Machine Learning to Automate the Visual Interpretation of Chemical Agglutination Tests

Sidharth Gupta[1,3], Alexandros A. Sklavounos[1,4], Joshua Dahmer[1,4], Anthony K.C. Yong[1,4], Mohammed A.A. Abdullah[1,4], Gilberto Camacho[1,5], Keith Morton[2,6], Matthew Shiu[2,6], Jean Labrecque[2,6], Teodor Veres[2,6], Aaron R. Wheeler[1,4], Alex Mariakakis[1,3]

University of Toronto[1], National Research Council of Canada[2], Department of Computer Science[3], Department of Chemistry[4], Institute of Biomedical Engineering[5], Life Sciences Division[6]

{sid.gupta, alexandros.sklavounos}@mail.utoronto.ca, josh.dahmer@utoronto.ca
anthony.yong@mail.utoronto.ca, mohammed.abdullah@utoronto.ca, gilberto.camacho@mail.utoronto.ca
{keith.morton, matthew.shiu, jean.labrecque, teodor.veres}@cnrc-nrc.gc.ca
{aaron.wheeler, alex.mariakakis}@utoronto.ca

2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) | 978-1-6654-1647-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/PerComWorkshops53856.2022.9767280

***Abstract*—Many point-of-care tests rely on visual changes in color, shape, and size to convey results that can be read by the naked eye. One category of such tests is an agglutination test (AT), which relies on the clumping of micro-particles or cells in the presence of a target analyte. Although visual inspection is convenient and fast, it is subjective, prone to errors, and limits decision-making to coarse-grained results. We present an open-source software framework designed to facilitate the development and interpretation of ATs. This framework includes a web-based annotation interface for curating new image datasets, a computer vision pipeline that extracts informative AT features, and a machine learning module that allows AT developers to study how an AT agglutinates over time during future experiments. We present two case studies of our framework being used to develop and interpret tests.**

***Index Terms*—agglutination, image processing, annotation**

I. Introduction

Pregnancy tests and pH strips are just some of the many examples of commodity chemical tests that present their results through visual changes. More complex tests have been developed for a variety of problems ranging from water testing to virus detection [1], [2]. In this work, we focus our attention on agglutination tests (ATs), in which clumps of either micro-particles or cells are formed in the presence of a target analyte [3]. ATs have been used in clinical pipelines to assist in disease diagnostics via the detection of viruses, bio-markers, and small molecules [4]–[6], as well as to perform blood-grouping for pre-transfusion compatibility testing [7].

Because these tests are fast, inexpensive, and easy to interpret, they have become valuable point-of-care solutions that are pervasive around the world. However, vast literature suggests that the convenience of visual inspection comes with its own limitations [8]–[11]. First, such checks are highly subjective and error-prone since people can have varying visual acuity and experience in reading the results. Second, visual inspection limits decision-making to at best a few levels (e.g., negative, slightly positive, very positive). Third, the reaction rate of some ATs can yield important information, yet it is impractical for people to continuously monitor their tests. These issues not only hinder non-expert end-users who use ATs for diagnostic purposes, but also the chemists who develop these tests. Being able to automatically and instantaneously quantify the state of an AT would accelerate their development and improve their accuracy in the field.

In this work, we present an open-source framework that we have developed to automate the interpretation of ATs[1]. This framework includes three components: (1) a web-based annotation interface that can be used to curate new labeled image datasets, (2) a computer vision pipeline that measures visual features specific to ATs, and (3) a machine learning module that allows AT developers to study how an AT agglutinates over time during future experiments. Our framework, which is illustrated in Figure 1, facilitates the creation of new ATs as it allows developers to quickly label, examine, and evaluate new image datasets for novel AT formulations. Once a refined model has been generated, AT developers can deploy it as part of an accompanying smartphone app with their AT to enable robust and reliable point-of-care testing for end-users. After we describe this framework in detail, we demonstrate its utility in two case studies: one involving blood grouping and another involving SARS-CoV-2.

