Project no. 248992

Project acronym: NEUNEU

Project title: Artificial Wet Neuronal Networks from Compartmentalised Excitable Chemical Media

Small or medium-scale focused research project (STREP)

Deliverable 4.5 - Software models of sorting and classification implemented in acyclic networks (published on website)

Start date of project: 1.2.2010 Duration: 36 months

Project coordinator name: Dr. Peter Dittrich
Project coordinator organisation name: Friedrich Schiller University Jena

Noname manuscript No.
(will be inserted by the editor)

Evolving Self-Assembling Droplet Computers for Classification

Alexandra Diem · Gerd Grünert · Bashar Ibrahim · Peter Dittrich

Received: date / Accepted: date

Abstract Personalized medicine is a reliable way of treating patients with respect to their individual genome. However, these methods require the extraction of functional genomics data from the patient for "offline" analysis, which is costly and time consuming. A different approach to this problem could be the development of "smart drugs" that are capable of identifying and treating diseased cells on the spot. In this study we use an evolutionary algorithm with self-assembly to develop possible designs of networks composed of lipid-covered microfluidics droplets, which have been shown to be capable of information processing and may provide a suitable model domain for the development of smart drugs as *in vivo* diagnostic and treatment tools. The results of this simulation study suggest that the droplet networks are able to perform simple classification tasks. While the classification was not perfect targetting only a quarter of the healthy cells is considered a valuable improvement to reduce side-effects of medication.

Keywords Droplet Computer · Self-Assembly · Evolutionary Algorithm · Classification

The research reported here is supported in part by Future and Emerging Technologies Grant FP7-248992 "NEUNEU" from the European Union.

A. Diem
Institute for Complex Systems Simulation
University of Southampton
SO17 1BJ
Southampton
United Kingdom
E-mail: A.K.Diem@soton.ac.uk

G. Grünert · B. Ibrahim · P. Dittrich
Bio Systems Analysis Group
University of Jena
Ernst-Abbe-Platz 2
07743 Jena
Germany

1 Introduction

Side effects of drugs are encountered every day on every patient information leaflet and still are a serious issue in the treatment of various fatal diseases such as cancer [14]. With the upcome of cheaper microarray techniques, the discovery of pharmacogenomic biomarkers for the analysis of Single Nucleotide Polymorphisms, which have shown to have significant effects on drug responses [6,23,24], made way for the first approaches in personalized medicine [17,18]. Customized therapies and drugs that minimize the risk of harmful side effects [7] are now on demand. However, the extraction of the relevant functional genomics data is a timely and costly task and has still some way to go before it can be widely avaiblable as a treatment option [3].

Taking the idea of having personalised drug therapies a step further it is desirable to develop "smart drugs" that are capable of identifying and targeting diseased cells *in vivo*. Recently, Douglas *et al.* [4] have implemented nanorobots for intelligent drug delivery, which provide a first step towards this direction. However, to the best of our knowledge, no approach towards developing *in vivo* true smart drugs, which classify cells into diseased or healthy cells and target only the diseased cells, has yet been reported in the literature. The development of such smart drugs has potential implications for the severity of side effects and treatment and could ultimately eliminate this problem entirely. In this study we aim at tackling this challenge by developing designs for smart drugs, which consist of classifier networks of lipid-covered microfluidics droplets. Inspired by Unconventional Computing and based on these droplets, which have been shown to be able to transmit information amongst networks of such droplets [20], we present possible droplet network designs and design principles, which shall be suitable for the use as *in vivo* disease classifiers.

Unconventional computing has brought up a range of fields that try to step away from the classical *von Neumann* architecture for computing, including DNA, bacterial and chemical computing [21]. The use of synthetic biology led to the development of bacterial computing, as bacteria are more flexible and robust to changes in conditions compared to DNA [2]. However, since this approach exploits living systems, which are potentially too large to be used as smart drugs, it may be less suitable for smart drugs. This can be overcome by the use of chemical computing, where relatively well known sets of chemical reactions are used to perform a computation computation. An example for a chemical computation would be the use of enzymatic dynamics to build chemical logical gates [25].

