Medico Multimedia Task at MediaEval 2019

Steven Hicks[1,3], Pål Halvorsen[1,3], Trine B. Haugen[3], Jorunn M. Andersen[3], Oliwia Witczak[3], Konstantin Pogorelov, Hugo L. Hammer[3], Duc-Tien Dang-Nguyen[4], Mathias Lux[2], Michael Riegler[1]
[1]SimulaMet, Norway; [2]Alpen-Adria-Universität Klagenfurt, Austria;
[3]Oslo Metropolitan University, Norway; [4]University of Bergen, Norway;

ABSTRACT

The *Medico: Multimedia for Medicine Task* is running for the third time as part of MediaEval 2019. This year, we have changed the task from anomaly detection in images of the gastrointestinal tract to focus on the automatic prediction of human semen quality based on videos. The purpose of this task is to aid in the assessment of male reproductive health by providing a quick and consisted method of analyzing human semen. In this paper, we describe the task in detail, give a brief description of the provided dataset, and discuss the evaluation process and the metrics used to rank the submissions of the participants.

1 INTRODUCTION

The 2019 *Medico: Multimedia for Medicine Task* tackles the challenge of predicting certain quality measurements of sperm using a multimodal dataset consisting of microscopic video recordings of human semen, associated sensor data, and participant-related data. Male infertility accounts for approximately 60% of cases involving couples having problems conceiving a child [1]. The first step to evaluating male fertility is often through the assessment of semen, where a clinician measures the number of living sperm and looks for any abnormalities that may be present in the spermatozoa (living sperm). This process is done through a microscope, where the clinician has to count each visible sperm manually to calculate specific metrics related to the movement and shape of the spermatozoa. Having a tool that could aid in the detection of sperm and calculating these metrics would not only reduce the time spent per sample but also lessen the large inter- intra-observer variability between and within clinics [3].

For the 2019 edition of Medico, we present three different tasks. The first two tasks aim to predict key indicators of good semen quality, specifically the motility and morphology of sperm. Motility and morphology are two of the most common metrics used to evaluate spermatozoa, and to predict these measurements gives the participants a good opportunity to utilize both the video and sensor data available in the dataset. The third task relates to looking at individual sperm to figure out which one moves the fastest. For all tasks, we require participants to measure and report the data processing performance in terms of time spent on each frame being analyzed. As we aim for real-time applications, the processing speed is an important factor, especially for the individual spermatozoon selection task during the in vitro fertilization procedure (a procedure where a sperm is injected into an egg outside of the body).

MediaEval'19, 27-29 October 2019, Sophia Antipolis, France

2 DATASET DETAILS

The VISEM dataset [2] contains more than 35 gigabytes of videos and related data from 85 male participants aged 18 years or older. For each participant, we include a set of measurements collected from an analyzed semen sample. This includes a video of the live spermatozoa, a sperm fatty acid profile, the fatty acid composition of serum phospholipids, some study participant-related data and World Health Organization (WHO) analysis data [4]. In total, the dataset consists of six CSV-files, five for data and one which maps video IDs to study participants IDs. The duration of each video varies between two to seven minutes and runs at approximately 50 frames-per-second. The resolution for each video is 640×480 and is stored in a ".avi" file. The name of each video file contains the videos ID, the date it was recorded, a small optional description, and ends with the code of the person who assessed the video. For the purpose of this task, we have split the videos within the video directory into three separate "folds". This was done in order to make it easier for participates to perform the required three-fold cross-validation.

For the analysis data, sensor data, and participant-related data, the following CSV-files are provided:

- *semen_analysis_data:* The results of standard semen analysis.
- *fatty_acids_spermatozoa:* The levels of several fatty acids in the spermatozoa of the participants.
- *fatty_acids_serum:* The serum levels of the fatty acids of the phospholipids (measured from the blood of the participant).
- *sex_hormones:* The serum levels of sex hormones measured in the blood of the participants.
- *study_participant_related_data:* General information about the participants such as age, abstinence time and Body Mass Index (BMI).
- *videos:* Overview of which video-file belongs to what participant.

VISEM is publicly available [1] without any restriction. Each study participant agreed to donate their data for research and provided the necessary consent to distribute it. It is also important to point out that all data is fully anonymized and follows the state of the art with respect to the privacy of medical information.

