A multi-site loiasis dataset of whole blood videos

Linda Djune Yemeli [1,*], Rella Zoleko-Manego [2], Charles B. Delahunt [3,4,*], Donald F. Fankam [1], Yannick Y. Nzeuhang [1], Jean G. Bopda [1], Mbakam S. Laetitia [1], Yves A. Balog [1], Steve M. Tchana [1], Ghyslain Mombo-Ngoma [2,6], Michael Ramharter [2,6,7], Maria Diaz de Leon Derby [5], Zaina L. Moussa [5], Ethan Spencer [4], Dipayan Banik [4], Matthew D. Keller [4], Karla N. Fisher [8,9], Daniel A. Fletcher [5], Isaac I. Bogoch [8,9,10], Anne-Laure Le Ny [4], Saskia Deda David [2,6], Joseph Kamgno [1]

1 Higher Institute for Scientific and Medical Research (ISM), Yoaoundé, Cameroon **2** Centre de Recherches Médicale de Lambaréné, Gabon (CERMEL) **3** University of Washington, Seattle, Washington **4** Formerly at Global Health Labs, Inc, Bellevue, Washington (ceased operations in 2025) **5** Department of Bioengineering, University of California, Berkeley, Berkeley, California **6** Center for Tropical Medicine Bernhard Nocht Institute for Tropical Medicine & Dept. of Medicine University Medical Center Hamburg-Eppendorf (BNITM) **7** German Center for Infection Research, Partner Site Hamburg-Lübeck-Borstel-Riems **8** Division of General Internal Medicine, Toronto General Hospital, University Health Network, Toronto, Canada **9** Division of Infectious Diseases, Toronto General Hospital, University Health Network, Toronto, Canada **10** Department of Medicine, University of Toronto, Toronto, Canada
* denotes equal contribution
joseph.kamgno@ismcm.org, rella.zoleko@cermel.org, delahunt@uw.edu

Abstract
Loiasis is a blood-borne filarial infection under consideration for inclusion in the WHO's priority list of Neglected Tropical Diseases. In addition to its direct impact on human health, loiasis frequently occurs as a co-infection with other filarial diseases, including onchocerciasis (river blindness) and lymphatic filariasis (LF, elephantiasis). This overlapping geographic distribution seriously hinders mass drug administration (MDA) with ivermectin (IVM) for onchocerciasis and LF control, because individuals with *Loa loa* microfilarial (mf) loads exceeding 30,000 mf/mL are at high risk of developing severe adverse events, including coma and death, following IVM treatment. Therefore, safe MDA in loiasis co-endemic areas requires active detection and exclusion of high *L. loa* mf individuals from treatment, the so-called Test and Not Treat (TaNT) strategy. Automated, AI-driven diagnostics have shown strong potential to enhance TaNT implementation and fill a crucial care need for underserved populations, yet important performance limitations remain. To support AI research for the pressing health care needs of filarial diseases, we release this first-of-its-kind dataset to enable development of AI algorithms for detecting *L. loa* in whole blood samples. The dataset includes over 33,000 videos of whole blood in capillaries, in 4721 sessions (7 videos per session) from 1940 patients, captured by NTDscope portable imaging devices during four studies in 2023 and 2025 in Cameroon and Gabon. The videos contain a range of *L. loa* microfilariae counts and negative cases, as well as some cases with *Mansonella perstans*, another tropical filarial parasite. The metadata include expert microfilariae counts from Giemsa-stained blood films for each patient.

Resumé La loase est une infection filarienne hématogène actuellement envisagée pour inclusion dans la liste prioritaire des maladies tropicales négligées de l'Organisation Mondiale de la Santé (OMS). En plus de son impact direct sur la santé humaine, la loase survient fréquemment en co-infection avec d'autres filarioses, notamment l'onchocercose (cécité des rivières) et la filariose lymphatique (FL, éléphantiasis). Cette distribution géographique chevauchante complique sérieusement les campagnes de traitement de masse à l'ivermectine (IVM) pour le contrôle de l'onchocercose et de la FL, car les individus portants des charges microfilariennes de *L. loa* supérieures à 30 000 microfilaires (mf)/mL courent un risque élevé de développer des effets indésirables graves, comprenant l'encéphalopathie, le coma et la mort, à la suite du traitement à l'IVM. Ainsi, l'implémentation du traitement de masse dans les zones co-endémiques à la loase nécessite la détection active et l'exclusion des individus ayant une forte charge microfilarienne, selon la stratégie dite Test and Not Treat (TaNT) . Les outils diagnostiques automatisés et fondés sur l'intelligence artificielle ont montré un fort potentiel pour améliorer la mise en œuvre du TaNT et répondre à un besoin crucial de soins dans les populations sous-représentées, bien que des limitates de performance importantes persistent. Pour soutenir le développement d'outils d'IA répondant aux besoins urgents de santé publique liés aux filarioses, nous mettons à disposition ce jeu de données afin de permettre le développement d'algorithmes d'IA capables de détecter *L. loa* dans des échantillons de sang total. Le jeu de données comprend plus de 33 000 vidéos de sang total dans des capillaires provenant de 4 721 sessions chez 1 940 patients, enregistrées à l'aide du NTDscope au cours de quatre études menées en 2023 et 2025 au Cameroun et au Gabon. Les vidéos couvrent un large éventail de charges microfilariennes de *L. loa*, ainsi que des cas négatifs et certains cas présentant *Mansonella perstans*, un autre parasite filarien tropical. Les métadonnées contiennent notamment les dénombrements experts de microfilaires issus de frottis sanguins colorés au Giemsa pour chaque patient.

