UNIVERSITY OF LIVERPOOL

Computer Vision Systems for Zebrafish Larvae Analysis

Thesis submitted in accordance with the requirements of

the University of Liverpool for the degree of Doctor in Philosophy

By

Bayan Abd AlKareem Alsaaidah

August 2019

Abstract

Biologists and pharmacologists commonly use zebrafish embryos during the testing of drugs due to their properties and high genetic similarity with mammals. The testing of these substances is a tedious and painstaking process, carried out manually by trained experts who determine whether the embryos have been deformed or killed as a result of administering the chemical.

Computer vision presents efficient solutions for such aquaculture research problems by the application of machine learning systems. However, having accurate and cost-effective monitoring and analysis systems still has challenges due to the small size and the high similarity of the samples and the unwanted position of the samples due to their free swimming. In this thesis, a set of novel and cost-effective systems are presented consisting of detection, classification, identification, and counting zebrafish samples.

In the first part, two applications are proposed to address several problems that the biologists face in their work. First one is a novel and accurate system for detecting and counting the number of transparent live eggs inside a petri dish that may contain hundreds of eggs either dead or live with other types of objects. The

second one is counting the adult fish inside a tank that confuses the biologists in their work without transferring the fish from tank to another as an attempt for counting them.

In the second part, an image analysis pipeline is presented from the data acquisition process to data analysis and classification system to get a final result about the embryo status. The proposed work aims to identify the health and detect the abnormalities of zebrafish embryos using scanned images and based on computer vision algorithms. The images are comprised of many features which could be extracted automatically or manually.

In the third part, a multi-label classification system for zebrafish embryo deformities based on microscopic images is presented. The proposed system aims to determine the numbers, health, and presence of abnormalities in zebrafish larvae using a classification system with high-throughput delivering results faster than the manual process, and assisting in the pharmacological and toxicological experiments. The novelty here is having a malformation classification system depending on different feature types and identifying all the deformity classes that may appear on a sample body at the same time using binary relevance multi-label algorithm.

According to the proposed application purposes and based on computer vision algorithms, this thesis successfully addressed most problems and provided efficient models for analysing, classifying, and counting the zebrafish larvae. The proposed models provided the biologists by a novel and cost-effective computer vision models.

Declaration of Authorship/Originality

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as full acknowledgement within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and in the preparation of the thesis itself has been acknowledged. I certify that all information sources and literature used are indicated in the thesis.

Signature of candidate

. . . . . .

Dedication

To my beloved husband, who's always pushed me to achieve my dreams. Without his love, encouragement, support and patience this thesis would be hardly completed.

This thesis is dedicated to him.

Bayan

Acknowledgments

First and foremost, all praise is due to Allah who guided me to this. This thesis would not have finished without the strength that God gave me.

I acknowledge my supervisor Dr Waleed Al-Nuaimy. He was always available with his invaluable advice, encouraging words, creative vision and supporting guidance. Many thanks and appreciations for Dr Iain Young for his support and advice. His knowledge and experience help me a lot during my PhD journey.

I want to thank Dr Lynn Mclaughlin from the Institute of Technology Directorate and the aquarium staff in the Institute of Integrative Biology laboratories in the University of Liverpool, UK, for their help and support during the experiments. Many thanks to my colleague from the biological science department Rui Zhang for her help during the experiments.

I am thankful to my husband Dr Mohammad Rasoul Alhadidi for his support, encouraging and loving over the years that helped me to complete this thesis.

I am thankful to my children, Lemar, Osama, and Ahmad for there

existing in my life, they encouraged me by their pure and innocent words.

I would like to thank all members of my family, my loving mother, my brothers and sisters for their support and encouragement.

I am thankful to all my friends, Shoroq Alfarajat, Ola Younis, Iman Olaymat, and Lubna Alharbi for their support and encouragement.

To All,,,I am pleased to say..***THANK YOU***

Bayan

28-6-2019

Contents

Abstract ii

Declaration of Authorship/Originality iv

Dedication v

Acknowledgments vi

1 Introduction 1

1.1 Background 1

1.2 Motivation 2

1.3 Research Aims and Objectives 5

1.4 Summary of Contributions 6

1.5 List of Publications 7

1.6 Thesis Outline 8

2 Background 11

2.1 Introduction 11
2.2 Biological Research 11
2.2.1 Zebrafish Larvae 13
2.2.2 Zebrafish Larva Screening 15
2.3 Computer Vision Technologies 17
2.3.1 Colour Space Models 22
2.3.2 Texture Features 26
2.3.3 Hough Transform 28
2.3.4 Classification and Regression Decision Tree . 29
2.3.5 Convolutional Neural Networks 32

3 Automatic Counting Systems 35

3.1 Introduction 35
3.2 Challenges and Objectives 35
3.3 Automatic Zebrafish Counting System 36
3.3.1 Experiments 38
3.3.2 Video Analysis 41
3.3.3 Statistical Analysis 42
3.3.4 Results and Discussion 44
3.4 Eggs Counting System 46
3.4.1 Data Collection 48

3.4.2 System Methodology 52
3.4.3 Results and Discussion 55
3.5 Summary 58

4 Image Analysis Pipeline For Zebrafish Larvae 60
4.1 Introduction 60
4.2 Challenges and Objectives 60
4.3 Hardware And Data Collection 62
4.3.1 Experiment protocol 62
4.3.2 Chemicals 63
4.3.3 Hardware 63
4.4 Data Preparation 66
4.4.1 Embryo Orientation 67
4.5 The System Methodology 72
4.5.1 Zebrafish Larva Identification and Classification Inside Petri Dish 74
4.5.2 Zebrafish Larva Identification Inside n-well Plate 82
4.6 Results and Discussion 87
4.7 Summary 92

5 Zebrafish Embryo Malformation Classification 95

5.1 Introduction 95
5.2 Challenges and Objectives 96
5.3 Data Set 97
5.4 Binary Relevance And Decision Tree 99
5.4.1 Image Analysis 100
5.4.2 Feature Extraction 101
5.4.3 Training classification models 112
5.4.4 Testing the trained models 113
5.5 Deep Learning 114
5.5.1 CNN architecture 116
5.5.2 Training and Validation 118
5.6 Results and Discussion 119
5.6.1 The Performance Equations 121
5.6.2 Results and discussion 122
5.7 Summary 128

