Explorations in Texture Based Classification for Bacterial Infection in Zebrafish

G.A. den Heijer[1] G.A.DEN.HEIJER@UMAIL.LEIDENUNIV.NL
P. van der Putten[1] PUTTEN@LIACS.NL
A.E. Nezhinsky[1] ANEZHINS@LIACS.NL
E.L. Benard[2] E.L.BENARD@BIOLOGY.LEIDENUNIV.NL
A.H. Meijer[2] A.H.MEIJER@BIOLOGY.LEIDENUNIV.NL
F.J. Verbeek[1] FVERBEEK@LIACS.NL

[1]Imaging & BioInformatics, Leiden Institute of Advanced Computer Science, Leiden University, Niels Bohrweg 1, 2333 CA, Leiden, the Netherlands
[2]Molecular Cell Biology, Institute Biology Leiden, Leiden University, Einsteinweg 55, 2333 CC, Leiden, the Netherlands

Keywords: tuberculosis, texture, zebrafish, classification

Abstract

Tuberculosis is a serious infectious disease caused by *Mycobacterium tuberculosis* bacteria that infect cells of the human immune system. These infected cells form clusters, called granulomas, which are the hallmark of this disease. Animal model systems have been used to understand the complex molecular interactions involved in tuberculosis infection. The zebrafish embryo model system is useful to study the early stages of tuberculosis infection and granuloma formation. In these transparent embryos, the clustering of infected immune cells into granulomas can be studied through fluorescence imaging and image analysis can generate parameters describing the progression of infection. Such analysis supports experiments that are conducted to reveal the underlying mechanisms of granuloma formation. For image analysis a number of different features need be probed. Next, the bias of these features and the subsequent data analysis need be understood. Texture can potentially be an important descriptor of infection patterns. Prior to large scale application we explore the analysis of groups of infected embryos with a number of standard texture features. In this manner we hope to obtain insight for the analysis of future experiments. Here we compare six groups of infected zebrafish, for which the start conditions were slightly different, using a two-sample t-test, clustering and classification methods. The results from our analysis provide further insight in balanced analysis of experimental data.

1. Introduction

Tuberculosis (TB) affects about a third of the world population and is caused by infection with the bacterial species *Mycobacterium tuberculosis* (Dye *et al.*, 1999). The exact mechanisms of the progression of this disease are, to date, subject of investigation (Lesley and Ramakrishnan, 2008; Russell, 2011). The progress of the infection is usually halted before any serious damage is done. In fact, roughly 90% of all persons infected with TB never experience any symptoms but develop a latent infection where the bacteria become dormant. However, under certain conditions that weaken the immune system, such as malnutrition, immunosuppressive therapies, or HIV co-infection, the bacteria can be reactivated from dormancy and cause active disease (Gengenbacher and Kaufmann, 2012). *M. tuberculosis* has a unique method of surviving and replicating in its host: it infects the macrophages that try to kill it. From there it attracts and infects other macrophages. This eventually results in structures called granulomas, which consist of tight aggregates of infected and uninfected immune cells (Saunders and Britton, 2007; Russell, 2011; Gengenbacher and Kaufmann, 2012). These granulomas are considered as the hallmark of TB.

There is a range of features that can be used to describe granulomas in microscopy images. Considering the texture of the granuloma might be an effective approach to reveal differences between groups/organisms. We therefore wish to study these parameters with real data. The zebrafish is a good model system for the study of Tuberculosis; they can be obtained in large numbers, they can be

[1] Appearing in *Proceedings of the 21st Machine Learning conference of Belgium and The Netherlands*.

manipulated easily and they are transparent in their embryonic phase, so that the disease progression can be monitored (Prouty *et al.*, 2003; Nezhinsky *et al.*, 2010). Moreover, their immune system is very similar to that of humans, i.e. genes found to be involved in TB infection and progression are likely to be involved in human TB infection and progression as well. The system can also be explored from the perspective of the bacteria. For example, Stoop *et al.* (2011) have recently used zebrafish to identify mycobacterial genes involved in the initial stages of granuloma formation.

