A Deep Learning-Based Evaluation for Microscopic Agglutination Test

Risa Nakano[1], Yuji Oyamada[1*], Ryo Ozuru[2*], Satoshi Miyahara[3], Michinobu Yoshimura[2], and Kenji Hiromatsu[2]

1 Graduate School of Engineering, Tottori University, Japan
2 Department of Microbiology and Immunology, Faculty of Medicine, Fukuoka University, Japan
3 Department of Microbiology, School of Medicine, University of Occupational and Environmental Health, Japan

* oyamada@tottori-u.ac.jp, ozuru@fukuoka-u.ac.jp

Abstract

Microscopic Agglutination Test (MAT) is widely used as the gold standard for diagnosis of a zoonosis leptospirosis. One of the biggest drawbacks of MAT is that the test result might be inconsistent because the examiners estimate the agglutination rates based on their subjective perception. To solve this issue, we propose a deep neural network that mimics MAT's subjective evaluation of the agglutination rate estimation. Using a pre-trained DenseNet121, we can efficiently optimize the network parameter in the training phase. The trained network was validated with our original dataset. The experimental results show that the proposed network estimates agglutination rates with some satisfying quality. Additionally, we apply a typical visualization technique that gives us clues to understand why and how deep learning methods make their decisions. The results indicate that the proposed network extracts image features representing the abundance of leptospires in images. All the results suggest that deep learning is helpful for the agglutination rate estimation and that fulfilling the accountability helps medical experts understanding how deep learning works.

Introduction

Leptospirosis is a zoonosis caused by the pathogenic bacterium *Leptospira* [1]. This zoonosis is regarded as one of the neglected tropical diseases because it is often spread in tropical and subtropical regions with high rainfall or typhoons but attracts less attention than other infectious diseases such as malaria and AIDS [2]. Prevention of the spread of leptospirosis requires serovar-dependent vaccination, and it is crucial to identify the serovars prevalent in the region from the more than 250 recognized serovars.

The Microscopic Agglutination Test (MAT) is the standard test for serovar identification and definitive diagnosis of leptospirosis by indirectly estimating the amount of antibodies in patient sera based on the strength of antigen-antibody reaction [3, 4]. The antigen-antibody reaction occurs only when the test serum is mixed with its corresponding serovar of leptospires as antigens. More antibodies bind more antigens, and agglutination formed by the binding of antigens precipitates over time. Less agglutination is formed with a more diluted test serum because dilution reduces the volume of antibodies, resulting in weak antigen-antibody reactions. In MAT, we

estimate the reduction of free leptospires by adding test serum as an indirect estimation of the strength of antigen-antibody reactions.

The MAT procedure is based on the above characteristics. Test serum is mixed with various serovars of leptospires in a step dilution. For each dilution factor, we measure the strength of the antigen-antibody reaction by comparing the mixed solution against the negative control (no-sera control). The measured strength ranges from 0 to 100% and is determined subjectively. MAT has several drawbacks. Subjective evaluation heavily relies on laboratory technicians' proficiency and its evaluation criteria may differ depending on the technicians. In addition, the test is time-consuming because we repeat the series of procedures as many times as the number of candidate serovars.

As a first step in the computerization of MAT, we have researched machine learning adaptation of MAT based on traditional machine learning techniques [5]. The proposed method is a binary classification that classifies a dark-field microscopy image into either agglutinated or not agglutinated according to free leptospires' appearance on the image. While this is a worthwhile study as the first application of machine learning to MAT, there are two areas for improvement. The extracted features cannot distinguish free leptospires and other obstacles as large as leptospires. In addition, the method ignores step dilution, and thus, highly diluted test serum is classified as agglutinated even when the test serum is mixed with the matched serovar leptospires.

Deep learning [6] has been actively introduced in medical image processing [7]. In contrast to traditional machine learning techniques, deep learning techniques are constructed on a single framework so that researchers can concentrate on designing network architecture and optimizing its parameters. The network architecture represents how networks obtain output from input data, effectively mimicking the cognitive processes of medical experts. The network parameters represent the internal variables that the network uses for its computations. Appropriate parameters are obtained through the training process.

This paper aims to build a deep-learning method for calculating the agglutination rate of MAT. For this purpose, we propose a deep neural network mimicking MAT's subjective evaluation. Furthermore, we utilize the visualization technique that gives us clues as to why and how deep learning methods make their decisions. It is frequently used when deep learning is applied for medical purposes [8]. Please note that all the experiments in this paper were conducted in a laboratory. Experiments and verification with diverse institutes are outside of the scope of this paper.

Materials and methods

Acquisition of MAT image data

MAT was performed using standard methods with rabbit sera obtained in the previous study [9] We studied the following *Leptospira* serovars: Poi, Losbanos, Manilae, or Icterohaemorrhagiae. For each test serum, we performed MAT with one matched and one unmatched serovar. Table 1 shows the eight serovar-serum combinations. The test sera were diluted in 2-fold increments from 1:10 to 1:10240 and incubated at 30°C for 3 hours after *Leptospira* was added so that the final concentrations were ∼107/mL. Each sample was placed on a glass slide followed by a cover glass and observed under a dark-field microscope (BX-50, Olympus, Tokyo, Japan) using a CCD camera (Basler ace acA800-510um, Basler AG, Ahrensburg, Germany). We took 2,880 ($8 \times 12 \times 30$) photographs, 30 from different fields of view per sample. All photographs were saved in TIFF format with a resolution of 512×512. Hereafter, the 11 images with different dilution magnifications are called *inspection images*. On the other hand, photographs capturing the negative control, which is no-sera control, are called *reference images*.

Table 1. Experimental serovar-serum combinations.

Match/Unmatch	Case name	Serovar	Test serum
Match cases	Poi-Poi	Poi	Poi
	Los-Los	Losbanos	Losbanos
	Man-Man	Manilae	Manilae
	Ict-Ict	Icterohaemorrhagiae	Icterohaemorrhagiae
Unmatch cases	Poi-Ict	Poi	Icterohaemorrhagiae
	Los-Poi	Losbanos	Poi
	Man-Los	Manilae	Losbanos
	Ict-Man	Icterohaemorrhagiae	Manilae

Data

Here, the following processing is applied to the photos to create a dataset for the experiments.

First, we made 79,200 image pairs from the 2,880 images of the eight serovar-serum combinations. For each serovar-serum combination, we made 9,900 image pairs by coupling the 30 reference images and 330 inspection images.

Next, we annotated the agglutination rates for each pair. The annotated agglutination rates were assigned to pairs of the same dilution factor. The same agglutination ratio was assigned to multiple pairs for two reasons: ideally, the amount of agglutination observed in photographs taken on the same glass slide would be perfectly consistent, and to reduce the cost of annotation. The annotated values were set to 11 levels, ranging from 0 to 100% in increments of 10% interval. The quantization levels were determined based on actual test values, but there is no medical or laboratory basis. Table 2 shows the annotated values for each dilution factor; of the eleven annotation values, seven were used (0, 20, 30, 50, 70, 90, and 100). These annotated values are used as the ground truth value for the experiments. When the serovar and test sera were correctly matched, the agglutination rate decreased with the value of the dilution factor increased. In the case of mismatched servos, the agglutination ratio was almost zero regardless of the dilution factor. The Ict-Man pair shows a high agglutination rate at low dilution factors despite the serovar-serum mismatch. This suggests the possibility of cross-reaction. Notably, this tendency may be attributed to individual differences, and there is no assurance that the Ict-Man pair will consistently exhibit such reactions.

Table 2. Serum type-annotated values of agglutination rate for each serum tested. These values are used as the ground truth in the experiments.

	Dilution factor	10	20	40	80	160	320	640	1280	2560	5120	10240
Match cases	Poi-Poi	90	90	100	100	100	100	100	100	70	50	50
	Los-Los	70	90	90	100	100	90	70	50	30	0	0
	Man-Man	70	90	90	100	100	100	100	100	100	100	100
	Ict-Ict	70	70	90	100	100	100	100	90	70	30	0
Unmatch cases	Poi-Ict	20	0	0	0	0	0	0	0	0	0	0
	Los-Poi	0	0	0	0	0	0	0	0	0	0	0
	Man-Los	0	0	0	0	0	0	0	0	0	0	0
	Ict-Man	90	90	70	30	20	0	0	0	0	0	0

Finally, the 79200 MAT image pairs were divided into subsets for training and evaluating the proposed neural network. The training dataset optimized the network parameters and UMAP, while the test dataset evaluated the trained network.