Keywords
Loa loa, Mansonella, neglected tropical disease, mobile microscopy, machine learning

1. Background

Loiasis is an emerging Neglected Tropical Diseases (NTDs) endemic in West and Central African rainforest areas (Kelly-Hope et al., 2012; Metzger and Mordmüller, 2014; Ramharter et al., 2024). The disease is caused by the filarial nematode *Loa loa*, transmitted to humans through the bite of the tabanid fly Chrysops (Boussinesq, 2006). Epidemiological data estimated that over 14 million individuals are at risk of infection, with an estimated 10 million cases (Metzger and Mordmüller, 2014). Loiasis causes a wide range of clinical symptoms, from none at all to severe complications, including transient angioedema (Calabar swelling), ocular migration of the adult parasite under the conjunctiva (eye worm) (Boussinesq, 2006), and encephalitis, associated with high microfilariae (mf) densities (>30,000 mf/mL) (Gardon et al., 1997). Other symptoms involving deep organs (lung, heart, kidney) and excess mortality have been reported (Buell et al., 2016; Chesnais et al., 2017). Loiasis is under consideration for inclusion in the WHO's priority list of Neglected Tropical Diseases (Jacobsen et al., 2022; WHO, 2025a, 2001).

In addition, loiasis is often co-endemic with onchocerciasis (river blindness). This overlapping geographic distribution seriously hinders mass drug administration (MDA) with ivermectin (IVM) for effective control of onchocerciasis. Notably, ivermectin treatment can cause catastrophic, even fatal, side effects in persons with high *L. loa* infections (>30,000 mf/uL). Therefore, safe MDA in co-endemic areas requires active detection and exclusion of high *L. loa* mf individuals from treatment, the so-called Test and Not Treat (TaNT) strategy (Gardon et al., 1997; Boussinesq et al., 2003; Boulle et al., 2025; Pion, 2020).

TaNT depends on accurate, point-of-care diagnostics, the development of more efficient tools has become an urgent priority (diagnosis via Giemsa-stained blood films is neither rapid nor point-of-care).

2. Summary

Automated, AI-driven diagnostics have shown strong potential to greatly improve TaNT and fill a crucial care need for underserved populations. However, to our knowledge no datasets exist for rapid diagnosis of loaisis (or filariasis) in whole blood, and the only (to our knowledge) existing automated diagnostic operating on whole blood has known performance gaps (Djune-Yemeli et al., 2026). Therefore, we release this AI-ready dataset to enable development of AI algorithms to better detect *L. loa* in whole blood samples.

The dataset contains over 33,000 mp4 videos of whole blood in capillaries from 4721 diagnostic sessions collected in four field studies in Cameroon and Gabon, as well as relevant metadata (e.g. filarial load ground truth). The videos were captured by the NTDscope (de Leon Derby et al., 2025), which is a next-generation version of the CellScope (D'Ambrosio et al., 2015).

We hope that AI researchers can use this dataset, plus the description of the clinical context, to develop clinically-deployable algorithms that pair with hardware such as the NTDscope to effectively address the long-neglected health care challenge of *L. loa* and other filarial infections.

3. Discussion

This dataset, the first of its kind, offers sufficient data (1,958 patients and over 33,000 videos) to train and validate current generation AI algorithms. Its four studies and five sites enable holdout testing. It has a full range of filarial loads, including over 60 patients with high-risk loads. It also includes *M. perstans* infections, enabling speciation algorithms.

Limitations: All videos are from one device type, and there are no lymphatic filariasis cases. There are no per-object annotations; the ground truth consists of (i) filarial counts from Giemsa-stained blood films and (ii) per-video manual counts.

4. Resource Availability

4.1 Clinical Use Cases

Effective algorithms could improve care in several important use cases, if the AI development is informed, from the beginning, by both clinicians and the clinical use case – their needs and constraints. The use cases and their target performance requirements are as follows:

1. TaNT (*L. loa*, Test and Not Treat): The crucial requirement of this use case is to distinguish between high microfilaremia infection (i.e. >20,000 mf/mL) vs low microfilaremia infection (including negative cases). False Negatives (i.e. a high microfilaremia mistakenly labeled as low microfilaremia and therefore eligible for ivermectin) create risk of a serious adverse event (SAE) with harm to the patient. False Positives (i.e. lower microfilaremia mistakenly labeled as microfilaremia >20,000) carry a cost in that patients who could receive ivermectin are instead routed into other care channels. Note that correct diagnosis of negatives (0 mf/mL) vs low- microfilaremia infection is irrelevant for this use case. However, certain subpopulations (e.g. males age 31-40) may need lower thresholds than 20,000 (Boulle et al., 2025). A result delivered within 2 or 3 minutes of video capture is needed to support clinical care workflow. Since compute hardware is constrained, this time-to-result limits model options and might force image downsampling.

2. Mapping The goal of this use case is to support spatial mapping of *L. loa*, to inform MDA programs. This task requires quantitation at all levels of filaremia, since in endemic areas a majority of infected individuals carry fewer than 8,000 mf/mL e.g. (Veletzky et al., 2024). It also requires distinguishing zero mf counts (i.e. True Negatives in the usual parasitemia sense). Community MicroFilarial Load (equivalently the Williams Geometric Mean) (Kamtchum and alia, 2014) uses the formula

$$CMFL = exp\left(\frac{\Sigma_i\ ln(x_i + 1)}{n}\right) - 1 \qquad (1)$$

where x_i = filaremia for case i, n = number of cases, and the sum is over all cases. The required accuracy depends on how sensitive the CMFL formula, in context, is to errors in the input counts x_i. An important caveat (not addressable with this dataset) for diagnostics that rely solely on microfilarial counts is that in endemic areas a large fraction of infected individuals exhibit no detectable microfilariae (Veletzky et al., 2024); these amicrofilaremic cases would be systematically missed. Separately, the RAPLOA method is mainly used to characterize areas as high or low risk for SAEs, and considers only the presence of adult worm in the eye (Chesnais et al., 2017; Ramharter et al., 2024).