6 Conclusion and Future Work **131**

6.1 Conclusion 131
6.2 Future Work 134

Appendix A: HP Scanjet G4050 Photo Scanner 160

Appendix B: 96-well Plate Technical Sheet 162

Appendix C: 48-well Plate Technical Sheet 165

Appendix D: Petri Dish 100mm Technical Sheet 168

List of Figures

1.1 Number of living animal procedures in UK 2

2.1 Zebrafish adult (Danio rerio) 14

2.2 Zebrafish embryo (3dpf) 15

2.3 Toxicological experiments 16

2.4 Hue Saturation Value space 23

2.5 CIELab colour space 25

2.6 Decision tree algorithm 30

2.7 General structure of CNN 34

3.1 Fish counting system 38

3.2 First setup 39

3.3 Second setup 40

3.4 Background removing: (a) Frame (b) Background (c) Subtraction result 43

3.5 Statistical values for each video (10 fish) 44

3.6 Statistical values for each video (5 fish) 45
3.7 Egg counting stages 49
3.8 Zebrafish eggs inside a petri dish (1200 dpi) 50
3.9 Live and dead eggs 51
3.10 Small region of the RGB petri dish image 53
3.11 Hue channel image of the HSV colour space 53
3.12 Gamma correction result 54
3.13 Binary image 55
3.14 Detected circles (live eggs) 56
3.15 Egg counting system results 57

4.1 Images using scanner with (a) 300dpi (b)1200dpi . 65
4.2 General methodology for orientation correction . . . 68
4.3 Embryo rotation: a) before b) after 69
4.4 Horizontal flip: a) before b) after 69
4.5 Vertical flip: a) before b) after 70
4.6 Fixing results: a) before b) after 71
4.7 General overview of the system (a) dish images (b) n-well plate images 72
4.8 Sample classes a) dead egg b) live egg c) chorion d) normal 73
4.9 Segmentation process (petri dish) 75

4.10 Samples roundness feature 76
4.11 Zebrafish larva eye detection 77
4.12 Pre-processing operations 79
4.13 Pre-processing operations: a) 96-well plate b) 48-well plate 83
4.14 Inner and outer well diameter 84
4.15 Wells segmentation 85
4.16 Wells labeling 86
4.17 Zebrafish larva segmentation 87
4.18 Zebrafish larva close to the well edge 88
4.19 Examples of well-plate classification 89
4.20 Classification results for the whole dish image . . . 89
4.21 Identification results for the whole dish image: a) Example 1 b) Example 2 c) Example 3 d) Example 4 91
4.22 Drawbacks of the classifier 92

5.1 Malformation Examples: a) Chorion b) Dead c) Down-Curved-Tail d) Edema e) Hemostasis f) Necrosed-yolk-sac g) Short-tail h) Up-curved-tail/fish i) Normal 98
5.2 Orientation fixing points 99
5.3 General structure of the system 100
5.4 Image pre-processing 101

5.5 Positive y-axis gradient of the cropped tail (a) normal tail (b) up-curved tail (c)down-curved tail . . . 103

5.6 Polynomial curve fitting for the y-axis gradient (a) up-curved tail (b) down-curved tail (c) normal tail . 104

5.7 Curvature angle points 105

5.8 Curvature angle for different classes 106

5.9 The convex hull of the larva body 107

5.10 The extreme points of the body 107

5.11 Example of the eye segmenting with the yolk 108

5.12 The segmented yolk 108

5.13 The vertical projection 109

5.14 Red spots detection 110

5.15 Hemostasis red spot detection 111

5.16 The most effective features for dead class 111

5.17 The extracted features for short class 112

5.18 Training stage (Binary relevance) 114

5.19 The testing stages 115

5.20 GoogleNet architecture 117

5.21 ResNet50 architecture 117

5.22 Training error and testing error for two network depths 118

5.23 Zebrafish larva with detected malformations 123

5.24 Misclassified: a) Dead larva b) Chorion larva 124

5.25 Misclassified hemostasis sample 124

5.26 Multi-label classification accuracy using GoogleNet 127

List of Tables

2.1 Summary of Methods 20

3.1 Fish counting systems 37

3.2 Counting error (10 fish) 46

3.3 Counting error (5 fish) 46

4.1 Data Set Division 82

5.1 Extracted features 102

5.2 Multi label classes 120

5.3 Multi-label evaluation 123

5.4 Evaluation comparison of the proposed system with the existing one 125

5.5 Testing accuracy of DL models 128

Chapter 1

Introduction

1.1 Background

Producing any new drug or developing an existing one needs to be tested before consuming by humans, this process is always done on an experimental model to check the effects of these substances and decide the suitability of this drug for human. Zebrafish prove that it is a great model for this type of experiment because of their fast responses that appear on their bodies [1].

Zebrafish becomes a fundamental animal model that is recently used in pharmacological and toxicological studies according to its high genetic similarity with mammals and also due to the fast response of their bodies either on the behaviour of the fish or any physical change that could be happened of any part of the body [2, 3]. Many biological studies found that by adding chemical substances with several concentrations, the zebrafish embryos may either stay normal or exhibit a change in any part(s) of their bodies according to the type of these materials [4–6].

According to the annual statistics of living animals that are involved in the scientific research in UK, 3.08 million procedures was carried on in UK using mice, rat, and fish in 2013 only. Using fish was 12% of these procedures and becomes the second popular animal model after mice since 2008 that registered rising in the number of the pharmacological and toxicological studies. Figure 1.1 shows these statistics in 2013 [7].

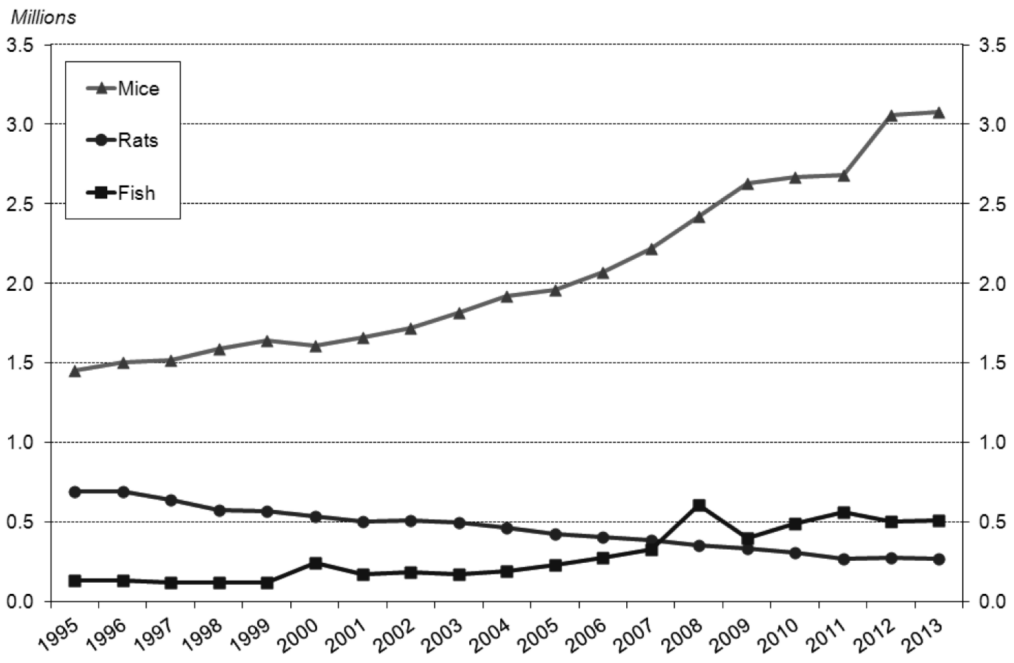


Figure 1.1: Number of living animal procedures in UK

1.2 Motivation

Larvae are typically bred in large quantities, inside the experimental container either n-well plates or the petri dish. The biologist needs to study and analyse each larva individually for the observa-

tion process which is tedious and costly, as it takes a long period to decide all samples status if it is normal or has been affected by the added chemicals. Manual monitoring and analysing is a time-consuming process.

