

Research proposal: Explainability methods for machine learning systems for multimodal medical datasets

Andrea M. Storås*
andrea@simula.no
SimulaMet
Oslo, Norway

Inga Strümke†
SimulaMet
Oslo, Norway

Michael A. Riegler‡
SimulaMet
Oslo, Norway

Pål Halvorsen*
SimulaMet
Oslo, Norway

ABSTRACT

This paper contains the research proposal of Andrea M. Storås that was presented at the MMSys 2022 doctoral symposium. Machine learning models have the ability to solve medical tasks with a high level of performance, e.g., classifying medical videos and detecting anomalies using different sources of data. However, many of these models are highly complex and difficult to understand. Lack of interpretability can limit the use of machine learning systems in the medical domain. Explainable artificial intelligence provides explanations regarding the models and their predictions. In this PhD project, we develop machine learning models for automatic analysis of medical data and explain the results using established techniques from the field of explainable artificial intelligence. Current research indicate that there are still open issues to be solved in order for end users to understand multimedia systems powered by machine learning. Consequently, new explanation techniques will also be developed. Different types of medical data are applied in order to investigate the generalizability of the methods.

CCS CONCEPTS

• **Applied computing** → **Health care information systems**; • **Human-centered computing**; • **Computing methodologies** → **Machine learning**; *Computer vision*; • **Information systems** → *Multimedia information systems*;

KEYWORDS

explainable artificial intelligence, medical imaging, personalized medicine

* Also affiliated with OsloMet, Norway

† Also affiliated with Norwegian University of Science and Technology, Norway

‡ Also affiliated with UiT - The Arctic University of Norway, Norway

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MMSys '22, June 14–17, 2022, Athlone, Ireland

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9283-9/22/06...\$15.00

<https://doi.org/10.1145/3524273.3533925>

ACM Reference Format:

Andrea M. Storås, Inga Strümke, Michael A. Riegler, and Pål Halvorsen. 2022. Research proposal: Explainability methods for machine learning systems for multimodal medical datasets. In *13th ACM Multimedia Systems Conference (MMSys '22)*, June 14–17, 2022, Athlone, Ireland. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3524273.3533925>

1 INTRODUCTION

The use of machine learning (ML) for solving medical tasks has increased significantly over the past years, and multimedia and medicine have been regarded teammates for some time already for ML-powered systems [24]. Some examples of medical multimedia datasets include the KVASIR series [3, 10, 12, 22] from gastrointestinal examinations and VISEM [9] from the area of human reproduction. Increasing amounts of health data, including images, videos and sensor data, that are available for analysis, development of efficient algorithms and powerful hardware are some important reasons that explain why ML has gained popularity in the medical field [23]. ML has among other things been applied for detection of polyps in the gastrointestinal tract [11, 25], classification of skin cancer [7] and prediction of hypoxemia during surgery [17]. The manual examination and interpretation of medical data can be time consuming and must typically be performed by an experienced medical expert. The interpretations might also be subjective and dependent on the specific operator. Consequently, ML could be useful, as it opens up for automatic, consistent and efficient analysis of medical data.

When the amount of medical data is large, deep learning can be a powerful tool [23]. Convolutional neural networks (CNNs) have for example been successful for analysis of images and videos [8], also within the medical field. Unfortunately, these algorithms are highly complex and not easily interpreted by humans. Lack of interpretability can be a barrier when it comes to generating trustworthy systems. Indeed, this has been identified as a challenge when it comes to implement and apply ML models in medicine and healthcare [6, 32]. Explainable artificial intelligence (XAI) aims to solve this challenge by providing explanations of the ML models and their predictions [1]. In order for healthcare providers to trust and decide to use ML models in their clinical practice, XAI methods should be applied [1].

