Research Article
Machine Learning-Based Individual Identification of Laboratory Mice

[1,2]Jun Ogasawara, [1]Nobuyoshi Matsumoto, [1]Tadashi Yokokawa, [2]Masato Yasui and [1,3]Yuji Ikegaya

[1]Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo 113-0033, Japan
[2]Department of Pharmacology, School of Medicine, Keio University, Tokyo, 160-8582, Japan
[3]Center for Information and Neural Networks, National Institute of Information and Communications Technology, Suita City, Osaka, 565-0871, Japan

First and second Author contributed equally to this work.

Abstract

Background and Objective: Individual identification of laboratory animals is essential in behavioral science. Conventional methods often involve invasive marking of the animals' bodies and may cause infectious diseases. The purpose of this study was to establish an accurate and non-invasive method for the individual identification of native laboratory mice. **Materials and Methods:** A total of 706 photographs of four mice were taken. Using the photographs, an open-source computing algorithm method was adopted to create a model that identified individual mice. **Results:** Using the high-resolution photographs and the open-source computing algorithm with deep learning, an accurate algorithm for individual identification of mice, which outperformed classical identification algorithms, was established. **Conclusion:** Compared with the other conventional methods, this model exhibited higher performance in individual identification of mice. It will provide a platform to automatically label individual mice for social behavioral experiments.

Key words: Machine learning, mouse, individual identification, object detection, behavior, precision, recall, F1 score

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Individual identification of laboratory rodents is crucial for behavioral neuroscience but is a difficult task for even highly experienced researchers because these rodents are usually inbred or closed colonized and thus are very similar in size and color. The current methods for identifying individual rodents are largely classified into two strategies: temporal marking and quasi-permanent marking. The temporal methods include (i) marking an ear or tail with an ink pen or spray, (ii) dyeing a patch of fur with food coloring, picric acid, etc. (only for albino or light-colored rodents) and (iii) shaving a patch of fur. The quasi-permanent methods include (iv) ear punches or tags[1,2], (v)tattoos on the ear, tail, or toe[3], (vi) subcutaneous implantation of wireless microchips[4,5] and (vii) radio frequency identification devices[6,7]. These methods work to some extent[8], but they may have practical disadvantages. For example, temporal methods cause changes in the body appearance of mice. Because mice perceive others at least in part by sight[9], changes in their appearance may impinge on social behavioral assays. Permanent methods are inevitably accompanied by damage to tissues or skin and may cause infectious diseases[2,10]. Thus, these conventional methods may affect physiological homeostasis and animal behavior *perse*. To tackle this issue, the purpose of this study was to establish an accurate and non-invasive method for the individual identification of native laboratory mice using an open-source machine learning algorithm.

MATERIALS AND METHODS

Animal ethics: Animal experiments were performed with the approval of the Animal Experiment Ethics Committee at The University of Tokyo (approval number: P29-11) and according to the University of Tokyo guidelines for the care and use of laboratory animals. These experimental protocols were carried out in accordance with the Fundamental Guidelines for Proper Conduct of Animal Experiment and Related Activities in Academic Research Institutions (Ministry of Education, Culture, Sports, Science and Technology, Notice No. 71 of 2006), the Standards for Breeding and Housing of and Pain Alleviation for Experimental Animals (Ministry of the Environment, Notice No. 88 of 2006) and the Guidelines on the Method of Animal Disposal (Prime Minister's Office, Notice No. 40 of 1995). This study was conducted in September 2019. All animals were housed under a 12-h dark-light cycle (light from 08:00 to 20:00) at $22\pm1^{\circ}$C with *ad libitum* food and water.

Animal preparation: Four postnatal 3-week-old male C57BL/6J mice (Japan SLC, Shizuoka, Japan) were separately bred in a single cage. The mice were named Hamlet, Othello, Lear and Macbeth.