Proposed deep neural network

Here we describe our proposed network and experiments to verify its accuracy.

The input data was a MAT image pair, an inspection image, and its corresponding negative control as a reference image. Given a MAT image pair, the proposed deep neural network estimated the agglutination rate, representing the ratio of agglutination in the inspection image compared to the reference image, as a percentage. The proposed network consists of three sub-networks (feature extraction network, feature comparison network, and regression network) as shown in Fig. 1.

The feature extraction network extracts discriminative image features from the input images independently. The network is a twin neural network with the same architecture and weight as shown in Fig. 1 and each branch takes one of the pair images. We utilized the feature extraction part of Densenet121 [10] and added a global pooling layer at the end. With this architecture, the output of a branch is a 1024-dimensional vector as an image feature. The output image features are supposed to represent the number of bacteria visible in the input images but each of their vector elements do not directly represent physical quantities.

The feature comparison network calculates a pair feature between the extracted image features. Assuming that image features are distributed along the number of bacteria in the feature space, we subtract the two image features as the pair feature.

The regression network estimates the agglutination rate of the image pair from the calculated pair feature as a real percentage. It is a feedforward neural network with three fully connected layers separated by ReLU activation functions. It takes a pair feature of 1024 dimensions, passes through two hidden layers with 256 and 64 units, respectively, and outputs an agglutination rate ranging from 0 to 100%.

Fig 1. The architecture of the proposed deep neural network.

Network parameters optimization

This section describes how we trained the network parameters. We first explain the training flow and then describe the settings of the training algorithm.

For the network parameter optimization, we took several training steps as shown in Table 3 shows how each sub-networks optimized. As mentioned above, the feature extraction network was pre-trained for natural image classification while the regression network was designed for our problem and initialized by He's method [11]. This means that the feature extraction network was optimized for natural image classification while the regression network was not optimized for any purposes in the initial state. In step 1, we optimized only the regression network so that the entire network can estimate the agglutination rate from natural image features. In step 2, we optimized only the feature extraction network so that the network can extract MAT image features. Finally, we optimized the entire network so that the network can extract MAT image features suitable for agglutination rate estimation and can estimate the agglutination rate from the extracted features.

Table 3. The optimization steps. Which type of images each sub network is optimized in each optimization step.

	Feature extraction network	Regression network
Initial state	Natural images	None
Step 1	Natural images	Natural images
Step 2	MAT images	Natural images
Step 3	MAT images	MAT images

Adam is used as the training algorithm for all three training steps. The learning rates were set to 1.0×10^{-3} when learning parameters of the regression network from the initialization state, while the learning rate for subsequent learning was set to 1.0×10^{-5} to keep the parameters fine-tuned. The number of epochs was set to 50 so the learning error would converge during the learning phase. The learning error is the absolute squared error between the ground truth, which is shown in Table 2, and its estimate. We also applied data augmentation to rotate each image at each epoch randomly.

Agglutination rate estimation experiment

We used the test dataset to evaluate the trained network's performance. For each image pair of the dataset, we compared the annotated agglutination rates and their estimated values.

Visualization

In this section, we describe a visualization technique to explain the rationale of the decisions of the proposed network. As previously mentioned, it is often challenging to interpret the feature vectors and comparison vectors derived by our proposed networks, as these do not directly correspond to physical quantities. To address this challenge, we introduced using UMAP visualization [12], a common technique in computer-aided diagnosis (CAD) when deep learning networks are employed. UMAP converts high dimensional vectors into corresponding two-dimensional spaces while keeping the global structure of the feature distribution. By coloring the distribution in the converted two-dimensional spaces based on various attributes, we expect to find important attributes for the proposed network. This visualization aids in providing a clearer understanding of how our proposed network perceives differences in individual images and image pairs.

Visualization of image features

For the visualization of feature vectors, we used only reference images and colored each converted vector based on its serovar-serum pair. Feature vectors are intended to capture the amount of agglutination, and therefore, negative control must have no difference in the feature space.

Visualization of image pair features

For feature comparison vector visualization, we used all the match image pairs and colored each feature vector based on its agglutination rate. All the unmatched image pairs are omitted for this visualization because reference and inspection images have no difference in theory. Agglutination rate is used as the coloring metric because feature comparison vectors are intended to capture the difference in agglutination rate while remaining unaffected by other factors.

Results

In this section, we present the results and discussions obtained from the experiments described in the previous chapter, as well as feedback from microbiology and infectious diseases experts. Through the experiments, the annotated values shown in Table 2 are used as the ground truth. The estimation error is defined as the absolute difference between a predicted value and its corresponding ground truth.

Experiment on Agglutination rate estimation

Here we present the results of an experiment to evaluate the accuracy of the proposed network.

Fig. 2 compares the ground truth agglutination rate annotated by a laboratory technician and the estimated ones.

Fig. 2.A shows the statistics of the estimated values relative to the true values as a box-and-whisker diagram. It shows a positive correlation between the true value and the estimated value. As the true value increases, the estimation error, the absolute difference between an estimated value and its corresponding ground truth, also tends to increase. The data in the boxed portion of the box-and-whisker diagram (from the second to the third quartile) are solid. In response to this result, we received feedback from the experts that (1) they were pleased to quantify and visualize the agglutination rate and (2) the error between the true value and the estimated value was about 10%, which is an acceptable error.

Fig. 2.B shows the frequency and bias of the estimated values relative to the true value. It shows that data with a true value of 0, which accounts for about half of the data, can be accurately estimated. On the other hand, data with a true value of 20 or more tend to estimate values lower than the true value, especially for data with the highest frequency of 50 or more, where the predicted value of the agglutination rate is about 20% smaller.

Fig 2. Agglutination rate estimation experiment. A: Boxplot showing the correlation between the annotated agglutination rate and the estimated agglutination rate. B: Stacked histogram showing the frequency of the predicted agglutination rate against the annotated agglutination rate.

The experts' feedback supports our claim that the proposed network can estimate the agglutination rate with a certain degree of accuracy. However, since some data have significant errors, it is necessary to analyze what kind of image pairs cause significant errors.

The leading causes of the larger error are the imbalance of the dataset and the imperfect annotation. As shown in Table 2, our dataset is so imbalanced that roughly half of the dataset is a 0% agglutination rate. We ran the network training by taking all image pairs equally, and thus, the proposed network estimates are less accurate for higher agglutination rate data.

Next, Fig. 3 shows the change in the agglutination rate with the dilution factor. The positive cases, plots framed by red border, show a high agglutination rate even at higher dilution factors. On the other hand, the negative cases, plots framed by a blue border, show that the agglutination rate remains low regardless of the dilution factor. In the case of Ict-Man, the agglutination rate is higher only at lower dilution factors, indicating the occurrence of cross-reaction. While the results are based on the limited amount of data, it is worth noting that the proposed network effectively visualizes the difference between the cross-reaction and positive infection.

The results imply that the proposed network can distinguish negative, positive, and cross-reactions based on the changes in agglutination rate in response to variations in dilution factor. Even though further experiments with more variety of data are necessary, this experiment shows the potential of CAD. With the capability to find cross-reaction, we can investigate the conditions of cross-reaction that can be applied to elucidate *Leptospira*.

Fig 3. Transition of agglutination rate against dilution factor for different serovar-test serum pairs. In each plot, the horizontal and vertical axis represent the dilution factor in logarithmic and the estimated agglutination rate. Each plot is colored by its annotated agglutination rate.

Experiment on visualization

Here, we present the results of an experiment to visualize the proposed network's decision-making basis. UMAP projects original data into corresponding lower-dimensional spaces, placing similar data points nearby while ensuring that dissimilar data points are well-separated or spread out.