In this PhD project, which lies at the intersection of artificial intelligence (AI), multimedia systems and health, we want to explore whether and how efficient XAI methods can be developed

and applied to increase the understanding and acceptance of AI systems in clinical practice or not. Established XAI methods will be applied for explaining ML models that are used for the analysis of medical multimedia data including tabular data, images and videos. Because current research indicate that existing XAI methods do not always meet the needs of the end users [13, 21], new explanation techniques will also be developed. To see how well the XAI methods generalize across the medical field, they will be applied on multimedia data from the medical domain, namely tabular data, images and videos, arriving from different sources. XAI methods are also usually only applied to one modality at the time and one question we also plan to explore is how XAI can be used to tackle multimodal data applications. Because the experiments are based on multimedia datasets, we believe the PhD project is a good fit for MMSys.

2 RELATED WORK

For our purposes, it is useful to divide XAI methods into intrinsic and extrinsic methods. Intrinsic explanations aim to explain the inner workings of the model. Extrinsic explanation methods treat the model as a 'black box' and present its behaviour [18]. Further on, global explanations aim to describe the model as a whole, while local explanations focus on explaining single predictions [18]. Shapley additive explanations (SHAP) [16] is an example of an established extrinsic XAI method. The method is model agnostic, meaning that it can be applied to any ML model. SHAP provides local model explanations and can be applied to any type of data [14]. When it comes to deep learning models for image analysis, intrinsic gradient-based methods such as GradCAM [26] are dominating [14]. However, the field of XAI is still in its beginning, and there are room for improving the explanation methods.

Because decisions made by ML models in a healthcare setting can affect treatments and thus patient outcomes, healthcare providers must have a high level of confidence in the models. If the models are too complex to understand, healthcare providers might refuse to use them. In our research, we will apply established XAI methods and develop new techniques to enable increased understanding of ML models in the medical field. Ideally, the use of ML can lead to faster and more accurate diagnoses and improve the outcome and quality of life for the patients.

3 DATA

Typical challenges associated with medical data are, besides privacy, data acquisition and labeling. It can be difficult to obtain medical data from a large population for several reasons. Firstly, regulatory aspects and privacy protection might restrict the access to the data. In the biomedical field, studies have shown that more than one third of research articles do not share the raw data [15, 34]. Secondly, if the disease or treatment is rare, the number of patients will be relatively low. This means that the dataset will be small or skewed towards healthy controls. Different regulations across countries and/or continents further complicates the sharing of medical data [4]. Labeling of medical data must usually be done by medical experts. Because the labeling tends to be time consuming and medical experts are busy, it can be challenging to obtain a large number of labeled medical data [23]. When the target variable is not

available, it becomes difficult to train ML models using supervised learning.

In this PhD project, different types of medical data will be taken into account. Available medical datasets include images from gastrointestinal examinations [3, 12], videos of *in vitro* fertilization procedures, tabular data from transplanted patients and images, proteomics and tabular data from patients with dry eye disease, respectively. The datasets are diverse and come with different challenges. Regarding the data from gastrointestinal examinations, a large number of the images are unlabeled. Consequently, self-supervised or unsupervised learning techniques must be applied in order to exploit the entire dataset. This is also true regarding the videos of fertilization procedures. The datasets from transplanted patients and patients with dry eye disease are not publicly available. This limits the ability to share the code and data, which again leads to less transparency regarding data analysis and model development. Moreover, the number of patients in these two datasets is relatively small and the patients come from a limited geographical area (Norway). This presents a potential challenge for developing ML models that can generalize well to new data. Because several of the datasets contain images or videos, they must typically be analyzed using complex ML models embedded in even more complex systems in order to get adequate results. In general, these models are not transparent and are difficult to explain to the end users.

4 PROPOSED APPROACH

The first part of the PhD project is to develop ML models based on the available medical data. The results will then be explained using existing XAI methods. Finally, we will try to improve the explanations by creating and testing new approaches. Below follows a description of the plans for each of the datasets:

- Images from endoscopy and colonoscopy: Develop a transparent method to detect and segment polyps in images from gastrointestinal examinations. Investigate semi-supervised ML methods to label unlabeled images based on a small amount of labeled images.
- Videos from *in vitro* fertilization procedures: Cluster video frames from the intracytoplasmic sperm injection (ICSI) procedure into groups that represent different stages of the procedure. This can make the examination of the videos more effective, and the medical personnel will save time.
- Tabular data from kidney transplanted patients: Develop ML models to predict the drug exposure based on patient characteristics and explain the results. Such systems can potentially assist clinicians to design personalized drug regimens.
- Images and tabular data from patients with dry eye disease: Develop ML methods for efficient analysis of meibography images, which can be used for diagnostic purposes. Feedback from eye doctors will be used to improve the system. We will also apply ML and XAI to explore the relationship between dry eye disease and the proteins and lipids that are present in the tears. This can increase the understanding of dry eye disease.

