Arjun Sarkar [1,2], Jan-Philipp Praetorius [1,2], Marc Thilo Figge [1,3]

[1] Applied Systems Biology, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Jena, Germany
[2] Faculty of Biological Sciences, Friedrich Schiller University, Jena, Germany
[3] Institute of Microbiology, Faculty of Biological Sciences, Friedrich Schiller University, Jena, Germany
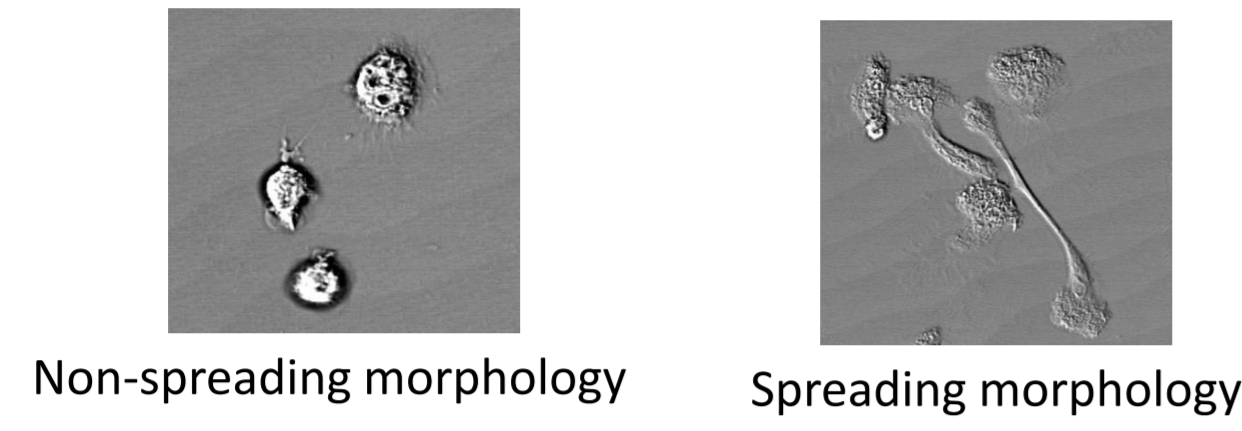
Single cell morphodynamics analysis

- Whole blood infection – 9 Donors [1]
- Time-lapse microscopy data
- 3 cases each (Mock, *C. albicans*, *C. glabrata*)
- Each video with 260 frames / images
- Video duration: 30 min, Time-step: 7 sec
- Neutrophils exhibit two morphological appearances