3. Lymphatic Filariasis (LF) Lymphatic filariasis (Goldin and Juergens, 2025; WHO, 2025b) also presents mf in blood similarly to *L. loa* (but only during night-time). The NTDscope device is potentially useful for detection of LF (unpublished data). However, LF has generally much lower filaremia (often $<$ 200 mf/mL), and thus requires diagnostics with very low LoD. This implies a different algorithm optimization target than for TaNT or *L. loa* mapping. Time-to-result: within 2 – 3 minutes of video capture.

4. Speciation It is valuable to distinguish *L. loa* vs *M. perstans*, for two reasons: To distinguish low-microfilaremia mono-infections of Loa loa from M. perstans (for accurate *L. loa* mapping purposes); and also to better study the two illnesses for their own sakes. To a first approximation, distinguishing the two species is not relevant to TaNT, because *M. perstans* filaria is in general low, and so does not materially affect quantitation at high *L. loa* microfilaremia (Hemilembolo and et al., 2023); however, unpublished data indicates that *M. perstans* loads can exceed 8000 mf/mL. The two filaria seem to have different patterns of movement in blood, which makes speciation in the videos potentially feasible. In addition, it is valuable to distinguish loiasis from LF because *L. loa* microfilariae can cause false positives in LF diagnostics that use microfilarial counts (however, this dataset contains no LF-positive cases).

5. Coagulation Once the blood capillary is loaded for imaging by the NTDscope device, the blood starts coagulating after a few minutes. Because this prevents filaria motion in the videos, it can lead to False Negatives for TaNT. Therefore a strong coagulation detection algorithm would improve safety in the TaNT use case.

4.2 Role of AI

Effective AI-driven automated diagnosis has potential to greatly improve the efficiency of diagnostics for all these use cases. First, it can mitigate the challenge of understaffing (insufficient numbers of trained microbiologists and laboratory personnel). Second, the NTDscope device enables rapid, on-the-spot diagnosis in the field, both eliminating the need to return blood samples to the lab for evaluation and also saving clinical teams much time. However, these benefits depend on effective detection algorithms tuned to the various use cases, and *L. loa* diagnostics (and NTD diagnostics generally) are currently underserved by the medical AI community. The purpose of this dataset is to enable development of AI solutions that can meet the stringent clinical requirements of the various *L. loa* use cases, in order to accelerate progress in loiasis health care as well as other filarial diseases such as LF.

A baseline algorithm (hereafter "2026 algorithm") for *L. loa* detection on these videos is described in (Djune-Yemeli et al., 2026). It performs reasonably well for TaNT and quantitation. However, it has a high LoD (750 – 1200 mf/mL), does not distinguish species, and does not detect coagulated blood samples; interestingly, it does not use any AI methods; rather, it leverages biophysical priors and applies image processing and statistical methods. Therefore an opportunity exists to apply current AI methods to develop improved *L. loa* detection algorithms.

4.3 Data/Code Location

The dataset will be freely available at Zenodo.

4.4 Licensing

The data is open source with the Creative Commons license: CC-BY 4.0: Attribution 4.0 International `https://creativecommons.org/licenses/by/4.0/deed`

4.5 Ethical Considerations

Studies in Cameroon received ethical clearance from the Centre Regional Ethical Committee for Research on Human Health (CRERSH-Ce) and administrative authorization from the Centre Regional Delegation for Public Health of the Ministry of Public Health (CEIs 0094/CRERSHC/ 2023, 00934/CRERSHC/2025. Study population was first informed through the local health system (Health District, Health areas and community health personnel, community leaders. . .). In each community, all eligible individuals were invited to a central point (health facility or chieftaincy). An

introductory speech was given by a team member and translated into a local language by community health personnel. All the potential participants were given the opportunity to ask questions before making the decision whether to participate or not. Consent/assent/parental consent forms were signed by all the participants.

The Gabon study was approved by the ethics committee of CERMEL under the number CEI-026/2022. Participants or their legal representatives gave informed consent before any study related procedures were performed.

4.6 Dataset acquisition

All videos in this dataset were collected with the NTDscope device. Site details are as follows:

Cameroon Sample collection involved both the NTDscope test and Calibrated Thick Blood Smears (CBTS) in parallel. NTDscope testing on finger-prick blood, followed procedures in (D'Ambrosio et al., 2015). In parallel, CTBS were prepared using 70 µL of finger-prick blood, following standardized protocols of the ISM laboratory. Briefly, non-heparinized finger-prick blood was collected and drawn onto a microscope slide. The slide was then allowed to dry and stained with 10% Giemsa using standard procedures (Walther and Muller, 2003). For CBTS, on each stained slide 50 µL (not 70 µL) of blood was examined under a light microscope at 10X objective for blood dwelling mf. All mf present on the smear were counted and reported. Mf of *M. perstans*, also endemic in the study area, were distinguished from those of *L. loa* by size and the absence of a sheath. The final result was expressed in microfilariae per milliliter of blood (mf/mL) (WHO, 1991).