Visualization of the distribution of image features

Fig. 4, 5, and 6 hows the distribution of image feature vectors with different color annotations. These plots indicate that the feature extraction network extracts image features representing the abundance of bacteria in the images. Reference images are plotted closer while inspection images are distributed. Most of the inspection images of unmatch cases are plotted closer to reference images, and ones of Ict-Man cases, which might cause a cross-reaction, are distributed broadly. Inspection images of match cases are distributed broadly, and the distance from the reference images is proportional to the dilution factor.

Fig 4. The distribution of all the image feature vectors. A: Reference images. B: Inspection images colored by dilution rate.

Fig 5. The distribution of image feature vectors of all the unmathced inspection images. A: Colored by dilution rates. B: Colored by test serum-serovar combinations.

Fig 6. The distribution of image feature vectors of all the matched inspection images. A: Colored by dilution rates. B: Colored by test serum-serovar combinations.

Visualization of the distribution of image pair features

Fig. 7, 8, and 9 show the distribution of pair feature vectors with different color annotations. Considering the distribution trend, the proposed neural network extracts features representing the difference in the number of bacteria between image pairs. Image pairs of unmatch cases have two clusters: the larger cluster consists of image pairs with low agglutination rates while the smaller cluster is Ict-Man cases, which might cause cross-reaction. Image pairs of match cases are distributed along with agglutination and dilution rates.

Fig 7. The distribution of all the pair feature vectors.

Fig 8. The distribution of all the unmatched pair feature vectors. A: Colored by agglutination rates. B: Colored by test serum-serovar combinations.

Fig 9. The distribution of all the matched pair feature vectors. A: Colored by agglutination rates. B: Colored by test serum-serovar combinations.

Discussion

This study aims to replace the subjective evaluation of skilled laboratory technicians with deep learning techniques. Many studies have been published on implementing deep learning in CAD, and this paper is based on the same idea. As previously mentioned, this study has only been validated on a small data set and thus faces many challenges. However, many of the experimental results of the proposed network are interesting and cannot be ignored as chance or unintentional bias. We have obtained valuable findings for efforts to introduce deep learning into MAT as a technology to understand and prevent the spread of Neglected Tropical Diseases. Based on these results, another contribution of this study is the clarification of the challenges of resolving the difficulties of animal experiments and annotation.

The most important issue is evaluating with a wider variety of data. We used limited data for the experiment, and hence, we cannot distinguish whether the obtained result is due to individual differences or universal characteristics. To address this issue, we need to increase the number of samples per serovar and the number of serovars. With a sufficient amount of data, we can verify the possibility of determining negative, positive, and cross-reaction from the transition of agglutination rate against dilution factors.

It is also important to handle the dataset's imbalance. Since our dataset has an equal number of positive and negative cases, data with a 0% agglutination rate dominates. Hence, the proposed network pays less attention to data with medium and high agglutination rates. For this issue, we need to apply a training algorithm designed for such imbalanced data or collect more data with medium and high agglutination rates.

Mitigating annotation bias is also a significant concern. The current annotation task requires subjective evaluation by mature laboratory technicians. To reduce the annotation cost, a more cost-effective annotation strategy, such as human-in-the-loop training [13], is favored.

Conclusion

Preventing the spread of a zoonotic disease necessitates speeding up and automating definitive diagnosis. In the case of leptospirosis, MAT is the standard test for definitive diagnosis. This paper aims to provide computer-aided diagnosis support for calculating agglutination rates based on subjective assessment of MAT. We estimated the agglutination rate using the proposed deep learning network. Additionally, we visualized the decision basis of the proposed network to prevent black boxes.

We conducted the agglutination rate estimation experiments and the visualization experiment as the basis for deep learning judgments. The agglutination rate estimation experiment indicates that the deep learning technique is helpful for agglutination rate estimation. We found that infection can be determined by looking at the agglutination rate. The visualization experiment shows the proposed method fulfills its accountability. The results of visualization experiments also showed a correlation with non-annotated serovars. We indicate that deep learning may have captured the characteristics of the leptospires themselves, such as shape differences among serovars. To validate this indication, further experiments with more variety of data are required.

Our results suggest from these experiments that deep learning is helpful for agglutination rate estimation and that fulfilling the accountability helps medical experts understand how deep learning works. Our proposed network estimates the agglutination

rate as ordinary laboratory technicians do. Although the estimation error is relatively higher for higher agglutination cases, this kind of objective quantification is still helpful for the technicians. Furthermore, we found that the sequence of the estimated agglutination rates visualizes the difference between positive reaction and cross-reaction. Additionally, the visualization using the dimensionality compression algorithm has solved the black box which is the biggest problem in deep learning, and we believe that we have fulfilled our accountability to non-specialists in computer science.

Acknowledgments

This work was supported by Fukuoka University Research Grant (Subject number: 221041) and JSPS KAKENHI Grant Numbers JP18K16174 and JP21K16320 to RO.

References

1. Faine S. Leptospira and leptospirosis.; 1994.
2. Farrar J, Hotez PJ, Junghanss T, Kang G, Lalloo D, White NJ, et al. Manson's Tropical Diseases E-Book. Elsevier health sciences; 2023.
3. Musso D, La Scola B. Laboratory diagnosis of leptospirosis: a challenge. Journal of Microbiology, Immunology and Infection. 2013;46(4):245–252.
4. Organization WH, et al. Human leptospirosis: guidance for diagnosis, surveillance and control. WHO/CDS/CSR/EPH 2002.23. World Health Organization; 2003.
5. Oyamada Y, Ozuru R, Masuzawa T, Miyahara S, Nikaido Y, Obata F, et al. A machine learning model of microscopic agglutination test for diagnosis of leptospirosis. Plos one. 2021;16(11):e0259907.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436–444.
7. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical image analysis. 2017;42:60–88.
8. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IW, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nature biotechnology. 2019;37(1):38–44.
9. Villanueva SY, Saito M, Tsutsumi Y, Segawa T, Baterna RA, Chakraborty A, et al. High virulence in hamsters of four dominant Leptospira serovars isolated from rats in the Philippines. Microbiology. 2014;160(2):418–428.
10. Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T, Keutzer K. Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:14041869. 2014;.
11. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision; 2015. p. 1026–1034.
12. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426. 2018;.
13. Monarch RM. Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI. Simon and Schuster; 2021.