Non-spreading morphology

Spreading morphology

Single cell morphodynamics – classifier on cell tracks
- AMIT_v3 [2] (Algorithm for Migration and Interaction Tracking) for tracking and splitting cells
- Split cells merged across 40 frames to form videos
- EfficientNetB0 [3] (Convolutional Neural Network - CNN) for feature extraction
- Transformer network for classification
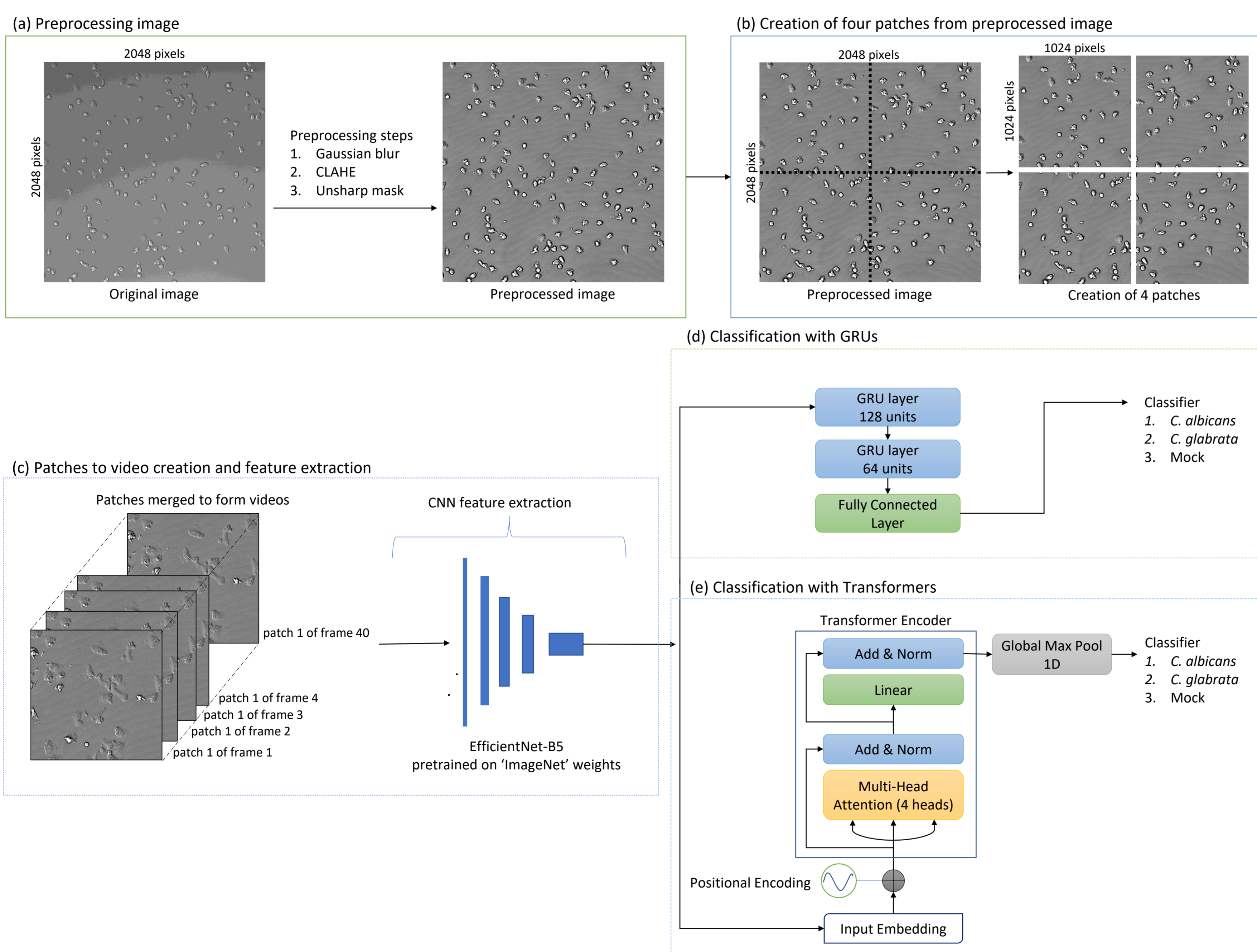- LOOCV - Leave One Out Cross-Validation applied
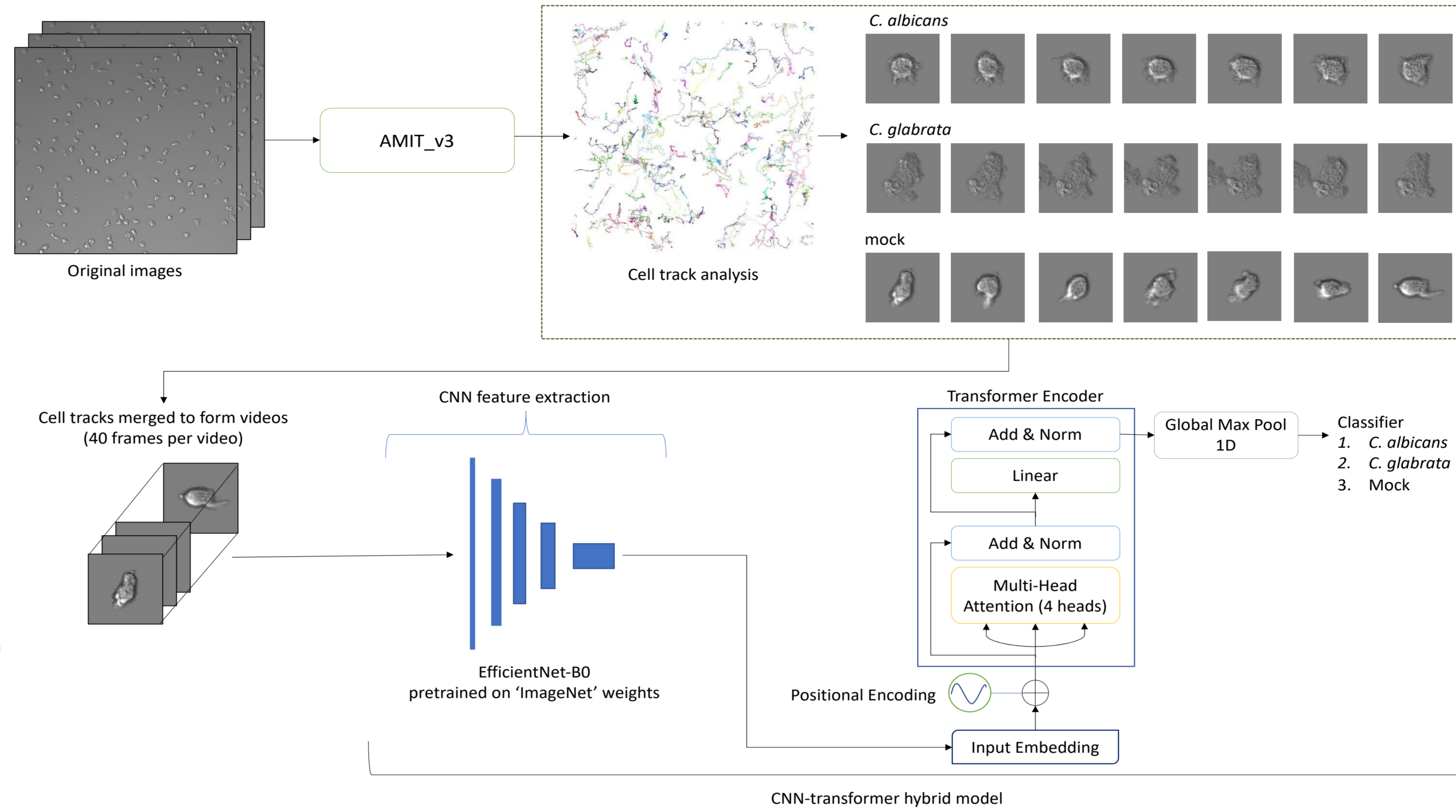
Results
- Mock classified correctly
- Network confuses between *C. albicans* and *C. glabrata*

