Forensic Science International: Genetics 56 (2022) 102602

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Sperm hunting on optical microscope slides for forensic analysis with deep convolutional networks – a feasibility study

Raffael Golomingi, Cordula Haas *, Akos Dobay, Sören Kottner, Lars Ebert

Zurich Institute of Forensic Medicine, University of Zurich, Switzerland

ARTICLE INFO

Keywords:
Machine learning
Convolutional neural networks
Artificial intelligence
Forensics
Microscopy
Sperm detection

ABSTRACT

Microscopic sperm detection is an important task in sexual assault cases. In some instances, the samples contain no or only low amounts of semen. Therefore, the biological material is transferred onto a glass slide and needs to be manually scanned using an optical microscope. This work can be very time consuming, especially when no spermatozoa is present. In such a case, the result needs to be validated. In this article we show how convolutional neural networks can perform this task and how they can reduce the scanning time by locating the sperm cells on images taken under the microscope. For this purpose, we trained a VGG19 network and a VGG19 variation with 1942 images, some containing sperm cells and some not.

1. Introduction

Body fluid identification constitutes one of the tasks that are carried out by forensic practitioners to help criminal investigations. Traditional methods for body fluid identification involve primarily enzymatic and immunological tests [1]. Especially in the case of sexual offences, a common question is, whether semen can be detected in a crime stain. Sperm traces can be recovered e.g. from the vagina, vulva, cervix, anus and skin of the victim or from objects or clothing. The most reliable and widely accepted technique for the detection of semen is the microscopic identification of sperm cells [2,3]. Usually the sperm-loaded swab is smeared onto a microscope slide, which is stained to make the sperm cells visible. Several staining methods are used, e.g. Christmas tree, hematoxylin and eosin, Baecchi's, Papanicolaou's and Wright's [1]. Other staining methods use a fluorescently labeled human sperm–specific antibody for the microscopic detection of human spermatozoa, e.g. SPERM HY-LITER™ [4–6], with the disadvantage that the fluorescent staining fades away and cannot be preserved. A drawback of the microscope technique is, if the semen donor is azoospermic or underwent a vasectomy, no sperm cells are detectable even if semen is present. In such cases other tests are available that detect seminal fluid proteins, e.g. enzymatic tests for seminal acid phosphatase (SAP) [7] and immunologic tests for prostate-specific antigen (PSA) [8] or semenogelin [9], albeit some of these lack specificity or sensitivity [1]. In recent years new techniques for the identification of body fluids emerged, e.g. RNA profiling [10], microbial 16S-rRNA sequencing [11], and DNA methylation analysis [12]. These techniques need a high input of starting material and/or are laborious and expensive. Therefore, microscopy is still a cheap and reliable method to detect the presence of sperm cells. In our laboratory we use Baecchi's staining, where acid fuchsine colors the sperm heads in pink and methyl blue colors the sperm tails in blue [13]. Documenting samples containing a large number of sperm cells is relatively straight-forward. However, in many sexual assaults, it is not clear whether a sexual act has taken place and an ejaculation occurred. In addition, no or only a low number of sperm cells might be present. In the case where only a low number of sperm cells are present, the microscope slide has to be scanned manually in a meandering manner, which is a very time consuming and tiresome task, and can lead to errors. Sometimes, only one or two sperm cells are present and it is crucial for the investigation not to miss them.

Newest technologies would allow to facilitate this work, e.g. by automated scanning of the microscope slide and afterwards image processing. Vandewoestyne et al. describe an automated screening method to detect spermatozoa stained with Sperm HYLITER™ [14]. They used the PALM MicroBeam system (P.A.L.M. Microlaser Technologies) for scanning of microscope slides and an optimized image analyzing software module (AxioVision Commander Script, Carl Zeiss) for automated specimen identification and image processing. Detected spermatozoa were then isolated using laser capture microdissection and robust DNA profiles could be obtained from as little as 30 spermatozoa.