To study the granuloma infection, zebrafish have been infected with *Mycobacterium marinum*, a close relative to *Mycobacterium tuberculosis*. After infection with the bacteria, images are acquired with bright field and fluorescence microscopy. The images from fluorescence microscopy are used for the texture analysis.

The data that are evaluated for this study are from six infection experiments. In the experiments, the conditions differ (injection needles, bacteria cultures). All considered the experiments provide a good scope of what one can be expected from the injection experiments.

The images need to be analyzed so that the parameters can be extracted. To that end an automated framework is used (Nezhinsky *et al.*, 2012). The analysis if this framework is focused on the spread and size of the granuloma. For this study, we consider the texture of the granuloma. This additional information about the granuloma may provide additional features for describing granuloma development and along that line contribute to better analysis of experimental data.

The aim of our exploration is to gain insight in the quality of the texture measures for the analysis of granulomas in zebrafish that arise from TB infection. Therefore, a framework and a statistical analysis strategy will be developed for simulating the tracking of differences in texture between groups (experiments). This setup can then be used for follow-up research. In addition the testing framework may prove useful for the assessment of additional types of texture descriptors.

The remainder of this paper is organized as follows. First the background of the research is explained. Next the experiments used will be explained and their results will be presented after that. Next, these results will be discussed, along with possible future work that might improve our insights in TB infection and progression. Finally, this paper will be summarized with a conclusion.

2. Related work

Zebrafish is used to gain insight in the infection progression and to obtain a better understanding of the exact mechanisms behind the disease. Until recently the analysis of the images from the infection was done manually; no suitable tools were available to conduct an analysis in a high-throughput fashion. An automated analysis will contribute to the objectivity of the results that are obtained.

A framework for automatic shape retrieval and granuloma detection was developed (Nezhinsky & Verbeek, 2012) so as to obtain objective, reproducible and robust analysis. This framework has been successfully applied to analyze the spatial distribution and size of the granulomas in zebrafish larvae over time. This work was focused on mutant bacteria strains and their effect on infection. Amongst other things it was observed that size and spread of the granulomas increases over time. To that end other features should be considered. In this study, therefore, the changes in granuloma texture in zebrafish infected with *Mycobacterium marinum*, are measured and analyzed.

3. Experiments and results

In this section, the experiments, as well as the observed results will be described. The focus of the exploration will be on texture descriptors and their value for classification of the groups. At the same time the groups are analyzed so that we can better understand the variation that is natural to these kinds of experiments.

3.1 Overall approach

In the study by Nezhinsky *et al.* (2012), the changes between mutant bacteria were considered and analyzed. In this study control groups are used to test and experiment on (cf. 3.2).

Zebrafish groups are regularly used to study the disease progression. One zebrafish group will be the reference group, used to observe the normal texture development in zebrafish infection. By classifying the texture of the granulomas we hope to learn more about the mechanism behind tuberculosis, as well as to establish whether texture is an appropriate feature for classification in this context. In a traditional classification approach a classifier would be built on known positive and negative examples, and then applied to the experiment data sets to identify the ones that are the least similar to the control group. However, in this study we use classification in a different manner. We build models for each reference-experiment combination, and a strong predictive power is seen as an indication of significant differences in texture. This is supplemented with basic univariate and cluster analysis for more explorative data mining.

3.2 Imaging

In each experiment, bright field and fluorescence images are taken from a batch of zebrafish infected with bacteria over consecutive days. For each experiment, a separate batch of zebrafish left uninfected. While no granulomas are present in these zebrafish, in the fluorescence images some background noise is observed. These images are used as an estimate of the level of background noise in the rest of the experiment.

The bright field images depict up to three zebrafish, aligned as in Figure 1. The software developed by Nezhinsky *et al.* (2012) detects the shape and location of the zebrafish. These are then used to distinguish actual spots in the corresponding fluorescence image from background noise.

Figure 1. Bright field image corresponding to Figure 2.