II. Data Collection and Annotation

The process for generating a labelled image dataset starts with using a camera to record AT experiments. Our framework only imposes two requirements on this procedure. The first requirement is that the experiments should be designed such that the AT exhibits the full range of agglutination scores, which is usually a trivial outcome of AT development. The second requirement is that the AT developer should be able to process the recording to generate tightly cropped images of the AT at various time points. An important consideration during the data collection process is determining which video

[1]https://github.com/sidguptacode/ML_AT_Interpretation

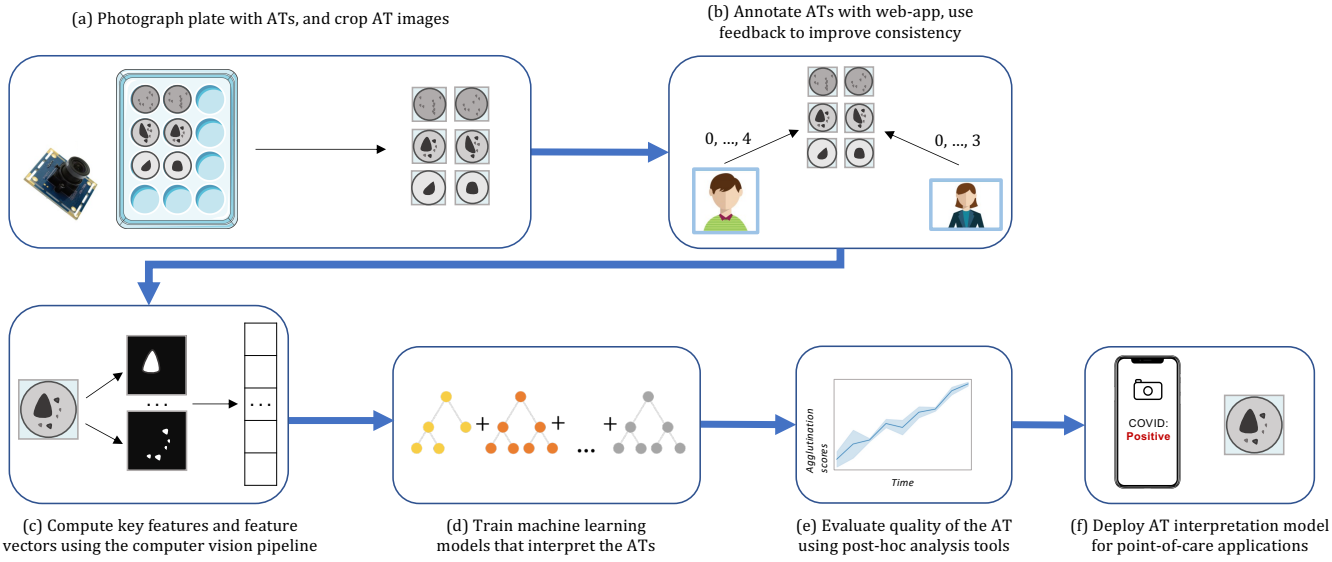

Fig. 1. Overview of the AT framework: (a) AT developers collect images of AT experiments, and if needed, each image is segmented or cropped; (b) annotators use our web annotation framework to assign agglutination scores; (c) images are processed to extract computer vision features specific to ATs; (d) machine learning models are trained to predict agglutination scores; (e) post-hoc analyses are conducted on the predictions to evaluate the quality of the AT during the various experiments; and (f) the AT and machine learning model can be deployed as a point-of-care system after sufficient accuracy is achieved Steps (a)-(e) are repeated as AT developers iterate on their assay's formulation.

frames to keep in the image dataset. Most cameras record at either 30 fps or 60 fps, yet most point-of-care tests take up to 30 minutes. This means that video recordings will contain many redundant frames for ATs with longer reaction times. Furthermore, many ATs change in a non-linear fashion — quickly at first, then gradually approaching its steady state — so labelling frames at a uniform interval over time can lead to an uneven distribution of labels in the dataset. We explain our frame selection procedure for our case studies later in this paper, but the optimal strategy for selecting which frames to select is highly application-specific.

Once the dataset has been generated, AT developers and recruited annotators can use our web-based interface to label the images with agglutination scores (Figure 2). The scores come from a standardized scale that ranges from 0 to 4, where a 0 indicates no reaction and a 4 indicates that a single large clump has formed [7]. The interface presents annotators with one image at a time in a randomized order to mitigate potential anchoring bias across consecutive frames from the same experiment [12]. Annotators can flag images for which they are uncertain about the label; these images are then escalated for other annotators to review. The interface also includes a variety of features in the periphery to assist annotators. At the bottom left-hand corner of the interface, annotators can read the formal criteria for the different agglutination scores. At the right side of the interface, annotators can cycle through prototypical example images for each score to facilitate side-by-side comparisons. These examples are uploaded by a lead investigator to encourage consistency across multiple annotators, and the examples for each score can be gradually expanded over time as the dataset gets labeled.