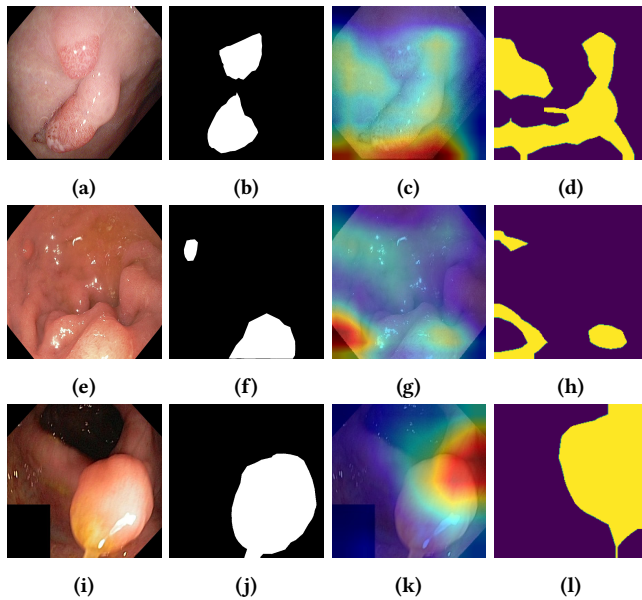


Figure 1: Polyp segmentation experiments. (a, e, i) Original images, (b, f, j) ground truth segmentation masks, (c, g, k) heat maps and (d, h, l) generated segmentation masks.

5 PROJECTS

This section describes the planned projects, including the results that have been obtained so far in the PhD work.

5.1 Polyp segmentation

Automatic detection and segmentation of polyps in images from endoscopies and colonoscopies can be useful for faster evaluation of the examination. However, it is often difficult to obtain a large number of labeled images. We therefore developed a ML pipeline that combines semi-supervised labeling of the gastrointestinal images with unsupervised segmentation [29]. Global features are extracted from the images, and clustering is used to label unlabeled data according to their proximity to the labeled data points. A deep neural network is then trained on all the labeled images to classify them as ‘polyps’ or ‘non-polyps’. GradCAM [26] is used to extract heat maps from the resulting model, and segmentation masks are generated based on the heat maps. The resulting models and source code can be found publicly online¹. Some examples of original images, ground truth segmentation masks, heatmaps extracted with GradCAM and the corresponding segmentation masks using the proposed method are provided in Figure 1. Despite sub-optimal performance on the segmentation task, the method has a high level of transparency. There are several possible directions of future work to improve the system.

5.2 Clustering videos from *in vitro* fertilization procedures

Examinations of the *in vitro* fertilization procedure called ICSI [20] are performed to evaluate technical aspects of the procedure. Videos

¹<https://github.com/kelkalot/Medico-2021-Team-Medical-XAI>

of the ICSI procedure can be applied for training purposes, internal quality control and refinement of internal procedures at the fertility clinic. We propose a method that automatically cluster video frames from the ICSI procedure. First, deep features are extracted from the video frames. The dimensionality of the features are then reduced before the data points are clustered. By reducing the dimensionality, it is possible to visualize the video frames in a scatter plot and inspect how they are located in relation to each other. The resulting clusters are shown in Figure 2. Our method is able to separate the video frames into different stages of the procedure. This could be useful in the fertility clinic in order to analyze ICSI videos more efficiently. The source code for the proposed method is publicly available²