Here, we use the Belousov-Zhabotinsky (BZ) [26] medium as a reaction-diffusion system and apply it to lipid-covered droplets for information processing to develop a wet lab model for smart drugs (see Figure 1). Note that we do not expect to use BZ medium within an actual drug, it is used for demonstrative purposes only. Excitable chemical media like BZ have been used for information processing approaches during the last decades [16,19]. One variant has recently successfully been used to create information processing lipid covered droplets [20] and will therefore be used for the wet lab implementation

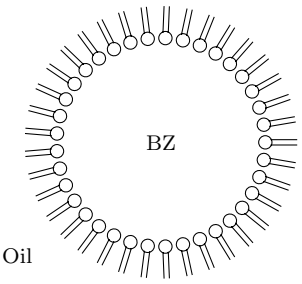
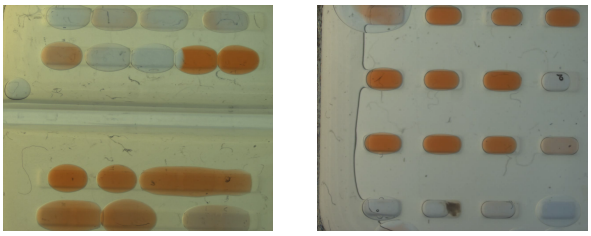


Fig. 1: Schematic representation of Belousov-Zhabotinsky (BZ) medium covered by a lipid monolayer. The lipid molecules form a single layer membrane around the BZ, with the hydrophile heads of the lipid molecules gathering densely around the BZ and the hydrophobic tails pointing towards the oil phase.

of the droplet networks that are proposed in this study. Similarly, Tinsley *et al.* [22] have applied BZ medium on porous beads to mimic quorum sensing in bacteria.

In this study we use an evolutionary algorithm and simulated BZ droplets to find reaction rules that could guide the self-assembly of these droplets into networks, which can classify a given dataset. This classifier represents a potential smart drug that could be engineered using microfluidics.

There are various options of how to prepare the BZ medium with respect to reactant concentrations or lipids, which all result in different characteristics of the oscillations, e. g. different starting times or frequencies. It is important to note that BZ medium oscillates spontaneously with a certain frequency. A droplet is considered to be activated by another droplet if it turns blue before it would on its normal osciallation frequency. Then different recipes, which express different oscillation properties, can be developed for the BZ medium and will be used to define different elements of the network. To discover these, various experiments have been carried out at the universities in Soutampton (Agents, Interactions and Complexity Group and Nano Research Group), Bristol (International Centre for Unconventional Computing) and Warsaw (Department of Complex Systems and Chemical Processing of Information) [20, 1]. The experiments consist of small droplet networks with various reactant concentrations and sizes. These experiments allow the analysis of self-excitation times of different BZ recipes using image processing tools. Example images of such experiments are shown in Figure 2.



(a) (b)

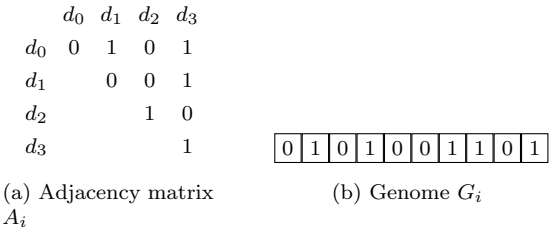
Fig. 2: Screenshots of droplet experiment videos. Blue colour indicates a droplet being in the state "active", while red colour indicates a droplet being in the state "inactive". a) Activation of an inactive droplet is possible within either the two top or the two bottom rows. It can be observed from the videos that waves of blue colour are travelling along a row of droplets, indicating a subsequent activation of inactive droplets via active droplets. b) Equally distanced droplets without being in close contact with other droplets do not allow propagation of oscillation waves and are useful to study the characteristic frequency for a spontaneous activation for different BZ recipes. The experiments have been conducted at the Centre for Hybrid Biodevices, University of Southampton.

2 Method

2.1 Evolutionary Algorithm

Creating networks of droplets which for classification purposes by hand is in most cases a very counter-intuitive task. Therefore we have used an evolutionary algorithm (EA) approach for the creation of the droplet networks. Our EA mostly implements features of genetic algorihtms [8,11], which include a binary genome and recombination. Mutations occur with a probability of $\frac{1}{l}$ with l being the length of the genome and recombination is a one-point crossover of two random individuals (no selection pressure here), creating two offsprings. The selection process is elitist between all individuals (old population plus offspring). An evolutionary run lasts for 1000 generations with a population size of 8 and 30 offspring that are created in each generation. As initial conditions we define all individuals to only exhibit one random reaction rule, which means the only thing it can do is assemble a network of two different types of droplets. Note that these two droplet types can also be the same. In order to develop a functioning droplet network the genomes of the individuals have to be modified via mutation and recombination and selected to exhibit the correct set of rules that will facilitate the assembly of a functional droplet network.

Individuals in the population are stored as bit strings, which represent the upper half of an adjacency matrix for different droplet types (*reaction rules*, see



(a) Adjacency matrix A_i (b) Genome G_i

Fig. 3: Conversion of an adjacency matrix to a genomic bitstring of an individual. a) The adjacency matrix defines the self-assembly rules (or reaction rules), where the d_j denote different types of droplets. A 1 in the adjacency matrix means that droplets of types d_k and d_l can form a connection. b) The adjacency matrix is translated into a linear genome simply by reading the adjacency matrix from top to bottom and left to right.