Gabon Recruitment for sampling was conducted at the Centre de Recherches Médicales de Lambaréné (CERMEL), in Moyen Ogooue (Ekouk, Nzong Mbang, petit Odavo) and Ngounie (Sindara) provinces of Gabon, from 13 August 2023 to 24 August 2024. The region is highly endemic for *L. loa* and a range of other parasitic infectious diseases (Veletzky et al., 2020; Zoleko Manego et al., 2017; Ramharter et al., 2021). Participants for this study were from Lambarene and surrounding villages and Sindara.

5. Methods: Dataset contents

5.1 General organization

The data structure follows a nested hierarchy: Site, session, video. Site: There are 5 sites (three Cameroon studies at different times, and two sites in Gabon), as well as two collections of deliberately coagulated samples. The session is the crucial unit. It contains 7 videos of a capillary of blood from a single patient, and is the basis for patient diagnosis. The videos capture successive non-overlapping Fields of View (FoVs) along the capillary. In this dataset, patients have 1 to 4 sessions depending on the study protocols. Videos are 5 seconds long, captured at 15 to 35 frames per second (FPS). FoV size is about 4.5 mm × 6.2 mm. The original image (i.e. frame) size is 1440 × 1080 × 3, with pixel pitch ≈ 4.3 μm / pixel. 2026 algorithm experiments indicated that frames can be downsampled by 3× (to 360 × 480 × 3) with no harm to algorithm performance but with great increase in processing speed. However, 3× downsampled frames are clearly different from the full sized frames. The videos provided in this dataset have been downsampled 2× by simple subsampling (to 720 × 540 × 3) to reduce file size. To access the collection of full-sized videos (≈280 GB) please contact the Fletcher Lab at University of California, Berkeley (fletch@berkeley.edu).

Each session has its own folder, named according to the following scheme: "site_patient_table" or "site_patient"

"site" can have the following values:
- "camJuly" (Cameroon July 2023)
- "camOct" (Cameroon October 2023)
- "camMay" (Cameroon May 2025)
- "sind[Sc, Pv]" (Sindara, Gabon, 2023). "Sc" and "Pv" refer to "screening" and "patient visits".
- "lamb[Ek, Nz, Po, Tr]" (Lambarene, Gabon, 2023). "Ek", "Nz", "Po" and "Tr" refer to local collection sites.

"patient" is a number with three digits (for camJuly, sind, and lamb); or with four digits (for camOct and camMay) including leading zeros as needed.

"table" Table tags exist for two sites only: camMay and camOct. The tags distinguish the multiple sessions from each patient. Details are in the "Per site details" section.

Each session folder contains seven .mp4 videos. Video filenames have the form:
"site_patient_video#_focusDepth_timestamp", where "site_patient" matches the session folder name. Video filenames do not include "table" information.

"video#" is 1 to 7 where "1" is imaged first. Video FoVs step along the capillary.

"focusDepth" is an integer 0 to 300. It's the distance in microns of the camera's focus plane below the top surface of the capillary. This might hold relevance because the magnification changes slightly depending on focal depth. Also, "0" or "300" could indicate out-of-focus images.

"timestamp" has the form "hhmmss" (the date has been removed). The timezone is local.

5.2 Per-site details

Summary statistics for all sites are shown in **Table 1**.

Site	Cameroon July 2023	Cameroon Oct 2023	Gabon 2023 Lambarene	Gabon 2023 Sindara	Cameroon May 2025
Site tag	camJuly	camOct	lamb	sind	camMay
# patients	44	1147	95	122	550
# sessions	46	3397	95	122	1011
# Giemsa reads per patient	1	2	2	2	2
# sessions per patient	1	4 *	1	1	2
# *loa*-negative patients	31	788	70	85	353
# loa*loa*-positive patients	13	352	19	24	197
loa count range (mf/mL)	380 – 27,400	20 – 133,180	20 – 77,980	40 – 13,940	20 – 123,980
# patients with loa $\geq$ 30,000 mf/mL	0	31	1	0	33
# *perstans*-negative patients	40	984	88 **	98	NA
# *perstans*-positive patients	4	156	1 **	11	NA
perstans count range (mf/mL)	40 - 400	20 - 7580	100	10 - 560	NA
# *loa* + *perstans* coinfections	0	79	0	5	NA
# video with quality notes	25 (out of 308)	many	all	all	0
# videos with manual mf counts	all	many	all	all	0

Table 1: Summary statistics for all sites.
* Some sessions are missing from this dataset, so 1 – 4 sessions are included per patient.
* 10 patients have missing *perstans* reads (likely 0).

5.2.1 Cameroon July 2023

Videos are in folder "july2023Cameroon" There are 50 sessions, one per patient, from 5 devices (a subset of those used in October Cameroon). The device ID for each session is recorded in "july2023CameroonMetadata.xlsx". The distribution of *L. loa* filaremias is shown in Fig 3, top left.

5.2.2 Cameroon October 2023

Videos are in folder "october2023Cameroon".

The October Cameroon study enrolled 1147 patients. It had a uniquely complex protocol that yielded four sessions per patient: two from a single capillary run through the same device twice (to get intradevice variability on a single capillary), and two from a pair of capillaries, run through two devices (to get interdevice variability from a single finger stick). The protocol is shown in Fig 1.

The 4 sessions per patient are notated by "tables" (in the sense of a workstation where the blood draw was done) A1, A2, BC1, BC2. Due to an upload glitch in the deployed devices, some sessions are missing from this dataset, as follows: Number of sessions per table (out of a possible 1147): A1 = 770; A2 = 723: BC1 = 1101; BC2 = 803 Number of sessions per patient (out of a possible 4): 19 patients have 0; 112 have 1; 286 have 2; 307 have 3; 523 have all 4 sessions.