The task to detect sperm cells or to confirm that no sperm cells are present can be formulated as a classification problem. The images can be

* Correspondence to: University of Zurich, Winterthurerstrasse 190/52, CH-8057 Zurich, Switzerland.
E-mail address: cordula.haas@irm.uzh.ch (C. Haas).

https://doi.org/10.1016/j.fsigen.2021.102602
Received 29 June 2021; Received in revised form 26 September 2021; Accepted 4 October 2021
Available online 9 October 2021
1872-4973/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

divided into two classes, one class of images containing "sperm cells" and the other one with "no sperm cells". In recent years convolutional neural networks (CNN) proved to be a good choice for image classification. In 2012, a CNN won the ImageNet challenge for the first time [15]. Since then, CNN became the dominant technology in the field of image classification and pattern recognition. One major advantage of CNN's is that the algorithm learns the features of an object by itself and no additional annotation is required. This makes them suitable for our purpose since the sperm cells can be masked by other objects, the coloring might be weak or they might be deformed. The strength to generalize is what makes CNN superior over other models for image classification. Rajpurkar et al. [16] demonstrated that CNN's are able to operate with the same accuracy as expert eyes in classification of pneumonia on chest X-Rays. Similarly, Esteva et al. [17] could show that the classification of skin cancer using deep neural networks (DNN) works with an accuracy comparable to 21 board certified dermatologists. Recently, there have been some studies involving deep learning with sperm cells in a clinical framework. For example Mohammad et al. successfully trained RetinaNet [18] and a modified version of the CSR-DCF algorithm to track sperm cells on gray scale video clips taken from phase-contrast microscopes to increase the accuracy of sperm detection [19]. Riordon et al. trained a deep neural network to classify sperm cells according to one of the World Health Organization's shape criteria. They used two data sets, one where the sperm cells were stained, the other consisted of gray scale images [20]. This study showed that CNN's can compete with previous machine learning methods. However, we did not find any previous research of sperm detection with CNN's in a forensic setting.

In this study, we trained and tested a CNN for the automatic detection of sperm cells on photos from microscope slides colored with Baecchi's staining. We suggest, that this technique could be used to automate the task of analyzing microscope slides, thus reducing time, costs and eliminating human error.

2. Materials and methods

We evaluated two different approaches using a CNN. The first one only detects the presence or absence of a sperm cell on a given image using the VGG19 CNN. The VGG network was chosen because of its simple architecture and good performance in classification tasks [21]. The second method (VGG19-CAM) is based on VGG19 but has an altered architecture to also create class activation maps (CAM) which can be used to highlight the region where the sperm was detected.

2.1. Data

The microscope slides from genuine sexual assault cases were manually scanned with a BX41 or BX43 optical microscope from Olympus (Wallisellen, Switzerland) using the 100x oil objective. The images were taken with an UC30 digital camera from Olympus. The data set for this study consisted of 1942 microscopy images, that were randomly and retrospectively selected from our database and manually divided into the "sperm" and "no sperm" class by an experienced technician. All the images were anonymized, meaning that the case ID was erased. A total of 923 images contained sperm cells and 1019 were without sperm cells. The images showed smears recovered from the vagina, vulva, cervix, anus, skin or from objects or clothing. The original images had a resolution of 2080 × 1544 pixels. For the training with VGG19 and VGG19-CAM the images were downscaled to 345 × 256 pixels and then center-cropped to have a final dimension of 256 × 256 pixels. Augmentation techniques were used to get more training data and prevent potential overfitting of the network. We used random horizontal and vertical flipping for data augmentation. As can be seen in Fig. 1, not all pictures have the same hue or saturation and contain other objects than sperm cells. This is due to the variation introduced by the sample collection procedure, the applied pressure to smear the swab on the microscope slide, and the manual staining procedure. Because this study was conducted using anonymized image data, the Cantonal Ethics Committee issued a declaration of no objection (KEK-ZH No. 2015–0686).

2.2. Network architectures

The architecture of the two CNNs is shown in Supplementary Figure 1. The base network for training was the 19 layers deep convolutional neural network architecture of the Visual Geometry Group (VGG19) where the last layer was altered with a sigmoid function.