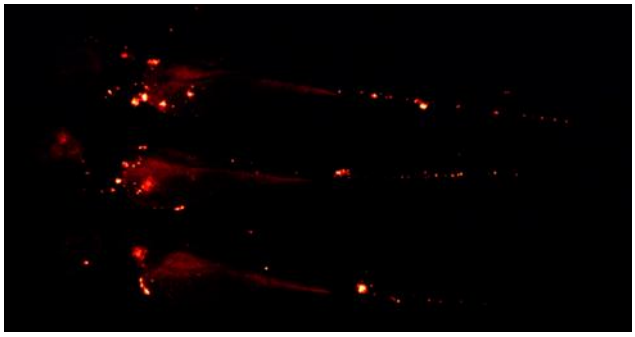

Figure 2. A fluorescence image; contrast and brightness have been adjusted for visualization purposes.

3.3 Framework

The analysis framework developed by Nezhinsky *et al.* (2012) was used to detect the zebrafish in the bright field images and the granulomas in their corresponding fluorescent images. This framework was integrated in a new framework in which the shape and location of the granulomas are not considered. Instead, six basic texture descriptors (cf. 3.4) are calculated. The output of the new framework is a matrix of values of the six texture descriptors for every granuloma that is detected. Table 1 shows the amount of granulomas detected for each experiment. As can be seen, the distribution of the granulomas is very imbalanced. This will be discussed later.

Table 1. Amount of granulomas detected for every group.

EXPERIMENT	A	B	C	D	E	F
GRANULOMAS	516	1398	5712	90	2164	2479

3.4 Texture descriptors

Six basic texture descriptors have been used to study changes in granuloma textures in experimental zebrafish groups (Gonzalez & Woods, 2008). These six descriptors will be explained briefly.

$$\max_{i,j}(p_{ij})$$

The maximum probability is the probability of the most frequent pixel intensity transition.

$$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{(i-m_r)(j-m_c)p_{ij}}{\sigma_r\sigma_c};\ \sigma_r \neq 0;\ \sigma_c \neq 0$$

where K is both the row and column length of G, m_r and m_c are the mean row and column values, and σ_r^2 and σ_c^2 are their variances, denoted as

$$m_r = \sum_{i=1}^{K} i \sum_{j=1}^{K} p_{ij};\quad m_c = \sum_{j=1}^{K} j \sum_{i=1}^{K} p_{ij}$$

$$\sigma_r^2 = \sum_{i=1}^{K}(i-m_r)^2 \sum_{j=1}^{K} p_{ij};\quad \sigma_c^2 = \sum_{j=1}^{K}(j-m_c)^2 \sum_{i=1}^{K} p_{ij}$$

Correlation is a descriptor for the correlation of each pixel to its neighbor. Note that the standard deviations may not be zero. However, this is only the case if the contrast in the image is zero.

$$\sum_{i=1}^{K}\sum_{j=1}^{K}(i-j)^2 p_{ij}$$

Contrast is a descriptor for the contrast for each pixel to its neighboring pixel. The value of this descriptor tends to be larger for larger images.

$$\sum_{i=1}^{K}\sum_{j=1}^{K} p_{ij}^2$$

This is the measure for uniformity in the image.

$$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{p_{ij}}{1+|i-j|}$$

Homogeneity measures the spatial closeness of the pixels to the diagonal.

$$-\sum_{i=1}^{K}\sum_{j=1}^{K} p_{ij}\log_2 p_{ij}$$

Entropy is a measurement for the variability of the image. In this equation, $\log_2 p_{ij}$ is defined as 0 if $p_{ij} = 0$. As with contrast, the value of this descriptor tends to be larger for larger images, albeit not nearly as dramatically due to its logarithmic nature.

In order to obtain these measurements easily, a co-occurrence matrix of the images of each individual spot is constructed. This co-occurrence matrix, referred to as G, will be filled with counts of pixel intensity transition from pixels and their neighbor to the right. Using G, one can now calculate the probability for a specific transition for two given pixel intensities denoted as $p_{ij} = g_{ij}/n$, where g_{ij} is the value in G at the ith row and the jth column and n is the amount of transitions (or the amount of pixels that have a neighboring pixel to the right) in the image. For example, if in an image of 5 by 5 pixels

the transition of a pixel intensity from 2 to 1 occurs twice, the probability of this transition is $p_{21} = g_{21}/(5*(5-1)) = 0.1$.