Three of the tables (A1, BC1, and BC2) can be assumed to be "normal" and equivalent. BC1 and BC2 sessions are from two capillaries filled from the same finger prick, both run concurrently. Table A1 is from a different finger prick (in the other hand), run concurrently with BC1 and BC2. Thus, the variation in a single patient's *loa* burdens in these three sessions is normal.

Table A2 sessions were the result of the A1 capillary being run a second time on the same device. Due to the time delay, A2 sessions have a markedly higher incidence of coagulation, which can begin after just a few minutes. The variation between a patient's A1 and A2 is likely due to coagulation.

October Cameroon metadata includes annotations of coagulation in videos and a dataset for training coagulation detection models. For more details see subsection "Known failure modes – coagulation".

There were 13 devices used. The device ID for each session, along with demographic metadata are found in "october2023CameroonMetadata.xlsx".

There are two expert microscopy reads, on Giemsa-stained blood films from different finger pricks. Intra-patient variation is expected [4], and is shown in Fig 2. Distributions of the *loa* and *perstans* filaremias (by Giemsa) are shown as cumulative distribution functions in Fig 3, bottom left and bottom center.

5.2.3 Cameroon May 2025

Videos for 1011 sessions from 550 patients are in folder "may2025Cameroon". Most patients have two sessions, with different capillaries and devices. These are labeled as tables "A" and "B". There were 14 devices (the same as used in October Cameroon, but not mappable due to naming differences). The device IDs per session and demographic metadata are in "may2025CameroonMetadata.xlsx". *M.*

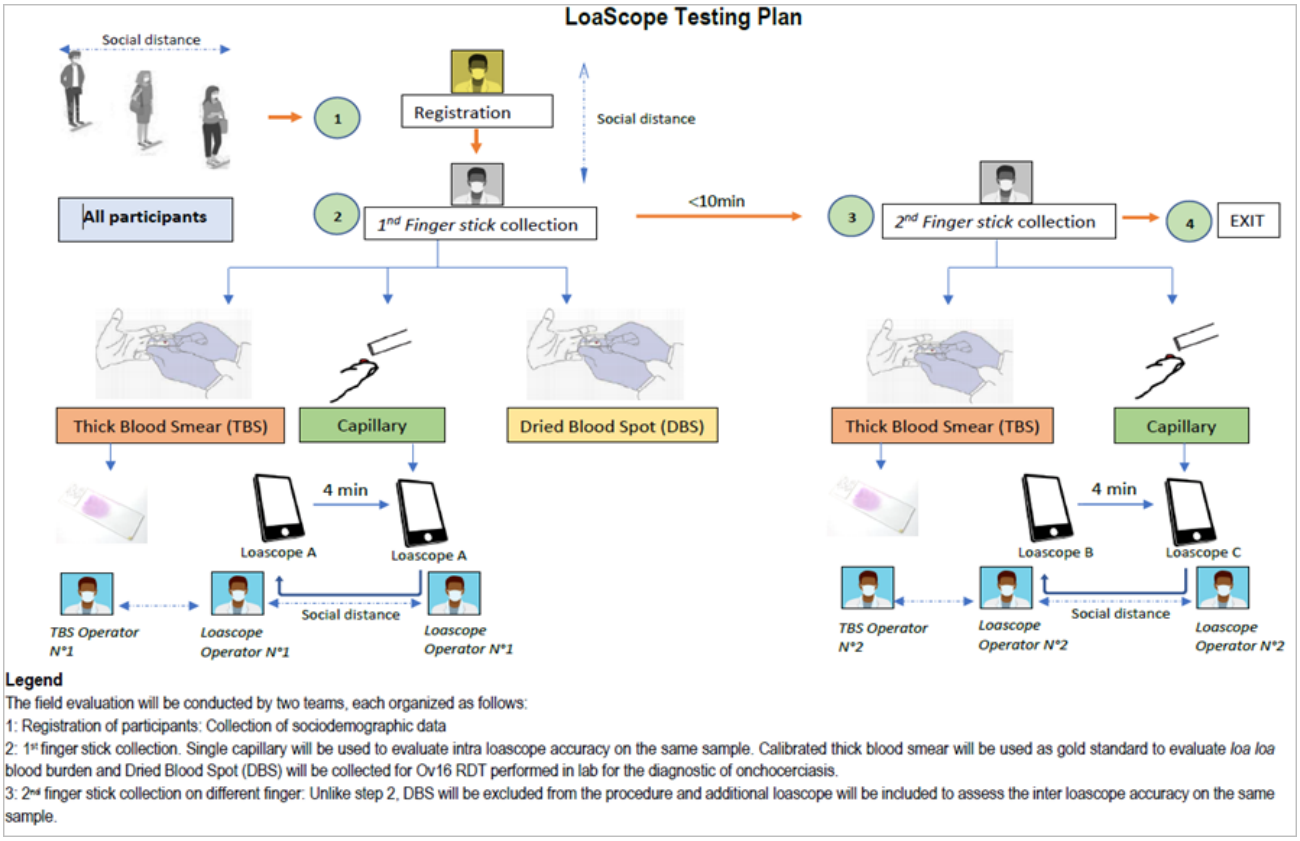

Figure 1: Protocol for October Cameroon ("camOct"). The path on the right resulted in two sessions, one from each of two capillaries (labeled as "tables" BC1 and BC2). The path on the left resulted in two consecutive sessions from a single capillary ("tables" A1 and A2).