Figure 3). Different types of droplets can for example vary in properties such as size, refractory times or frequencies. For this study we assumed that the following types exist: input (at least two different), connector, and-function and one-way. Input droplets are externally stimulated droplets, connectors carry an oscillation wave straight through, and-function droplets only get excited if they experience an excitation from two of their neighbours and one-way droplets can only carry oscillations in one specific direction. A 1 in the reaction rules matrix means that the corresponding droplets are able to form a connection. In the experiments such different droplet types with different binding properties can be realised via special connecting membrane proteins that are inserted into the lipid layer.

2.2 Self-Assembly of Droplet Networks

The evaluation of an individual is carried out on droplet networks. This means that we need a genotype-phenotype mapping that translates the set of reaction rules into a droplet network. This is achieved via self-assembly on a 20×20 grid. It is seeded with a droplet of a random type and further droplets can bind to the existing network according to the reaction rules. In each step, an empty position x neighbouring a filled position y is randomly chosen. From the reaction rules adjacency matrix we then randomly choose a droplet type to place on the empty space, such that $A_{d_x d_y} = 1$, where d_x is the droplet type at position x and d_y is the droplet type at position y. We assume a *von Neumann* neighbourhood for the self-assembly process and for simplicity we assume that every droplet type is infinitely available. An example of such a self-assembled droplet network is shown in Figure 4. Note that our self-assembly process differs from the typical self-assembly reported by other authors [12]. Here, we do not restrict the self-assembly in any way to provide the system with a large degree of freedom in order to be able to fully explore the space of possibilites.

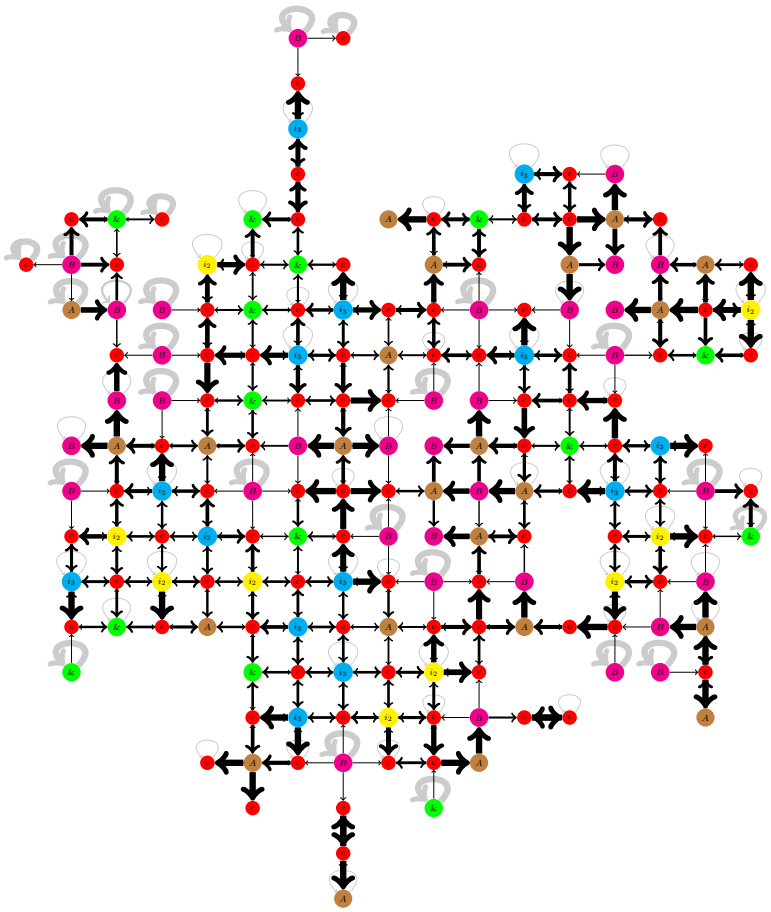


Fig. 4: Example of a droplet network created via self-assembly. Different colours resemble different droplet types. The different droplet types are denoted by different colors and with the symbols c:connector droplet, A, B: directed forwarding of signals only from A to B, &: less excitable droplet, only forwarding signal if two signals arrive at the same time, i_0, i_1, i_2: input droplets. The arrows indicate the predominant direction of signal propagation, where an arrow from a droplet to itself indicates a high level of self-excitation.