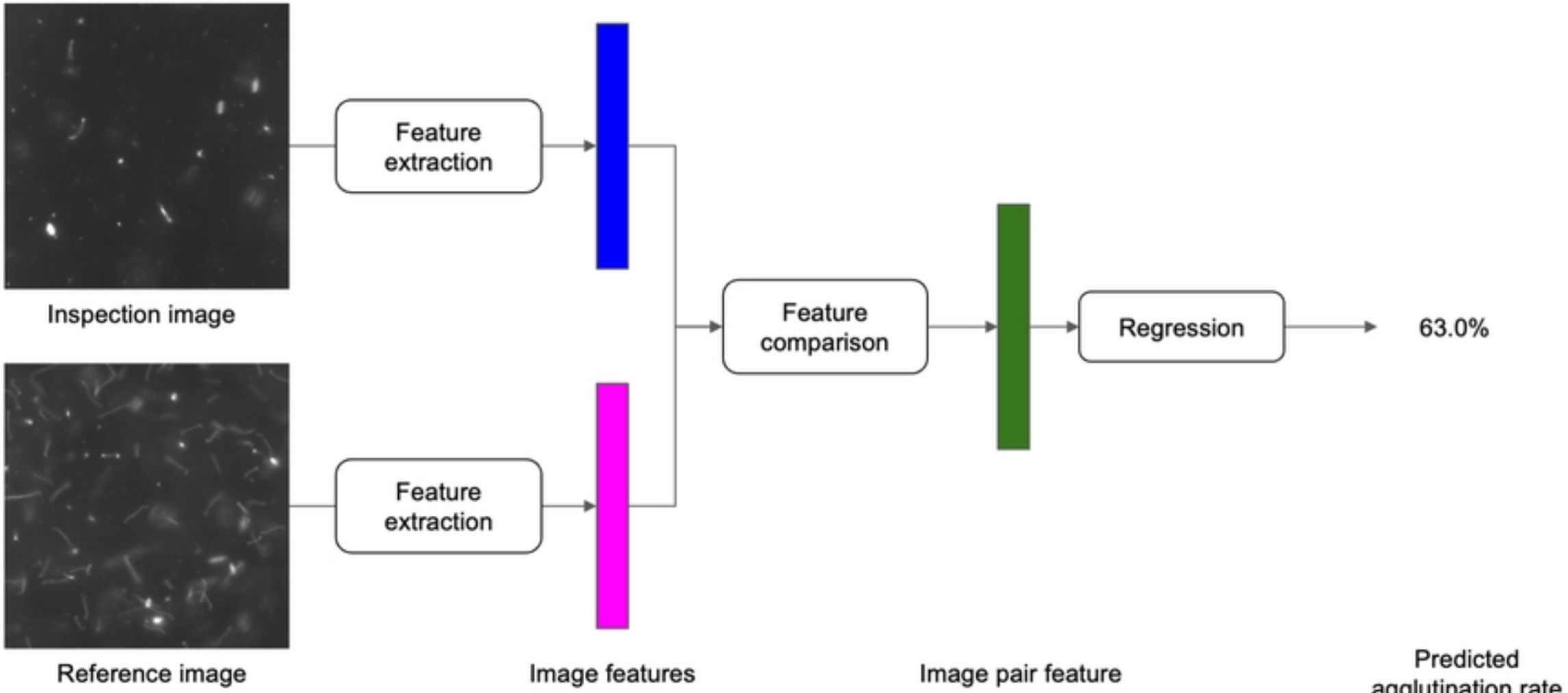

Inspection image
Feature extraction
Reference image
Feature extraction
Image features
Feature comparison
Image pair feature
Regression
63.0%
Predicted agglutination rate

Figure1

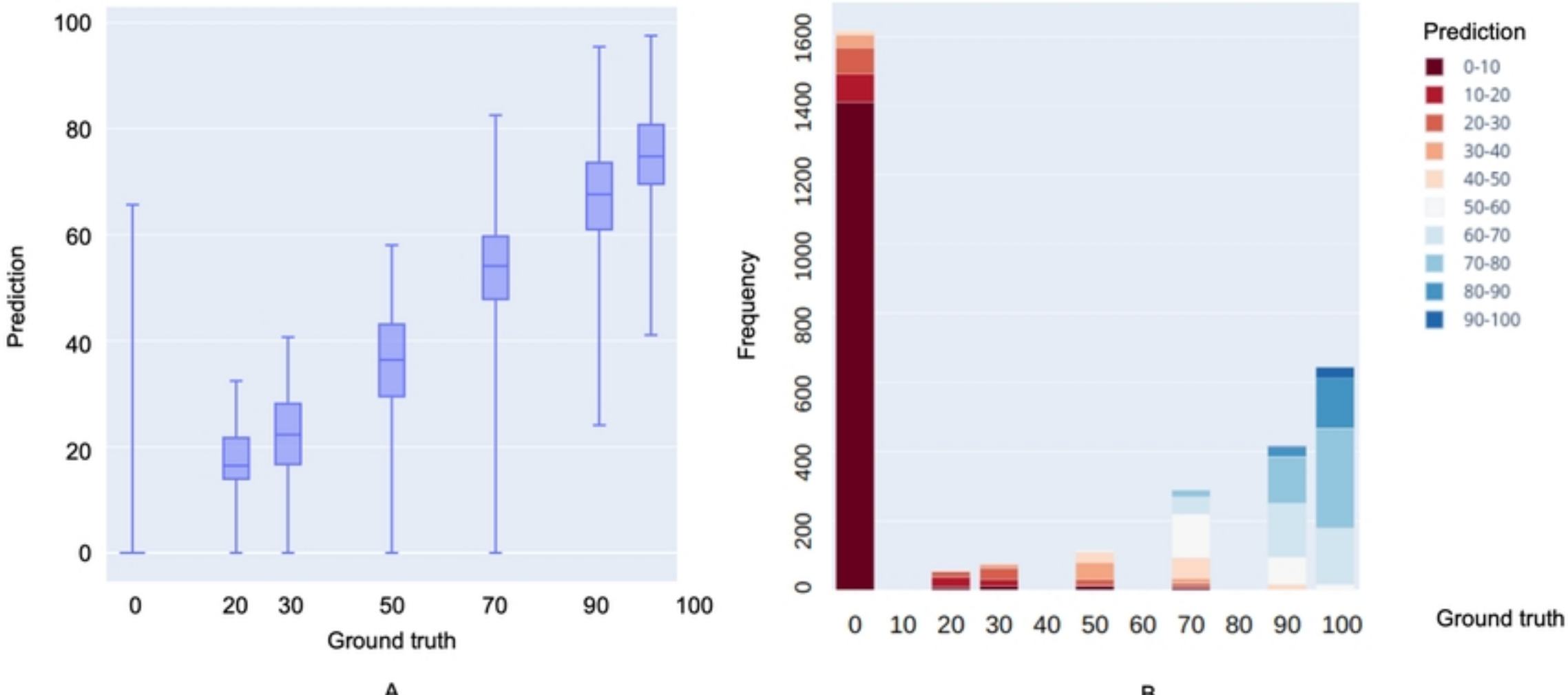

Prediction
100
80
60
40
20
0
0
20
30
50
70
90
100
Ground truth
A
Frequency
1600
1400
1200
1000
800
600
400
200
0
0
10
20
30
40
50
60
70
80
90
100
Ground truth
B
Prediction
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100