3.5 Univariate analysis

All six texture descriptors have been calculated for the granulomas detected in the images. The results have been statistically analyzed.

The Shapiro-Wilk test for normality has been performed for every texture descriptor in all experiments in order to assess whether the data found is normally distributed. None of the groups passed this test for any of the descriptors. However, when calculating contrast, the resulting value tends to be higher for larger images. Since this value has a quadratic growth with respect to size, the contrast values are not likely to be normally distributed. The same applies to entropy, albeit to a much lesser extent, because the values have a logarithmic growth with respect to size. It would be desirable to correct both these data before doing a normality test. This is not yet completed for this dataset.

To get an impression of what we can expect in the multivariate analysis, univariate analysis is performed.

After the normality test, a two-sample t-test was done between the reference group and each individual experimental group. If the t-test indicates that the reference group and an experimental group have different means, then this is an indication that the experimental group is different from the reference group.

Table 2 shows that the reference group has a significantly different mean than D and F for all attributes. C differs significantly from A for all descriptors but correlation. B and E are only significantly different with respect to A for correlation, contrast and entropy. It is, however, also possible that the contrast and entropy alone are sufficient for good clustering and classification results. These observations suggest that the data can be clustered and classified with good results.

Table 2. Results of a two-sample t-test between the reference group A and each individual experimental group. Y(es) indicates a significant difference between the means with p=0.05, while N(o) indicates there is not.

	Hypothesis for experiment				
descriptor	B	C	D	E	F
Max. prob.	N	Y	Y	N	Y
Correlation	Y	N	Y	Y	Y
Contrast	Y	Y	Y	Y	Y
Uniformity	N	Y	Y	N	Y
Homogeneity	N	Y	Y	N	Y
Entropy	Y	Y	Y	Y	Y

3.6 Clustering experiments

Next, we attempted to cluster the values of the texture descriptors for the reference group and each experimental group in two groups. As with the two-sample t-test, sufficient separation between the reference and experimental groups would indicate an effect on granuloma development in the experiment.

We first used cluster algorithms to divide all data into two and three clusters. Clustering to two clusters can tell us which groups are similar to other groups. Clustering to three groups can tell us how the data is spread further. The results of this can be seen in Table 3 and 4.

Table 3. Clustering all data simultaneously to two clusters using Expectation Maximization (E.M.) and simple K-means.

	E.M.		Simple K-means	
	% Instances assigned to cluster			
Exp.	1	2	1	2
A	68	32	68	32
B	60	40	61	39
C	32	68	32	68
D	39	61	38	62
E	65	35	65	35
F	69	31	69	31

Table 4. Clustering all data simultaneously to three clusters using Expectation Maximization (E.M.) and simple K-means.

	E.M.			Simple K-means		
	% Instances assigned to cluster					
Exp.	1	2	3	1	2	3
A	49	51	0	48	51	1
B	44	55	1	43	54	3
C	24	48	28	23	40	37
D	29	28	43	30	24	46
E	51	45	4	51	42	7
F	59	38	3	58	36	6

As can be seen in Table 3, the resulting clusters for the two cluster algorithms are remarkably similar. The clusters in Table 4 are also very similar, albeit to a lesser extent.

Note that, in both Table 3 and 4, groups C and D have approximately the same number of instances assigned to them. The same is true for the other groups. The fact that groups C and D stand out from the rest suggests that the textures of these groups differ from those of reference group A, while those from groups B, E and F do not differ from the control group. In Table 4 we also see that the number of instances in clusters 2 and 3 are different for groups C and D, suggesting that these groups are also different from each other.

Subsequently, all experiments were clustered individually with the reference group.