2.3 Objective Function

The evaluation of an individual is conducted via our simulation tool Dropsim[1]. It implements event-based modelling for BZ via spatial and temporal propagation of excitation waves within droplet networks [10]. Stimulation patterns of various frequencies representing the input values for the classification are

[1] `http://www.chemicalneuronalnet.uni-jena.de/Results/Project+Media.html`

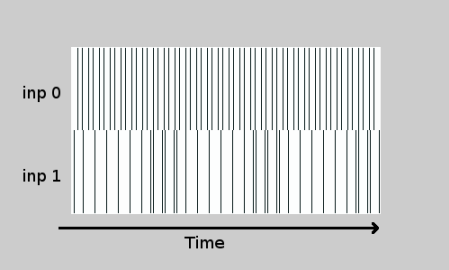


Fig. 5: Example of an input stimulus for a network with two different input droplets.

applied to the network's input droplets. An example of a stimulation pattern for two different input droplets is shown in Figure 5. The result of droplet network computation is the peak sequence of a particular droplet, which is analyzed for the number of oscillations for each stimulus pattern. We then find a threshold which best distinguishes between the two output classes. For example, for one output class it is desired to have less oscillations than for the other output class. We test each droplet in the network as a possible output droplet so that at the end we can select the droplet that best classifies the dataset.

3 Results

We have used our EA to classify different kinds of datasets. First we have created various linearly separable artificial datasets, with four of them exhibiting a wide separation distance and one of them being in very close range. These datasets are used to test the EA and get an upper bound of how good it could perform on a real dataset and to get an idea of the dynamics and behaviour of droplet networks. Finally, we have used our EA to classify the cancer dataset from the Proben1 [15] benchmark datasets.

3.1 Artificial Datasets

We have used five different artificial datasets to evaluate the performance of our evolutionary algorithms. The datasets are shown in Figure 6. We have used linear classification problems here to get an upper bound for the classification that can be carried out by our droplet networks.

3.2 Cancer Dataset

The cancer dataset we have used is a reduced version of the cancer dataset within the Proben1 benchmark dataset, which had been introduced by Prechelt

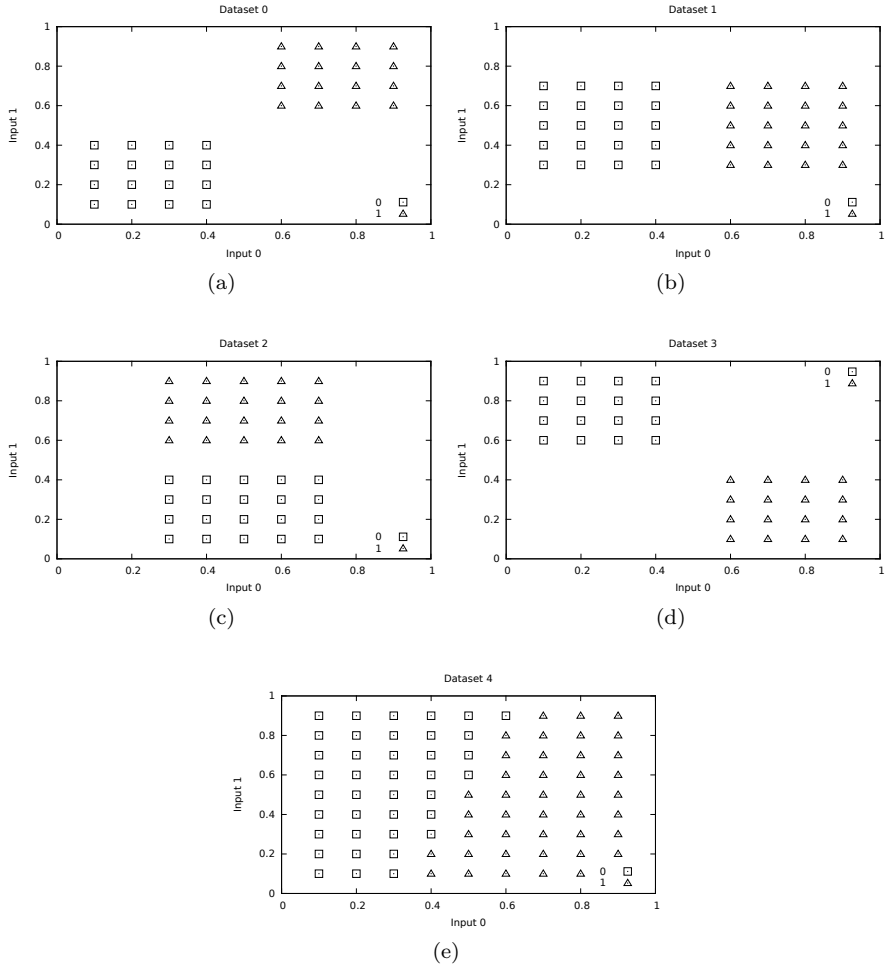


Fig. 6: Artificial datasets to evaluate our evolutioary algorithm. The different shapes represent different classes (0 and 1), which have to be recognised by the EA. Each of these datasets is linearly separable, which makes them suitable benchmark datasets for our EA. The classification results for these artificial datasets are expected to be in any case better than the results for the real cancer dataset. Datasets a) - d) should be easier to classify than dataset e) as the class separation is much more distinct in these datasets.