Photography: After the mice were fully habituated to experimenters, RGB-colored photographs ($3,936\times2,624$ pixels, 24-bit intensity) of them were take using a commercially available digital camera (DSC-RX1R, Sony, Tokyo, Japan), which was set at a distance of approximately 50 cm above mice. A total of 706 photographs (i.e., 160, 218, 181 and 147 images for Hamlet, Othello, Lear and Macbeth, respectively) were obtained. Camera settings were fixed throughout the experiments; that is, the F-number, the ISO speed and the exposure time were 5.0, ISO-500 and 1/800 sec, respectively. A mouse was surrounded with whiteboards to increase the contrast between the animal and the background and LED panel lights were set above the whiteboards to increase photosensitivity. The luminous intensity at the positions of the mouse's eyes was kept under $500\times$ to reduce light-induced retinal damage.

Data preprocessing: Each photograph was labeled with a mouse ID. The mouse in each image was automatically enclosed in a rectangle using Open CV, an open-source computer vision library[11], implemented in the Python environment[12]. The automatically defined rectangle was slightly smaller than the actual body size of the mouse, so the size of the rectangle was expanded by 20% to ensure the rectangle covered the entire body of the mouse, including its whiskers and tail.

Machine learning-based algorithm: For the automatic individual identification of mice, the Google Cloud service Cloud AutoML Vision Object Detection was adopted, which can be trained to evaluate and deploy models based on various image datasets (https://cloud.google.com/vision/automl/object-detection/docs/). This service was designed to simultaneously classify and localize the mouse in each image by drawing the bounding boxes of the object (i.e., the mouse). To create a model for object classification and localization, 706 images with the correctly labeled, ground-truth bounding boxes (i.e., the coordinates of four vertices) were imported into a dataset in the cloud. The images were randomly divided into training, validation and test datasets at a ratio of 8:1:1; thus, there were ten distinct training datasets and thereby ten models corresponding to the test data sets were trained.

Each model was trained using a cloud computing engine. One node represented the computing power equivalent to a machine equipped with an NVIDIA Tesla V100 GPU. The training was automatically canceled when either no further performance improvement was expected or the computational resource reached a limitation of 40 node hours. After the training, the model outputted many bounding boxes for each image of a given test dataset. For all the bounding boxes of the test images, the confidence score and the intersection over the union (IoU) were calculated. The confidence score (%) ranges between 0 and 100 and indicates the extent to which a model is confident that a given bounding box contains a target object[13]. The IoU signifies the ratio of the overlapping area of the predefined ground-truth bounding box and the predicted bounding box to the area of the union of the two boxes[14]. All the bounding boxes were sorted in descending order of the confidence score and classified a given predicted bounding box as true positive or false positive depending on whether the actual IoU was higher or lower than the preset IoU threshold, respectively. The precision and the recall were calculated as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

where, TP, FP, FN represents the true positive, false positive, and false negative, respectively. The precision signified the ratio of the correctly predicted bounding boxes to all the predicted bounding boxes with confidence scores over the threshold. The recall represented the ratio of the correctly predicted bounding boxes to all the ground-truth bounding boxes. The model performance was evaluated using the precision and the recall while varying the confidence and IoU thresholds and calculated the area under the precision-recall curve as a function of the IoU threshold.

The model also returned the mouse ID for each test image depending on the confidence and IoU thresholds. Then created a 4×4 confusion matrix and calculated the micro-precision, macro-recall and macro-F1 score as the mean of the four precisions of the predicted classes, the mean of the four recalls of the actual classes and the harmonic mean of the micro-precision and macro-recall, respectively[15]. The same process was repeated for the other nine datasets. The macro-precision, macro-recall and macro-F1 scores were then calculated based on each of the ten test datasets to evaluate the performance of the model using randomly labeled training and validation datasets and truly labeled test datasets.