Because in forensic casework a second expert has to approve the findings, it is helpful to highlight and flag regions with potential sperm cells. To achieve this the network was altered to match the architecture for class activation mapping (VGG19-CAM), meaning that the last pooling layer was replaced with an average pooling layer and the fully connected layers with the exception of the last one have been removed [22]. This alteration is used to visualize which features in an image were relevant for the prediction. With the use of this information a heat map can be generated and superimposed on the microscopy image or, as it is

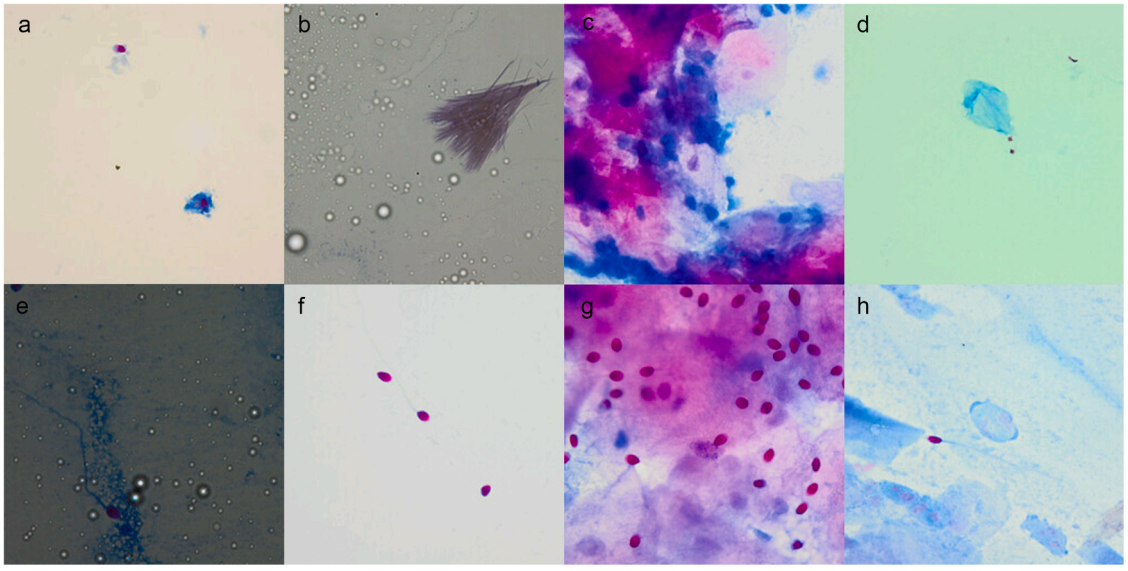


Fig. 1. Exemplary microscopic images contained in the data set. Pictures a-d contain no sperm cells and pictures e-h contain many or few sperm cells.

in our case, a bounding box around the most active regions as can be seen in Fig. 2.

2.3. Implementation

Both CNN's were implemented in Python with the use of the Keras network API [23] and compared with the use of the scikit-learn library [24]. Keras already provides a VGG19 implementation which we altered to match the image dimensions and the VGG19-CAM implementation.

2.4. Training

Twenty percent of the data were randomly selected and retained for later testing. The other 80% were used for training with a validation split of 20%. Both networks were trained for about 500 epochs using a binary cross-entropy optimizer and stochastic gradient descent (SDG) as loss function. Both networks were trained from scratch and without pre-trained weights. The reason was, that the pre-trained version was trained using the ImageNet dataset, where no histological images were present. The training was done on a Nvidia GeForce RTX 2080 Ti graphic processing unit (GPU) with 11 gigabyte of random access memory (GDDR6) [25].

2.5. Bounding box generation

The bounding boxes for the VGG19-CAM network were generated by passing through the image, calculating the CAM and then applying a threshold to extract the most active region. Around this region, a bounding box was generated. Some examplary results can be seen in Fig. 2.

2.6. Testing

In order to see how the CNN's will perform on new data, they were tested with the test set which was retained from the training and fine tuning process. This was done to test how well the network generalized from the training data to unknown data. The accuracy was measured by dividing the sum of the true positive (TP) and true negative (TN) predictions by the sum of the true positive, true negative, false positive (FP) and false negative (FN) predictions as seen in Eq. 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

To see what fraction of all positive classified instances were actually correctly predicted, one can calculate the precision according to Eq. 2. The recall (Eq. 3) tells what fraction of all real positive instances are recognized by the model.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

However, there is a tradeoff between precision and recall, since increasing one value leads to a reduction of the other. A precision recall curve can be created to better visualize this tradeoff and help to choose an appropriate decision threshold. The F_1 score (Eq. 4) is the harmonic mean between precision and recall [26]. To get a high score both the precision and recall results need to be high.