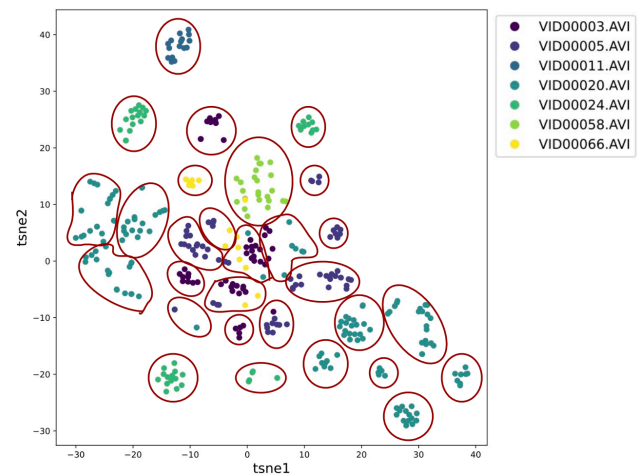


Figure 2: Results from clustering of video frames from the ICSI procedure. The frames are colored after which video they belong to, and the clusters are indicated with red circles.

5.3 Predicting drug exposure in kidney transplanted patients

Following organ transplantation, the patient needs life-long immunosuppression to avoid rejection of the transplanted organ. Tacrolimus is among the most commonly used immunosuppressive drugs following kidney transplantation at centers around the world [5]. In order to decide the drug regimen for a given patient, therapeutic drug monitoring is applied, where clinicians estimate the systemic drug exposure based on drug concentration measurements [5]. If the drug exposure is too high, the dose is reduced, while an exposure that is too low suggests an increased dose. The systemic drug exposure is highly variable both between patients and within the same patient over time [27, 33]. This makes it challenging to decide the optimal drug regimens, even for experienced doctors. Consequently, an efficient way of predicting the exposure would be valuable in the clinic. We are currently developing ML models that predict the drug exposure based on patient characteristics. The XAI library SHAP [16] is applied to investigate which features that are important for the model predictions. The results

²https://github.com/AndreaStoraas/UnsupervisedClustering_ICSI

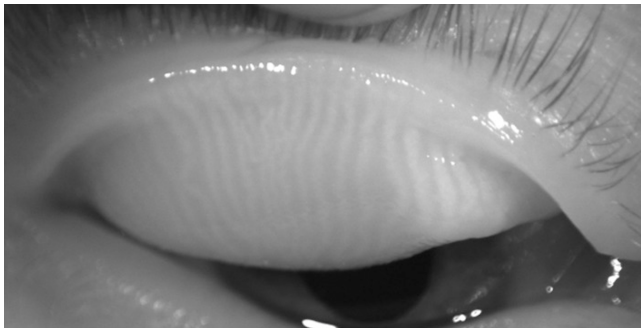


Figure 3: A meibography of the upper eye lid. The meibomian glands are observed as light vertical lines on the eye lid. The image is from a figure published in [2].

following internal model evaluation and prospective testing on new data are promising. In the future, more work is needed to see how well the ML models perform compared to today's standard of care and other existing methods that do not apply ML.

5.4 Dry eye disease

Dry eye disease is a common, yet underdiagnosed and undertreated condition. The prevalence ranges between 5 and 50%, depending on the population studied and diagnostic criteria applied [28]. To get an overview of existing methods using AI related to dry eye disease, we performed a systematic literature review [30]. We found that the overall results are promising, but that more work is needed when it comes to the model development process, prospective testing in the clinic and standardisation.

One of our datasets include images of meibomian glands from patients with dry eye disease. Meibomian gland dysfunction can give rise to evaporative dry eye disease [31]. Lack of protective meibum makes the tear film evaporate quickly from the surface of the eye, leading to dry eyes. Examination of meibographies can be used to identify meibomian gland dysfunction [31] and target the treatment. Figure 3 provides an example of a meibography and is a modification of a figure originally published in [2]. Results from manual examination of meibographies can vary between different observers [19], and the procedure is time consuming. Hence, an automatic method could lead to more consistent results, as well as alleviate the eye doctors. We will develop a computer vision tool to automatically analyze meibography images. In collaboration with eye doctors, the system will be evaluated prospectively in the clinic. Moreover, feedback from the doctors will be used to improve the system using reinforcement learning.