Other algorithms: To compare the performance of the currently used our method with the performances of other methods, identify the mouse ID based on three similarity indices, the correlation coefficient[16,17], color histograms[18,19] and feature point distance[20], for all possible image pairs. For the correlation coefficient-based algorithm, Pearson's product-moment correlation coefficient "r" was calculated for two grayscale digital images[16,17] and defined the value "r" as the similarity index. The value "r" ranges between -1 and 1 and takes values of 1, 0, or -1 if two images are completely identical, uncorrelated, or anti-correlated, respectively. Another similarity index was calculated using the color histograms of multicolored images because the correlation coefficient was defined for monochrome images and signified the degree of the linear relationship between two measured quantities[17]. Given a pair of color histograms H_1 and H_2, the intersection of the color histograms was defined as follows:

$$\frac{\sum_{i=1}^{n} \min(H_1(i), H_2(i))}{\sum_{i=1}^{n} H_1(i)}$$

where i signifies the color components; note that $n = 2^{24}$ if an image is 24-bit-colored. The denominator indicated the total number of pixels in each of the images[19]. This value was used as the similarity index. Moreover, A-KAZE local features were calculated to extract key points on two given images[21] and superimposed one image on the other. Then calculated the mean of all of the Euclidian distances of the pairs of corresponding feature points and defined this value as the similarity index. The value was zero when the two images were completely identical. For one image out of 706 images, correlation coefficients were calculated between the given image and the other images (i.e., 705) and averaged the value for each mouse ID. The mouse ID with the highest value was defined as the predicted ID of the image. This procedure was repeated for the other two indices and all the images; note that for the feature point distance, the mouse ID with the lowest value was defined as the predicted ID. Confusion matrices were calculated based on the four algorithms for the representative test dataset (Fig. 3).

Analytical software and statistics: The data were analyzed using custom-made MATLAB (MathWorks, Natick, MA, USA) or Python routines. The summarized data are reported as the mean± the standard deviation (SD). p<0.05 was considered statistically significant. When multiple pairwise comparisons were required, the original p-values were corrected with Bonferroni's correction and the corrected p-values were compared with 0.05.

RESULTS

Machine learning-based model for individual mouse identification: To develop a new automatic method to identify images of animals that can be barely distinguished by human eyes, a total of 706 still photographs of four male mice was taken using a portable consumer digital camera in a handheld style (Fig. 1). For each image, a mouse was enclosed by a rectangle (Fig. 2a, blue) slightly larger than the circumscribed rectangle that was defined automatically by OpenCV (Fig. 2a, green). The expanded rectangle was labeled with the mouse ID and was used as a ground-truth bounding box for machine learning.

Using a set of these annotated images, trained a custom-made deep learning model in the open-source cloud computing service Google Cloud AutoML Vision Object Detection. After training, the model detected the predicted box (Fig. 2b). Using the ground-truth box and the predicted box, the IoU was calculated by dividing the intersection of the two boxes by the union (Fig. 2c). The quality of the new model

was evaluated with precision and recall scores, which were plotted for various IoU thresholds as a function of the confidence score (Fig. 2d-e). The precision and the recall increased and decreased, respectively, as the confidence score threshold was increased, while they were both higher for lower IoU thresholds(Fig. 2d-e). As the recall was increased, the precision was decreased (Fig. 2f). Consistent with this, the area under the precision-recall curve (AuPRC) reached the maximum or minimum when the IoU threshold was below 0.47 or was beyond 0.93, respectively (Fig. 2g).

The model was then deployed to predict the mouse ID in each image. The automatic labeling for a total of 10 different test datasets was repeated. The macro-precision (Fig. 2h), macro-recall (Fig. 2i) and macro-F1 score (Fig. 2j) were calculated and all of them were significantly higher than those calculated from randomly annotated training datasets (Fig. 2h-j; macro-precision: $P = 2.5 \times 10^{-10}$, $t_9 = 30.0$; macro-recall: $P = 5.0 \times 10^{-10}$, $t_9 = 27.8$, macro-F1 score: $P = 3.1 \times 10^{-10}$, $t_9 = 29.3$, n = 10 test datasets, paired t-test).