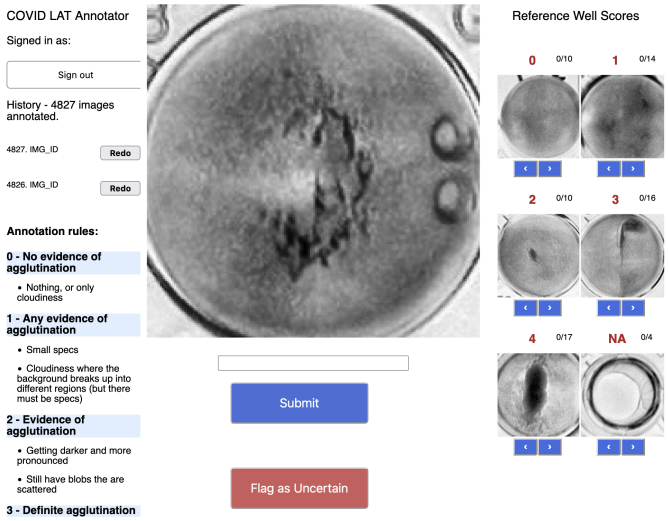

Fig. 2. Our web-based annotation interface provides annotators with the formal criteria for different agglutination scores and reference images to facilitate side-by-side comparisons.

Our interface also allows annotators to cycle through images with conflicting labels, and computes confusion matrices for the labels between all pairs of annotators. These features allow teams of annotators to identify troublesome images and converge on their criteria for future labelling.

III. Image Processing

The second component of our framework is a computer vision pipeline for automatically calculating visual features that characterize the degree of agglutination in an AT. The pipeline starts with a pre-processing stage for removing noise

TABLE I
The list of visual features automatically computed in our framework to characterize the degree of agglutination in an AT.

Region of interest	Feature Name
Entire image	Relative standard deviation of brightness
All clumps	Total area
	Total perimeter
	Mean brightness
	Standard deviation of brightness
	Relative standard deviation of brightness
Largest clump	Area
	Perimeter
	Mean light intensity
	Standard deviation of brightness
	Relative standard deviation of brightness
Background	Mean brightness
	Standard deviation of brightness
	Relative standard deviation of brightness

and isolating the important elements in each image. We first normalize the brightness across each image using contrast-limited adaptive histogram equalization (CLAHE). We then use global image thresholding to create a binary mask that locates any clumped agglutinates in the AT. After the mask is smoothed using morphological opening and closing operations, it is applied onto the contrast-enhanced image to isolate the largest clump (if one exists). We separately apply the inverted mask onto the same image to isolate the background. It is at this point that we compute the visual features listed in Table I, which are designed to characterize the scatteredness, cloudiness, and structural integrity of clumps within an AT.

Although we use our case studies to demonstrate how our image processing pipeline generalizes across AT designs, AT developers may still be interested in adapting or extending it for their own purposes. Our framework provides multiple user-friendly ways for developers to do so. First, we provide developers with a configuration file that allows them to add, remove, and re-order image processing steps. The file also allows developers to adjust parameters associated with those image processing steps, such as the threshold for image binarization. Second, our system creates and saves visualizations after each pre-processing step so that developers can ensure that the parameters are suitable for their particular ATs.

IV. Machine Learning and Analysis

The third component of our framework is a machine learning module that allows an AT developer to automatically score and study the agglutination of future experiments for the sake of test development. This module first requires training a model that processes the aforementioned computer vision features and returns an estimated agglutination score (0 to 4). Although the agglutination score scale is based on discrete integer levels, we use regression to preserve the ordering between consecutive scores and to encapsulate some notion of uncertainty; for example, it is equally likely that a 3.5 should be scored as 3 or as a 4. Converting predictions to integer scores on the scale is as simple as rounding. We use gradient boosted trees [13] in this work since they are able to avoid nonsensical interactions between unrelated features, yet other architectures can be used if need be.

Before training a single model for prediction on future experiments, we first use 10-fold cross-validation to compute the model's accuracy across the labeled dataset. The folds are split such that frames from the same AT are kept within a single fold to avoid potential information leak. Assuming the AT developer is satisfied with the cross-validated accuracy, a final model is trained across all of the labeled data. The AT developer can then use that trained model to process future experiments as long as the mechanism for agglutination does not change. Even if the AT developer changes their assay's formulation such that the AT is more sensitive to lower concentrations, the same model can be used since it predicts an agglutination score rather than a concentration.