Figure2

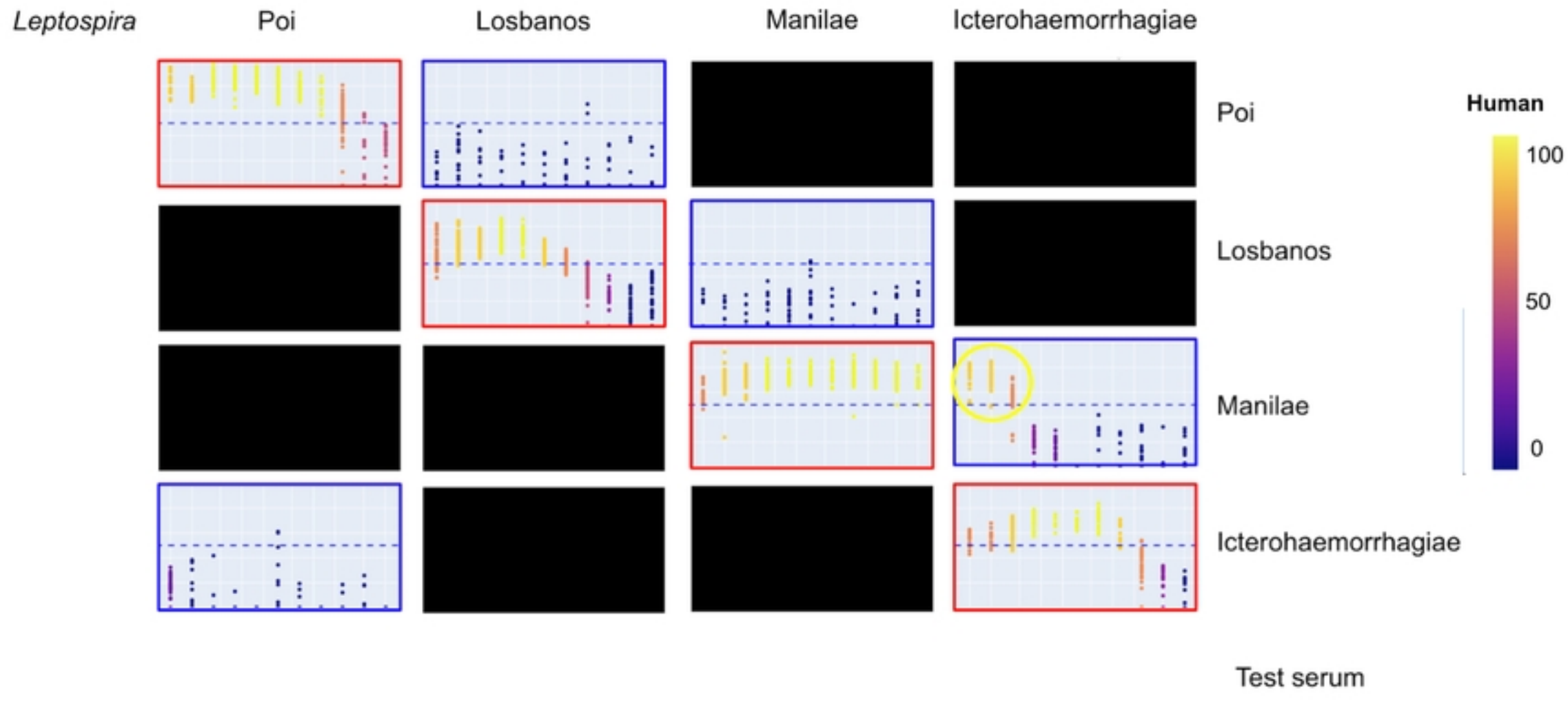
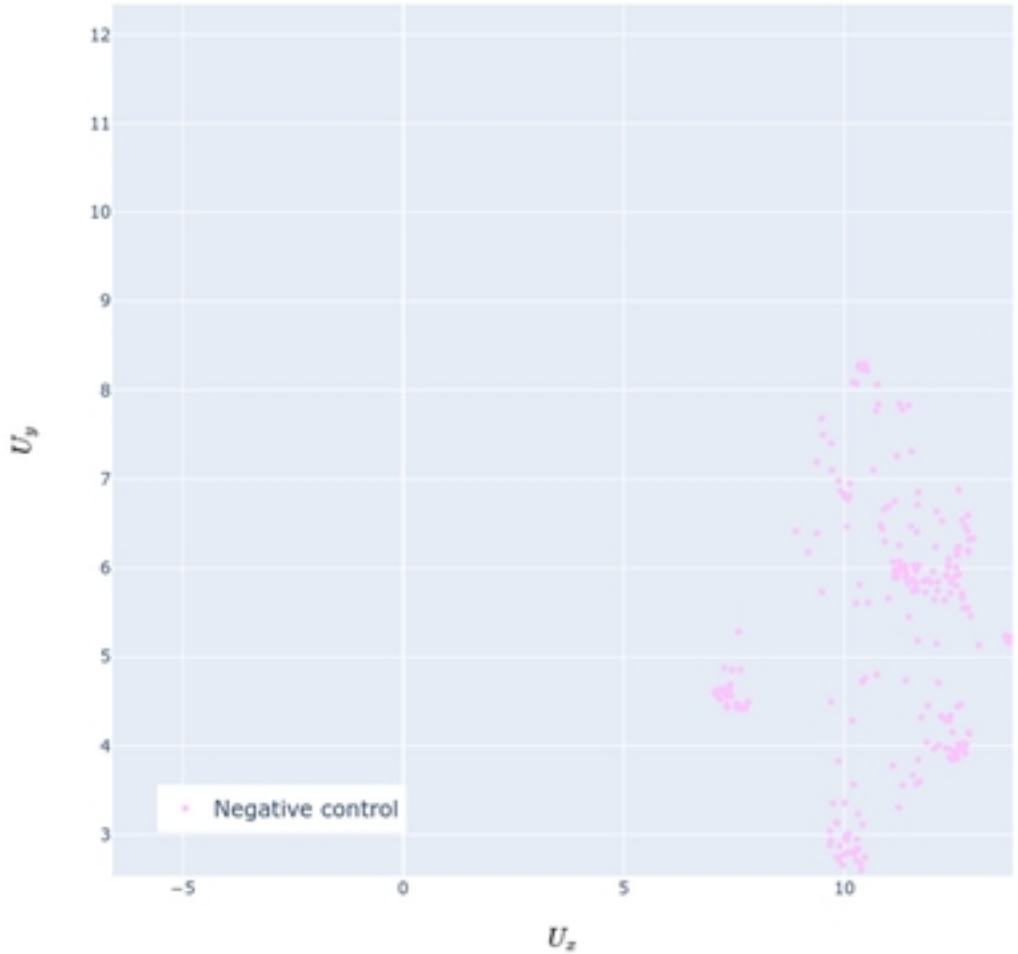
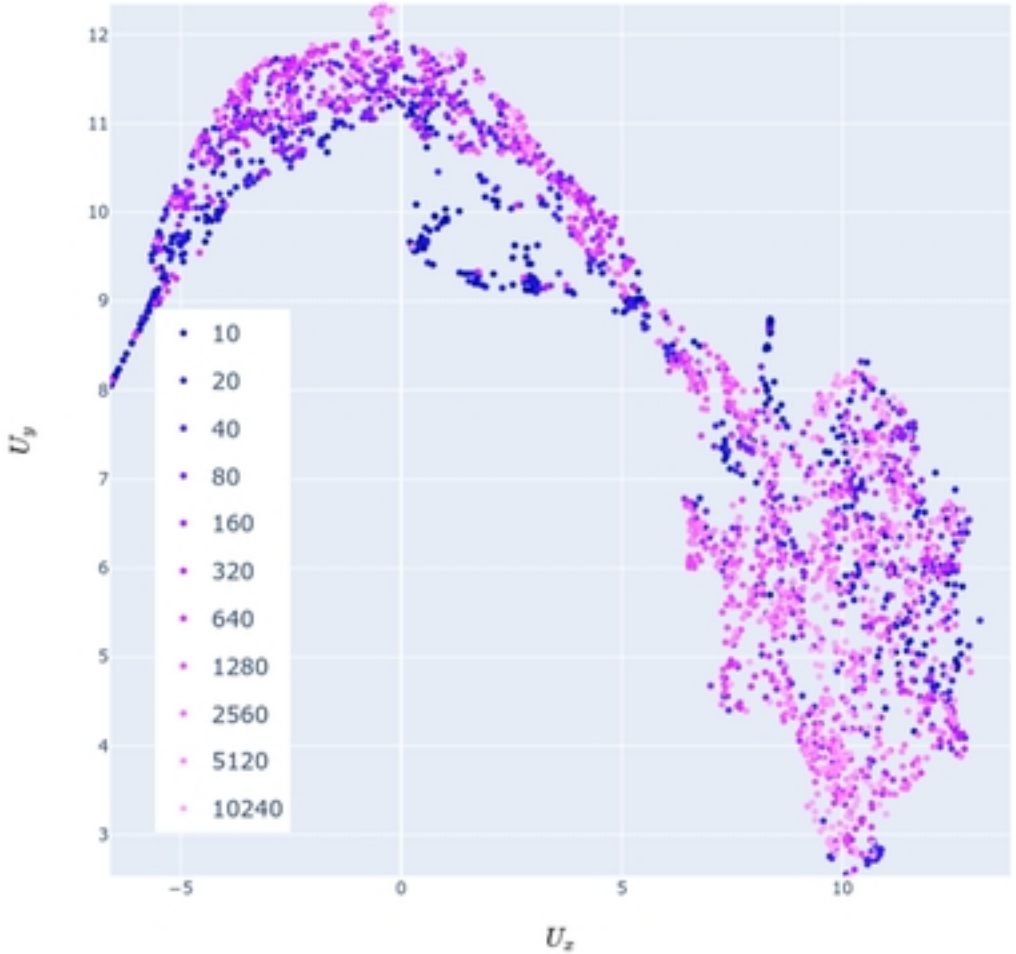

Leptospira
Poi
Losbanos
Manilae
Icterohaemorrhagiae
Poi
Losbanos
Manilae
Icterohaemorrhagiae
Human
100
50
0
Test serum

