (CESIMAR), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Puerto Madryn 9120, Argentina; mfrixione@cenpat-conicet.gob.ar (M.G.F.); madami@cenpat-conicet.gob.ar (M.A.A.); bertello@cenpat-conicet.gob.ar (M.B.); damico@cenpat-conicet.gob.ar (V.L.D.)
- ² School of Biological Sciences, University of Utah, Salt Lake City, UT 84112, USA
- ³ Instituto de Ciencias e Ingeniería de la Computación (ICIC), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and Departamento de Ingeniería Eléctrica y Computadoras, Universidad Nacional del Sur (UNS), Bahía Blanca 8000, Argentina; facundo.roffet@cs.uns.edu.ar (F.R.); cad@uns.edu.ar (C.D.)
- ⁴ Escuela de Producción, Ambiente y Desarrollo Sostenible, Universidad del Chubut, Puerto Madryn 9120, Argentina
- ⁵ Laboratorio de Investigación en Informática (LINVI), Departamento de Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB), Comodoro Rivadavia 9005, Argentina
- * Correspondence: pollicelli@cenpat-conicet.gob.ar

Abstract: Recently, nuclear abnormalities in avian erythrocytes have been used as biomarkers of genotoxicity in several species. Anomalous shapes are usually detected in the nuclei by means of microscopy inspection. However, due to inter- and intra-observer variability, the classification of these blood cell abnormalities could be problematic for replicating research. Deep learning, as a powerful image analysis technique, can be used in this context to improve standardization in identifying the biological configurations of medical and veterinary importance. In this study, we present a standardized deep learning model for identifying and classifying abnormal shapes in erythrocyte nuclei in blood smears of the hemispheric and synanthropic Kelp Gull (*Larus dominicanus*). We trained three convolutional backbones (ResNet34 and ResNet50 architectures) to obtain models capable of detecting and classifying these abnormalities in blood cells. The analysis was performed at three discrimination levels of classification, with broad categories subdivided into increasingly specific subcategories (level 1: “normal”, “abnormal”, “other”; level 2: “normal”, “ENAs”, “micronucleus”, “other”; level 3: “normal”, “irregular”, “displaced”, “enucleated”, “micronucleus”, “other”). The results were more than adequate and very similar in levels 1 and 2 (F1-score 84.6% and 83.6%, and accuracy 83.9% and 82.6%). In level 3, performance was lower (F1-score 65.9% and accuracy 80.8%). It can be concluded that the level 2 analysis should be considered the most appropriate as it is more specific than level 1, with similar quality of performance. This method has proven to be a fast, efficient, and standardized approach that reduces the dependence on human supervision in the classification of nuclear abnormalities in avian erythrocytes, and can be adapted to be used in similar contexts with reduced effort.

Keywords: deep learning; genotoxicity; avian erythrocytes; kelp gulls



Citation: Frixione, M.G.; Roffet, F.; Adami, M.A.; Bertellotti, M.; D'Amico, V.L.; Delrieux, C.; Pollicelli, D. Integrating Deep Learning into Genotoxicity Biomarker Detection for Avian Erythrocytes: A Case Study in a Hemispheric Seabird. *Math. Comput. Appl.* **2024**, *29*, 41. <https://doi.org/10.3390/mca29030041>

Academic Editor: Alexandre Souto Martinez

Received: 3 April 2024
Revised: 22 May 2024
Accepted: 24 May 2024
Published: 28 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Detection of nuclear abnormalities in erythrocytes has been performed in the last decade as the main procedure to assess genotoxicity in birds of different taxa [1–3]. Increases in nuclear abnormalities can be triggered by exposure of birds to different types of contaminants [4,5]. Synanthropic species can be considered part of a group of urbanized

species that could be good indicators of genotoxicity. Several species of different taxa can exhibit abnormalities in blood cells without showing deterioration in body condition [4–6]. In particular, some species of seagulls (*Larus* sp.), which in many cases are closely associated with anthropogenic activities and polluted environments, have been reported to have high abnormality rates in blood cells [7]. In contrast, less tolerant species such as terns may be affected by contaminants and show an increasing frequency of red blood cell abnormalities and consequent health deterioration [8]. Genotoxicity can lead to DNA damage during mitosis and eventually develop into cancer [9,10]. High frequencies of micronuclei in red blood cells have been found as a genotoxicity biomarker in different species [1], as well as high frequencies of other nuclear shapes defined as erythrocyte nuclear abnormalities (ENAs) [5].

The Kelp Gull (*Larus dominicanus*) is a large-sized gull [11] that is widely distributed throughout South America, southern Africa, Australia, New Zealand, and Antarctica [12]. This gull is an opportunistic and generalist species that uses several types of anthropogenic food sources during the breeding and non-breeding seasons, and is considered a good monitor to track environmental changes [13–15].

In general, abnormal forms in erythrocytes can be detected by means of microscopy examination. However, these tests are scored according to the judgment of visual inspection by experts, which makes meaningful comparison between samples difficult because of uncontrolled differences in scoring within and among observers. Furthermore, other genotoxicity tests are expensive in terms of the equipment, time, and expertise required. In this context, automated identification and detection techniques are being explored as a means to objectively and reproducibly standardize the detection of nuclear anomalies. Artificial intelligence (AI) has been broadly used in the last decade in human and veterinary medicine, specifically in image analysis as a tool for diagnosing diseases [16–18], to identify various features of veterinary importance, such as the detection of reticulocytes in blood samples of cats [19], the identification of skin tumor types in dogs [16], or the detection of hemoparasites (*Plasmodium gallinaceum*) in the blood of chickens [17]. In particular, deep learning (DL) has recently garnered significant attention due to its successful application in several image analysis contexts, in particular the recognition and identification of complex sets of shapes and objects at multiple levels of abstraction [20]. For this reason, DL is increasingly used in medical and veterinary image analysis [16,21], and to improve diagnostic accuracy [22].