Comparison of the performance with other methods: For comparison of the prediction performance, mouse IDs were predicted via using another classification strategy based on three different parameters, correlation coefficients, color histograms and feature point distances. For example, 71 still images of one dataset, including 16, 22, 18 and 15 images of Hamlet, Othello, Lear and Macbeth, were categorized into four classes to predict mouse IDs based on the machine learning-, correlation coefficient-, color histogram- and feature point

Fig. 1: Photographs of individual mice
Representative images of four mice. Each column contains photos of the identical mouse taken from different views
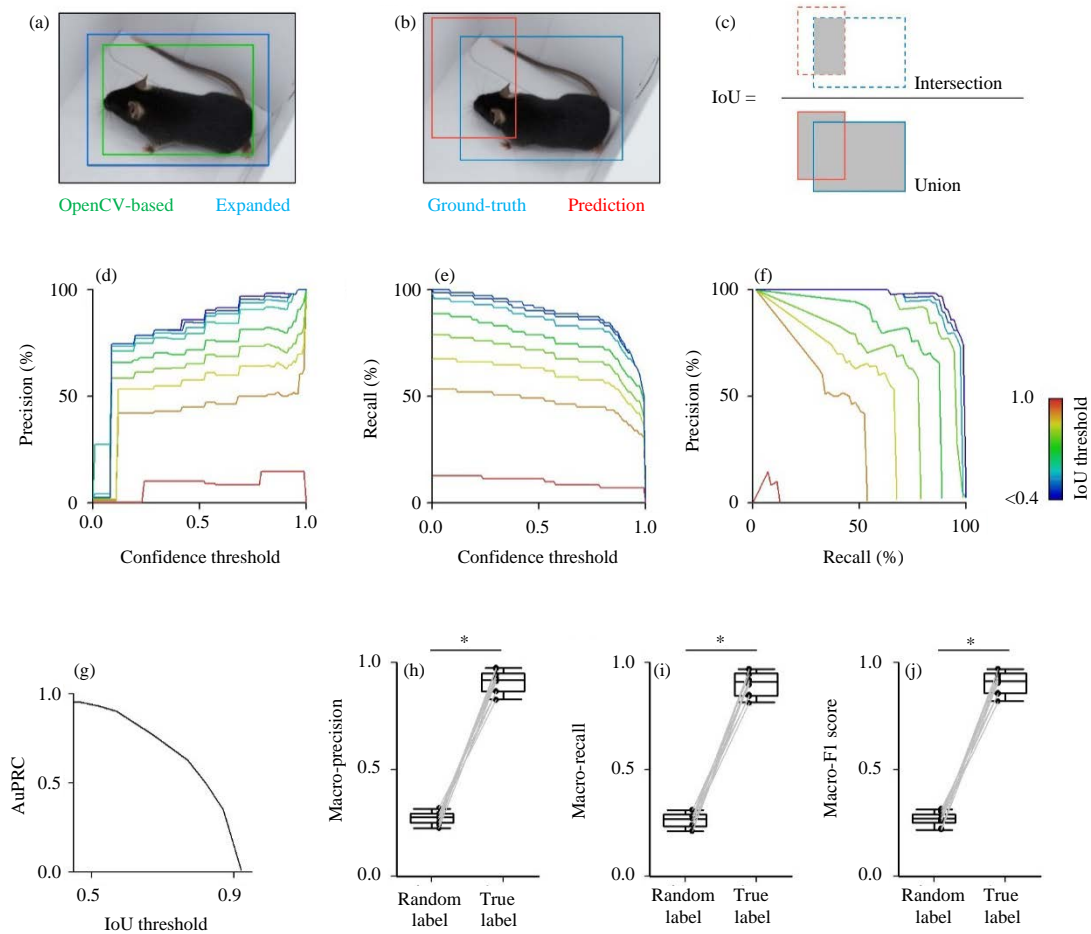
Fig. 2(a-j): Machine learning-based individual identification of four mice
The current model identified the mouse in each photograph, The performance of the model was characterized in terms of the macro-precision, the macro-recall and the macro-F1 score. *p<0.05, paired t-test

distance-based algorithms. In detail, out of 16 Hamlet images, 14 images were predicted to be Hamlet by the machine learning-based model, whereas only 7, 2 and 1 image were accurately predicted by a correlation coefficient-, color histogram- and feature point distance-based models, respectively (Fig. 3a). The same calculation was repeated for the other nine datasets and the micro-precision, macro-recall and macro-F1 scores were calculated for each dataset (i.e., 10 datasets in total) and each algorithm (i.e., 4 algorithms in total). The macro-precision of the deep learning-based model was significantly higher than that of the other methods (Fig. 3b; $P = 3.7 \times 10^{-6}$, $t_9 = 12.3$ for machine learning vs. correlation coefficient, $P = 8.1 \times 10^{-7}$, $t_9 = 14.7$ for machine learning vs. color histogram, $P = 2.0 \times 10^{-6}$, $t_9 = 13.2$ for machine learning vs. feature point distance; for the other