Figure3

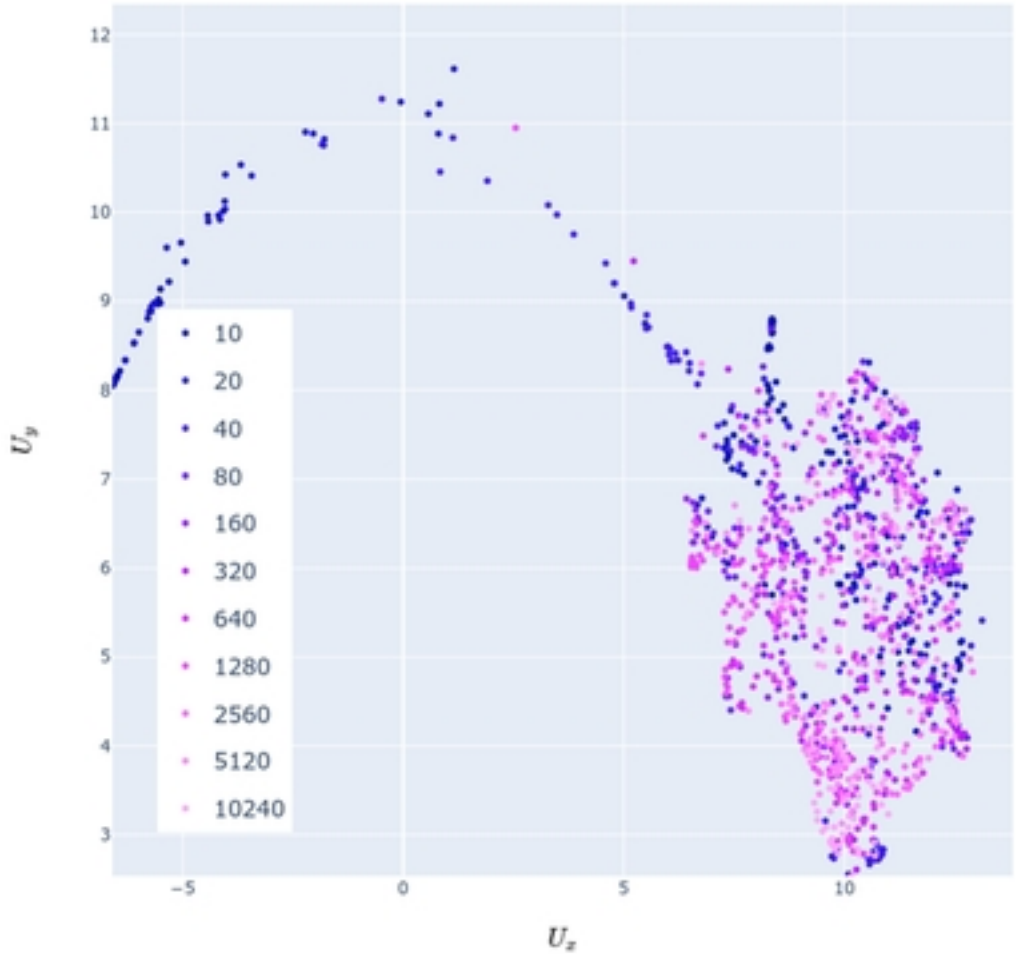

Negative control
U_y
U_x

A

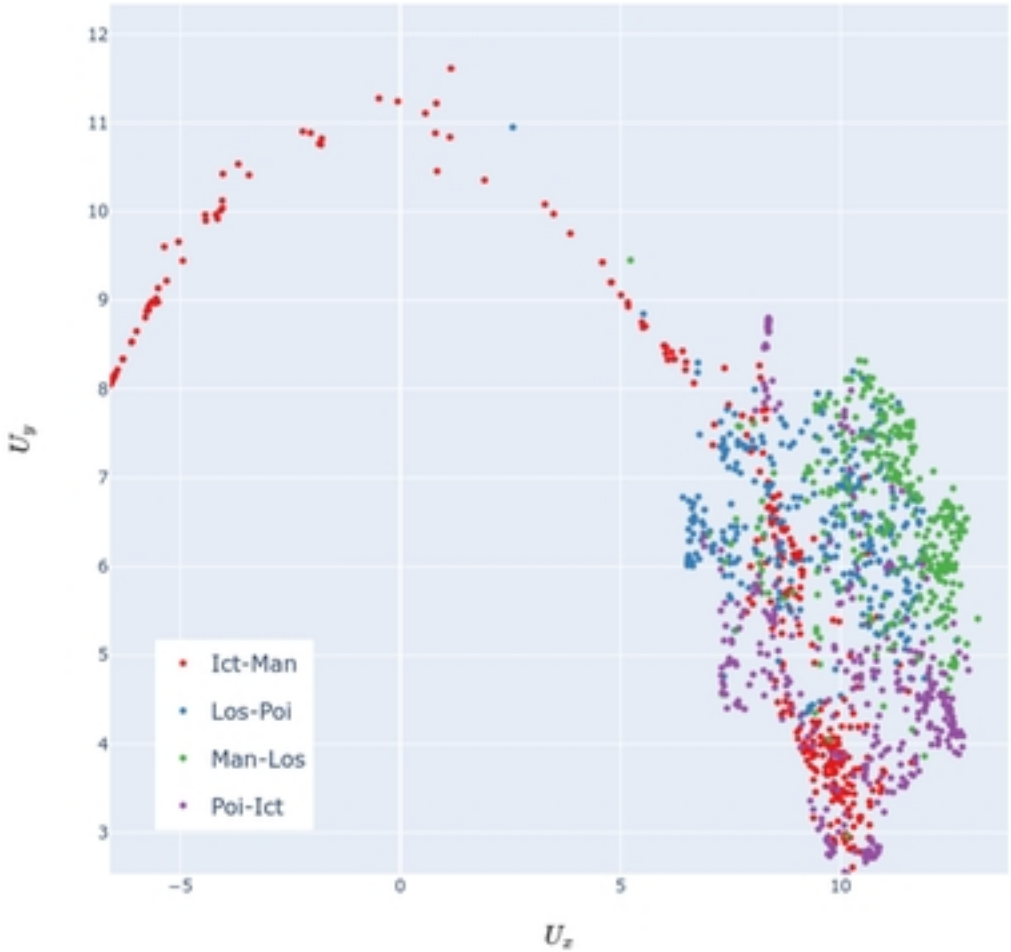

10
20
40
80
160
320
640
1280
2560
5120
10240
U_y
U_x

B

Figure4

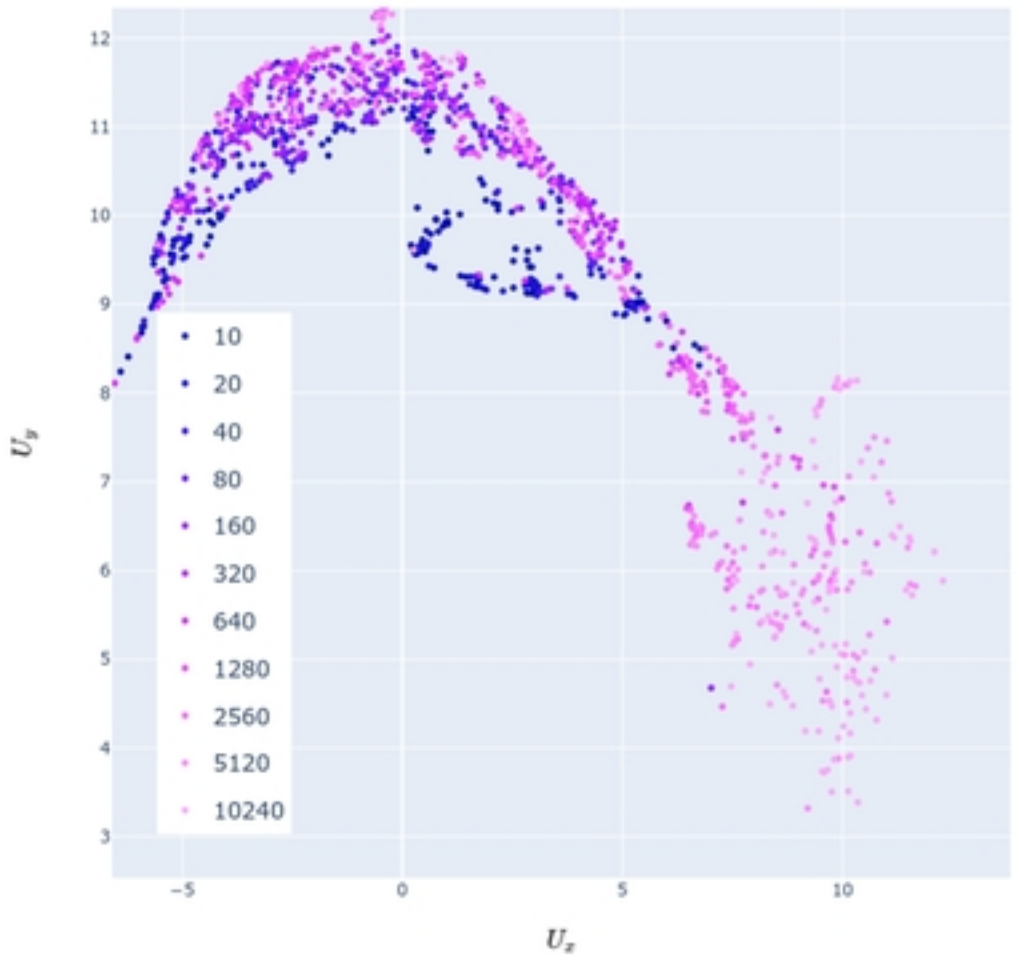

10
20
40
80
160
320
640
1280
2560