To the best of our knowledge, DL has not been used to detect nuclear abnormalities in avian erythrocytes. The aim of this study is to present a standardized DL model for identifying and classifying abnormal shapes in erythrocyte nuclei, which can be used to monitor genotoxicity in synanthropic seabirds. The model can be used to evaluate the presence of nuclear abnormalities in this species under different environmental conditions and to establish uniform criteria for identifying and classifying nuclear abnormalities.

2. Materials and Methods

For the analyses, we used images of blood smears from adult kelp gulls collected during previous studies in northeastern Patagonia [7]. Blood samples were collected from the ulnar vein and a thin blood smear was made using a fresh drop of blood [23], which was air-dried (between 1 and 3 min), and fixed in ethanol for 3 min. Once smears were dried they were stained with a kit for differential quick stain (Tinción 15—Biopur SRL, Rosario, Argentina; [24]). The staining kit consists of three steps, fixative (5 dips of 1 s each), solution 1 (Xanthenes; 5 dips of 1 s each), and solution 2 (Thiazines; 5 dips of 1 s each), draining between each step and at the end. Blood smears were photographed under a 100× magnification objective with oil immersion [1], using a Leica DM500 binocular microscope with Leica ICC50 W modular digital camera (Leica Microsystem, Wetzlar, Germany). Our approach involves manually annotating full-sized images by identifying and delineating regions of interest (ROIs) corresponding to different categories. These ROIs consist of individual elements in the form of bounding boxes that have been previously

classified, labeled, and annotated through consensus by three (3) biologists experienced in categorizing erythrocyte abnormalities. Images analyses were performed on a subset of 214 digital images randomly selected from 51 blood smears of kelp gulls. For global classification (level 1) we defined 3 categories of features in the blood smear images: (1) “normal” was indicated if the nucleus had an elliptically defined shape; (2) “abnormal” if the nucleus had a micronucleus, if the erythrocyte was enucleated [8], or if the nucleus was displaced [4] or had an irregular shape (budded, segmented, notched, tailed [8]); and (3) all other objects in the blood smears (white blood cells, platelets, broken erythrocytes, and unknown objects) as “other”. As a result, we obtained a total of 3,431 ROIs samples from all categories (Figure 1).

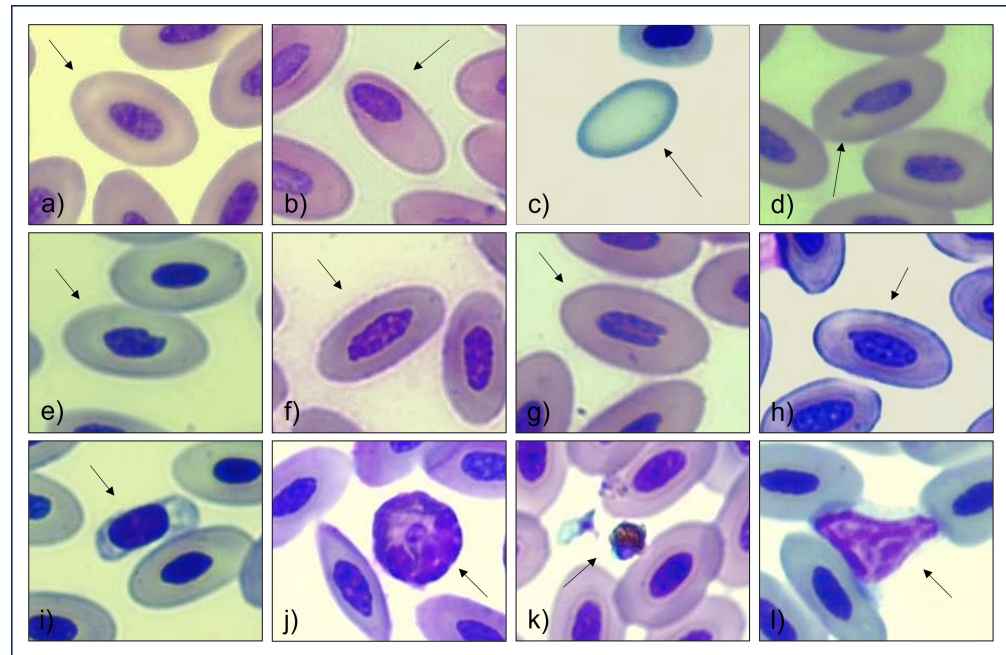


Figure 1. (a) Normal, (b) displaced, (c) enucleated, (d) micronucleus, (e–h) irregular, (i) other (thrombocyte), (j) other (heterophil), (k) other (artifact), and (l) other (lymphocyte).

2.1. Data Preparation

The analysis employed a disaggregated approach from this global categorization, breaking down the categories progressively into more specific subcategories, encompassing a total of six (6) subcategories in the last level of analysis (level 3): “normal erythrocytes”, “micronucleus”, “irregular”, “displaced”, “enucleated”, and “other” (Table 1). In the process of subcategorization, the finer-grained or more specific categories exhibited a reduced number of samples. The significant differences in sample quantities among subcategories show the imbalanced nature of this dataset, characterized by a highly uneven distribution of examples across the subcategories. To train models, ROI samples of subcategories were extracted, cropping the bounding boxes of each annotation from the larger full-sized image.

Because of the significant imbalance in the number of samples for each subcategory, and to address the problem of model generalization, prevent overfitting, and ensure a more representative dataset in terms of variability in chromaticity, luminance, and geometry, both downsampling and data augmentation techniques [25] were used for each model in the training set. Also, the images show color variations due to the staining procedure used to stain the blood smear samples. Therefore, we accounted for this factor during data augmentation, by means of the ColorJitter and RGBShift methods. By considering color variation as part of the augmentation process, we aimed to improve the robustness of the model and make it more invariant to possible differences in stain. In addition, considering the natural variability in the spatial position of elements in smears, we considered rotation and flip methods, as well as variations in focus, noise, occlusion, and sharpness.

