Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 65 (2015) 643 – 651

Procedia Computer Science

International Conference on Communication, Management and Information Technology (ICCMIT 2015)

Towards an Automated Zebrafish-based Toxicity Test Model Using Machine Learning

Alaa Tharwat[a,f], Tarek Gaber[b,d,f], Mohamed Mostaf Fouad[c,d,f], Vaclav Snasel[d], Aboul Ella Hassanien[e,f,g]

[a]*Faculty of Engineering, Suez Canal University, Ismailia, Egypt*
[b]*Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt*
[c]*Arab Academy for Science, Technology, and Maritime Transport, Cairo, EGYPT*
[d]*Faculty of Electrical Engineering and Computer Science, VSB-TU of Ostrava, Czech*
[e]*Faculty of Computers and Information, Cairo University, Egypt*
[f]*Faculty of Computers and Information, BeniSuef University, Egypt*
[g]*Scientific Research Group in Egypt (SRGE), http://www.egyptscience.net*

Abstract

Zebrafish animal is considered as one of the most suitable animals to test toxicity of compounds due many features such as transparency and a large number of embryos produced in each mating. The main problem of the zebrafish-based toxicity test is the manual inspection of thousands of animals images in different phases and this is not feasible enough for the analysis, i.e. it is slow and may be inaccurate process. To help addressing this problem, in this paper, an automated classification of alive (healthy) and coagulant (died because of toxic compounds) zebrafish embryos are proposed. The embryos' images are used to extract some features using the Segmentation-based Fractal Texture Analysis (SFTA) technique. The Rotation Forest classifier is then used to match between testing and training features (i.e. to classify alive and coagulant embryos). The experiments have proved that choosing threshold value of SFTA technique and the size of the rotation forest classifier have a great impact on the classification accuracy. With accuracy around 99.98 %, the experimental results have showed that the proposed model is a very promising step toward a fully automated toxicity test during drug discovery.

Peer-review under responsibility of Universal Society for Applied Research
Keywords: Fish embryo toxicity test; Zebrafish; Classification; Rotation Forest Classifier; Segmentation-based Fractal Texture Analysis (SFTA)

1. Introduction

Toxicity is an important research topic due to its vital role in the process of developing a new drug. Determining the toxicity level of thousands of compounds and comparing their toxicity levels have become a hot topic in recent

E-mail address: engalaatharwat@hotmail.com

Peer-review under responsibility of Universal Society for Applied Research
doi:10.1016/j.procs.2015.09.008

studies. There are many investigation methods to test the toxicity of any compound. One of the most famous methods is th animal to test the safety of a new drug. Food and Drug Administration (FDA) have reported that predicting drug failures before clinical trials save US $100 million per drug [1,2,3].

To choose a suitable animal for such safety test, there are many challenges, e.g. ensuring the reproducibility and the comparability. One of the animals that gained acceptance in drug discovery is the zebrafish. This is because the zebrafish enjoys many features supporting the test of the toxicity of any compound. Examples of these features include (a) the speed and the transparency [4], (b) the high production as they produce a large number (200-250 eggs) of embryos per mating [5].

Despite the fact that zebrafish embryos are the most suitable and promising model, but a number of challenges are still to be addressed. Firstly, the zebrafish model needs to be fully automated system (this is one of the biggest challenges because manual inspection of thousands of images is not accurate, slow, and not feasible for the data analysis). Secondly, this model should achieve a high level of reliability (this contributes toward a wide acceptance, i.e. becoming a standard model toxicity test).

To address the automation challenge, there should be automated process for the entire fish embryo test including the evaluation of data produced during a microscope screening process. However, to have a fully automated systems, data (i.e. embryo image) needs to be collected, preprocessed, and then features are extracted and then classified. However, collecting data and assign a true class label to each sample is considered a difficult step in the system. This step is currently handled manually because screening the microscope is performed in 24 or 48 hours after fertilization. During this period, the embryo is able to move within the chorion. Hence, detecting the actual position of the embryo in the image is difficult due to its movement and different orientation [4].

Many researchers have reported the use of the zebrafish embryos to investigate the toxicity of different compounds. Reimers *et. al* have used the zebrafish to compare the toxicity resulting from either ethanol or acetaldehyde exposure [6]. Also, using the zebrafish model, Hallare *et. al* have tried to evaluate the toxicity of diclofenac, anti-rheumatic drug [7]. Furthermore, R*ü*diger *et. al* have reported the first fully automatic approach for evaluating the zebrafish embryos without dechorionating or aligning them. They classified the healthy and coagulated embryos. They built their own data set which consists of 187 images of living embryos, 190 images of coagulant embryos, and seven erroneous images. To build their automated model, they first applied many preprocessing steps, dilation and erosion, to extract the image of the embryo clearly. They have then extracted and used eight geometric features for classification step. Their system has achieved an excellent accuracy rate, reached to 99.47% [8].

In this paper, a fully automated machine learning-based approach is proposed to classify zebra fish embryos to healthy and coagulated embryos. This approach aims to a fully automated evaluation of the toxicity of compounds based on zebrafish. The proposed approach makes use of the SFTA technique to extract features from zebrafish embryos images. These features are then classified using the Rotation forest ensemble classifier to match between the unknown and trained images.

The rest of this paper is organized as follows, Section (2) presents the methodology that is used in the proposed model; Section (3) explains the two phases of the proposed model and its detailed steps; Experimental scenarios are introduced in Section (4). Discussion of the experimental scenarios are discussed in Section (5). Finally, conclusions and future works are presented in Section (6).

2. Preliminaries

2.1. Segmentation-based Fractal Texture Analysis (SFTA)

SFTA feature extraction algorithm consists of two main steps. In the first step, the input grayscale image is decomposed into different binary images based on multi-level threshold algorithm. There are different methods to decompose the grayscale image into a set of binary images, but in this paper, Two-Threshold Binary Decomposition (TTBD) method is used. In the second step of SFTA, the features are extracted from each binary image. SFTA features consist of fractal dimension, mean, and size (pixel count), which are computed from the region's boundary of each binary image [9]. More details about the two steps of SFTA are explained in the next sections.