To help AT developers with their decision-making as they test various assay formulations, we provide them with the tools they need to easily generate a graph showing the agglutination score over time for a set of AT experiments. Developers can compare experiments at the same concentration with different assay formulations in order to decide which formulation is most sensitive. Meanwhile, they can also compare experiments with the same assay formulation at different concentrations in order to calculate the AT's limit-of-detection (LOD); an example of such a graph is shown in Figure 5.

V. Evaluation

We now present two examples of our framework being used to accelerate point-of-care test development: one for a blood-grouping test and another for a SARS-CoV-2 immunoassay.

A. Case Study 1: Blood Grouping

Blood grouping is a process that classifies human blood into categories (e.g., ABO and Rhesus) in order to determine the compatibility between two people for a blood transfusion or organ transplant. Although these procedures are often conducted in hospitals with sophisticated equipment, more accessible point-of-care blood grouping is needed in rural clinics that have to rely on a distant laboratory for blood grouping results. Emergency personnel who deliver care in response to roadside accidents would also benefit from point-of-care testing since they require results within minutes to deliver timely treatment.

Prior work from Sklavounos et al. [14] presented an AT specifically designed to determine the grouping of a blood sample. In short, blood sample droplets are mixed with droplets containing either anti-A, anti-B, anti-AB, or anti-D antibodies in separate chambers within the microfluidic cartridge, and the clumping of the red blood cells is used to identify the blood's typing. Each chamber utilizes the same agglutination mechanism, allowing us to mix images from different chambers for the sake of model training. An example of the AT is shown in Figure 3a.

Our goal was to develop software that could automatically interpret the test results presented by this system. To do this, we first used the AT to analyze blood samples from volunteers

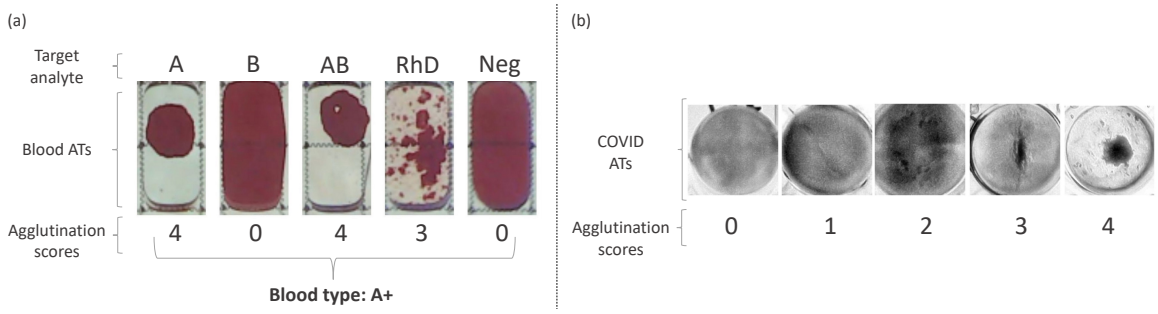

Fig. 3. (a) Blood grouping AT images with associated agglutination scores that map to a blood type group; (b) SARS-CoV-2 AT images with agglutination scores. Note that a large clump area denotes a lower score for the blood-grouping AT but a higher score for the SARS-CoV-2 AT.

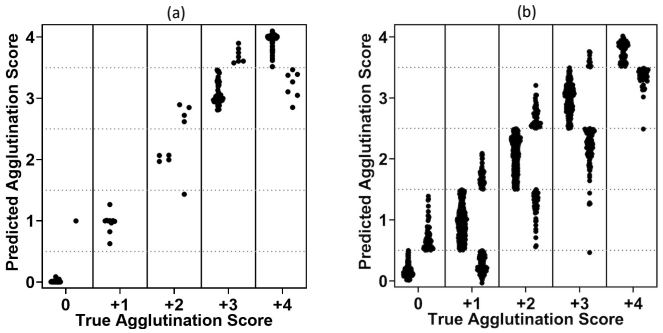

Fig. 4. Swarm plot showing the performance of our model on the (a) blood grouping dataset and (b) SARS-CoV-2 dataset in predicting agglutination scores against the ground truth.

that spanned all 8 blood types. Each sample was photographed after the test's required 2-minute wait time using a low-cost web camera (3264×2448 px) mounted at a fixed position relative to the digital microfluidics cartridge. Each image was cropped automatically around the individual chambers corresponding to each antibody and the negative control. After some images were removed due to device failures, we were left with 567 images of individual chambers. Two annotators assigned agglutination scores to the images, with all disagreements being resolved until a unanimous decision was reached. This procedure left us with an inter-rater reliability score of 1.00 according to Cohen's κ and the following distribution for labels 0 through 4: 235, 9, 14, 48, and 261 images.