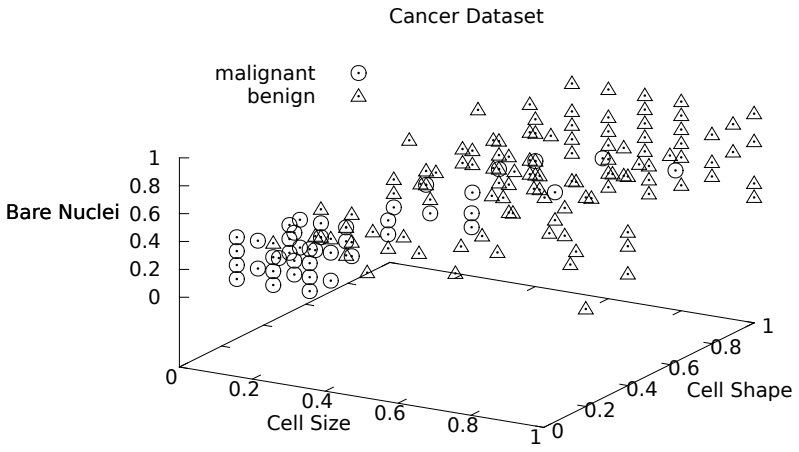


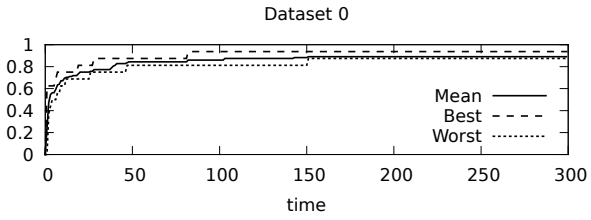
Fig. 7: Reduced cancer dataset. This is a modified version of the Wisconsin Breast Cancer Dataset to suit the implementation of the resulting droplet networks in the wet lab.

[15]. All datasets contained in Proben1 are not linearly separable and therefore harder to classify than the linearly separable artificial test datasets we have introduced in the section before.

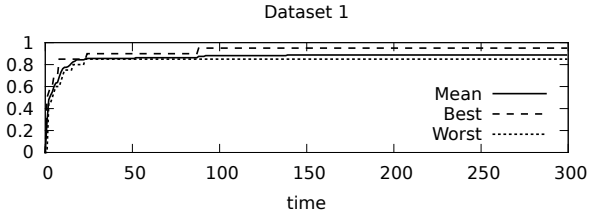
To reduce the dataset we have conducted a Principal Component Analysis [13] to identify the three most important variables. This reduction is necessary as we wanted to keep the simulations simple enough to be able to implement it in the wet lab at a later stage. The three selected variables were Uniformity of Cell Size, Uniformity of Cell Shape and Bare Nuclei. The reduced dataset is shown in Figure 7

3.3 Data Analysis and Classificiation

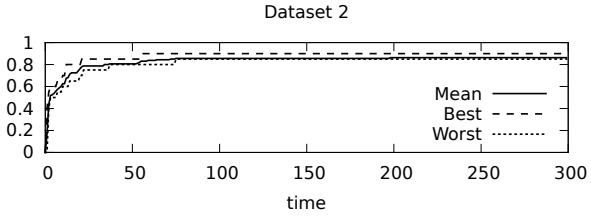
As described above we have first attempted to classify the artificial datasets in order to obtain an evaluation of classification via droplet networks. The results for these datasets are shown in Figure 8. The EA did not find a perfect solution to the problem within the given time frame of 300 generations. This may be due to the randomness that is involved in the creation of the droplet networks from a set of reaction rules, such that the same set of rules may lead to different droplet networks. Again, we have not restricted the network generation to be deterministic in order to keep the simulations as realistic as possible. Additionally, the stochasticity of Dropsim may add to the error. This can be seen in dataset 3, which is in general very easy to solve as low values



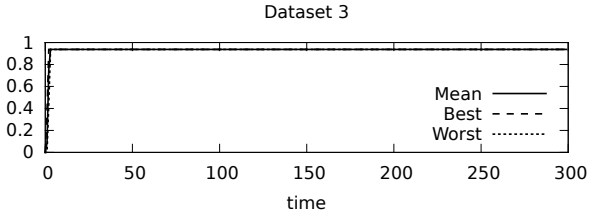
(a)



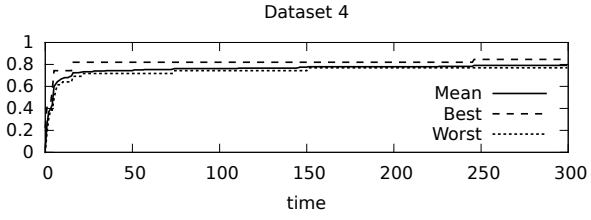
(b)



(c)



(d)



(e)

Fig. 8: Fitness of the droplet networks for the classification of the artificial datasets. a) - d) Linear classification problems, e) nonlinear classification problem.