Table 1. Analyses conducted at different scales depending on whether from grouped or individual categories. In parentheses, the number of ROIs used for each subcategory in each analysis (80% training/20% validation).

Analysis 1	Analysis 2	Analysis 3	Total
Normal (323/77)	Normal (321/79)	Normal (320/80)	400
	Micronucleus (14/4)	Micronucleus (14/4)	18
		Irregular (227/58)	285
Abnormal (294/75)	ENAs (280/71)	Displaced (46/12)	58
		Enucleated (6/2)	8
Other (236/60)	Other (237/59)	Other (238/58)	296
			1065

As a result, the “normal” category was randomly downsampled from 2871 to 400 samples, and the following data augmentation methods from Albumentations python library [26] were applied: HorizontalFlip, VerticalFlip, Rotate, Sharpen, GaussianBlur, GaussianNoise, RandomSizedCrop, ColorJitter, and RGBShift. Both the methods and the parameter settings applied are summarized in Table 2. Each augmentation method has an associated independent probability of being applied, usually set to a specific percentage (as a predefined parameter), which allows for the cumulative application of them to the original images. Then, for each training epoch the data augmentation pipeline involves sampling each image and sequentially applying a random combination of the selected transformations. It is worth noting that transformed images do not need to be stored on disk. No augmentation methods were used during the validation process.

Table 2. Augmentation types and parameters.

Augmentation Type	Parameters
HorizontalFlip	$p = 0.5$
VerticalFlip	$p = 0.5$
Rotate	$p = 0.5$
Sharpen	$p = 0.5$
ColorJitter	brightness = 0.3, contrast = 0.5, saturation = 0.5, hue = 0.0, $p = 0.5$
RGBShift	$p = 0.5$
GaussianBlur	$p = 0.5$
GaussianNoise	$p = 0.5$
RandomSizedCrop	min_max_height = (120,140), height = 150, width = 150, $p = 0.5$

2.2. Deep Learning Model Setting

Three models were developed, each for a different subcategorization analysis ($n = 3$). The deep learning models were constructed and trained using the FastAI API [27]. ResNet34 and ResNet50, two widely used convolutional neural network architectures in the field of computer vision, were employed as the base models [28]. A comprehensive hyperparameter tuning was previously performed to find the optimal parameters, which are shown in Table 3; this table describes the different architectures and hyperparameters of the CNN models that showed the best F1-score for each analysis. In all cases, the models were pre-trained with the ImageNet dataset [29], as given by the library Torchvision [30]. The models were trained for 31 epochs: the first with all the layers frozen except for the last one, and the rest with all layers unfrozen (if a layer is frozen, it means that its parameters cannot

be trained in that epoch). The amount of epochs was automatically determined using early stopping. Further training may achieve marginal improvements, at the expense of possible overfitting. All models were set with the cross-entropy loss function for optimization, which is commonly used in multi-class classification problems. The cropped sample images were resized with the “squish” method to fit the specified dimensions of 150×150 pixels.

For each trained model, the cropped images were divided into a training set (80%) and a validation set (20%), ensuring the representation of all categories in the overall dataset and maintaining the natural proportion of samples for each category. This helps ensure a fair and accurate evaluation of the models, allowing them to learn and generalize effectively for all categories rather than being biased towards the most represented ones.

Table 3. Model configurations and relevant hyperparameters for each analysis.

	Analysis 1	Analysis 2	Analysis 3
Architecture	ResNet50	ResNet34	ResNet50
Batch size	16	8	16
Pre-trained weights	yes		
Image size	150×150		
Resize method	squish		
Epochs	31		
Loss function	cross-entropy		

3. Results

Table 4 provides the accuracy and F1-score metrics for the three models at different levels of class analysis, labeled as analysis 1, analysis 2, and analysis 3 (Table 1). These metrics demonstrate the performance of the model as it progresses from broader to more detailed class analysis. The confusion matrices in Figure 2 offer a detailed breakdown of the model’s classification performance for individual classes. Simultaneously, the matrices highlight areas where the models may benefit from fine-tuning to reduce the occurrence of false positives or false negatives in specific classes, ultimately enhancing their performance in those areas. In Figures A1 and A2 of the Appendix A, we show details of the metrics and the evolution of the loss function during training and validation throughout epochs, and the five images with higher associated losses.

The model for analysis 1 exhibits exceptional performance, achieving an overall accuracy of 88.21% and an F1-score of 88.73%. This model delivers accurate and well-balanced classification results across the three classes. The associated confusion matrix shows a consistent identification of true positives across all classes, achieving an accuracy ranging from 84% to 91.67% in classifying instances across the three different classes.

In analysis 2, the classes from analysis 1 are further subdivided into more detailed categories. Specifically, “ENAs” and “micronucleus” were separated from the “abnormalities” category in the first level of analysis. This new division resulted in a slight decrease in the overall accuracy (less than 1%) but a reduction in the F1-score (close to 6%). The confusion matrix for this model reveals a tendency to misclassify instances, primarily as the “normal” class. Notably, the model achieves high accuracy in the “normal” class, with an approximate rate of 97.47%. However, in classes such as “ENAs” and “other” the accuracy is considerably lower, indicating that the model frequently misclassifies these instances as “normal”. This pattern of misclassification explains the decrease in the F1-score.

Finally, in analysis 3, the model further refines the classification by introducing even more detailed categories. In this level, the “ENAs” category is subdivided into “displaced”, “enucleated”, and “irregular”. At this stage, both accuracy and F1-score show a decline, of 4.1% and 11.7%, respectively, compared to the metrics of the model at the first level of analysis. Similar patterns of misclassification persist in this subsequent classification level. Notably, the confusion matrix reveals that classes such as “displaced”, “irregular”,

and “micronucleus” exhibit lower rates of true positive classifications. In contrast, the “enucleated”, “normal”, and “other” classes are almost perfectly classified. Furthermore, there is a tendency for false positives to be misclassified as either “irregular” or “normal” in certain cases.