Figure 4a shows the predicted agglutination scores plotted against the agglutination scores assigned by the annotators. The mean-squared error of the predictions was 0.026, and the inter-rater reliability between the predictions and the annotators was 0.94. Grading hemagglutination reactions has several additional benefits, such as identifying cases of uncertainty, detecting abnormal or weak reactions, and measuring dosage effects of antigens. Nevertheless, most end-users utilize these tests to make binary decisions where positive samples correspond to any non-zero agglutination score. Using a decision rule in which all non-zero scores correspond to positive results, our approach yielded an accuracy of 97.0% with a false negative rate of 0.0% and a false positive rate of 0.2%.

B. Case Study 2: SARS-CoV-2 Detection

The COVID-19 pandemic has had global societal and economic impact over the past two years. One of the best countermeasures against the spread of COVID-19 has been rapid testing since it helps identify people who are in need of care in a timely fashion. Accessible testing is also important so that people can know that they should isolate themselves in order to reduce the chances that they will infect others.

In response to this crisis, our team began developing an AT that detects the presence of SARS-CoV-2 — the virus causing COVID-19. The end goal of this effort was to deploy a low-cost and accessible SARS-CoV-2 immunoassay suitable for use by the general public. The AT analyzes a sample of bodily fluids (e.g., saliva) by mixing it with latex particles that were designed to bind to SARS-CoV-2. An example of this AT is shown in Figure 3b.

The experiments were conducted in a 96-well plate, with each well containing either a different concentration of SARS-CoV-2 or a different assay formulation. The experiments were recorded using a Raspberry Pi camera (4064×3040 px, 30 fps) mounted below the transparent, round-bottomed well plate. Frames from every 5-minute mark of the test were annotated since that spacing led to a reasonable distribution of agglutination scores. Since the wells were circular, we were able to automatically crop the images using the Hough circle detection algorithm. Four annotators were responsible for scoring the images. After collectively annotating a shared set of 40 images to establish agreement, each annotator was assigned a set of images to label on their own. Each set was constructed in a way such that 20% of the images overlapped with the other annotators so that disagreements could be identified; however, only score differences greater than 2 were resolved since this dataset had significantly more images. This procedure left us with an inter-rater reliability score of 0.79 according to Cohen's κ and the following distribution for labels 0 through 4: 693, 507, 537, 491, and 174 images.

Figure 4b shows the predicted agglutination scores plotted against the agglutination scores assigned by the annotators. The mean-squared error of the predictions was 1.15, and the inter-rater reliability between the predictions and the annotators was 0.71. Using the same binary decision scheme mentioned earlier, our approach yielded an accuracy

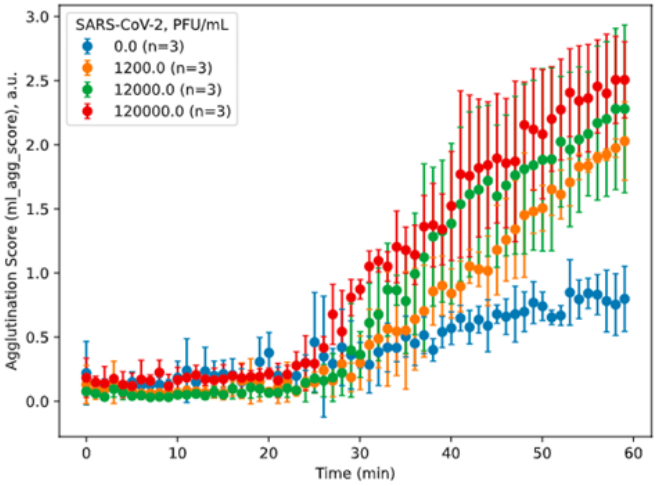

Fig. 5. Post-hoc analysis tool showing the separability of a SARS-CoV-2 AT at different viral concentrations. In this case, the AT developer tested four loads in duplicate: a sample with zero concentration of SARS-CoV-2 (in blue), and three samples with concentrations of SARS-CoV-2 increasing by a factor of 10 (in orange, green, and red, respectively).

of 82.0%. However, unlike with the blood-grouping AT, a non-zero agglutination score does not necessarily correspond to a sample that is positive with SARS-CoV-2. Figure 5 shows an example of how ATs challenged with different viral concentrations agglutinated over time. Even the control wells with no viral concentration slowly began to agglutinate over time, illustrating the importance of having continuous-valued predictions that evolve over time.