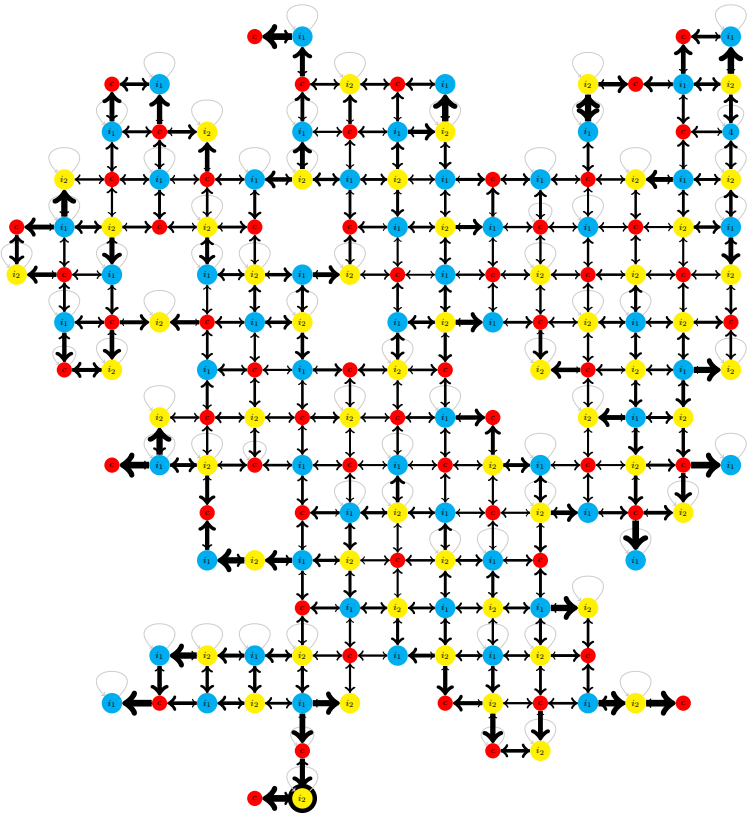


Fig. 9: Network that was self-assembled for solving the example problem displayed in Figure 6 (d). The different droplet types are denoted by different colors and with the symbols c:connector droplet, A, B: directed forwarding of signals only from A to B, &: less excitable droplet, only forwarding signal if two signals arrive at the same time, i_0, i_1, i_2: input droplets. The arrows indicate the predominant direction of signal propagation, where an arrow from a droplet to itself indicates a high level of self-excitation. In this network, the droplet that is used as an output of the computation (displayed in Figure 10) in the lowest row was marked with a black circle.

belong to class 0 and high values to class 1. However, even here we do not get a perfect score. A network evolved and assembled for the artificial dataset 4 from Figure 6 is shown in Figure 9. Its high fitness is further illustrated by comparing the output spike frequencies for inputs from both output classes over time in Figure 10.

A typical result for the classification of the cancer dataset is shown in Figure 11. Here, the slope of the fitness curve is shallower after the initial steep rise than the curves for the artificial datasets. This indicates that for the real dataset it is harder to overcome local optima. The observation that the

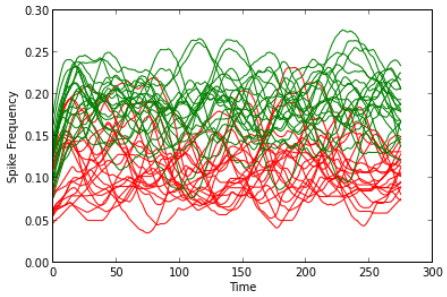


Fig. 10: Droplet network activity patterns for different stimuli, measured at second droplet in the lowest row of Figure 9. 20 simulation runs of each of both output classes 0 and 1 are ploted in red and green, respectively. Although the separation of the spike frequency is not complete, the spike patterns for the output class 1 generate a higher output acticity on average.