A large number of produced embryos that need to be classified within the period that has been specified by European law for animal protection which is 5 dpf (days post fertilisation) [8]. With all these factors, the need for a fully automated system still a demand for a high-throughput classification system.

Despite the rapid growth of zebrafish embryo using as an experimental model, there is still a lack of automated classification systems. Screening the zebrafish egg and larva abnormalities within one system where the zebrafish embryos grow in a fast way and hatch within a few hours has not been reached yet upon to our knowledge.

The lack of automatic capturing systems limits the development of fully automated systems. The traditional imaging tool consists of a microscope with a high-resolution camera which needs awareness of several factors, provides the system by one sample image and it is a time consumer. Using the proposed imaging tool which is affordable and easy to use, the images are collected automatically, and the biologist only needs to place the dish or the plate on the scanning area.

According to [9–11], designing an automated system to clas-

sify the deformations of zebrafish larvae is a considerable challenge. In 2012, Jeanray et al. attempted an automated system to classify two phenotypes then they extended their work in 2015 [12, 13]. However, it is not a fully automated system, and also there were insufficient results with some malformation classes, and it was used with a 6-well plate where the data acquisition was carried out individually for each well.

The demand for highly accurate and fast image acquisition and analysis processes presents another limitation for many works and studies [14]. Also, detecting the actual position of the embryo inside the well is also another problem [15].

The manual fish counting process is harmful to the fish where the scientists need to transfer the fish from tank to another using small nets. This work proposes a solution for fish counting problems such as noises, the reflection of the fish image on the tank sides. This system analyses a video using several statistical analysis that overcomes the tank reflection problem and the overlapping fish issue.

During the biologist experiments, they need to count the live and dead zebrafish eggs inside the petri dish regarding their research study. The manual detecting and counting process of the dead and live eggs is tedious and takes much time where the dish may contain hundreds of eggs. To address the manual process problem, a novel identification and counting system is proposed to detect the dead and live eggs then count them efficiently.

1.3 Research Aims and Objectives

The main goal of this research is to provide novel high-throughput systems to support biologists in their research and experiments. This research aims to develop computer vision techniques to address and overcome the biological experiment problems.

This research provides the aquaculture field with novel, high performance, and cost-effective applications consisting of zebrafish larva identification and classification based on two types of images (microscopic and scanned images) and zebrafish egg/adult counting based on computer vision techniques. The main objectives of this research are summarised below:

- Using efficient and more accurate machine intelligence algorithms for detection and classification of the malformations of zebrafish larvae. Identify all the deformity classes that may appear on a one sample body at the same time.

- Present an affordable and time-saving imaging tool which is a flatbed scanner with high resolution, low cost, and friendly use and introduce a fixed set up for video recording process to overcome the manual procedure.

- Design an image analysis pipeline to identify the status of the zebrafish egg/larva inside a petri dish and n-well plate based on computer vision techniques and based on the collected scanner images. Investigate the effects of the chemicals and drug and

recognise the survival rate of the samples.

- Propose a counting system to identify and count the live eggs inside the petri dish using image analysis techniques based on scanned images.

- Design an accurate and high-throughput system for fish counting inside a tank based on recording videos for free-swimming fish and using image processing algorithms.

1.4 Summary of Contributions

In this thesis, a range of problems and challenges were successfully addressed and overcome, and the proposed objectives have been achieved. The research contributions are presented below:

1. Present a high-performance image analysis pipeline to segment, detect, and identify the zebrafish larvae interaction with stimuli and find out the survival rate of the experimental samples.

2. Develop a novel multi-label classification system to classify zebrafish larva malformations using machine learning and deep learning models and focusing on the main features resulting in deep analysing of all classes to raise the reliability of the system.

3. Collect, organise, and analyse data set of zebrafish larva samples using a cost-effective tool which is the flatbed scanner.

4. Determine the numbers, health and presence of abnormalities in zebrafish larvae using a high-throughput system that is validated by measuring the well-known effects of several chemical substances that were tested in the Institute of Integrative Biology laboratories such as alcohol and ammonia, also testing the reliability of the system by comparing the automated system output with the biologist observations.

5. Present two counting systems which successfully count the swimming zebrafish inside a tank and count the zebrafish eggs inside a petri dish and addressed the traditional counting process problems.

1.5 List of Publications

During the research journey and to document the work concepts and results, several articles were published as follows:

1. **Bayan Al-Saaidah**, Waleed Al-Nuaimy, Moh'd R. Al-Hadidi, and Iain Young. Zebrafish Larvae Classification based on Decision Tree Model: A Comparative Analysis. Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 4, 347-353 (2018).

2. **Bayan Al-Saaidah**, Waleed Al-Nuaimy, Iain Young, and Moh'd R. Al-Hadidi. Multi-label classification for zebrafish larva malformations using binary relevance and decision tree. Image and

Vision Computing, Elsevier, (2019). (Submitted)

3. **Bayan Al-Saaidah**, Waleed Al-Nuaimy, Moh'd R. Al-Hadidi, and Iain Young. Automatic counting system for zebrafish eggs using optical scanner. In Information and Communication Systems (ICICS), 9th International Conference on (pp. 107-110). IEEE, Irbid, Jordan, 2018.

4. **Bayan Al-Saaidah**, Waleed Al-Nuaimy, Majid Al-Taee, Iain Young , and Qussay Al-Jubouri. Identification of Tail Curvature Malformation in Zebrafish Embryos. In International Conference on Information Technology (ICIT), 8th International Conference, IEEE, Amman, Jordan, 2017.

5. **Bayan Al-Saaidah**, Waleed Al-Nuaimy, Majid Al-Taee, Ali Al-Ataby, Iain Young , and Qussay Al-Jubouri, Analysis of Embryonic Malformations in Zebrafish Larvae, Developments in eSystems Engineering (DeSE), Liverpool, UK, 2016.

1.6 Thesis Outline

This thesis is organised into six chapters. The following present a brief description of the remaining chapters of this thesis.

1. Chapter 2: in this chapter, most concepts in the biological and engineering fields that were concerned and used in this work are illustrated. Background about the biological and toxicological

experiments is introduced. Overview of the zebrafish and its samples of eggs/larvae is provided. A brief description of the biological experiments is presented. The computer vision and the artificial intelligence fundamentals and theories that were used and benefited in this work are explained.

2. Chapter 3: this chapter introduces two counting systems to solve significant problems in the aquaculture. Background of the existing counting systems are reviewed. The challenges and limitation of this work are presented. The gathering data tools of the two systems are discussed. The main parts of the system and the detailed analysis of both systems are explained. The results are presented, considered, and compared with the existing systems.