pairwise comparisons, p>0.05, n = 10 test datasets for each of 4 methods, paired t-test with Bonferroni's correction). Similarly, the macro-recall of the current model was significantly higher than that of the other models (Fig. 3c; $p = 1.3 \times 10^{-7}$, $t_9 = 18.0$ for machine learning vs correlation coefficient, $P = 1.4 \times 10^{-7}$, $t_9 = 18.0$ for machine learning vs. color histogram, $P = 6.8 \times 10^{-8}$, $t_9 = 20.0$ for machine learning vs feature point distance; for the other pairwise comparisons, p>0.05, n = 10 test datasets for each of 4 methods, paired t-test with Bonferroni's correction). The current model also exhibited significantly higher performance in the macro-F1 score (Fig. 3d; $P = 3.9 \times 10^{-7}$, $t_9 = 16.0$ for machine learning vs. correlation coefficient, $P = 1.6 \times 10^{-7}$, $t_9 = 17.7$ for machine learning vs. color histogram, $P = 8.7 \times 10^{-9}$, $t_9 = 24.6$ for machine learning vs. feature point distance; for the other
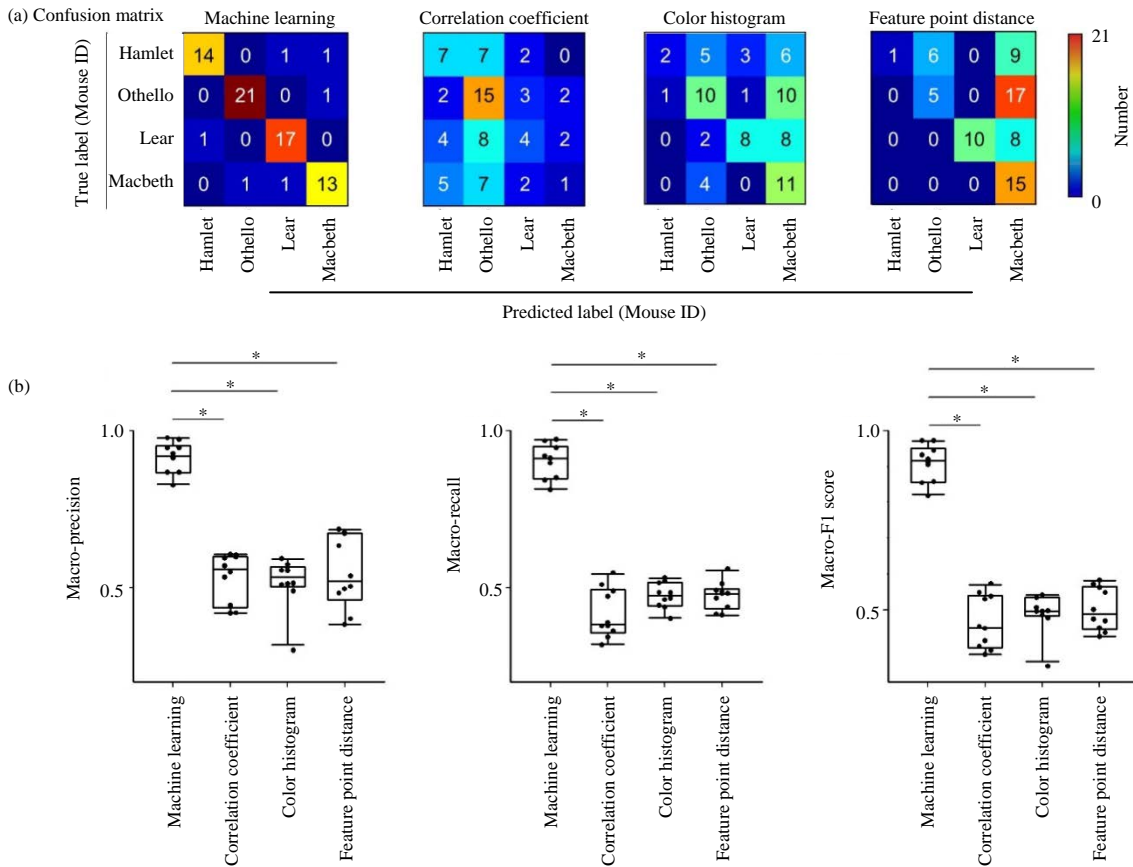
Fig. 3(a-b): Comparison of the individual identification performance between our algorithm and others
The performance of the current model was significantly higher than that of other classical methods.*p<0.05, paired t-test with Bonferroni's correction