Further on, we plan to examine the relationship between patient characteristics, tear proteomics and the severity of dry eye disease. Analyses using ML and XAI techniques might improve our understanding of the process of the disease and identify potential risk factors. The analyses will be performed on tabular data and potentially on multi-modal data by also including the meibographies.

6 CONCLUSION

This PhD project aims to explore the use of XAI techniques for medical applications in order to increase transparency and trust

of medical multimedia systems powered by ML. Different types of multimodal medical data from various medical domains will be applied. The development of transparent multimedia systems for automatic analysis of medical data can potentially be extended to other domains.

REFERENCES

- [1] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- [2] Muhammed Yasin Adil, Jiaxin Xiao, Jonatan Olafsson, Xiangjun Chen, Neil S. Lagali, Sten Ræder, Øygunn A. Utheim, Darlene A. Dartt, and Tor P. Utheim. 2019. Meibomian Gland Morphology Is a Sensitive Early Indicator of Meibomian Gland Dysfunction. *American Journal of Ophthalmology* 200 (2019), 16–25. <https://doi.org/10.1016/j.ajo.2018.12.006>
- [3] Hanna Borgli, Vajira Thambawita, Pia H Smedsrud, Steven Hicks, Debesh Jha, Eskeland Sigrun L, Kristin Ranheim Rand I, Konstantin Pogorelov, Mathias Lux, Duc Tien Dang Nguyen, Dag Johansen, Carsten Griwodz, Stensland Håkon K, Enrique Garcia-Ceja, Peter T Schmidt, Hugo L Hammer, Michael A Riegler, Pål Halvorsen, and Thomas de Lange. 2020. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data* 7, 1 (2020), 283. <https://doi.org/10.1038/s41597-020-00622-y>
- [4] Laura Bradford, Mateo Aboy, and Kathleen Liddell. 2020. International transfers of health data between the EU and USA: a sector-specific approach for the USA to ensure an 'adequate' level of protection. *Journal of Law and the Biosciences* 7, 1 (2020). <https://doi.org/10.1093/jlb/lxaa055>
- [5] Mercè Brunet, Teun Van Gelder, Anders Åsberg, Vincent Haufroid, Dennis A Heselink, Loralie Langman, Florian Lemaitre, Pierre Marquet, Christoph Seger, Maria Shipkova, Alexander Vinks, Pierre Wallemaçq, Eberhard Wieland, Jean Baptiste Woillard, Markus J. Barten, Klemens Budde, Helena Colom, Maja-Theresa Dieterlen, Laure Elens, Kamisha L. Johnson-Davis, Pawel K. Kunicki, Iain MacPhee, Satoshiro Masuda, Binu S. Mathew, Olga Millán, Tomoyuki Mizuno, Dirk-Jan A. R. Moes, Caroline Monchaud, Ofelia Noceti, Tomasz Pawinski, Nicolas Picard, Ron van Schaik, Claudia Sommerer, Nils Tore Vethe, Brenda de Winter, Uwe Christians, and Stein Bergan. 2019. Therapeutic drug monitoring of tacrolimus-personalized therapy: second consensus report. *Therapeutic drug monitoring* 41, 3 (2019), 261–307. <https://doi.org/10.1097/FTD.0000000000000640>
- [6] Federico Cabitza, Raffaele Rasoini, and Gian Franco Gensini. 2017. Unintended Consequences of Machine Learning in Medicine. *JAMA* 318, 6 (08 2017), 517–518. <https://doi.org/10.1001/jama.2017.7797>
- [7] Andre Esteve, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature (London)* 542 (2017), 115–118. <https://doi.org/10.1038/nature21056>
- [8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. 326 – 366 pages. <http://www.deeplearningbook.org>.
- [9] Trine B. Haugen, Steven A. Hicks, Jorunn M. Andersen, Oliwia Witeczak, Hugo L. Hammer, Rune Borgli, Pål Halvorsen, and Michael Riegler. 2019. VISEM: A Multimodal Video Dataset of Human Spermatozoa. In *Proceedings of the 10th ACM Multimedia Systems Conference (Amherst, Massachusetts) (MMSys '19)*. Association for Computing Machinery, New York, NY, USA, 261–266. <https://doi.org/10.1145/3304109.3325814>
- [10] Debesh Jha, Sharib Ali, Krister Emanuelsen, Steven A Hicks, Vajira Thambawita, Enrique Garcia-Ceja, Michael A Riegler, Thomas de Lange, Peter T Schmidt, Håvard D Johansen, et al. 2021. Kvasir-instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy. In *Proceedings of the International Conference on Multimedia Modeling (MMM)*. Springer, 218–229. https://doi.org/10.1007/978-3-030-67835-7_19
- [11] Debesh Jha, Pia H. Smedsrud, Dag Johansen, Thomas de Lange, Håvard D. Johansen, Pål Halvorsen, and Michael A. Riegler. 2021. A Comprehensive Study on Colorectal Polyp Segmentation With ResUNet++, Conditional Random Field and Test-Time Augmentation. *IEEE Journal of Biomedical and Health Informatics* 25, 6 (2021), 2029–2040. <https://doi.org/10.1109/JBHI.2021.3049304>
- [12] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. 2020. Kvasir-seg: A segmented polyp dataset. In *Proceedings of the International Conference on Multimedia Modeling (MMM)*. Springer, 451–462. https://doi.org/10.1007/978-3-030-37734-2_37
- [13] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. *Questioning the AI: Informing Design Practices for Explainable AI User Experiences*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3313831.3376590>
- [14] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. 2021. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* 23, 1 (2021). <https://doi.org/10.3390/e23010018>
- [15] Maria Littmann, Katharina Selig, Liel Cohen-Lavi, Yotam Frank, Peter Hönigschmid, Evans Kataka, Anja Mösch, Kun Qian, Avihai Ron, Sebastian