3. Chapter 4: in this chapter, a new classification model is introduced. The imaging tool and the data set collection are explained. The automatic detection, segmentation, and classification of the samples healthiness are briefly clarified. The system challenges and limitation are discussed. The system results are provided and analysed.

4. Chapter 5: in this chapter, an overview of the manual observation problems are presented. The system objectives are pointed out. The system works with all stages from data collection to feature extraction and classification are briefly explained. The result was provided, analysed, discussed, and compared with the existed systems.

5. Chapter 6: in this chapter, the research journey motivations and results are concluded. Some recommendations for future work are suggested and introduced.

Chapter 2

Background

2.1 Introduction

Recently, the zebrafish and specifically the larva has become the second animal model that is widely used in toxicological and pharmacological experiments. This chapter introduce the biological research idea, using the zebrafish larvae in the research, and how the zebrafish larva screening process is carried out. This chapter introduces a summary of the theories and concepts of the fundamental computer vision algorithms that were used in this research and provide a brief explanation of the zebrafish larva screening process.

2.2 Biological Research

Biology can be defined as the science of life which concerns all the living organisms. It studies their physical structure, cells, inner physiological processes and mechanisms, and the organism's development stages [16].

The cell is the fundamental unit of all living things. The whole organism cells have been produced from a single cell by cell division process. Basically, the cell presents the basic unit in many pathological processes [17]. Moreover, the cells contain the genetics information holder (DNA) which is passed from cell to another during the cell division process.

The biology science is like an umbrella that covers many branches which are more specific sciences. The biology includes the ecology, ethology, pharmacology, and genetics [18]. Each type has different characteristics and concerns.

Pharmacology is one of the biology science branches aims to study drug action using animal models [19]. The pharmacology science is overlapped with the toxicology which concerned and involved in drug practising and studying [20]. Diagnosing and testing the doses is carried out to find out the dose's effect depending on the relation between the treatment and the animal model.

Over recent years the zebrafish has become one of the most common animal models due to many factors including a high degree of genetic similarity with humans, short generation times, transparent larval stages, extensively annotated genome and simple husbandry [21, 22].

Zebrafish are now widely used in drug development, to measure the impact of environmental changes, of toxins and pollutants

and many other applications. However, using mammals in the biological experiment is expensive and laborious, it also led to an increasing number of ethical issues for toxicological research, and that have been limited in large-scale screening efforts [23].

Screening zebrafish larvae development and evaluating the effects of the chemical compounds is started from the early ages of the samples. This process is carried out manually for a large number of samples. After each mating time, the single female can produce hundreds of eggs [24] that differ depending on their hormones [25].

2.2.1 Zebrafish Larvae

Zebrafish is a freshwater fish which is called Danio rario. The zebrafish are belonging to the minnow family [26]. The zebrafish species have a sexual maturity approximately three months after being hatched. The single female of zebrafish can produce hundreds of eggs in each matting time which can be happened over the whole year [27]. After 24hpf (hours post fertilisation), most parts of the body are present and easily seen including the eyes, ventricles, and the heart [28]. Figure 2.1 shows an example of the adult zebrafish [29].

In particular, zebrafish larva is widely used in scientific research as an animal model according to its characteristics. The zebrafish larvae have been used in toxicity assessment [30–32], in medical research such as the cancer and the infection disease [33–

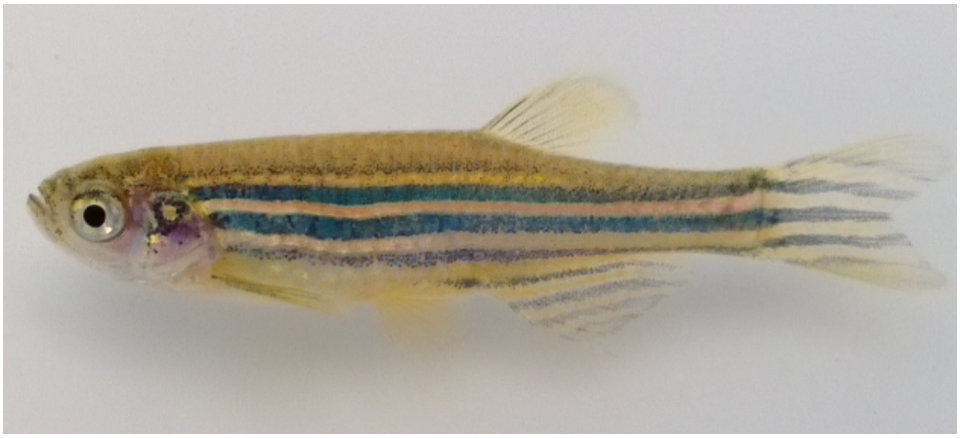


Figure 2.1: Zebrafish adult (Danio rerio)

35], and in development studies [36, 37].

The zebrafish larvae achieved high popularity according to many advantages that have been noticed and benefited by the researchers. Biologists use zebrafish in their experiments instead of other animals for many reasons. It can propagate over the year, not in a specific time, the time period between fertilisation to hatching is short (72 hours), the cheap husbandry, the high genes similarity between the zebrafish and the mammals, the shape of its body which is transparent help the biologists in their observations, and the fast response to the chemicals [38–41]. Figure 2.2 shows an example of 3dpf (days post fertilisation) embryo [42].

According to standard protocols, laws, and regulations zebrafish embryos have been submitted to use in the experiment instead of using the adult and also with a specific age depending on the degree of the acuteness of the added materials, as only larvae less than five days after fertilisation are exempt from this legislation [8, 43–45]. These animal protection laws require the researchers to

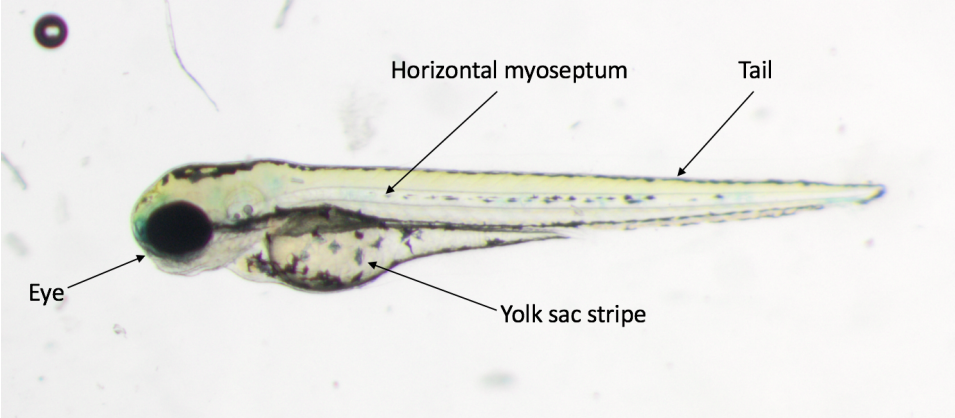


Figure 2.2: Zebrafish embryo (3dpf)

accelerate the screening experiments especially with a large number of samples. The artificial intelligence systems can save the researcher time and overcome the delays of manual processes.