3 EVALUATION AND TASK DESCRIPTIONS

Medico 2019 presents three different tasks, each meant to target a specific use-case within male fertility assessment. The three tasks for this year's Medico is the *prediction of morphology* task, the

[1]https://datasets.simula.no/visem

prediction of motility task, and the *unsupervised sperm tracking* task. Of these three tasks, the first two are mandatory in order to participate in the challenge. The third task is optional, but highly recommended, as it is an important problem within assisted reproductive technology. For the performance evaluation, the same dataset will be used for development and testing, but for testing, we ask the participants to perform three-fold cross-validation. For the processing speed evaluation, we will use minimum, average and maximum frame processing times in seconds measured by the participants as a time interval from the moment when an image has been completely loaded into memory to the moment of the final decision has been made by the corresponding task analysis algorithm. In the following, we will give a more detailed description of each task and discuss how each will be evaluated and ranked.

3.1 Prediction of Motility Task

Motility is the ability of an organism to move independently. In semen analysis, this is commonly split into three distinct categories, namely progressive, non-progressive, and immotile spermatozoa. A progressive spermatozoon is one that is able to move forward at a slow or fast pace, a non-progressive spermatozoon is a sperm that moves without forward progression, and immotile sperm do not move at all. The purpose of this task is to automatically find the percentage of progressive, non-progressive, and immotile spermatozoa for a given semen sample. We urge participants to perform multi-frame analysis over single-frame analysis. This is important due to the fact that single-frame analysis loses most of the temporal information needed to accurately assess the movement of the spermatozoa.

In order to participate in this task, task participants will submit a ".csv" file containing one prediction for each of the 85 study participants. Each line should be made up of five comma-separated values; ID of the study participant, percentage of progressive sperm, percentage of non-progressive sperm, percentage of immotile sperm, average frame processing time. To evaluate this task we will use root mean squared error and mean absolute error to evaluate the submissions. Submissions will be ranked based on the achieved mean absolute error as it shows the improvement of the automatic prediction compared to that of the manual assessment.

3.2 Prediction of Morphology Task

Evaluating the morphology of sperm involves looking for any abnormalities that may be present in the three parts that make up a spermatozoon, namely the head, midpiece, and tail. This task should predict the percentage of sperm with head defects, midpiece defects, and tail defects for a single semen sample.

As with the *prediction of motility* task, participants have to submit a ".csv" file containing one prediction of morphology assessment for each of the 85 study participants. Each line should be made up of the following five comma-separated values; ID of the study participant, percentage of sperm with tail defects, percentage of sperm with midpiece defects, percentage of sperm with head defects, average frame processing time. Similar to the *Prediction of motility* task, we will use mean squared error and mean absolute error to evaluate the submissions. Submissions will be ranked based on the achieved mean absolute error.

3.3 Unsupervised Sperm Tracking Task

This task is about finding the spermatozoon that moves fastest compared to all others, and requires that task participants track individual spermatozoon in order to evaluate the speed at a given point in time. Within this task, we defined the "fastest" spermatozoon in two different ways:

(1) *Fastest average speed:* The one that moves the longest distance during a video defined by the total distance divided by the length of the video.
(2) *Highest top speed:* the one that has the highest intermediate speed.

One specific challenge of this third task is that the video also changes due to the sample being moved while under the microscope to observe the sample in its entirety. We suggest that the participants track each sperm per viewpoint rather than per video. To evaluate this task, we will use manual evaluation with the help from three different experts within human reproduction.

4 DISCUSSION AND OUTLOOK

We believe the area of automatic semen analysis is an important, yet overlooked, area of research which has the potential of being very beneficial for those working to improve reproductive health. From a computer science point of view, the task is compelling as it features more than just simple image analysis as it requires the participants to analyze the temporal information present from one frame to another in order to make quality predictions. We hope that this task will encourage the multimedia community to aid in the development of computer-assisted reproductive health, and discover new and clever ways of analyzing multimodal datasets.

REFERENCES

[1] Sandro C. Esteves, Ricardo Miyaoka, and Ashok Agarwal. 2011. An update on the clinical assessment of the infertile male. *Clinics* 66 (00 2011), 691–700. https://doi.org/10.1590/S1807-59322011000400026
[2] Trine B. Haugen, Steven A. Hicks, Jorunn M. Andersen, Oliwia Witczak, Hugo L. Hammer, Rune Borgli, Pål Halvorsen, and Michael A. Riegler. 2019. VISEM: A Multimodal Video Dataset of Human Spermatozoa. In *Proceedings of the 10th ACM on Multimedia Systems Conference (MMSys'19)*. ACM, New York, NY, USA. https://doi.org/10.1145/3304109.3325814
[3] RW Manderscheid, CD Ryff, EJ Freeman, L McKnight-Eily, S Dhingra, and TW Strine. 2009. Evolving definitions of mental illness and wellness. *Preventing chronic disease* 7, 1 (dec 2009), A19–A19. https://www.ncbi.nlm.nih.gov/pubmed/20040234https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2811514/
[4] World Health Organization et al. 2010. WHO laboratory manual for the examination and processing of human semen. (2010).