- Schmid, et al. 2020. Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. *Nature Machine Intelligence* 2, 1 (2020), 18–24. <https://doi.org/10.1038/s42256-019-0139-8>
- [16] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc., 4765–4774. <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>
- [17] Scott M Lundberg, Bala Nair, Monica S Vavilala, Mayumi Horibe, Michael J Eisses, Trevor Adams, David E Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, and Su-In Lee. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature biomedical engineering* 2 (2018), 749–760. <https://doi.org/10.1038/s41551-018-0304-0>
- [18] Christoph Molnar. 2022. *Interpretable Machine Learning* (2 ed.). <https://christophm.github.io/interpretable-ml-book>
- [19] Jason J Nichols, David A Berntsen, G Lynn Mitchell, and Kelly K Nichols. 2005. An assessment of grading scales for meibography images. *Cornea* 24, 4 (2005), 382–388. <https://doi.org/10.1097/01.icc.0000148291.38076.59>
- [20] G. Palermo, H. Joris, P. Devroey, and A.C. Van Steirteghem. 1992. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. *The Lancet* 340, 8810 (1992), 17–18. [https://doi.org/10.1016/0140-6736\(92\)92425-F](https://doi.org/10.1016/0140-6736(92)92425-F) Originally published as Volume 2, Issue 8810.
- [21] Jeremy Petch, Shuang Di, and Walter Nelson. 2022. Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. *Canadian Journal of Cardiology* 38, 2 (2022), 204–213. <https://doi.org/10.1016/j.cjca.2021.09.004>
- [22] Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Concetto Spampinato, Duc-Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, Michael Riegler, and Pål Halvorsen. 2017. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. In *Proceedings of the ACM on Multimedia Systems Conference (MMSys)* (Taipei, Taiwan). ACM, New York, NY, USA, 164–169. <https://doi.org/10.1145/3083187.3083212>
- [23] Daniele Ravi, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang. 2017. Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics* 21, 1 (2017), 4–21. <https://doi.org/10.1109/JBHI.2016.2636665>
- [24] Michael Riegler, Mathias Lux, Carsten Griwodz, Concetto Spampinato, Thomas de Lange, Sigrun L. Eskeland, Konstantin Pogorelov, Wallapak Tavanapong, Peter T. Schmidt, Cathal Gurrin, Dag Johansen, Håvard Johansen, and Pål Halvorsen. 2016. Multimedia and Medicine: Teammates for Better Disease Detection and Survival. In *Proceedings of the ACM International Conference on Multimedia (Amsterdam, The Netherlands) (MM '16)*. ACM, New York, NY, USA, 968–977. <https://doi.org/10.1145/2964284.2976760>