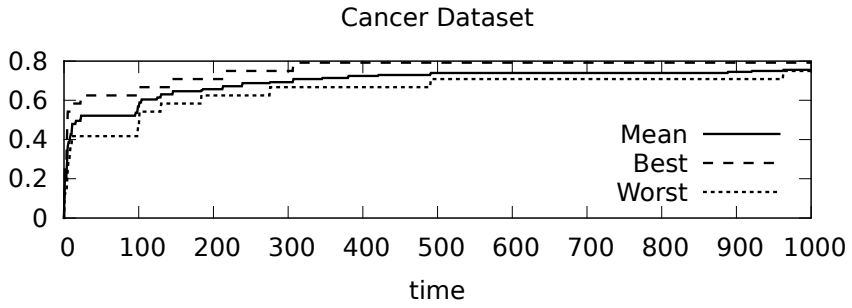


Fig. 11: Fitness of the droplet networks for the classification of the cancer dataset.

mean takes much longer to catch up with the highest fitness value compared to the artificial datasets also supports this assumption.

3.4 Evolutionary Dynamics

In this section we are going to describe the dynamics of the evolution of the droplet networks. For a set of 30 sample evolutionary runs with a population size of 8 and 30 offspring over 1000 generations the mean best fitness value that was reached for the cancer dataset was 0.623 with a standard deviation of 0.137. The mean of the first appearance of the highest fitness score was 281.767 with a standard deviation of 335.530.

As according to the initial conditions each individual in generation 0 has only one reaction rule present. Hence, most of them will not initially be able to

	CON	DIR A	DIR B	AND	IN 0	IN 1	IN 2
CON	0	1	1	1	1	1	1
DIR A		1	0	0	0	0	0
DIR B			0	0	0	0	0
AND				0	0	0	0
IN 0					0	0	0
IN 1						0	0
IN 2							0

Fig. 12: Reaction rules of the best individual of the sample evolutionary run shown in Figure 11. The different droplet types fulfill the following functions. CON: connector droplet, simple forwarding of signals, DIR: oneway forwarding of signals (direction A or B), AND: forwarding of signals only if two signals come in at the same time, IN: input droplets (types 0, 1, 2).

connect input droplets with anything else. Being able to connect at least one of the input droplets with other droplets is essential for any kind of information processing that is supposed to occur within the droplet networks. The data shows that the ability of connecting input droplet types to other droplets accounts for a rise in the fitness score of about 0.2.

The reaction rules for the best individual from the run shown in Figure 11, is depicted in Figure 12. Here, all input droplet types can be connected in the network. Connector droplets seem to be used as "spacer" between other kinds of droplets as they can be connected to any droplet type, except for other connector droplets. The only droplet type that can connect to itself is DIR A suggesting that there is a selection pressure for oneway signal transduction.

4 Conclusion

Our aim was to create a set of reaction rules that can guide the process of self-assembly for the droplets, resulting in a droplet network that can classify a given dataset. With a medical background explicitly in mind, this classifier shall act as a smart drug that can identify diseased cells and specifically attack these while leaving healthy cells to avoid side effects.

We have presented an evolutionary algorithm that finds reaction rules for the self-assembly of droplets into whole networks. We have first tested this evolutionary algorithm on linear classification problems in order to find an upper limit for the capabilities of the droplet networks. The results show that the networks do not reach the highest possible fitness scores, meaning that the droplet networks cannot fully classify the dataset. But given the amount of randomness and complexity in the self-assembly process this result is not surprising and it shows that it may even not be possible to let droplet networks grow on its own in a self-assembly process.

The results for the actual dataset are very similar. The EA progresses in the search space until it reaches its optimum, which correctly classifies about

80 % of the samples in the dataset. In this case the complexity of the problem is even further increased by adding nonlinearity to the dataset. Nevertheless, we have shown that these networks can process information and that they could potentially be a guiding structure for self-assembling smart drugs in the future. A smart drug that self-assembles out of its components and correctly identifies the majority of diseased cells would be a major breakthrough in medical research.

Note that we are not proposing a smart drug based on a BZ system. But BZ here serves as a promising model chemical and our theoretical work is accompanied by work in the wet lab [9,10,5]. It is likely that the dynamics of any potential medium for a smart drug will be hard to understand. For this reason it could be beneficial to rely on approaches that do not necessarily require a full understanding of the medium to be able to use it.

Acknowledgements The research reported here is supported in part by Future and Emergent Technologies Grant FP7-248992 "NEUNEU" from the European Union.