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

To visualize where the networks made more classification errors, a confusion matrix for both was generated. The confusion matrix tells which class generated most problems, when making a prediction [27].

3. Results

After testing both networks on the independent test set the following results could be achieved, that are represented in Tables 1, 2 and Fig. 3.

Table 1
This table displays the metrics of the VGG19 and VGG19-CAM (CAM) networks which resulted from the 387 test images. Support denotes the number of images for each class.

	Precision	Recall	F1-Score	Support
	VGG19 / CAM	VGG19 / CAM	VGG19 / CAM	
No Sperm	0.90 / 0.93	0.98 / 0.98	0.94 / 0.95	201
Sperm	0.97 / 0.97	0.88 / 0.92	0.92 / 0.95	186
Accuracy			0.93 / 0.95	387

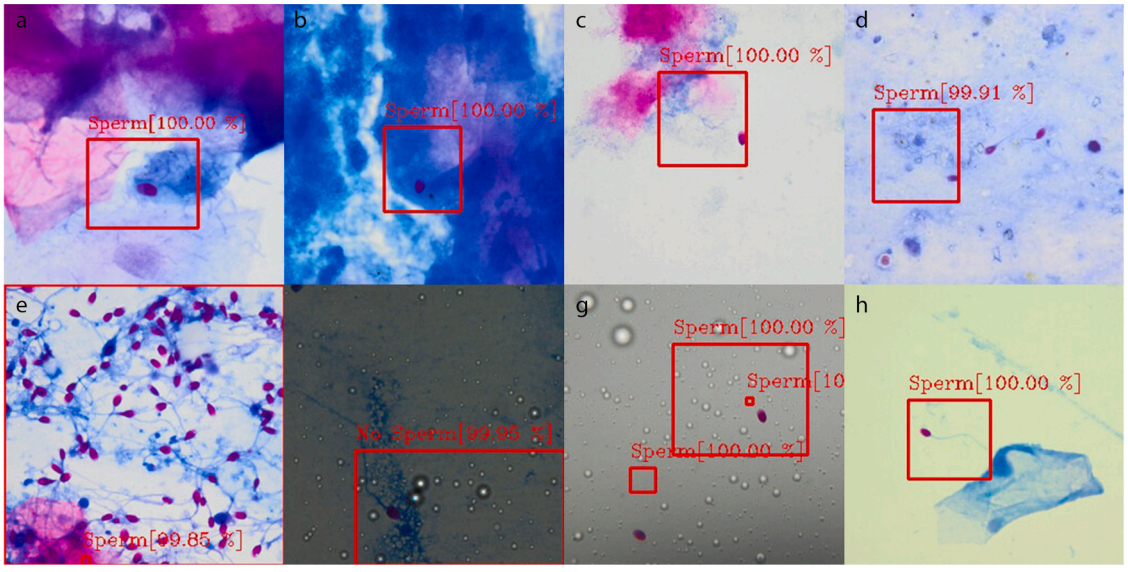


Fig. 2. These pictures show how the created bounding boxes of the VGG19-CAM network look like. The bounding boxes are not always accurate since they are generated from the regions the network is the most active (see pictures c, d, e, g).

Table 2
Confusion matrices of the VGG19 and VGG19-CAM networks. fp denotes the false positives, fn the false negatives.

	Predicted VGG19		Predicted VGG19-CAM	
	No Sperm	Sperm	No Sperm	Sperm
No Sperm	196	5 (fp)	196	5 (fp)
Sperm	22 (fn)	164	14 (fn)	172

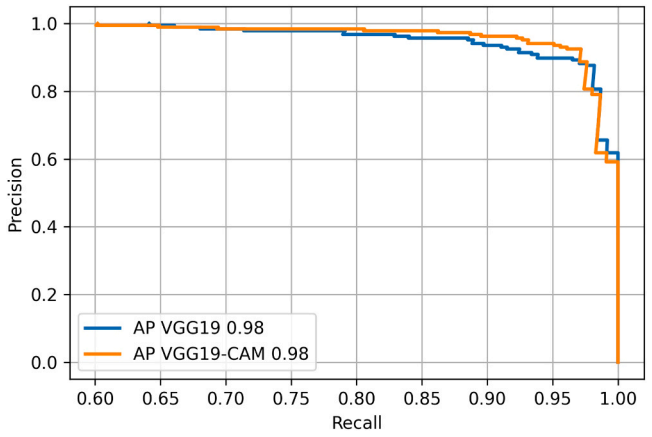


Fig. 3. The precision recall curves for the VGG19 and the VGG19-CAM network. AP = average precision.