pairwise comparisons, p>0.05, n = 10 test datasets for each of 4 methods, paired t-test with Bonferroni's correction). Thus, the current method enabled us to identify mice more accurately than classical methods.

DISCUSSION

In the present study, Google Cloud AutoML Vision Object Detection works as an accurate, non-invasive and markerless method for the individual identification of mice. Temporal or permanent markers are conventionally used for individual identification of laboratory animals, but these methods may cause behavioral or physiological problems. The performance of the current simple method was superior to the performances of the three classical methods tested here[16-20]. Compared with human faces[22], laboratory mice are almost indistinguishable, at least to human vision. A hypothesis was synthesized that the body appearances of these mice contain enough specific features to allow the discrimination of individuals. Because there was a significantly positive correlation between the model accuracy and the image resolution[23],in-focus and high-resolution images were acquired with caution. Consistent with this, in our preliminary study, we couldn't identify individual mice when out-of-focus and low-resolution images were used for machine learning training.

For computational object recognition, it is recommended that the object in the image is outlined by hand, but manual detection is time-consuming and can be inconsistent among images and among experimenters. Such inaccurate and inconsistent annotations may exacerbate the model performance[24]. To overcome this concern, an automatic conventional algorithm was adopted instead of manual annotation. Consequently, the mouse IDs in the still photographs were successfully identified. At present, the current model does not enable us to identify group-housed mice in time-lapse videos. However, in the near future, this model would be broadly applicable, even to social behavioral

studies, if the identification of group-housed mice was achieved. The current method is superior to the previous studies in that it identifies mice more accurately and non-invasively. This method is of significance in terms of the wide range of applicability for behavioral science.

CONCLUSION

Compared with the other conventional methods, our model exhibited higher performance in individual identification of mice, which was previously hard to achieve with either human eyes or classical computing algorithms. Our model will provide a platform to automatically label individual mice for social behavioral experiments.

SIGNIFICANCE STATEMENT

This study discovered that the machine learning-based model precisely identified apparently identical mice that are beneficial for all the researchers involved in social behavioral experiments. This study will help the researchers to uncover the critical areas of the sociality of animals that many researchers were not able to explore previously without invasive marking or careful monitoring. Thus, a new method of individual identification of animals may be arriving at the field of behavioral science.