VI. Related Work

The notion of utilizing cameras and machine learning to automate the analysis of point-of-care tests has been explored across test designs that exhibit various visual changes. Perhaps the simplest test design to be integrated with image processing is lateral flow rapid diagnostic tests (RDTs), which are particularly used in low- and middle-income countries to screen for diseases like malaria [15], [16], influenza [17], [18], and HIV [19]. Lateral flow RDTs present their results as horizontal lines along a thin strip of paper. Although RDTs are easy to read, prior work has revealed that many interpretation errors are made in the field [20]. To mitigate these issues, Park et al. [21], [22] developed RDT interpretation software that looks for changes in color intensity along the RDT strip.

Another application at the intersection of image processing and chemical testing is in the domain of colorimetric analysis. Point-of-care tests like pH strips and urinalysis dipsticks change their hue and saturation in response to target analytes and often come with a reference color chart that illustrates how the range of colors for a given reagent pad should be interpreted. Although reference charts are normally intended to aid end-users during visual inspection, researchers have leveraged them to train or calibrate models that account for ambient lighting [23]–[26]. Such approaches typically either use the chart to generate a parametric curve with color features or a nearest-neighbor model in a predetermined color space.

Agglutination tests (ATs) are more visually complex than the aforementioned test formats since both spatial and intensity-based information is important to how they should be read. Prior work has demonstrated that computer vision can be used to automate the interpretation of ATs [10], [11], [27], often by training models to predict analyte concentration based on the visual characteristics of the test after a predetermined amount of time. Since ATs change their appearance over time but not their concentration, using concentration as the model output requires relabelling a new dataset and retraining a new model whenever the underlying reaction changes. This is true not only for different types of reactions, but also for new assay formulations of the same reaction.

Our contribution is driven by the fact that chemists rely on a standardized scoring scale when assessing the degree of agglutination in an AT. The scale must be adapted for different agglutinates (e.g., micro-particles, red blood cells), meaning that a new image dataset and model will always be required for new AT form factors. However, a model that is able to predict agglutination scores for a given reaction can still be used without retraining when the assay's formulation changes, thereby enabling AT developers to more quickly assess the sensitivity and accuracy of their tests across experiments.

VII. Discussion and Conclusion

Point-of-care testing is quickly becoming an accessible method for convenient diagnosis and health-related testing, with ATs being one of the main form factors in this domain. In this paper, we presented an open-source framework for facilitating the development and interpretation of new ATs. Our contribution consisted of a web-based annotation interface, a computer vision pipeline for extracting AT-specific visual features, and a machine learning module to facilitate the interpretation of future experiments. Although ATs are designed to be easy to read, automating this process can yield significant benefits for multiple stakeholders. Our framework gives AT developers the chance to objectively compare the performance of their tests across experiments.

Our machine learning approach to this problem could be considered traditional in the sense that we trained models with 14 hand-crafted features inspired by domain expertise. Recent work has shown that deep learning models like convolutional neural networks (CNNs) can have even greater success for image classification, even in the context of image analysis for chemistry [28]. CNNs have the advantage of automatically learning features on the developer's behalf to yield potentially higher accuracy, with the downsides being a greater need for large training datasets and a lack of interpretability. We tried several CNN architectures and found that they did not perform as well as classical machine learning models, likely because of our small dataset sizes. Nevertheless, our framework makes it easy for researchers to try different CNN model configurations and even run a grid search for hyperparameter optimization.

Another limitation of our machine learning approach is that each frame is treated independently, hindering the model's ability to leverage information from previous and future

frames. Since agglutination tends to be gradual, it is unlikely for consecutive frames to have considerably different scores. One way to mitigate potentially noisy predictions is through post-hoc smoothing (e.g., moving average, low-pass filter). Another technique would be to use a model like a hidden Markov model (HMM) or a recurrent neural network (RNN) to explicitly track the state of the reaction over time.