Table 5. The percentages correctly clustered instances after clustering the control group and each individual experimental group to two clusters using E.M. and simple K-means.

	E.M.	Simple K-means
Exp.	% correctly clustered instances	
B	53	53
C	70	68
D	68	69
E	50	51
F	56	56

Table 6. The percentages of instances clustered to the control group and to each individual experimental group to two clusters after clustering using E.M. and simple K-means.

	E.M.		Simple K-means	
	% Instances assigned to cluster			
Exp.	1	2	1	2
A	22.52	26.02	22.99	25.55
B	26.54	24.92	26.96	24.50
A	11.61	38.54	16.09	34.06
C	31.84	18.02	33.51	16.35
A	47.52	0.17	46.37	1.32
D	31.35	20.96	29.21	23.10
A	22.76	25.82	23.66	24.93
E	24.40	27.01	23.81	27.61
A	25.21	23.87	25.21	23.87
F	20.27	30.65	19.93	30.99

Table 5 suggests that clustering the control group with experimental groups C and D results in a relatively low error rate. However, due to the imbalanced amount of data points, these numbers may not be reliable. In order to observe the within-spread of the clusters, we also clustered the reference group with each individual experiment. The results of this can be seen in Table 6, which suggests a clear distinction in cluster composition for clustering the control group with the experimental groups C and D.

Taken together, the results of the clustering experiments suggest that experimental groups C and D are similar and different from reference group A. Also, the experimental groups B, E and F do not seem to be clearly distinguishable from the reference group. These results do not correspond fully to that of the univariate analysis, where the reference group has a significantly different mean compared to D and F for all attributes, and C for all attributes except correlation. According to the results from clustering, F is not different from the control group.

3.7 Classification experiments

Lastly, an attempt was made to classify the values of the texture descriptors for the control group and each experimental group. Once again, successful classification on experimental groups would indicate that the experiment had effect on the granuloma development. The number of correctly classified instances is taken as the measure of how successful the classification was.

Directly applying classification on the data negatively influenced the results, because the number of granulomas than in experimental groups differed. Therefore, before classifying the reference and the other groups, we downsampled the larger of the two.

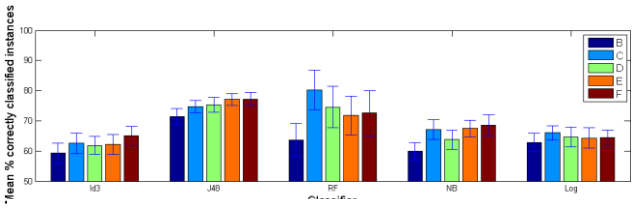

Figure 3. Results for classification of the reference group with each experimental group using Id3, J48, Random Forest, Naïve Bayes and a logarithmic classifier. The larger group was downsampled.

We repeated this experiment, but this time we upsampled the smaller group instead.

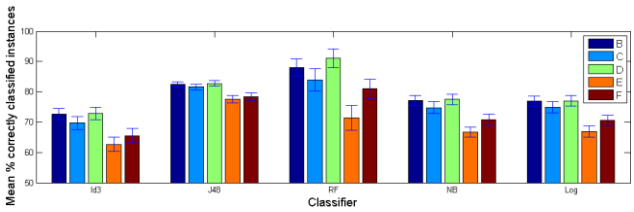

Figure 4. Results for classification of the control group with each experimental group using Id3, J48, Random Forest, Naïve Bayes and a logarithmic classifier. The smaller group was upsampled.

As can be seen in Figure 3 and 4, the classification results using downsampled and upsampled data differ considerably. While classification on upsampled information uses all available information, it might give results that are too optimistic due to repeated data points. On the other hand, downsampled information does not use all available information, but the results are reliable. We therefore chose to focus on the results for the downsampled information.

Table 7. The mean and standard deviations of the percentages correctly classified instances for classification of the control group with each experimental group using Id3, J48, Random Forest, Naïve Bayes and a logarithmic classifier.