2.2.2 Zebrafish Larva Screening

Using mammals in biological experiments is considered expensive and laborious which increase the ethical issues during the toxicological experiments [23]. In contrast, the zebrafish provide the biologists with many benefits including the low cost of the laboratory care and husbandry of the samples and the high throughput of the pharmacological and toxicological experiments [46].

According to all advantages of using zebrafish larva as an animal model, it was widely used in behaviour monitoring, malformation detection, and development screening. As a result, providing and developing new techniques becomes a demand to analyse and help the biologists in their work. New techniques have been provided by researchers to assist the biological and the pharmacol-

ogist in their work and present such as behaviour tracking [47–49], malformation detection [50, 51], species counting [52, 53], heart screening [54–56], and sizing [57, 58].

The toxicity assessments aim to identify and analyse the adverse effects of any drug or dose after adding them on the experimental organisms [59]. Figure 2.3 shows the general procedure of biological experiments.

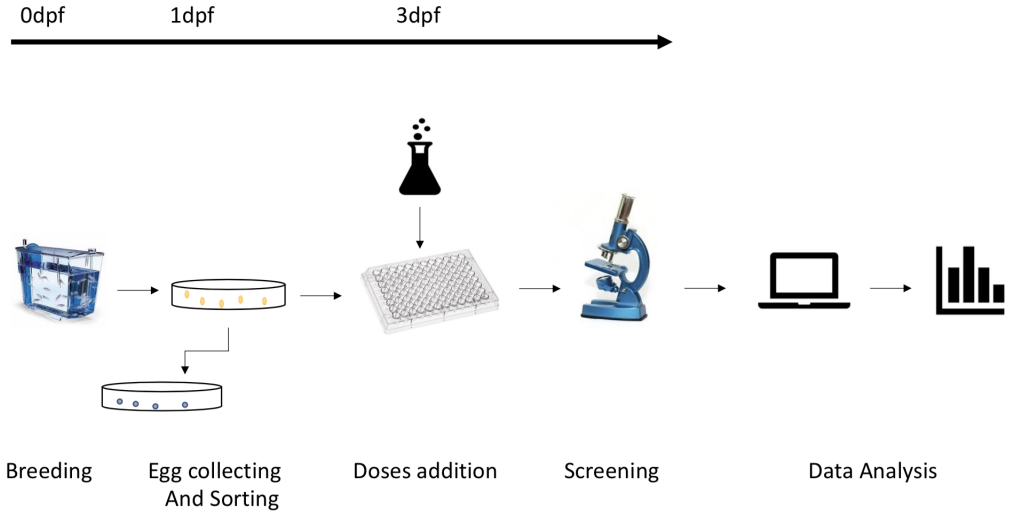


Figure 2.3: Toxicological experiments

The experiments start with breeding zebrafish males with females in a specific ratio between the males and the females like 1:1 or 3:2. After waiting overnight, the eggs will be collected from the tank and put inside a container for the sorting step where the dead eggs are removed, and the live eggs will be moved to the n-well plate for the experimental purposes. The sorting process is carried out using the pipette and under the microscope. The chemicals

and doses are added to the samples according to the experiment requirements. The final step before analysing the data, the samples are screened and checked to register the dose's effects that have been observed by expert [60, 61]. All mentioned steps are carried out manually by an expert biologist.

The zebrafish embryos respond to the stimuli by either changing the behaviour during their movement or as physical changes that are occurred of any part of their bodies. Many biological studies found that by adding chemical substances with several concentrations, the zebrafish embryos may either stay normal or exhibit a change in any part(s) of their bodies according to the type of these materials [4–6]. These malformations are registered and analysed to get reliable results of any medicine and find out the drug effects. The observation process takes a long period to decide the larva status, rather than the decision of this status should be accompanied by a biological expert and in most cases, this process is harmful to the zebrafish.

2.3 Computer Vision Technologies

Computer vision "is the science and technology of making machines that see" [62]. Computer vision is a multidisciplinary scientific field which deals with digital data (images/videos) using computational algorithms. Using the engineering aspects helps the other areas to automate their tasks which is usually executed by humans [63].

Computer vision concerns the data acquisition, processing,

extracting the required information, and analysing the obtained information according to the goal. The input data should be digitised to be comprehensible by the machines[64].

For a high-throughput zebrafish malformation screening, an automated system for zebrafish embryo phenotypic detection using machine learning algorithm and fluorescent images was presented by Vogt et al.. The zebrafish embryo fluorescent imaging is considered as a straight forward way for object detection and segmentation. However, the proposed system discarded the images with dorsal or ventral position objects. The proposed system used the Cognition Network Technology (CNT) for image analysis to analyse and measure the phenotypes of the experimental samples [11].

Alshut et al. [15] segmented the embryos from the collected images and classified the sample into living or coagulated using Bayes classifier. However, the image gathering process was carried out using a camera over the microscope; this way needs to be aware of several factors like illumination and the focal point for high performance capturing process. The capturing process is carried out using different focal points to reach a suitable one. The proposed classification algorithm was Bayes model using colour and shape features.

The same data set of [15] was used by Tharwat et al. [14] to classify the images into healthy and coagulated eggs. These texture features were extracted using Segmentation-based Fractal Texture Analysis (SFTA) with the rotation forest classifier.

Using video analysis for zebrafish embryos was proposed in [65] to classify if the sample is live or dead. By detecting the zebrafish embryo heart position and analysing the intensity variation of the heart, if the cyclic motion areas are zero, then the sample will be considered as a dead embryo.

Computer vision systems for zebrafish larva malformation detection and classification were limited and had several challenges. Recognising two classes of the zebrafish embryos was proposed in [12], where the authors designed a classification system to classify two types of zebrafish embryo malformations automatically. This process came after treating the eggs by adding some chemical substances. The authors proposed a classifier based on a tree learning algorithm for evaluating malformations of the larvae during their growth that are caused by adding chemical substances. They used a supervised learning method with image processing to identify only two types of the malformations, edema and curved tail.

This work was then extended in [13] by adding other classes to the previous system, including hemostasis, necrossed yolk sac, edema, and tail malformations. They used a 6-well plate and a classification model similar to that reported in [12], the liquid was removed, and the imaging was carried on individually for each larva and under the microscope which can be considered as a complicated imaging procedure. The designed system by Jeanray et al. [13] is not fully automated, and the accuracy of many classes in their system is between 53%-89% which need to be upgraded.

Using video analysis for zebrafish embryos was proposed in

[65] to classify if the sample is alive or dead. By detecting the zebrafish embryo heart position and analysing the intensity variation of the heart, if the cyclic motion areas are zero then the sample will be considered a dead embryo.