Table 4. Accuracy and F1-score metrics registered by CNN analyses conducted for erythrocyte images of blood smears of the Kelp Gull.

	Analysis 1	Analysis 2	Analysis 3
Accuracy	88.21%	87.79%	84.11%
F1-score	88.73%	82.88%	77.03%

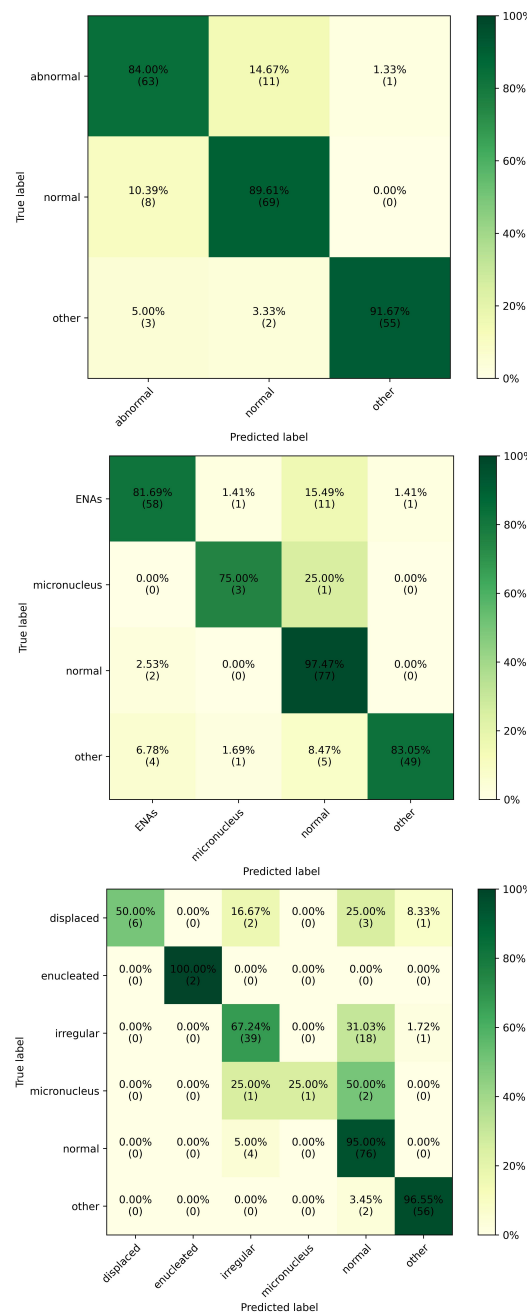


Figure 2. Confusion matrices obtained with the validation set for the three models. Each subfigure represents a different level of analysis achieved by the CNN for classifying erythrocyte images from blood smears of the Kelp Gull.

4. Discussion

In the last decades, several authors have focused on the classification of abnormal erythrocyte shape (mainly in human erythrocytes), some of them using automated deep learning [31–33]. However, the use of this tool for the detection of abnormalities in red blood cells of other taxa has been poorly explored. Birds, reptiles, amphibians, and fish (unlike mammals) have nucleated erythrocytes that could be affected by environmental pollution and produce a higher frequency of different abnormal erythrocyte types in the presence of pollutants [34–38]. The current study presents the first automated deep learning approach to classify blood cell abnormalities in nucleated erythrocytes in wild birds.

The CNN models showed higher values of F1-score and accuracy for the first level and reasonably good results for the second and third levels of analysis. The variability in accuracy among categories might be expected to increase as subcategories are disaggregated and become less frequent. Nevertheless, given the balance between the number of categories and the minimal difference in metrics between models for analysis 1 and analysis 2, we suggest that the model resulting from the latter analysis can be considered the best option for classifying abnormalities in avian erythrocytes. The classification quality metrics of these two models reflect their strong discrimination capabilities and their reliability in accurately categorizing instances within the specified classes. The models are demonstrated to be powerful tools for automated erythrocyte classification into grouped categories and offer faster performance compared to manual smear processing techniques and visual inspection. Notably, the CNN in our validation set classified over 210 ROIs from blood smear images in just two seconds, while the full consensus of visual inspection for the same number of ROIs by three experts could take almost an hour.

The results of analysis 3, in which category ENAs was disaggregated as “irregular”, “displaced”, and “enucleated”, showed that the categories “displaced” and “micronucleus” were classified as “normal” or “irregular” in a high proportion of cases. In fact, these categories were the ones with the least available training examples. These categories represent a potential double-identification condition that could be challenging to detect effectively by the CNN, since “displaced” and “micronucleus” could additionally have an irregular or a normal nucleus shape (see Figure 1). In this context, the apparent deviation of the nucleus from the cell axis or a separated material from the nucleus could be discarded by the model as the first decision for categorization if the focus of the model is on the shape of the central target.

In Figures A3 and A4 of the Appendix A, we present the five most misclassified items (during training and validation) for the three models. In the case of the most misclassified examples in the validation stage for analysis 3, three of these cases are “displaced” and “micronucleus”, incorrectly classified as “normal”. These misclassifications are less prevalent in analysis 1 and 2, in which the most frequent confusion arises between “normal” and “abnormal”, and “ENAs” between the other classes. In this regard, the number of “displaced” training items is one of the most numerous in the “ENAs” class ($n = 58$), while the amount of “micronucleus” items ($n = 18$) could be a problem for the CNN to identify the category during validation if 20% of the sample size is used (i.e., only four validation instances in this case). In analysis 3, the classes are subdivided into more specific cases, and thus, the small number of training examples increases their mutual confusion. On the other hand, the performance of the “enucleated” class, for which only eight instances were available in our dataset, is remarkable. The distinctive feature of these cases is lack of a nucleus (see Figure 1), which was effectively captured by the model. This implies that, in addition to class prevalence, also the actual form and shape of the diverse instances in a class exert influence in the performance of the classifier.