	% Correctly classified instances	
Exp.	Mean	σ
B	63.44	4.83
C	70.08	7.20
D	68.03	6.40
E	68.56	5.95
F	69.55	5.40

Table 7 shows that all experimental groups are classified better than at random. Group C is classified slightly better than the rest, except for group B,

which is classified clearly less well. This, again, does not correspond fully to previous results; most notably, we see that group F is clearly different from the control group, and the univariate analysis showed this as well, but clustering did not.

4. Future work

For this study we have analyzed the texture changes between several groups of infected zebrafish. The statistical analyses suggest that groups C, D and F are more different from the reference group. The results between these analyses are not corresponding and conclusive in all cases. To that end more research is required as well and analysis of the composition of the groups so as to determine the optimal combination of analyses for the most reliable results.

The next step is to further test this framework with a larger body more experimental sets; ones that have known differences with the reference set. The results of the texture analysis have been developed independent from that of analysis based on spread and size; a comparison might reveal interesting differences and correspondences. The texture analysis can also be expanded with advanced texture measurements.

A next step is to study the texture changes over time. To that end, new data will be made available that enable this study. Repeating the experiments and analyses with this new data will be beneficial for the outcomes yet obtained.

5. Conclusion

In support of a better understanding of the mechanisms behind TB, we have developed a new framework to obtain granuloma texture measurements from infected zebrafish. The textures of the granulomas of the reference group were compared to other groups in order to test the experimental analysis.

The comparison was done using a univariate analysis, clustering and classification. The results of these comparisons do not correspond in all cases. Experiments with induced differences, e.g. gene knock-out experiments will help to gain insight in using texture and texture change to study/detect changes in granuloma formation. The research can also be expanded with additional texture descriptors.

Acknowledgements

This research is partially supported by the SmartMix Program and by the Cyttron Program.

References

Dye, C., Scheele, S., Dolin, P., Pathania, V., Raviglione, M. C. (1999). *Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country.* WHO Global Surveillance and Monitoring Project.

Gengenbacher, M. & Kaufmann, S. H. (2012). *Mycobacterium tuberculosis: success through dormancy.* FEMS Microbiology Reviews. doi: 10.1111/j.1574-6976.2012.00331.x.

Gonzalez, R. C. & Woods, R. E. (2008). *Digital image processing.* 827-836.

Lesley, R., Ramakrishnan, L. (2008). *Insights into early mycobacterial pathogenesis from the zebrafish.* Current Opinion in Microbiology, 11, 277-283.

Nezhinsky, A. E., Verbeek, F. J. (2010). *Pattern recognition for high throughput zebrafish imaging using genetic algorithm optimization.* 5th IAPR Conference on Pattern Recognition in BioInformatics, Lecture Notes in Bioinformatics 6282: 302-312, Springer.

Nezhinsky, A. E., Stoop, E., van der Sar, A., Verbeek, F. J. (2012). *Numerical Analysis of Image Based High Throughput Zebrafish Infection Screens: Matching Meaning with Data.* Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms. Vilamoura, Algarve, Portugal 1 - 4.

Prouty, M. G., Correa, N. E., Barker, L. P., Jagadeeswaran, P., Klose, K. E. (2003). *Zebrafish-Mycobacterium marinum model for mycobacterial pathogenesis.* FEMS Microbiology Letters, 225, 177–182.

Russell, D. G. (2011), *Mycobacterium tuberculosis* and the intimate discourse of a chronic infection. Immunological Reviews, 240: 252–268. doi: 10.1111/j.1600-065X.2010.00984.x.

Saunders, B. M., Britton, W. J. (2007). *Life and death in the granuloma: immunopathology of tuberculosis.* Immunology and Cell Biology (2007), 85, 103–111.

Stoop E. J., Schipper T., Rosendahl Huber S. K., Nezhinsky A. E., Verbeek F. J., Gurcha S. S., Besra G. S., Vandenbroucke-Grauls C. M., Bitter W., van der Sar A. M. (2011). *Zebrafish embryo screen for mycobacterial genes involved in the initiation of granuloma formation reveals a newly identified ESX-1 component.* Disease Model Mechanisms.