Our second case study shows that going from agglutination scores to a concentration prediction is not always a trivial process. The advantage of training a model that predicts agglutination scores rather than concentration directly is that AT developers do not have to retrain a new model after they adjust their formulation for the sake of improved sensitivity or reaction time. One way to extend our approach would be to generate a second model that relies on hand-crafted temporal features from the agglutination-over-time graph to predict concentration. That being said, a sequence model could automatically identify important temporal features when it learns to predict concentration directly.

VIII. Acknowledgements

This research was supported by the Pandemic Response Challenge program of the National Research Council of Canada.

References

[1] J. Chen, M. Jin, Z. Yu, H. Dan, A. Zhang, Y. Song, and H. Chen, "A latex agglutination test for the rapid detection of avian influenza virus subtype h5n1 and its clinical application," *Journal of Veterinary Diagnostic Investigation*, vol. 19, no. 2, pp. 155–160, 2007, pMID: 17402609.

[2] M. Shoaib, K. Furqan, S. ur Rahman, A. Naveed, A. I. Aqib, M. F.-e.-A. Kulyar, Z. A. Bhutta, and M. S. Younas, "Detection of rotavirus in sewage and drinking water by latex agglutination test."

[3] F. J. Gella, J. Serra, and J. Gener, "Latex agglutination procedures in immunodiagnosis," *Pure and Applied Chemistry*, vol. 63, no. 8, pp. 1131–1134, 1991.

[4] G. Desmonts and J. S. Remington, "Direct agglutination test for diagnosis of toxoplasma infection: method for increasing sensitivity and specificity," *Journal of Clinical Microbiology*, vol. 11, no. 6, pp. 562–568, 1980.

[5] Y. Ye, P. Wang, Y. Zhou, F. Chen, and X. Wang, "Evaluation of latex agglutination inhibition reaction test for rapid detection of aflatoxin b1," *Food Control*, vol. 22, no. 7, pp. 1072–1077, 2011.

[6] W. Z. Abdullah, "Conventional rapid latex agglutination in estimation of von willebrand factor: Method revisited and potential clinical applications," *Journal of Immunology Research*, vol. 2014, p. 10, 12 2014.

[7] A. A. Sklavounos, J. Lamanna, D. Modi, S. Gupta, A. Mariakakis, J. Callum, and A. R. Wheeler, "Digital Microfluidic Hemagglutination Assays for Blood Typing, Donor Compatibility Testing, and Hematocrit Analysis," *Clinical Chemistry*, 09 2021, hvab180.

[8] L. A. Olopoenia and A. L. King, "Widal agglutination test - 100 years later: still plagued by controversy," *Postgraduate Medical Journal*, vol. 76, no. 892, pp. 80–84, 2000. [Online]. Available: https://pmj.bmj.com/content/76/892/80

[9] J. Piper, T. Hadfield, F. McCleskey, M. Evans, S. Friedstrom, P. Lauderdale, and R. Winn, "Efficacies of rapid agglutination tests for identification of methicillin-resistant staphylococcal strains as staphylococcus aureus," *Journal of Clinical Microbiology*, vol. 26, no. 9, pp. 1907–1909, 1988.

[10] T.-F. Wu, Y.-C. Chen, W.-C. Wang, Y.-C. Fang, S. Fukuoka, D. T. Pride, and O. S. Pak, "A rapid and low-cost pathogen detection platform by using a molecular agglutination assay," *ACS central science*, vol. 4, no. 11, pp. 1485–1494, 2018.

[11] Y. Oyamada, R. Ozuru, T. Masuzawa, S. Miyahara, Y. Nikaido, F. Obata, M. Saito, S. Y. A. M. Villanueva, and J. Fujii, "A machine learning model of microscopic agglutination test for diagnosis of leptospirosis," *bioRxiv*, pp. 2020–12, 2021.

[12] Y. Berzak, Y. Huang, A. Barbu, A. Korhonen, and B. Katz, "Anchoring and agreement in syntactic annotations," 2016.

[13] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[14] A. A. Sklavounos, J. Lamanna, D. Modi, S. Gupta, A. Mariakakis, J. Callum, and A. R. Wheeler, "Digital microfluidic hemagglutination assays for blood typing, donor compatibility testing, and hematocrit analysis," *Clinical Chemistry*, 2021.

[15] C. C. Azikiwe, C. C. Ifezulike, I. M. Siminialayi, L. U. Amazu, J. C. Enye, and O. E. Nwakwunite, "A comparative laboratory diagnosis of malaria: Microscopy versus rapid diagnostic test kits," *Asian Pacific Journal of Tropical Biomedicine*, vol. 2, no. 4, pp. 307–310, apr 2012.