With the available training data, the resulting model for analysis 2 can be considered as the best trade-off between accuracy and class disaggregation. This model is valuable for distinguishing between general categories used in several studies of genotoxicity, such as “ENAs”, “micronucleus”, and normal erythrocytes [39,40]. However, some kinds of specific nuclear abnormalities have been demonstrated to be more frequent in particular

pollution environments and are useful as biomarkers. For instance, De Souza (2017) [4] conducted experimental studies in Australian parakeets (*Melopsittacus undulates*) finding higher frequencies of erythrocytes with displaced nuclei in individuals exposed to tannery effluents. In this sense, it is important to improve the performance of classification for more specific categories, mostly in the cases of erythrocyte abnormalities evidencing a complex pattern to be identified, such as the “displaced” and “micronucleus” categories.

5. Conclusions

We developed a deep-learning-based analysis tool that provides a faster, more efficient, and standardized approach to laboratory analysis for classifying nuclear abnormalities in avian erythrocytes. Our method reduces the reliance on human interpretation and enables the identification and classification of abnormalities independent of human supervision. This not only saves valuable time but also improves the precision and reliability of the assessments. In addition, the model is able to generalize effectively across different staining conditions, which is important for real-world applications where variations in staining protocols may occur.

Possible improvements were also analyzed, using alternative approaches such as breakdown levels of class categorization, tuning settings in training, and data augmentation, among others. Further studies will consider a wider perspective on nuclear features, with potential dual- or multiple-abnormality characterization, thus turning the context into a multilabel classification problem, which requires different CNN architectures. The model has some limitations, particularly with respect to classifying more specific and less prevalent categories due to the lack of an adequate amount of training examples, a fact that can be circumvented upon availability of larger datasets. The advances presented may also serve as a foundation for future deep learning research on similar problems related with classification of nucleated erythrocytes, and thus, enable more efficient and accurate assessment of genotoxicity in birds, as well as environmental and conservation issues.

Author Contributions: M.G.F.: conceptualization, methodology, investigation, writing—original draft. F.R.: conceptualization, methodology, investigation, writing. M.A.A.: methodology, writing, M.B.: writing, investigation, funding acquisition, V.L.D.: methodology, writing, investigation, funding acquisition, C.D.: conceptualization, methodology, investigation, writing, D.P.: conceptualization, methodology, investigation, writing. All authors have read and agreed to the published version of the manuscript.

Funding: The field and laboratory work were supported by the Fund for Scientific and Technological Research-National Agency for Scientific and Technological Promotion, FONCyT [PICT 2018-02178] awarded to Marcelo Bertelotti and Verónica L. D’Amico; Doctoral scholarship CONICET awarded to Facundo Roffet, Doctoral scholarship CONICET and Province of Chubut awarded to Miguel A. Adami; and postdoctoral scholarships CONICET awarded to Martín G. Frixione and Débora Pollicelli. This work conforms to national, local, and institutional laws and requirements (N° 15/2021-DFyFS-MAGIyC).

Data Availability Statement: The data are publicly available at: <https://github.com/ImageLabUNS/erythrocytes>.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

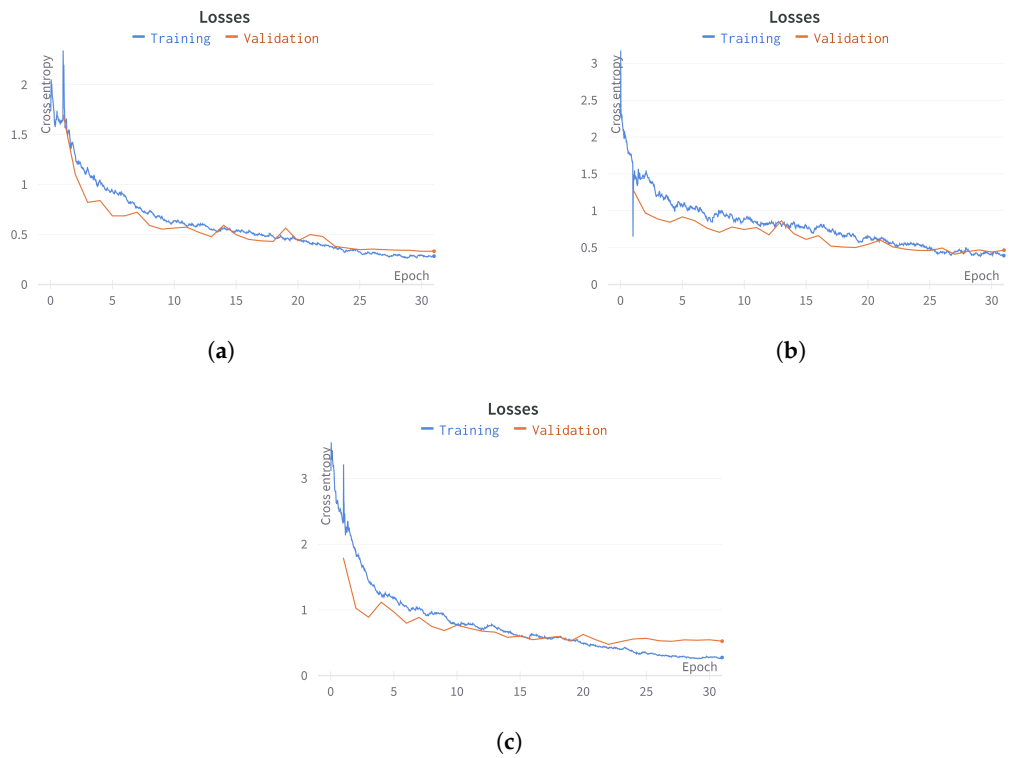


Figure A1. Evolution of loss in training and validation throughout epochs. (a) Analysis 1. (b) Analysis 2. (c) Analysis 3.