3.1. VGG19 results

The VGG19 network achieved an overall accuracy of 93% on the test data (Table 1). The F_1 score for the detection of a sperm cell was 92% (Table 1). The network had a high precision for the cost of a lower recall rate.

The confusion matrix in Table 2 shows that the network misclassified 5 of the 201 sperm negative test images as sperm cells and confused 22 of the 186 sperm positive images as negative.

Fig. 3 shows the precision recall curve. The network achieved an average precision (AP) of 0.98. The precision of the network decreases sharply after 95% recall.

3.2. VGG19-CAM results

VGG19-CAM network achieved an overall accuracy of 95% on the test data (Table 1). The metrics showed a F_1 score of 95% in detecting sperm cells (Table 1). For the sperm class the network had a precision of 97% and a recall rate of 92%.

The confusion matrix shows that it classified 14 of the 186 sperm images as not containing sperm cells and 5 of the 201 sperm negative images as being of the sperm class (Table 2).

Fig. 3 shows the precision recall curve with an average precision of 0.98 as well. Again the precision decreases after 95% of recall, however, the model seems to be able to keep a higher overall precision until this point compared to the VGG19 network.

3.3. False negatives and false positives

From the total of 24 false negatives, 12 were misidentified by both CNNs, 10 only by VGG19, 2 only by VGG19-CAM. Most of the 24 images were indeed difficult to classify, due to unusual color, very few sperm cells or overlapping debris. However, 6 of the images contained apparent sperm cells, most of them misclassified by VGG19 only.

From the total of 8 false positives, 2 were misidentified by both CNNs, 3 only by VGG19 and another 3 only by VGG19-CAM. Six of the images were difficult to classify, 2 images (one misclassified by VGG19, the other by VGG19-CAM) would be clear for a trained expert.

Overall, we do not see a systematic problem behind these misclassifications, and we cannot really explain the failure of the CNNs.

4. Discussion

In this article we tested the feasibility of CNNs for detecting sperm cells on images taken from microscope slides. The accuracy and the F_1 score for both networks are higher than 90%. With this accuracy we have achieved a good baseline for a reliable automated sperm detection system. By dividing the images into training, validation and test images, we could establish that the accuracy of the network will be stable when presented with new images.

The confusion matrices (*2*) show a weakness of these CNNs. VGG19 misclassified 22 sperm images to contain no sperm cells (false negatives). In contrast the VGG19- CAM made slightly less errors and only classified 14 of the 186 sperm images as having no sperm cells. From a forensic perspective the false positives are less of an issue, because they can easily be checked in VGG19-CAM with help of the bounding boxes. However, the false negatives are an issue, because they will be missed.

The networks use a sigmoid function to make a prediction on how certain they are if an image contains a sperm cell. The values of the prediction lie between zero and one and with the use of a threshold the labels "no sperm" or "sperm" are applied. In this case, a value ≥ 0.5 means the image contains a sperm cell. This threshold can be adjusted so that more images would fall into the class "sperm cells". This would increase the recall of the networks and lower the risk of missing sperm cells (false negatives). However, there is a trade-off between precision and recall, which is visualized in Fig. 3. Decreasing the precision would increase the recall, meaning that more false positives would occur. In a forensic context missing a sperm cell is more fateful than having an object misclassified as a sperm cell, because the latter can be verified by an expert. Therefore, when used in an automated process where an expert verifies detected sperm cells, it would be acceptable to increase the recall and taking into account more false positives, rather than missing a sperm cell. We believe that this still would lower the assessment time without decreasing the quality. Fig. 3 shows that with a recall of 95% both networks would still be able to correctly classify an object as sperm in over 80% of the cases.

The use of the VGG19-CAM network to generate bounding boxes as additional aid for the expert to quickly see which object has been identified as sperm can work but has some shortcomings. Because the bounding box is generated for the most active region of the network, they are not always perfectly placed as can be seen in Figs. 2c, 2d, 2e and 2g. Lowering or increasing the threshold would result in either the creation of bounding boxes with no sperm in it like in Fig. 2g or having a bounding box which is too big and thus not informative like in Fig. 2e. If a reliable object detection is needed, other network architectures are required and also annotation work on the images have to be conducted.