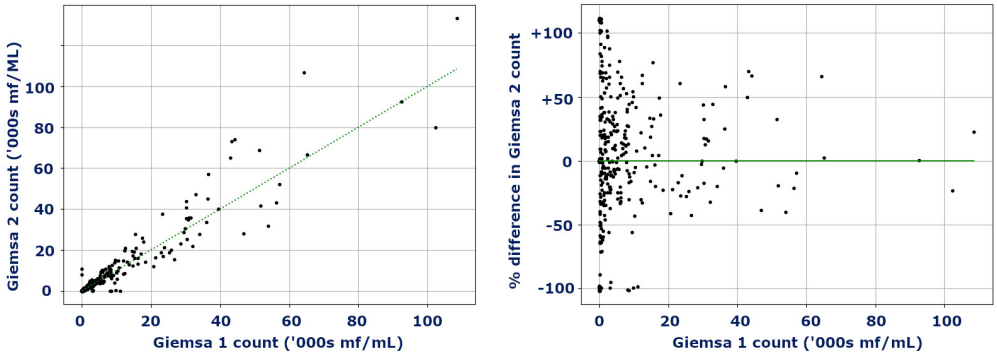

Figure 2: Intra-patient Giemsa read variability, *loa* only (*perstans* counts ignored), from October Cameroon. Left: scatterplot. Right: percent difference of Read 2 from Read 1 as reference, i.e. $100\frac{(read2-read1)}{read1}$. Values >100 were clipped to 100. Then all y-axis values were jittered.

perstans was not counted separately. The distribution of the *loa* + *perstans* filaremias (by Giemsa) is shown in Fig 3, bottom right.

5.2.4 Gabon 2023

Videos for 217 sessions are in the folder "gabon2023". There were two main sites: Lambarene and Sindara. Lambarene had four sub-sites: Ekouk, Nzemba, Petit Odavo, and Tranquille. Also, a few patients returned for later visits. Thus the tags for Lambarene sessions are lambEk, lambNz, lambOd, lambTr, and lambPv (for "Patient Visit"). There is no "table" tag. Sindara has one session per patient and no "table" tag. The distributions of Lambarene and Sindara *loa* filaremias are shown in Fig 3, top center and top right.

5.2.5 Deliberately coagulated sessions

Ten patients each had a capillary run through the device repeatedly to study coagulation. These sessions are found in the folder "coagulationExperiments". There is no metadata file. The timestamps on these videos offer valuable insight into coagulation times. Two patients from July Cameroon have sessions in "coagulatedJuly2023Cameroon". Eight patients from May 2025 Cameroon have sessions in "coagulatedMay2025Cameroon".

5.3 Failure modes

Some videos show certain failure modes, listed below. Quality and Coagulation annotations identify some of these videos (see section 5.4), especially in October Cameroon.

Bulk motion Ideally, the blood in the capillary is motionless, except as disturbed by swimming mf, and the only perturbation of pixel values from frame to frame is due to mf. However, in the Cameroon 2023 dataset, several videos contained regions of flowing blood ("bulk flow") due presumably to leaky capillaries. This flow creates perturbations from frame to frame, creating risk of erroneous mf detections. In a similar effect from different cause, bumping the device during imaging can induce bulk motion (e.g. a ripple-in-pond effect).

Flickering illumination In a few cases, videos appeared to have a flickering effect. This would cause pixel value differences and thus apparent motion which might risk erroneous mf detections.

Coagulation If the capillary is not imaged soon enough, the blood can begin to coagulate, which prevents movement by the mf (heparinized blood or capillaries are not logistically practical). This leads to False Negative samples in the TaNT use case. A common visual difference between coagulated and liquid samples is, roughly speaking, that liquid samples have a more uniform orange color, while coagulated samples have a more "checkerboard" pattern of white and red, since as the blood coagulates it pulls away from small regions in the capillary (see Fig 4). This was mainly seen in table A2 of the October Cameroon study, due to the unique protocol.

5.4 Ground truth data

There are four types of ground truth for sessions and videos:

1. Patient-level Giemsa counts (all sessions)
2. Manual counts of mf per video (some videos)
3. Manual quality annotations per video
4. Manual coagulation labels per video (some videos).

5.4.1 Giemsa counts per patient

The most important ground truth, corresponding to current clinical practice, is mf count (mf/mL) on Giemsa-stained blood films. All patients have an expert mf count by microscopy on 50 μL of Giemsa-stained blood from the same finger prick as the capillary fill, but not the actual blood that entered the capillaries.

An important aspect of this ground truth is variability: Intersample variability in mf counts from synchronous samples (e.g., from one finger prick, from two finger pricks