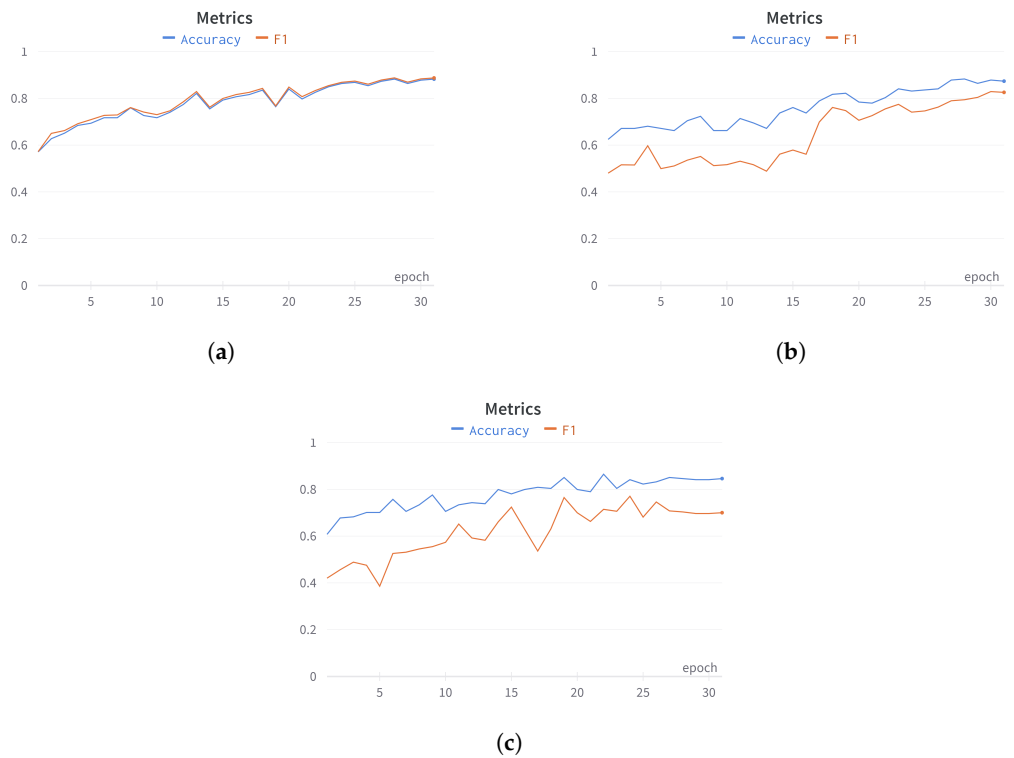

















Figure A2. Analysis of accuracy and F1-score metric trends in training and validation throughout epochs. (a) Analysis 1. (b) Analysis 2. (c) Analysis 3.

image	ground_truth	prediction	loss	↓	confidence
	abnormal	normal	3.852		0.9772
	abnormal	normal	3.137		0.9547
	abnormal	normal	2.976		0.9434
	abnormal	other	2.928		0.9449
	abnormal	normal	2.88		0.9429

(a)






image	ground_truth	prediction	loss	↓	confidence
	ENAs	micronucleus	10.55		1.
	other	normal	7.123		0.7385
	normal	micronucleus	5.356		0.9928
	micronucleus	ENAs	3.867		0.5315
	other	ENAs	3.577		0.9679

(b)






image	ground_truth	prediction	loss	↓	confidence
	other	normal	5.437		0.9784
	displaced	normal	4.714		0.744
	irregular	normal	3.901		0.9764
	irregular	normal	3.637		0.963
	irregular	normal	3.485		0.9671

(c)






Figure A3. The five most misclassified items according to the loss metric during training. (a) Analysis 1. (b) Analysis 2. (c) Analysis 3.

image	ground_truth	prediction	loss	↓	confidence
	other	abnormal	5.813		0.5451
	other	normal	5.396		0.5339
	abnormal	normal	4.072		0.9819
	abnormal	normal	3.059		0.9525
	abnormal	normal	2.751		0.8966

(a)

image	ground_truth	prediction	loss	↓	confidence
	other	micronucleus	10.138		0.9971
	other	ENAs	7.663		0.9994
	micronucleus	normal	4.325		0.8128
	ENAs	micronucleus	3.4		0.9184
	ENAs	normal	2.444		0.8994

(b)

image	ground_truth	prediction	loss	↓	confidence
	micronucleus	normal	6.503		0.8439
	micronucleus	normal	5.166		0.732
	displaced	normal	5.156		0.9295
	irregular	normal	4.339		0.981
	irregular	normal	4.262		0.9832

(c)

Figure A4. The five most misclassified items according to the loss metric during validation. (a) Analysis 1. (b) Analysis 2. (c) Analysis 3.