5120
10240
U_y
U_x

A

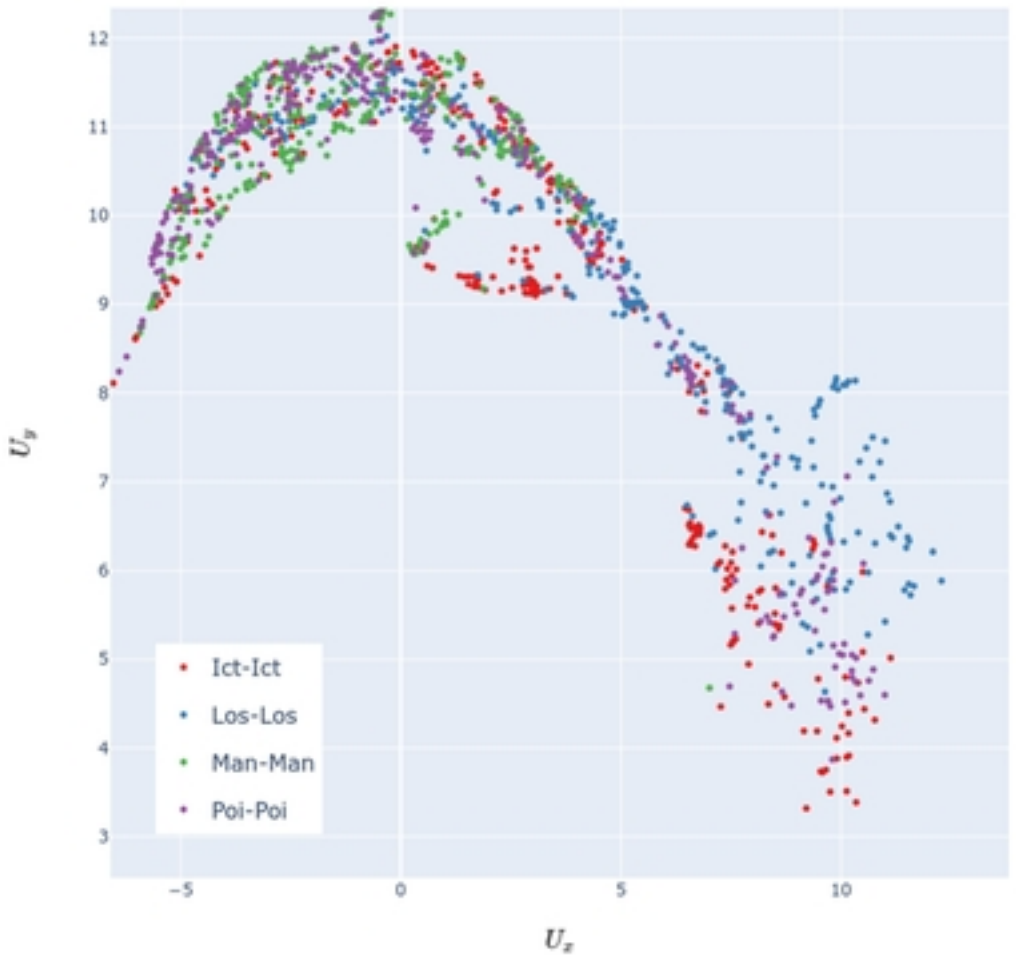

Ict-Man
Los-Poi
Man-Los
Poi-Ict
U_y
U_x

B

Figure5


10
20
40
80
160
320
640
1280
2560
5120
10240
U_y
U_x

A


Ict-Ict
Los-Los
Man-Man
Poi-Poi
U_y
U_x

B

Figure6

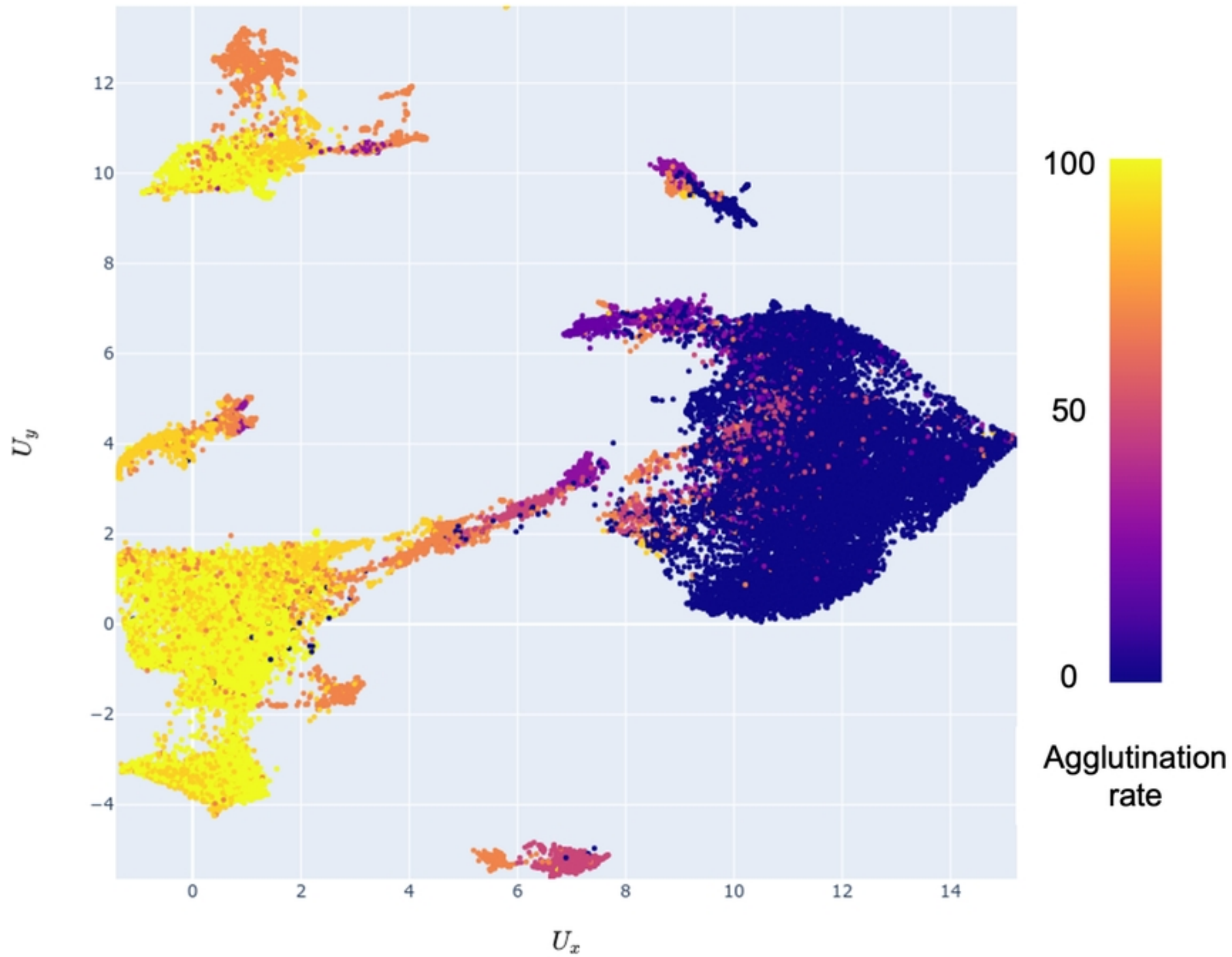