[16] C. Manyando, E. M. Njunju, J. Chileshe, S. Siziya, and C. Shiff, "Rapid diagnostic tests for malaria and health workers' adherence to test results at health facilities in Zambia," *Malaria Journal*, vol. 13, no. 1, p. 166, may 2014.

[17] S. Huang, K. Abe, S. Bennett, T. Liang, P. D. Ladd, L. Yokobe, C. E. Anderson, K. Shah, J. Bishop, M. Purfield, P. C. Kauffman, S. Paul, A. E. Welch, B. Strelitz, K. Follmer, K. Pullar, L. Sanchez-Erebia, E. Gerth-Guyette, G. Domingo, E. Klein, J. A. Englund, E. Fu, and P. Yager, "Disposable Autonomous Device for Swab-to-Result Diagnosis of Influenza," *Analytical Chemistry*, vol. 89, no. 11, pp. 5776–5783, jun 2017.

[18] S. W. Ryu, J. H. Lee, J. Kim, M. A. Jang, J. H. Nam, M. S. Byoun, and C. S. Lim, "Comparison of two new generation influenza rapid diagnostic tests with instrument-based digital readout systems for influenza virus detection," *British Journal of Biomedical Science*, vol. 73, no. 3, pp. 115–120, sep 2016.

[19] E. Kakalou, V. Papastamopoulos, P. Ioannidis, K. Papanikolaou, O. Georgiou, and A. Skoutelis, "Early HIV diagnosis through use of rapid diagnosis test (RDT) in the community and direct link to HIV care: a pilot project for vulnerable populations in Athens, Greece," *Journal of the International AIDS Society*, vol. 17, p. 19619, nov 2014.

[20] S. A. Harvey, L. Jennings, M. Chinyama, F. Masaninga, K. Mulholland, and D. R. Bell, "Improving community health worker use of malaria rapid diagnostic tests in Zambia: package instructions, job aid and job aid-plus-training," *Malaria Journal*, vol. 7, no. 1, p. 160, dec 2008.

[21] C. Park, A. Mariakakis, J. Yang, D. Lassala, Y. Djiguiba, Y. Keita, H. Diarra, B. Wasunna, F. Fall, M. S. Gaye *et al.*, "Supporting smartphone-based image capture of rapid diagnostic tests in low-resource settings," in *International Conference on ICTD*, 2020, pp. 1–11.

[22] C. Park, H. Ngo, L. R. Lavitt, V. Karuri, S. Bhatt, P. Lubell-Doughtie, A. H. Shankar, L. Ndwiga, V. Osoti, J. K. Wambua *et al.*, "The design and evaluation of a mobile system for rapid diagnostic test interpretation," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, pp. 1–26, 2021.

[23] M. Anthimopoulos, S. Gupta, S. Arampatzis, and S. Mougiakakou, "Smartphone-based urine strip analysis," in *IEEE Imaging Systems and Techniques (IST)*, Oct 2016, pp. 368–372.

[24] R. Yang, W. Cheng, X. Chen, Q. Qian, Q. Zhang, Y. Pan, P. Duan, and P. Miao, "Color Space Transformation-Based Smartphone Algorithm for Colorimetric Urinalysis," *ACS Omega*, vol. 3, no. 9, pp. 12 141–12 146, sep 2018.

[25] J. I. Hong and B.-Y. Chang, "Development of the smartphone-based colorimetry for multi-analyte sensing arrays," *Lab Chip*, vol. 14, pp. 1725–1732, 2014.

[26] M. Ra, M. S. Muhammad, C. Lim, S. Han, C. Jung, and W. Kim, "Smartphone-based point-of-care urinalysis under variable illumination," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1–11, 2018.

[27] Y. Luo, H.-A. Joung, S. Esparza, J. Rao, O. Garner, and A. Ozcan, "Quantitative particle agglutination assay for point-of-care testing using mobile holographic imaging and deep learning," *arXiv preprint arXiv:2106.01837*, 2021.

[28] O. Zhao, N. Kolluri, A. Anand, N. Chu, R. Bhavaraju, A. Ojha, S. Tiku, D. Nguyen, R. Chen, A. Morales, D. Valliappan, J. Patel, and K. Nguyen, "Convolutional neural networks to automate the screening of malaria in low-resource countries," *PeerJ*, vol. 8, p. e9674, 08 2020.