ACKNOWLEDGMENT

REFERENCES

1. Castelhano-Carlos, M.J., N. Sousa, F. Ohl and V. Baumans, 2010. Identification methods in newborn C57BL/6 mice: a developmental and behavioural evaluation. Lab. Animals, 44: 88-103.
2. Kitagaki, M., T. Suwa, M. Yanagi and K. Shiratori, 2003. Auricular chondritis in young ear-tagged Crj:CD (SD) IGS rats. Lab. Animals, 37: 249-253.
3. Kasanen, I.H.E., H.M. Voipio, H. Leskinen, M. Luodonpaa and T.O. Nevalainen, 2011. Comparison of ear tattoo, ear notching and microtattoo in rats undergoing cardiovascular telemetry. Lab. Animals, 45: 154-159.
4. Elcock, L.E., B.P. Stuart, B.S. Wahle, H.E. Hoss and K. Crabb *et al.*, 2011. Tumors in long-term rat studies associated with microchip animal identification devices. Exp. Toxicol. Pathol., 52: 483-491.
5. Ball, D.J., G. Argentieri, R. Krause, M. Lipinski, R.L. Robison, R.E. Stoll, G.E. Visscher, 1991. Evaluation of a microchip implant system used for animal identification in rats. Lab. Animal Sci., 41: 185-186.
6. Howerton, C.L., J.P. Garner and J.A. Mench, 2012. A system utilizing radio frequency identification (RFID) technology to monitor individual rodent behavior in complex social settings. J. Neurosci. Methods, 209: 74-78.
7. De Chaumont, F., E. Ey, N. Torquet, T. Lagache and S. Dallongeville *et al.*, 2019. Real-time analysis of the behaviour of groups of mice via a depth-sensing camera and machine learning. Nat. Biomed. Eng., 3: 930-942.
8. Dahlborn, K., P. Bugnon, T. Nevalainen, M. Raspa, P. Verbost and E. Spangenberg, 2013. Report of the Federation of European Laboratory Animal Science Associations Working Group on animal identification. Lab. Animals, 47: 2-11.
9. Sakaguchi, T., S. Iwasaki, M. Okada, K. Okamoto and Y. Ikegaya, 2018. Ethanol facilitates socially evoked memory recall in mice by recruiting pain-sensitive anterior cingulate cortical neurons. Nat. Commun., 10.1038/s41467-018-05894-y
10. Cover, C.E., C.M. Keenan and G.E. Bettinger, 1989. Ear tag induced *Staphylococcus* infection in mice. Lab. Animals, 23: 229-233.
11. Pulli, K., A. Baksheev, K. Kornyakov and V. Eruhimov, 2012. Real-time computer vision with OpenCV. Commun. ACM, 55: 61-69.
12. Oliphant, T.E., 2007. Python for scientific computing. Comput. Sci. Eng., 9: 10-20.
13. Redmon, J., S. Divvala, R. Girshick and A. Farhadi, 2016. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, IEEE, Las Vegas, Nevada, USA., ISBN:978-1-4673-8852-8, pp: 779-788.
14. Rezatofighi H., N. Tsoi, J. Gwak, A. Sadeghian, I. Reid and S. Savarese, 2019. Generalized intersection over union: a metric and a loss for bounding box regression. arXiv.
15. Li, R., W. Liu, Y. Lin, H. Zhao and C. Zhang, 2017. An ensemble multilabel classification for disease risk prediction. J. Healthcare Eng., 10.1155/2017/8051673
16. Rodgers, J.L. and W.A. Nicewander, 2012. Thirteen ways to look at the correlation coefficient. Am. Statistician, 42: 59-66.
17. Kaur, A., L. Kaur and S. Gupta, 2012. Image recognition using coefficient of correlation and structural similarity index in uncontrolled environment. Int. J. Comp. Appl., 59: 32-39.
18. Stricker, M.A. and M. Orengo, 1995. Similarity of color images. Proceedings of the Storage and Retrieval for Image and Video

Databases, February 9, 1995, San Jose, CA., USA., pp: 381-392.

19. Swain, M.J. and D.H. Ballard, 1991. Color indexing. Int. J. Comput. Vision, 7: 11-32.

20. Zhou, J. and J. Shi. 2002. A robust algorithm for feature point matching. Comp. Graphics, 26: 429-436.

21. Alcantarilla, P., J. Nuevo and A. Bartoli, 2014. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In Proceedings British Machine Vision Conference 2013, 2014 BMVA Press, Pages: 13.1-13.11.

22. Taigman, Y., M. Yang, M.A. Ranzato and L. Wolf, 2014. Deepface: Closing the gap to human-level performance in face verification. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, IEEE, Columbus, Ohio, ISBN:978-1-4799-5118-5, pp: 1701-1708.

23. Schofield D., A. Nagrani, A. Zisserman, M. Hayashi, T. Matsuzawa, D. Biro and S. Carvalho, 2019. Chimpanzee face recognition from videos in the wild using deep learning. Sci. Adv., 10.1126/sciadv.aaw0736

24. Hao, z., L. Yao and Y. Wang, 2014. The influence of inconsistent data on cost-sensitive classification using prism algorithms: an empirical study. J. Comp. Sci., 9: 1880-1885.