Computer vision algorithms have been used in many automatic applications for zebrafish specimen screening. The screening for one part of the zebrafish body or the whole body, use images or videos, and analyse the data features according to the system goals. Table 2.1 shows different applications of zebrafish specimen screening.

Table 2.1: Summary of Methods

Author	**Application Name**	**Imaging Tool**	**Target part**	**Fish Age**
Alshut et al. 2009 [15]	Identification and classification	microscope	Whole body	48hpf
Alshut et al. 2010 [9]	Toxicological screening	microscope	Whole body	48hpf
Berghmans et al. 2007 [66]	Behaviour tracking	Camera	Whole organism	6dpf
Bhat et al. 2009 [67]	Heartbeat detection	Microscope	Tissue (heart,blood cells)	48hpf

Chan et al. 2009 [68]	Heartbeat detection	Microscope	Tissue	52hpf
Cachat et al. 2011 [69]	Behaviour patterns and fish tracking	Camera	Whole body	Adult
Chen et al. 2011 [50]	Phenotype detection	Microscope	Tissue	n.s.
Cario et al. 2011 [70]	Behaviour patterns and fish tracking	Camera	Whole organism	Adult
Kanungo et al. 2011 [5]	Phenotype detection	Microscope	Cells	28hpf
Kato et al. 2004 [71]	Behaviour patterns	Camera	Whole organism	3dpf-adult
Liu et al. 2006 [72]	Phenotype detection	Microscope	Whole organism	24hpf
Liu et al. 2008 [6]	Cell patterns	Microscope	Cells	24hpf
Liu et al. 2012 [73]	Phenotype detection	Camera	Whole body	48hpf
Mandrell et al. 2012 [74]	Toxicity screening	Camera	Whole body	6hpf
Peravali et al. 2012[75]	Automatic screening	Camera	Whole organism	24-120hpf
Spomer et al. 2012 [76]	Heartbeat and tissue detection	Camera	Tissue	60-72hpf
Vogt et al. 2009 [11]	Phenotype detection	Microscope	Whole body	26-48hpf
Xu et al. 2010 [77]	Phenotype detection	Microscope	Whole body	3-5dpf

Zanella et al. 2010 [78]	Cell detection	Microscope	Cells	3-7hpf
Jeanray et al. 2015 [12, 13]	Phenotype detection	Camera and Microscope	Whole body	3dpf
Bayan [79]	Phenotype detection	Scanner	Whole organism	1-5dpf
Bayan [80]	Phenotype detection	Public data set (microscope)	Whole organism	3dpf

According to the research objectives, several concepts and algorithms have been used and adapted due to the system goals. A brief explanation and an extensive background about the algorithms and theories that were used in this research are presented in the next sections.

2.3.1 Colour Space Models

The colour space is defined as a mathematical model that is produced from the organisation of the colours in a specific way. The primary colours where all the other colours are produced are Red, Green, and Blue which present the channels of the colour space RGB. Every pixel in an image is presented by the three main channels which are the red, green, and blue. The RGB colour space is the standard colour system for the digital images [81]. However, the RGB colour space can be converted into different colour spaces like the Hue Saturation Value (HSV) [82].

The HSV colour space is different from the RGB colour space where the HSV separates the image intensity from the colour information (chromaticity) [83]. The intensity of the colour is represented by the value channel where the hue and saturation channels

represent the chromaticity; the hue channel represents the colour as observed by the human vision, and the saturation is related to the purity of the hue and the white channels mixture [84].

The HSV has a cylindrical geometry with angular dimensions starting from the red colour, passing the green and blue to represent more colours such as yellow, magenta, and cyan. Figure 2.4 shows the HSV colour space[85].

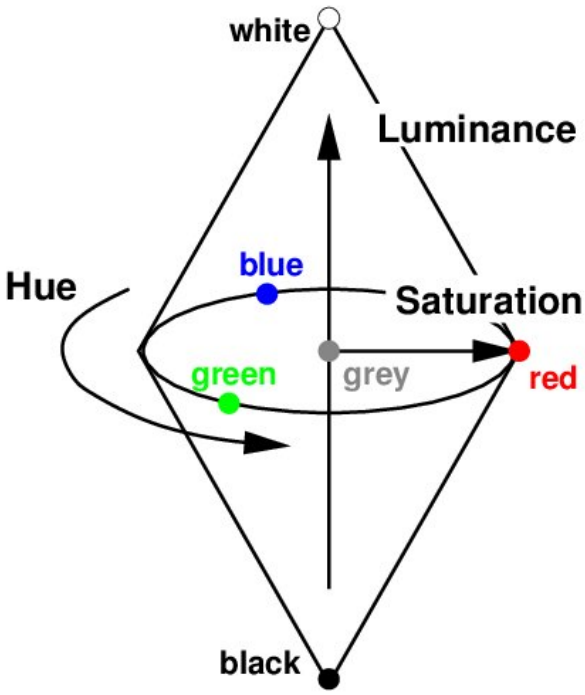


Figure 2.4: Hue Saturation Value space

The conversion from the primary colour space RGB to the HSV colour space is carried out using several calculations are shown in the equations below. The range of the RGB values is from 0-255 to 0-1 values [86].

R' = R/255, G' = G/255, B' = B/255

C_{max} = max(R',G',B')
C_{min} = min(R',G',B')
$\Delta = C_{max}$ - C_{min}

$$H = \left\{ \begin{array}{l} 60^o x \left(\frac{G'-B'}{\Delta} \bmod 6 \right), C_{max} = R' \\ 60^o x \left(\frac{B'-R'}{\Delta} + 2 \right), C_{max} = G' \\ 60^o x \left(\frac{R'-G'}{\Delta} + 4 \right), C_{max} = B' \end{array} \right\} \tag{2.1}$$

$$S = \left\{ \begin{array}{ll} 0 & , \Delta = 0 \\ \Delta \ / \ C_{max} & , \Delta <> 0 \end{array} \right. \tag{2.2}$$

$$V = C_{max} \tag{2.3}$$

The CIELab is a colour space that is designed for perceptual uniformity concerning the human colour vision. CIE is defined as a commission on illumination internationally in 1976 [87]. It represents the colours as three numerical values (channels) L*a*b. L channel is related to the lightness, a and b represent yellow-blue and red-green combinations. The red-green (*a) has been used in marine sediments as an indicator of iron-manganese content [88].

Each channel of the CIELab colour space has a different values according to each channel properties. L* channel values are between 0 (black) to 100 (white), the values of channel a* varies from -60 (green) to +60 (red), and b* channel values are between -60 (blue) and +60 (yellow). Figure 2.5 shows the CIELab colour space [89, 90].