References

1. Baesse, C.Q.; Tolentino, V.C.d.M.; Silva, A.M.d.; Silva, A.d.A.; Ferreira, G.Â.; Paniago, L.P.M.; Nepomuceno, J.C.; Melo, C.d. Micronucleus as biomarker of genotoxicity in birds from Brazilian Cerrado. *Ecotoxicol. Environ. Saf.* **2015**, *115*, 223–228. [[CrossRef](#)] [[PubMed](#)]
2. Stocker, J.; Morel, A.P.; Wolfarth, M.; Dias, J.F.; Niekaszewicz, L.A.B.; Cademartori, C.V.; da Silva, F.R. Basal levels of inorganic elements, genetic damages, and hematological values in captive *Falco peregrinus*. *Genet. Mol. Biol.* **2022**, *45*, e20220067. [[CrossRef](#)] [[PubMed](#)]
3. Tomazelli, J.; Rodrigues, G.Z.P.; Franco, D.; de Souza, M.S.; Burghausen, J.H.; Panizzon, J.; Kayser, J.M.; Loiko, M.R.; Schneider, A.; Linden, R.; et al. Potential use of distinct biomarkers (trace metals, micronuclei, and nuclear abnormalities) in a heterogeneous sample of birds in southern Brazil. *Environ. Sci. Pollut. Res.* **2022**, *29*, 14791–14805. [[CrossRef](#)] [[PubMed](#)]

4. de Souza, J.M.; Montalvão, M.F.; da Silva, A.R.; de Lima Rodrigues, A.S.; Malafaia, G. A pioneering study on cytotoxicity in Australian parakeets (*Melopsittacus undulatus*) exposed to tannery effluent. *Chemosphere* **2017**, *175*, 521–533. [[CrossRef](#)] [[PubMed](#)]
5. Santos, C.S.; Brandão, R.; Monteiro, M.S.; Bastos, A.C.; Soares, A.M.; Loureiro, S. Assessment of DNA damage in *Ardea cinerea* and *Ciconia ciconia*: A 5-year study in Portuguese birds retrieved for rehabilitation. *Ecotoxicol. Environ. Saf.* **2017**, *136*, 104–110. [[CrossRef](#)] [[PubMed](#)]
6. Brandts, I.; Cánovas, M.; Tvarijonaviciute, A.; Llorca, M.; Vega, A.; Farré, M.; Pastor, J.; Roher, N.; Teles, M. Nanoplastics are bioaccumulated in fish liver and muscle and cause DNA damage after a chronic exposure. *Environ. Res.* **2022**, *212*, 113433. [[CrossRef](#)] [[PubMed](#)]
7. Frixione, M.G.; D’Amico, V.; Adami, M.A.; Bertellotti, M. Urbanity as a source of genotoxicity in the synanthropic Kelp Gull (*Larus dominicanus*). *Sci. Total Environ.* **2022**, *850*, 157958. [[CrossRef](#)] [[PubMed](#)]
8. Oudi, A.; Chokri, M.A.; Hammouda, A.; Chaabane, R.; Badraoui, R.; Besnard, A.; Santos, R. Physiological impacts of pollution exposure in seabird’s progeny nesting in a Mediterranean contaminated area. *Mar. Pollut. Bull.* **2019**, *142*, 196–205. [[CrossRef](#)] [[PubMed](#)]
9. Fenech, M. Cytokinesis-block micronucleus cytome assay. *Nat. Protoc.* **2007**, *2*, 1084–1104. [[CrossRef](#)]
10. Valko, M.; Izakovic, M.; Mazur, M.; Rhodes, C.J.; Telser, J. Role of oxygen radicals in DNA damage and cancer incidence. *Mol. Cell. Biochem.* **2004**, *266*, 37–56. [[CrossRef](#)]
11. Torlaschi, C.; Gandini, P.; Esteban, F.; Peck, R.M. Predicting the sex of kelp gulls by external measurements. *Waterbirds* **2000**, *23*, 518–520. [[CrossRef](#)]
12. de Almeida Santos, F.; Morgante, J.S.; Frere, E.; Millones, A.; Sander, M.; de Abreu Vianna, J.; Dantas, G.P.d.M. Evolutionary history of the Kelp Gull (*Larus dominicanus*) in the southern hemisphere supported by multilocus evidence. *J. Ornithol.* **2016**, *157*, 1103–1113. [[CrossRef](#)]
13. Bertellotti, M.; Yorio, P.; Blanco, G.; Giaccardi, M. Use of Tips By Nesting Kelp Gulls At a Growing Colony in Patagonia. *J. Field Ornithol.* **2001**, *72*, 338–348. [[CrossRef](#)]
14. Frixione, M.G.; Alarcón, P.A. Composicion de la dieta post-reproductiva de la gaviota cocinera (*Larus dominicanus*) en el lago Nahuel Huapi, Patagonia Argentina. *Ornitol. Neotrop.* **2016**, *27*, 217–221. [[CrossRef](#)]
15. Frixione M.G.; Lisnizer, N.; Yorio, P. Year-round use of anthropogenic food sources in human modified landscapes by adult and young Kelp Gulls. *Food Webs* **2023**, *35*, e00274. [[CrossRef](#)]
16. Bertram, C.A.; Aubreville, M.; Donovan, T.A.; Bartel, A.; Wilm, F.; Marzahl, C.; Assenmacher, C.A.; Becker, K.; Bennett, M.; Corner, S.; et al. Computer-assisted mitotic count using a deep learning-based algorithm improves interobserver reproducibility and accuracy. *Vet. Pathol.* **2022**, *59*, 211–226. [[CrossRef](#)] [[PubMed](#)]
17. Kittichai, V.; Kaewthamasorn, M.; Thanee, S.; Jomtarak, R.; Klanboot, K.; Naing, K.M.; Tongloy, T.; Chuwongin, S.; Boonsang, S. Classification for avian malaria parasite *Plasmodium gallinaceum* blood stages by using deep convolutional neural networks. *Sci. Rep.* **2021**, *11*, 1–10. [[CrossRef](#)] [[PubMed](#)]
18. Marzahl, C.; Aubreville, M.; Bertram, C.A.; Stayt, J.; Jasensky, A.K.; Bartenschlager, F.; Fragoso-Garcia, M.; Barton, A.K.; Elsemann, S.; Jabari, S.; et al. Deep Learning-Based Quantification of Pulmonary Hemosiderophages in Cytology Slides. *Sci. Rep.* **2020**, *10*, 9795. [[CrossRef](#)]
19. Vinicki, K.; Ferrari, P.; Belic, M.; Turk, R. Using Convolutional Neural Networks for Determining Reticulocyte Percentage in Cats *arXiv* **2018**, arXiv:1803.04873. <https://doi.org/10.48550/arXiv.1803.04873>.
20. Bengio, Y. *Learning Deep Architectures for AI*; Now Publishers Inc.: Boston, MA, USA, 2009; Volume 2, pp. 1–27. [[CrossRef](#)]
21. Greenspan, H.; Van Ginneken, B.; Summers, R.M. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Trans. Med. Imaging* **2016**, *35*, 1153–1159. [[CrossRef](#)]
22. Aubreville, M.; Bertram, C.A.; Marzahl, C.; Gurtner, C.; Dettwiler, M.; Schmidt, A.; Bartenschlager, F.; Merz, S.; Fragoso, M.; Kershaw, O.; et al. Deep learning algorithms out-perform veterinary pathologists in detecting the mitotically most active tumor region. *Sci. Rep.* **2020**, *10*, 16447. [[CrossRef](#)] [[PubMed](#)]
23. Blanco, G.; Rodríguez-Estrella, R.; Merino, S.; Bertellotti, M. Effects of spatial and host variables on hematozoa in white-crowned sparrows wintering in Baja California. *J. Wildl. Dis.* **2001**, *37*, 786–790. [[CrossRef](#)] [[PubMed](#)]
24. D’Amico, V.L.; Fazio, A.; Palacios, M.G.; Carabajal, E.; Bertellotti, M. Evaluation of Physiological Parameters of Kelp Gulls (*Larus dominicanus*) Feeding on Fishery Discards in Patagonia, Argentina. *Waterbirds* **2018**, *41*, 310–315. [[CrossRef](#)]
25. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
26. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and Flexible Image Augmentations. *Information* **2020**, *11*, 125. [[CrossRef](#)]
27. Howard, J.; Gugger, S. Fastai: A layered api for deep learning. *Information* **2020**, *11*, 108. [[CrossRef](#)]
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
29. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
30. Marcel, S.; Rodriguez, Y. Torchvision the Machine-Vision Package of Torch. In Proceedings of the 18th ACM International Conference on Multimedia (MM ’10), New York, NY, USA, 25–29 October 2010; pp. 1485–1488. [[CrossRef](#)]