The future goal would be to combine convolutional neural networks with automated microscopy. That requires a microscope with a fully automated scanning system and either image or video recording. Instead of using a class activation map for object detection, other network typologies like "You only look once" (YOLO) [28] or the above mentioned RetinaNet, which are specifically created for this task, could be trained. This would require to define bounding boxes for the sperm cells. Furthermore, the whole pipeline of taking the image from the slide, resizing and feeding it through the network to generate a report for the expert has to be evaluated. Because the error rate of a human for doing this task is not known in real-life scenario, our system should be tested against trained experts.

5. Conclusions

In this article we showed how convolutional neural networks can be used for automatic detection of sperm cells on images taken with an

optical microscope. We trained and tested two networks based on the VGG19 architecture and achieved an accuracy of over 90%. With the use of class activation mapping an additional simple visual guide can be provided to an expert to verify the classification. However, an expert is required to make a final verification on the decision of the model.

Acknowledgements

We thank Annette Brunner for sorting out the 1942 images and both Annette Brunner and Cristina Stadler for the discussions on practical issues regarding sperm detection. We thank the Emma Louise Kessler Foundation for funding our AI projects.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2021.102602.

References

[1] K. Virkler, I.K. Lednev, Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene, Forensic Sci. Int. 188 (1) (2009) 1–17, https://doi.org/10.1016/j.forsciint.2009.02.013. URL ⟨http://www.sciencedirect.com/science/article/pii/S0379073809000826⟩.

[2] B.S. Astrup, J.L. Thomsen, J. Lauritsen, P. Ravn, Detection of spermatozoa following consensual sexual intercourse, Forensic Sci. Int. 221 (1) (2012) 137–141, https://doi.org/10.1016/j.forsciint.2012.04. 024. URL ⟨https://www.sciencedirect.com/science/article/pii/ S0379073812001958⟩.

[3] T. Magalhães, R.J. Dinis-Oliveira, B. Silva, F. Corte-Real, D. Nuno Vieira, Biological evidence management for dna analysis in cases of sexual assault, Sci. World J. 2015 (2015) 1–11, https://doi.org/10.1155/2015/365674. URL ⟨https://doi.org/10.1155/2015/365674⟩.

[4] K.W.P. Miller, J. Old, B.R. Fischer, B. Schweers, S. Stipinaite, K. Reich, Developmental validation of the sperm hy-litertm kit for the identification of human spermatozoa in forensic samples*,†, J. Forensic Sci. 56 (4) (2011) 853–865, https://doi.org/10.1111/j.1556-4029.2011.01796.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j. 1556-4029.2011.01796.x.

[5] C. Westring, M. Wiuf, S. Nielsen, J. Fogleman, J. Old, C. Lenz, K. Reich, N. Morling, Sperm hy-literTM for the identification of spermatozoa from sexual assault evidence, in: Forensic Sci. Intern.: Genet., 12, 2014, pp. 161–167, https://doi.org/10.1016/j.fsigen. 2014.06.003.

[6] A. Takamura, K. Watanabe, T. Akutsu, Development of a quantitative validation method for forensic investigation of human spermatozoa using a commercial fluorescence staining kit (sperm hy-literTM express), Int. J. Leg. Med. 130 (6) (2016) 1421–1429, https://doi.org/10.1007/s00414-016-1381-9 (URL), ⟨https://doi.org/10.1007/s00414-016-1381-9⟩.

[7] P.L. Kirk, ss Kind, in: a s curry (Ed.), The Acid Phosphatase-test In Methods Of Forensic Science, (vol. 3), 11, john wiley & sons, new york and interscience publishers, 1965, p. 267.

[8] M.N. Hochmeister, B. Budowle, O. Rudin, C. Gehrig, U.V. Borer, M. Thali, R. Dirnhofer, Evaluation of prostate-specific antigen (psa) membrane test assays for the forensic identification of seminal fluid, J. Forensic Sci. 44 (5) (1999) 1057–1060.