Confusion matrix – transformers (LOOCV) on single cell videos

GRU (Gated Recurrent Unit) and Transformer based analysis of video data

(a) Preprocessing image

(b) Creation of four patches from preprocessed image

(c) Patches to video creation and feature extraction

(d) Classification with GRUs

(e) Classification with Transformers

Workflow
1. Preprocessing for perfect separation between background and neutrophils
   - Gaussian blur - reduce noise from the images
   - CLAHE (contrast-limited adaptive histogram equalization) - contrast enhancement
   - Unsharp Mask – sharpen image
2. Creating patches to increase dataset size
3. Patches merged across 40 frames to form videos
4. CNN used for feature extraction
   - CNN used - Pretrained EfficientNetB5 [3]
5. Classifier network
   - Extracted features through GRU [4] and Transformer [5] network

Results
- Pro: Transformer [5] classifies all videos correctly
- Con: Long inference time due to time-consuming video formation and feature extraction steps (up to 3 hours for each case)
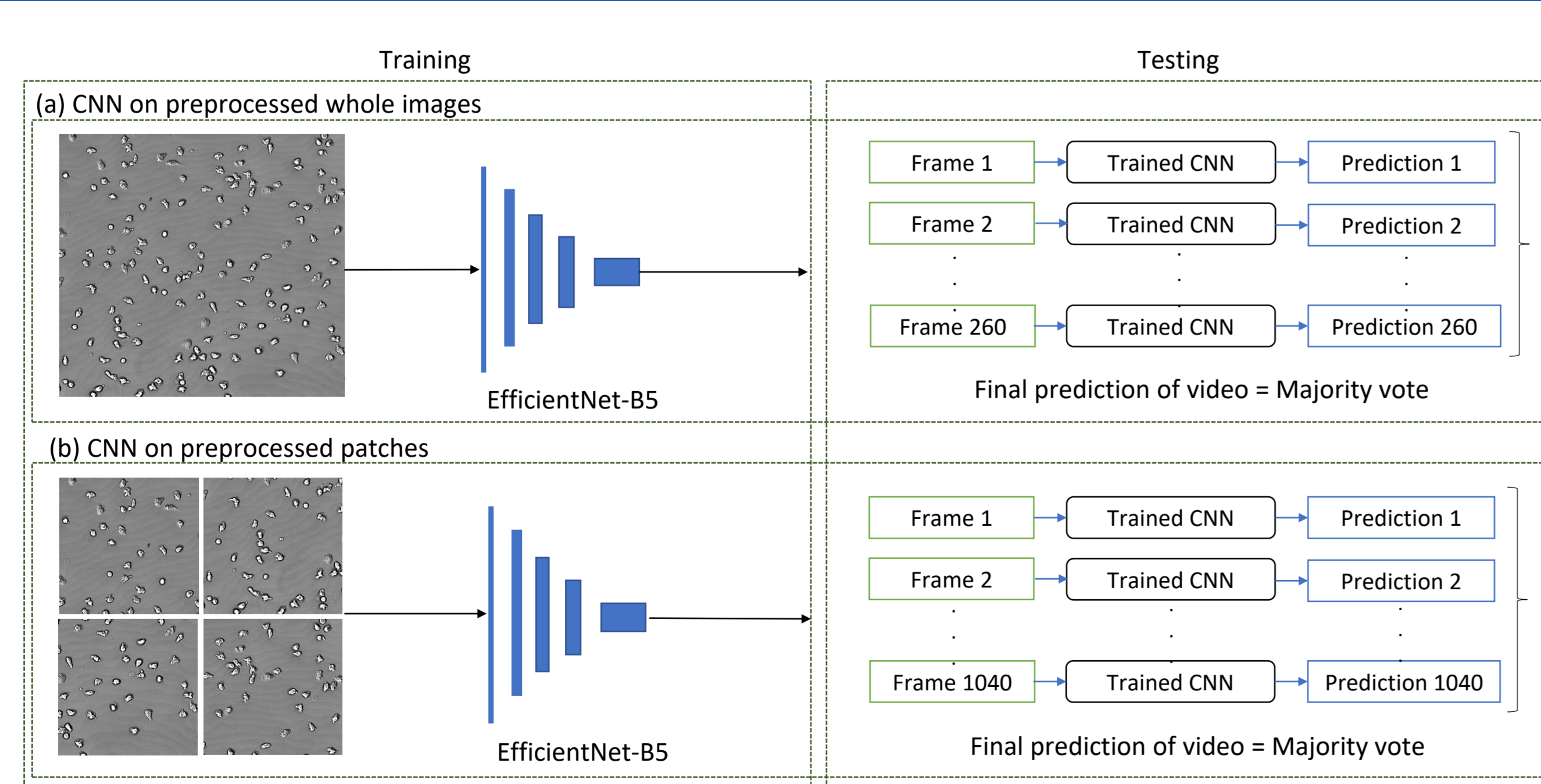
Confusion matrix – (d) GRUs (LOOCV) on preprocessed patch videos