- [25] Michael Riegler, Konstantin Pogorelov, Jonas Markussen, Mathias Lux, Håkon Kvale Stensland, Thomas de Lange, Carsten Griwodz, Pål Halvorsen, Dag Johansen, Peter T. Schmidt, and Sigrun L. Eskeland. 2016. Computer Aided Disease Detection System for Gastrointestinal Examinations. In *Proceedings of the ACM on Multimedia Systems Conference (MMSys)* (Klagenfurt, Austria). ACM, New York, NY, USA, Article 29, 4 pages. <https://doi.org/10.1145/2910017.2910629>
- [26] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 618–626. <https://doi.org/10.1109/ICCV.2017.74>
- [27] Christine E Staatz and Susan E Tett. 2004. Clinical pharmacokinetics and pharmacodynamics of tacrolimus in solid organ transplantation. *Clinical Pharmacokinetics* 43 (2004), 623–653. <https://doi.org/10.2165/00003088-200443100-00001>
- [28] Fiona Stapleton, Monica Alves, Vatinée Y Bunya, Isabelle Jalbert, Kaevalin Lekhanont, Florence Malet, Kyung-Sun Na, Debra Schaumberg, Miki Uchino, Jelle Vehof, et al. 2017. TFOS DEWS II epidemiology report. *The ocular surface* 15, 3 (2017), 334–365. <https://doi.org/10.1016/j.jtos.2017.05.003>
- [29] Andrea M. Storås. 2021. Unsupervised Image Segmentation via Self-Supervised Learning Image Classification. In *Proceedings of MediaEval 2021 CEUR Workshop*.
- [30] Andrea M. Storås, Inga Strümke, Michael A. Riegler, Jakob Grauslund, Hugo L. Hammer, Anis Yazidi, Pål Halvorsen, Kjell G. Gundersen, Tor P. Utheim, and Catherine J. Jackson. 2022. Artificial intelligence in dry eye disease. *The Ocular Surface* 23 (2022), 74–86. <https://doi.org/10.1016/j.jtos.2021.11.004>
- [31] Alan Tomlinson, Anthony J. Bron, Donald R. Korb, Shiro Amano, Jerry R. Paugh, E. Ian Pearce, Richard Yee, Norihiko Yokoi, Reiko Arita, and Murat Dogru. 2011. The International Workshop on Meibomian Gland Dysfunction: Report of the Diagnosis Subcommittee. *Investigative Ophthalmology & Visual Science* 52, 4 (03 2011), 2006–2049. <https://doi.org/10.1167/iov.10-6997f>
- [32] Alfredo Vellido. 2020. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural computing and applications* 32, 24 (2020), 18069–18083. <https://doi.org/10.1007/s00521-019-04051-w>
- [33] Raman Venkataramanan, Arun Swaminathan, Tata Prasad, Ashok Jain, Sheila Zuckerman, Vijay Warty, John McMichael, Jacqueline Lever, Gilbert Burckart, and Thomas Starzl. 1995. Clinical pharmacokinetics of tacrolimus. *Clinical Pharmacokinetics* 29, 6 (1995), 404–430.
- [34] Joshua D. Wallach, Kevin W. Boyack, and John P. A. Ioannidis. 2018. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. *PLOS Biology* 16 (11 2018), 1–20. <https://doi.org/10.1371/journal.pbio.2006930>