[9] B. Pang, B. Cheung, Identification of human semenogelin in mem- brane strip test as an alternative method for the detection of semen, Forensic Sci. Int. 169 (1) (2007) 27–31, https://doi.org/10.1016/j.forsciint.2006.07.021 (doi:https: //doi.org/), ⟨http://www.sciencedirect.com/science/article/pii/ S037907380600524X⟩.

[10] C. Haas, E. Hanson, J. Ballantyne, Mrna and microrna for body fluid identification, in: J.A. Siegel, P.J. Saukko, M.M. Houck (Eds.), En- cyclopedia of Forensic Sciences, (Second edition), second edition edi-,, Academic Press, Waltham, 2013, pp. 402–408, https://doi.org/10.1016/B978-0-12-382165-2.00069-6. URL http://www.sciencedirect.com/science/article/pii/ B9780123821652000696.

[11] J.L. Metcalf, Z.Z. Xu, A. Bouslimani, P. Dorrestein, D.O. Carter, R. Knight, Microbiome tools for forensic science, Trends Biotechnol. - ogy 35 (9) (2017) 814–823, special Issue: Environmental Biotechnology. doi:⟨https://doi.org/10.1016/j.tibtech.2017.03.006⟩.

[12] A. Vidaki, M. Kayser, Recent progress, methods and perspectives in forensic epigenetics, Forensic Sci. Int.: Genet. 37 (2018) 180–195, https://doi.org/10.1016/j.fsigen.2018.08.008 (URL), ⟨http://www.sciencedirect.com/science/article/pii/ S1872497318304204⟩.

[13] B. Baecchi, Neue methode zum nachweis der spermatozoen in zeugflecken, DMW - Dtsch. Med. Wochenschr. 35 (25) (1909) 1105–1106, https://doi.org/10.1055/s-0029-1201548. URL https://doi.org/10.1055%2Fs-0029-1201548.

[14] M. Vandewoestyne, D. Van Hoofstat, F. Van Nieuwerburgh, D. Deforce, Au- tomatic detection of spermatozoa for laser capture microdissection, Interna- tional J. Leg. Med. 123 (2) (2008) 169–175, https://doi.org/10.1007/s00414-008-0271-1. ⟨https://doi.org/10.1007/s00414-008-0271-1⟩.

[15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. (IJCV) 115 (3) (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.

[16] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M.P. Lungren, A.Y. Ng, CheXNet: Radiol. - Lev. Pneumonia Detect. Chest X-Rays Deep Learn. (2017) arXiv:1711.05225v3.

[17] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118, https://doi.org/10.1038/nature21056 (URL), ⟨https://doi.org/10.1038/nature21056⟩.

[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss dense Object Detect. (2018) arXiv:1708.02002.

[19] M. reza Mohammadi, M. Rahimzadeh, A. Attar, Sperm Detect. Track. phase-Contrast Microsc. Image Seq. Using Deep Learn. Modif. csr-dcf (2020) arXiv: 2002.04034.

[20] J. Riordon, C. McCallum, D. Sinton, Deep learning for the classification of human sperm, Comput. Biol. Med. 111 (2019), 103342, https://doi.org/10.1016/j.compbiomed.2019.103342 (URL), ⟨https://www.sciencedirect.com/science/article/pii/ S0010482519302112⟩.

[21] K. Simonyan, A. Zisserman, Very Deep convolutional Netw. Large-Scale Image Recognit. (2014) arXiv:1409.1556.

[22] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep fea- tures for discriminative localization, Comput. Vis. Pattern Recogni- tion (2016).

[23] F. Chollet, et al., Keras, ⟨https://keras.io⟩ (2015).

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[25] Nvidia, NVIDIA GeForce RTX 2080 Ti-Grafikkarte. URL ⟨https://www.nvidia.com/de-de/geforce/graphics-cards/ rtx-2080-ti/⟩.

[26] C. Sammut, G.I. Webb (Eds.), F1-Measure, Springer US, Boston, MA, 2010, https://doi.org/10.1007/978-0-387-30164-8_298. URL https://doi.org/10.1007/978-0-387-30164-8_298.

[27] K.M. Ting, Confusion Matrix, Springer US, Boston, MA, 2010, https://doi.org/10.1007/978-0-387-30164-8_157. URL https://doi.org/10.1007/978-0-387-30164-8_157.

[28] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, arXiv (2018).