31. Albertini, M.C.; Teodori, L.; Piatti, E.; Piacentini, M.P.; Accorsi, A.; Rocchi, M.B. Automated Analysis of Morphometric Parameters for Accurate Definition of Erythrocyte Cell Shape. *Cytom. Part A* **2003**, *52*, 12–18. [[CrossRef](#)]
32. Durant, T.J.; Olson, E.M.; Schulz, W.L.; Torres, R. Very deep convolutional neural networks for morphologic classification of erythrocytes. *Clin. Chem.* **2017**, *63*, 1847–1855. [[CrossRef](#)] [[PubMed](#)]
33. Egelé, A.; Stouten, K.; van der Heul-Nieuwenhuijsen, L.; de Bruin, L.; Teuns, R.; van Gelder, W.; Riedl, J. Classification of several morphological red blood cell abnormalities by DM96 digital imaging. *Int. J. Lab. Hematol.* **2016**, *38*, e98–e101. [[CrossRef](#)]
34. Campana, M.A.; Panzeri, A.M.; Moreno, V.J.; Dulout, F.N. Micronuclei induction in *Rana catesbeiana* tadpoles by the pyrethroid insecticide lambda-cyhalothrin. *Genet. Mol. Biol.* **2003**, *26*, 99–103. [[CrossRef](#)]
35. Cavalcante, D.G.; Martinez, C.B.; Sofia, S.H. Genotoxic effects of Roundup® on the fish *Prochilodus lineatus*. *Mutat. Res.-Genet. Toxicol. Environ. Mutagen.* **2008**, *655*, 41–46. [[CrossRef](#)] [[PubMed](#)]
36. Morita, T.; Hamada, S.; Masumura, K.; Wakata, A.; Maniwa, J.; Takasawa, H.; Yasunaga, K.; Hashizume, T.; Honma, M. Evaluation of the sensitivity and specificity of in vivo erythrocyte micronucleus and transgenic rodent gene mutation tests to detect rodent carcinogens. *Mutat. Res.-Genet. Toxicol. Environ. Mutagen.* **2016**, *802*, 1–29. [[CrossRef](#)] [[PubMed](#)]
37. Suárez-Rodríguez, M.; Montero-Montoya, R.D.; Garcia, C.M. Anthropogenic nest materials may increase breeding costs for urban birds. *Front. Ecol. Evol.* **2017**, *5*, 233573. [[CrossRef](#)]
38. Zapata, L.M.; Bock, B.C.; Orozco, L.Y.; Palacio, J.A. Application of the micronucleus test and comet assay in *Trachemys callirostris* erythrocytes as a model for in situ genotoxic monitoring. *Ecotoxicol. Environ. Saf.* **2016**, *127*, 108–116. [[CrossRef](#)] [[PubMed](#)]
39. Barbosa, A.; De Mas, E.; Benzal, J.; Diaz, J.I.; Motas, M.; Jerez, S.; Pertierra, L.; Benayas, J.; Justel, A.; Lauzurica, P.; et al. Pollution and physiological variability in gentoo penguins at two rookeries with different levels of human visitation. *Antarct. Sci.* **2013**, *25*, 329–338. [[CrossRef](#)]
40. De Mas, E.; Benzal, J.; Merino, S.; Valera, F.; Palacios, M.J.; Cuervo, J.J.; Barbosa, A. Erythrocytic abnormalities in three Antarctic penguin species along the Antarctic Peninsula: Biomonitoring of genomic damage. *Polar Biol.* **2015**, *38*, 1067–1074. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.