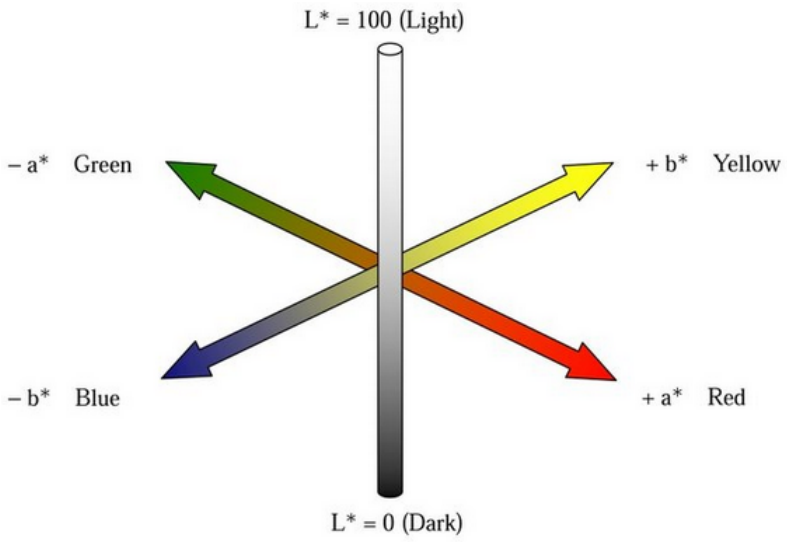


Figure 2.5: CIELab colour space

The transformation from RGB colour space to the CIELab model, it is done after passing through the RGB-CIE XYZ transformation. The transformation process is carried out using the following calculations.

$$L* = 116f\left(\frac{Y}{Y_n}\right) - 16 \tag{2.4}$$

$$a* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \tag{2.5}$$

$$b* = 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \tag{2.6}$$

where:

$$f(t) = \begin{cases} \sqrt{3}t & ,t > \delta^3 \\ \frac{t}{3\delta^2} + \frac{4}{29} & ,otherwise \end{cases}$$

The values of X_n, Y_n, and Z_n are found from the RGB-CIE

XYZ transformation. $\delta = 6/29$.

The colour transformation has been used and benefited in many image processing applications. However, this technique has not been used yet for zebrafish embryos recognition relating to the literature.

2.3.2 Texture Features

Basically, the feature extraction process is defined as a dimensions reduction process. Where the most valuable information will be extracted for the large data and describe the original data [91].

The image texture is the information of the spatial arrangement of the colours or intensities of a specific region in an image [92]. The average grey level and the spatial frequency are the most accurate and useful features for object segmentation [93].

The grey level relations are represented by the co-occurrence matrix which is defined as the angular relationships between the neighbour resolution cells which is interpreted by the distance between these cells [94].

The co-occurrence matrix is used to extract multiple numerical features that are used for texture identification and classification. The second statistics are calculated depending on a matrix $C_{\theta,d}(I_{p1}, I_{p2)}$ of the relative frequencies that describes how often the

two pixels (I_{p1}, I_{p2}) of different or similar gray levels N_g appear as a pair in the image matrix concerning the distance d and the direction θ. The value of this parameter N_g is 8 levels.

By using the co-occurrence matrix, many texture features can be extracted where the smoothness, coarseness, and the image texture information are described and quantified. The image contrast, correlation, cluster shade, cluster prominence, energy, homogeneity, entropy, and variance are measured as following [94, 95]:

1. Auto correlation:

$$f_1 = \sum_i \sum_j (ij)(C(i,j)) \tag{2.7}$$

2. Contrast:

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} C(i,j) | \, |i-j| = n \} \tag{2.8}$$

3. Correlation:

$$f_3 = \frac{\sum_i \sum_j (ij) C(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{2.9}$$

Where:

$$\mu_x = \sum_i i C_x(i) \tag{2.10}$$

$$\mu_y = \sum_j j C_y(j) \tag{2.11}$$

$$\sigma_x = \sum_i C_x(i) - \mu_x(i) \tag{2.12}$$

$$\sigma_y = \sum_j C_y(j) - \mu_y(j) \tag{2.13}$$

4. Cluster prominence:

$$f_4 = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 C(i,j) \tag{2.14}$$

5. Energy:

$$f_5 = \sum_i \sum_j C(i,j)^2 \tag{2.15}$$

6. Entropy:

$$f_6 = \sum_i \sum_j C(i,j) log(C(i,j)) \tag{2.16}$$

7. Homogeneity:

$$f_7 = \sum_i \sum_j \frac{1}{1 + (i-j)^2} C(i,j) \tag{2.17}$$

2.3.3 Hough Transform

Hough transform is a feature extraction technique used in computer vision to extract objects depending on their shape and the geometric features [96, 97].

Hough transform started with line shape detection concerning the orientation and the scale [98] then it was improved to detect the circles and ellipses [99].

Basically, the Hough transform works on a two-dimensional data space e.x images. The set of pixels with mapped points (intersection) are grouped, and the slope will be calculated to detect the line features [100].

The Hough Circle Transform (HCT) is a feature extraction technique that aims to detect and identify the circular objects with different dimension and to concern the overlapping and occlusion problems [101].

The HCT algorithm start checking each edge point (i, j) by applying the circle equation: $(i-a)^2+(j-b)^2 = r^2$. For each value of a, all possible values of b are found according to the circle equation and the radius value. The resulted cells present the detected circles in the image using this algorithm. If the radius is unknown, a third dimension will be added to a, b which is r which has a higher computationally expensive [102].

The Hough transform has been used in many applications in different fields. It has been used for shape-based tracking [103]. It was used and benefited for lane detection in transport [104, 105], medical applications [106], and in the industrial applications [107].

It was used for object recognition purposes as proposed in [108] by designing an automated injection system and utilised the HT to detect the nuclei of the embryo. Nuclei recognition using HT for tracking system was proposed in [109].

2.3.4 Classification and Regression Decision Tree

The decision tree algorithm is one of the most popular predictive models that has been widely used in machine learning and statistics [110].

The decision tree partition the feature space into smaller

regions recursively by capturing the relationships between features and the labels as shown in Figure 2.6 [111].