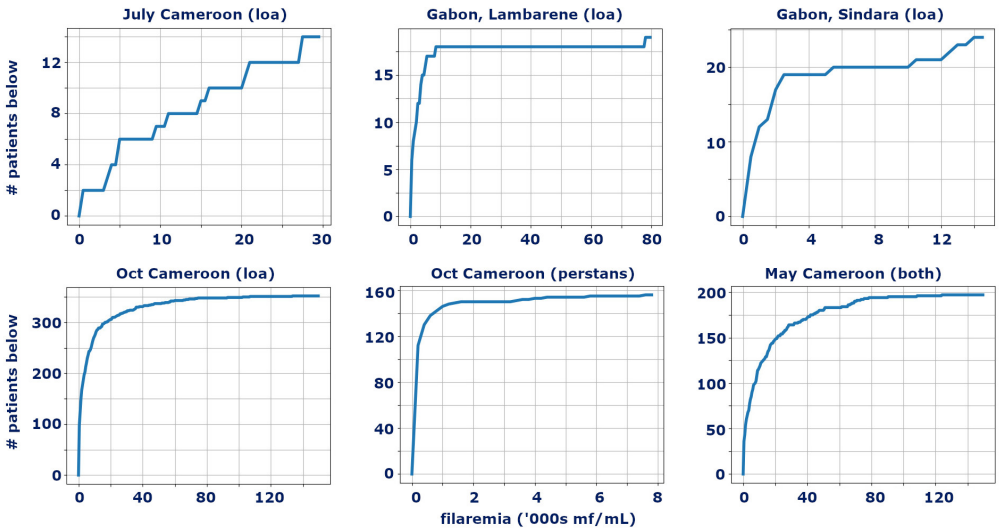

Figure 3: Distribution of positive filaremias shown as cumulative distribution functions. (Top row, L to R): July Cameroon (*loa*), Gabon Lambarene (*loa*), Gabon Sindara (*loa*). (Bottom row, L to R): October Cameroon (*loa*), October Cameroon (*perstans*), May 2025 Cameroon (*loa* + *perstans*). x-axis = filaremia; y-axis = number of patients below that value.

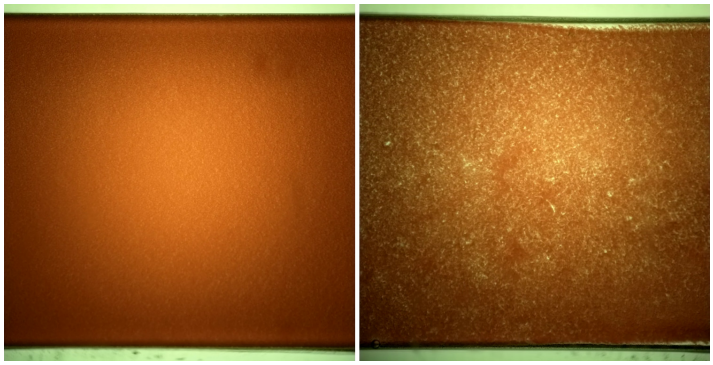

Figure 4: Normal (left) and Coagulated (right) video frames. Most videos in the dataset look normal, but Table "A2" of October Cameroon has videos spread between these two extremes.

from the same different hands, etc) is not well characterized but is known to be potentially high (Gardon et al., 1997). The October Cameroon study, uniquely, had two Giemsa counts from synchronous finger pricks from different hands. This enables insight into the variability between Giemsa counts, as shown in Fig 2. The Giemsa counts are found in (spreadsheet / sheet):

- july2023CamMetadata.xlsx / fieldMetadata
- oct2023CamMetadata.xlxs / fieldMetadata, also / algorithm_mfPerML_perTable
- gabon2023Metadata.xlsx / *_Giemsa
- may2025CamMetadata.xlsx / fieldMetadata

5.4.2 Manual counts per video

For some of the videos, humans watched the videos and counted the mf with visible paths (they did not mark the x-y locations). These counts are not perfect, especially when the video contains over 30 mf since accurate tracking becomes difficult. The manual counts are significantly lower per volume of blood than the Giemsa counts. The capillaries appear to have a sieving effect, so that the count of mf inside a capillary is about 50% of the count in the original blood sample. The per-video counts are found in:

- july2023CamMetadata.xlsx / humanMfPerVid_Quality_Device
- oct2023CamMetadata.xlsx / humanMfPerVideo_qualityNotes
- gabon2023Metadata.xlsx / *_humanMfPerFov

5.4.3 Quality annotations per video

Many videos in the October Cameroon and Gabon studies were manually inspected for quality issues, such as the failure modes described above. These notes are found in:

- july2023CamMetadata.xlsx / humanMfPerVid_Quality_Device
- oct2023CamMetadata.xlsx / humanMfPerVideo_qualityNotes
- gabon2023Metadata.xlsx / *_Quality

5.4.4 Coagulation labels per video

If there is a time delay (roughly >3 - 4 minutes) between blood draw and imaging, blood can coagulate in the capillary before or during imaging. Identifying coagulation matters to: (i) properly evaluate undercounting errors in filaria detector algorithms (mf are undetectable in coagulated videos due to lack of motion); and (ii) assemble train-test sets for coagulation detector dev. There are five sources for coagulated videos:

1. 62 complete sessions from October Cameroon were manually annotated for coagulation with labels "normal", "coagulated", or "interim". The videos in a given session can have different labels (if the blood was coagulating during imaging). "Interim" labels should be omitted from the train set for a video classifier because their status is unclear. For realistic validation, all 7 of a session's videos should be checked, and the session should receive a "coagulated" label if even one of its videos is labeled as coagulated. This coagulation dataset is listed in coagulationOct2023Cam.xlsx. The train-test split is stratified by session. This training set can be readily expanded by applying criteria 2 - 5 below.

2. There are deliberately coagulated sessions (see 5.2.5).

3. Compare algorithm vs Giemsa mf counts: If a session's algorithm count is much lower than its Giemsa count, it is a suspect for coagulation. Conversely, for positive patients, if a session's algorithm count is higher or the same as the Giemsa count, then it is likely uncoagulated (unless the session exhibits bulk motion or illumination flicker, as described in "Session failure modes" above, since these can produce erroneous high counts by the 2026 algorithm).

4. Table A2 of the October Cameroon study are good candidates for coagulation, due to the time delays required by the protocol.

5. A few Gabon videos exhibit coagulation, and are listed in gabon2023Metadata.xlsx / *_Quality

6. Validation

The videos are the raw .mp4 data, downsampled using Python and openCV to read and write .mp4s; 2× decimation) (Python Software Foundation; Bradski, 2000).