References

1. Adamatzky, A., Holley, J., Dittrich, P., Gorecki, J., De Lacy Costello, B., Zauner, K.P., Bull, L.: On architectures of circuits implemented in simulated Belousov-Zhabotinsky droplets. Bio Systems **109**(1), 72–77 (2012)
2. Baumgardner, J., Acker, K., Adefuye, O., Crowley, S.T., Deloache, W., Dickson, J.O., Heard, L., Martens, A.T., Morton, N., Ritter, M., Shoecraft, A., Treece, J., Unzicker, M., Valencia, A., Waters, M., Campbell, A.M., Heyer, L.J., Poet, J.L., Eckdahl, T.T.: Solving a Hamiltonian Path Problem with a bacterial computer. Journal of Biological Engineering **3**, 11 (2009)
3. Crews, K.R., Hicks, J.K., Pui, C.H., Relling, M.V., Evans, W.E.: Pharmacogenomics and individualized medicine: Translating science into practice. Clinical Pharmacology & Therapeutics **92**(4), 467–475 (2013)
4. Douglas, S.M., Bachelet, I., Church, G.M.: A logic-gated nanorobot for targeted transport of molecular payloads. Science **335**(6070), 831–834 (2012)
5. Escuela, G., Gruenert, G., Dittrich, P.: Symbol representations and signal dynamics in evolving droplet computers. Natural Computing pp. 1–10 (2013). DOI 10.1007/s11047-013-9384-7. URL `http://dx.doi.org/10.1007/s11047-013-9384-7`
6. Evans, W.E., Relling, M.V.: Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics. Science **286**(5439), 487–491 (1999)
7. Evans, W.E., Relling, M.V.: Moving towards individualized medicine with pharmacogenomics. Nature **429**(6990), 464–468 (2004)
8. Goldberg, D.E., Deb, K.: A comparative analysis of selection schemes used in genetic algorithms. In: Foundations of Genetic Algorithms, pp. 69–93. Morgan Kaufmann, San Francisco (1991)
9. Gruenert, G., Dittrich, P., Zauner, K.P.: Artificial Wet Neuronal Networks from Compartmentalised Excitable Chemical Media. ERCIM News **85**, 30–32 (2011)
10. Gruenert, G., Szymanski, J., Holley, J., Escuela, G., Diem, A., Ibrahim, B., Adamatzky, A., Gorecki, J., Dittrich, P.: Multi-scale modelling of computers made from excitable chemical droplets. International Journal of Unconventional Computing **9**, 237–266 (2013)
11. Holland, J.H.: Genetic Algorithms. Scientific American **267**, 66–72 (1992)
12. Johnston, I.G., Ahnert, S.E., Doye, J.P.K., Louis, A.A.: Evolutionary dynamics in a simple model of self-assembly. Physical Review E **83**(6), 066,105 (2011)
13. Jolliffe, I.T.: Principal Component Analysis, 2nd edn. Springer, New York (2002)

14. Monsuez, J.J., Charniot, J.C., Vignat, N., Artigou, J.Y.: Cardiac side-effects of cancer chemotherapy. International Journal of Cardiology **144**(1), 3–15 (2010)
15. Prechelt, L.: Proben1: A set of neural network benchmark problems and benchmarking rules. Tech. rep. (1994)
16. Rambidi, N.G.: Information-processing Capabilities of Chemical Reaction-Diffusion Systems. 2: Finding Paths in a Labyrinth Based on Reaction-Diffusion Media. Advanced Materials for Optics **34**, 27–34 (1999)
17. Sadee, W.: Drug therapy and personalized health care: pharmacogenomics in perspective. Pharmaceutical Research **25**(12), 2713–2719 (2008)
18. Sadée, W., Dai, Z.: Pharmacogenetics/genomics and personalized medicine. Human Molecular Genetics **14**(2), R207–R214 (2005)
19. Steinbock, O., Toth, A., Showalter, K.: Navigating complex labyrinths: optimal paths from chemical waves. Science **267**(5199), 868–871 (1995)
20. Szymanski, J., Gorecka, J.N., Igarashi, Y., Gizynski, K., Gorecki, J., Zauner, K.P., de Planque, M.: Droplets with information processing ability. International Journal of Unconventional Computing **7**, 185–200 (2011)
21. Teuscher, C., Nemenman, I., Alexander, F.J.: Novel computing paradigms: Quo vadis? Physica D: Nonlinear Phenomena **237**(9), v–viii (2008)
22. Tinsley, M., Taylor, A., Huang, Z., Showalter, K.: Emergence of Collective Behavior in Groups of Excitable Catalyst-Loaded Particles: Spatiotemporal Dynamical Quorum Sensing. Physical Review Letters **102**(15), 1–4 (2009)
23. Weinshilboum, R., Wang, L.: Pharmacogenomics: bench to bedside. Nature Reviews: Drug discovery **3**(9), 739–48 (2004). DOI 10.1038/nrd1497
24. Wilkinson, G.R.: Drug metabolism and variability among patients in drug response. The New England Journal of Medicine **352**(21), 2211–21 (2005)
25. Zauner, K.P., Conrad, M.: Enzymatic computing. Biotechnology Progress **17**(3), 553–9 (2001)
26. Zhabotinsky, A.M., Zaikin, A.N.: Autowave processes in a distributed chemical system. Journal of Theoretical Biology **40**(1), 45–61 (1973)