Confusion matrix – (e) Transformers (LOOCV) on preprocessed patch videos

CNN based analysis on image data

Training

Testing

(a) CNN on preprocessed whole images

(b) CNN on preprocessed patches

Final prediction of video = Majority vote

Confusion Matrix

Confusion matrix – (a) CNN (LOOCV) on preprocessed whole images

Confusion matrix – (b) CNN (LOOCV) on preprocessed patches
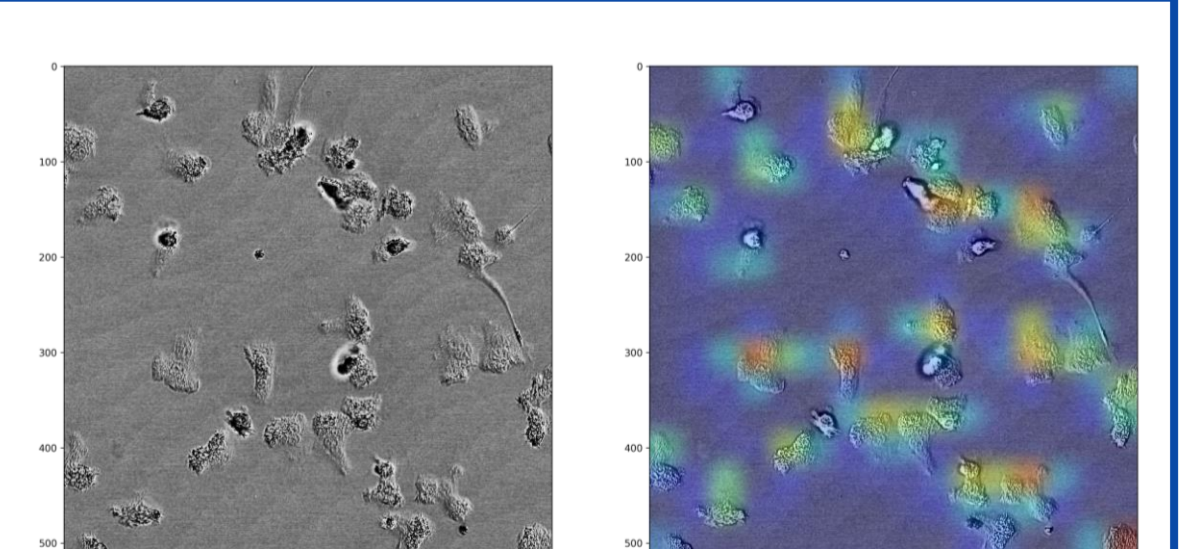
Grad-CAM on *C. glabrata* shows CNN focusing on the spreading cells to make predictions

Workflow
1. Preprocessed images and patches trained using CNN – EfficientNetB5 [3]
2. Majority vote across 260 frames gives final prediction

Results
- CNN on patches classifies all videos correctly
- Very fast inference time (15 minutes)
- Grad-CAM [6] for explainability
- Needs fewer images for inference
- Ensemble learning for best results

arjun.sarkar@leibniz-hki.de

www.leibniz-hki.de

References
[1] Belyaev, I., & Praetorius, JP. *et al*. 2022. Comput struct biotechnol j.,20, (2297-2308)
[2] Belyaev, I. *et al*. 2021. Cytom Part A. cyto.a.24466
[3] Tan, M., & Le, Q. 2019. ArXiv, abs/1905.11946.
[4] Chung, J., *et al*. 2014, ArXiv:1412.3555
[5] Vaswani, A., *et al*. 2017, ArXiv:1706.03762
[6] Selvaraju, RR., *et al*. 2017, ArXiv:1610.02391