For ground truth QC, see "Data acquisition" and "Ground truth data" subsections.

Acknowledgments

Anonymized.

Conflicts of Interest

Anonymized.

References

C. Boulle, S.D. Pion, M. Boussinesq, and alia. Revisiting the *Loa loa* microfilaremia thresholds above which serious adverse events may occur with ivermectin treatment. *PLoS Neglected Tropical Diseases*, 2025.

M. Boussinesq. Loiasis. *Annals of Tropical Medicine & Parasitology*, 100:715–731, 2006. .

M. Boussinesq, J. Gardon, N. Gardon-Wendel, and J-P. Chippaux. Clinical picture, epidemiology and outcome of Loa-associated serious adverse events related to mass ivermectin treatment of onchocerciasis in Cameroon. *Filaria Journal*, 2003.

G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

K. G. Buell, C. Whittaker, M. Boussinesq, and alia. Atypical clinical manifestations of loiasis and their relevance for endemic populations. *Open Forum Infectious Diseases*, 6:ofz417, 2016. .

C. B. Chesnais, I. Takougang, M. Boussinesq, and alia. Excess mortality associated with loiasis: A retrospective population-based cohort study. *Lancet Infectious Diseases*, 17:108–116, 2017. .

M. V. D'Ambrosio, M. Bakalar, D.A Fletcher, and alia. Point-of-care quantification of blood-borne filarial parasites with a mobile phone microscope. *Science Translational Medicine*, 2015.

M. de Leon Derby, Z.L. Moussa, D.A. Fletcher, and alia. NTDscope: A multi-contrast portable microscope for disease diagnosis, 2025.

L. Djune-Yemeli, C.B. Delahunt, J. Kamgno, and alia. Automated detection of loa loa: a field trial in cameroon. *In Review*, 2026.

J. Gardon, N. Gardon-Wendel, M. Boussinesq, and alia. Serious reactions after mass treatment of onchocerciasis with ivermectin in an area endemic for Loa loa infection. *Lancet*, 1997.

J. Goldin and A. L. Juergens. Filariasis. In *StatPearls*. StatPearls Publishing, Treasure Island, FL, 2025.

M. C. Hemilembolo and et al. *Loa loa* and *Mansonella perstans* microfilaremia in the department of Lékoumou, Republic of Congo. *Parasites & Vectors*, 2023.

K.H. Jacobsen, B.C. Andress, A.S. Williams, and alia. A call for loiasis to be added to the WHO list of neglected tropical diseases. *Lancet Infectious Diseases*, 22(10): e299–e302, 2022. .

T. J. Kamtchum and alia. Epidemiology of *Loa loa* and *Mansonella perstans* filariasis in the Akonolinga Health District, Centre Region, Cameroon. *Health Sciences and Diseases*, 2014.

L. A. Kelly-Hope, M. Bockarie, and D. H. Molyneux. *Loa Loa* ecology in Central Africa: Role of the Congo River System. *PLoS Neglected Tropical Diseases*, 6:e1605, 2012. .

W. G. Metzger and B. Mordmüller. *Loa Loa*—does it deserve to be neglected? *Lancet Infectious Diseases*, 14: 353–357, 2014. .

S. D. S. Pion. Implications for annual retesting after a test-and-not-treat strategy for onchocerciasis elimination in areas co-endemic with *Loa loa* infection: an observational cohort study. *Lancet Infectious Diseases*, 2020.

Python Software Foundation. URL `https://www.python.org/`).

M. Ramharter, S. T. Agnandji, A. A. Adegnika, P. G. Mremsner, and alia. Development of sustainable research excellence with a global perspective on infectious diseases: Centre de recherches médicales de lambaréné (cermel), gabon. *Wiener klinische Wochenschrift*, 2021.

M. Ramharter, J. Butler, R. Zoleko Manego, and alia. The african eye worm: current understanding of the epidemiology, clinical disease, and treatment of loiasis. *Lancet Infectious Diseases*, 2024. .

L. Veletzky, J. Hergeth, D. R. Stelzl, M. Ramharter, and et alia. Burden of disease in Gabon caused by loiasis: a cross-sectional survey. *Lancet Infectious Diseases*, 2020.

L. Veletzky, K. A. Eberhardt, M. Ramharter, and alia. Analysis of diagnostic test outcomes in a large loiasis cohort from an endemic region: Serological tests are often false negative in hyper-microfilaremic infections. *PLoS Neglected Tropical Diseases*, 18(3):e0012054, 2024. .

M. Walther and R. Muller. Diagnosis of human filariases (except onchocerciasis). In *Advances in Parasitology*. Academic Press, 2003.

WHO. *Basic Laboratory Methods in Medical Parasitology*. World Health Organization, 1991. ISBN 92-4-154410-4.

WHO. Rapid assessment procedures for loiasis: Report of a multi-centre study. `http://54.213.233.103/static/uploaded_files/loiasis.pdf`, 2001.

WHO. Loiasis: A Very Neglected Disease in Central Africa. `https://www.who.int/news-room/events/detail/2025/12/04/default-calendar/loiasis--a-very-neglected-disease-in-central-africa`, 2025a.

WHO. Lymphatic filariasis. `https://www.who.int/health-topics/lymphatic-filariasis`, 2025b.

R. Zoleko Manego, G. Mombo-Ngoma, M. Witte, P. B. Matsiegui, and alia. Demography, maternal health and the epidemiology of malaria and other major infectious diseases in the rural department Tsamba-Magotsi, Ngounié Province, in central African Gabon. *BMC Public Health*, 2017.