Figure7

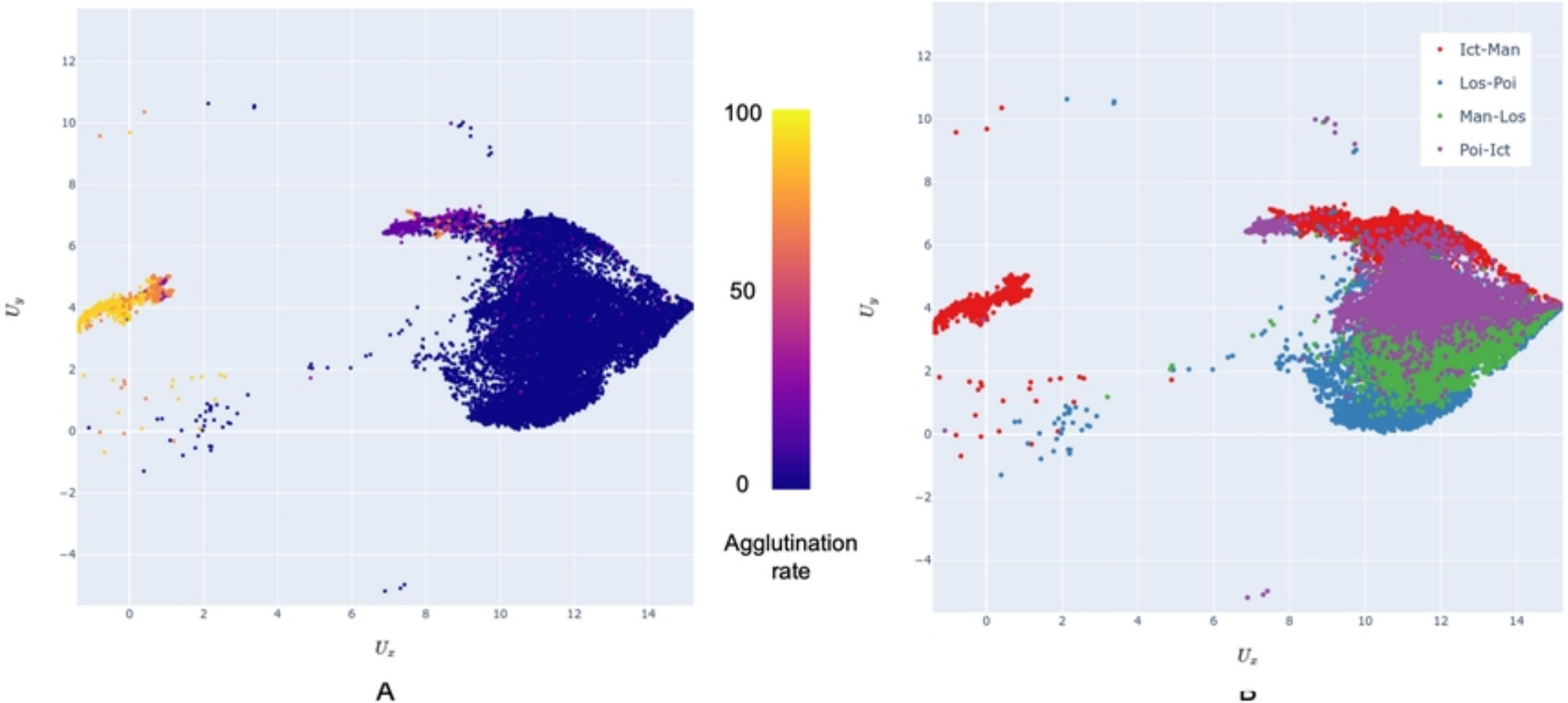

100
50
0
Agglutination rate
Ict-Man
Los-Poi
Man-Los
Poi-Ict
U_x
U_y
A
B

Figure8

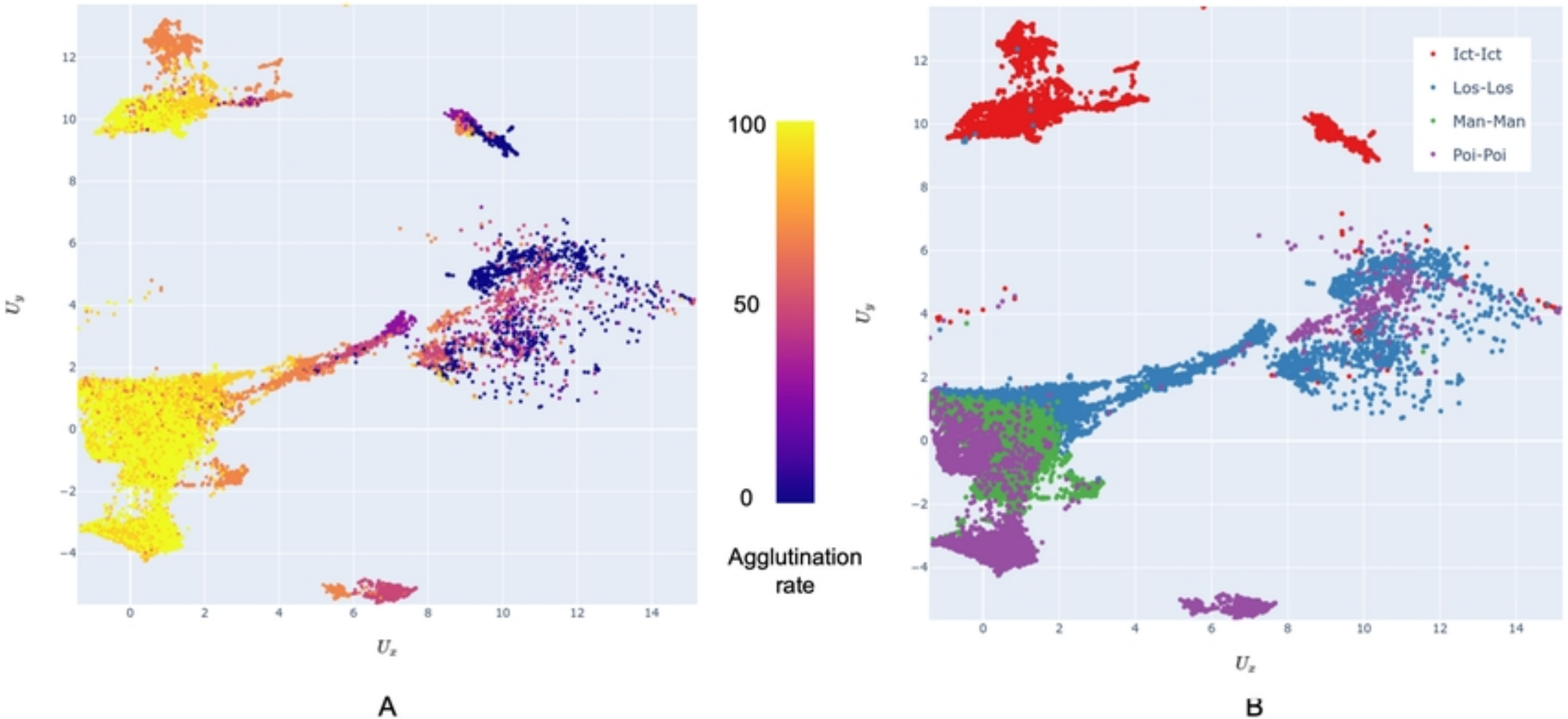

100
50
0
Agglutination rate
Ict-Ict
Los-Los
Man-Man
Poi-Poi
U_x
U_y
A
B

Figure9