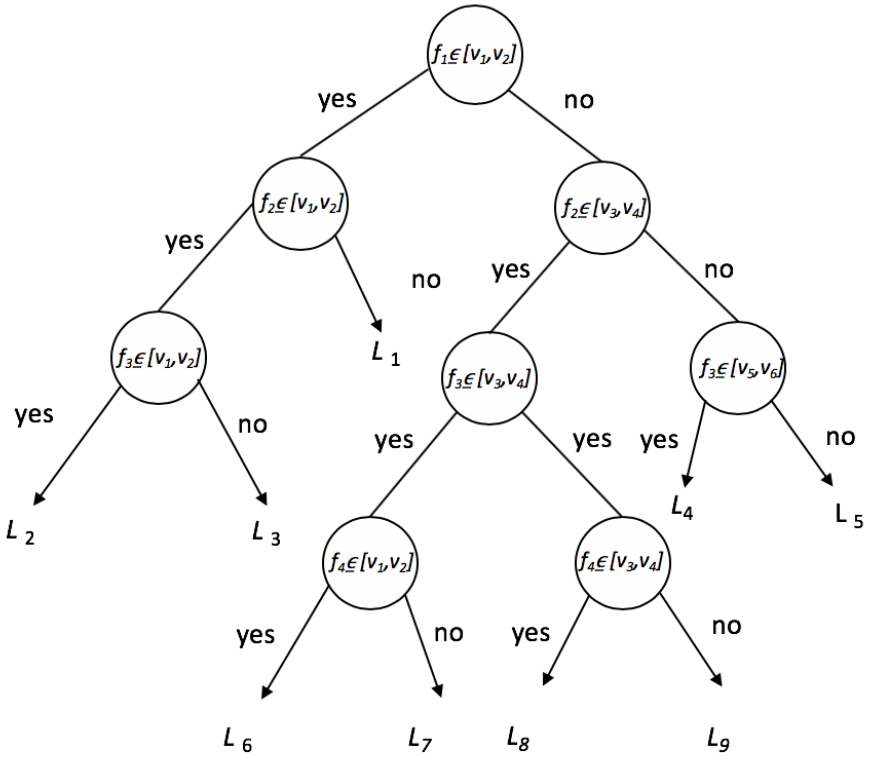


Figure 2.6: Decision tree algorithm

The main advantages of the decision tree algorithm, the simplicity of the algorithm mathematics and it is easy to understand and interpret. Furthermore, it is considered as a useful tool with the small data size [112]. Moreover, the data does not require a lot of preparation and no need for normalisation.

The decision tree has a challenge with the overfitting problem when the training and the testing performance are not incorporated.

The model is adequately trained, but with new data, performance is worse [113]. The decision tree cannot be very robust, where the model efficiency can be affected by any small change of the training data [114].

The main concept of the decision tree is the conditions, several questions of ***If-Then*** type questions have been answered by the model depending on the feature values. The separated rules for each category are combined using the logical operations (And, Or) statements. The high numbers of rules can be reduced using the heuristics implementation [115] and without affecting the system performance.

There are two main types of the decision tree in data mining:

- Classification: In this type, the system predict discrete or categorical classes.
- Regression: This type able to predict the continuous real numbers like the weather temperatures.

The Classification And Regression Tree (CART) was introduced by Breiman et al. to cover both types in one algorithm [116].

The decision tree is constructed by a multiple of rules based on feature values of the training data set. The rules are selected depending on the variable (feature) values to reach the best splitting of tree nodes starting from the origin node to the leaves recursively. The terminal nodes of the tree represent the observation (categorical, numerical). The node splitting represents a challenge due to

the overfitting problem; this process can be done according to the equation below [111].

$$(j^*, s^*) = \underset{\substack{j \in 1,...,p \\ s \in S_j}}{arg\ min}\ \mathrm{L}(s) \tag{2.18}$$

The CART algorithm has been used by many researchers for automatic zebrafish screening purposes. Sardine eggs have been identified and recognised automatically using the CART algorithm depending on the egg features such as the shape, the size and the shade [117]. On the other hand, the CART algorithm has been used for species classification for fish type (fillet or not) [118] and larva malformation observations [13].

2.3.5 Convolutional Neural Networks

The deep learning (DL) or hierarchical learning is a part of the machine learning approaches. The deep learning concept based on the data representation [119]. The deep learning is presented by a set of algorithms which are typically neural networks are learned as multiple levels corresponding to the abstraction levels [120].

The DL is a biological-inspired by the nervous system with the communication patterns and the signal processing [121]. The most known deep learning models are based on the artificial neural

networks (ANNs).

The first model learning approach was introduced in 1965 by Oleksiy Ivakhnenko [122] who is named as "Father of Deep Learning" [123]. He presented a multi-layer approach in which the input data properties are automatically filtered [124]. The DL term was introduced in 1986 by Rina Dechter [125].

In DL, the most commonly used algorithm in image analysing is the Convolutional Neural Network (CNN) which is a multi-layer of perceptrons with the least pre-processing requirements [126]. The main difference between DL and the traditional machine learning methods can be summarised by the feature extraction process [127]. The DL network structure is designed to extract the most essential information using filters which is traditionally done using feature engineering process, and it is considered as expensive and challenging process [128].

The CNN consists of a multi-layers input layer, output layer, and hidden layers. The hidden layers are composed of convolutional layers, ReLU layers, pooling layers, fully connected layers, and normalisation layers [129]. Figure 2.7 shows the general structure of the CNN [130].

1. **Convolutional layer:** this layer is considered as the core block of the deep CNN [131]. The results of this layer are presented by feature maps that are produced by convolving the kernel filters the input image dimensions. The number of the convolutional

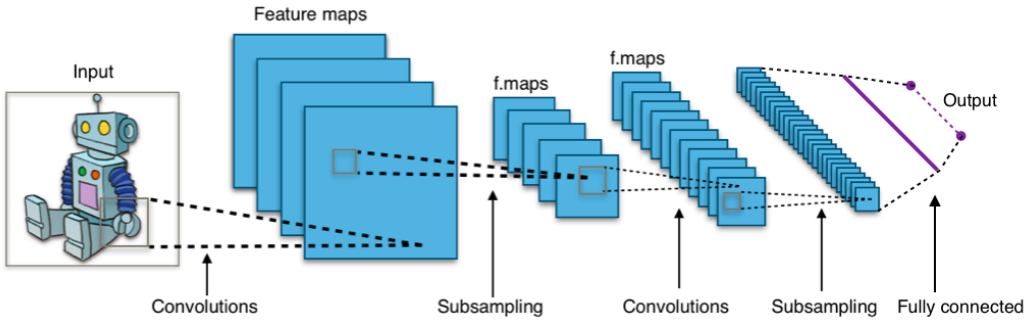


Figure 2.7: General structure of CNN

layer (depth) depends on the image size and the filter size.

2. **Max-pooling layer:** this layer received the feature maps and reduced their resolution to achieve spatial variance. The pooling windows can have random sizes and overlap with each other [132].

3. **Activation function layer:** Rectified Linear Unit (RELU) is the most common activation function that is used to avoid the gradient problems [133].

4. **Fully connected layer:** this is the final layer of the CNN where all neurons in this layer are connected with all activations of the previous layer. The number of neurons in this layer is hyper-parameter to be chosen [134].

Recently, the convolutional neural network (CNN) is widely used in many computer vision applications such as image segmentation [135, 136], image recognition [137], and image classification [138, 139], drug discovery [140, 141], and video analysis [142].

Chapter 3

Automatic Counting Systems

3.1 Introduction

Dealing with zebrafish larvae requires dealing with both eggs, larva, and adults. Counting eggs and identifying whether they are alive or dead is considered a time-consuming process due to the large number of the eggs that are produced after each mating. In this chapter, two automatic counting systems for zebrafish species are presented and explained. One for fish counting inside the tank based on recorded videos and the other for Counting the live eggs inside a petri dish based on scanned images. The research challenges and results are discussed.

3.2 Challenges and Objectives

The most important challenge with fish counting is the overlapping of the fish inside the tank, which would be predicted as one object

and the tank material (glass) reflection. The main challenge in the egg counting system is the high similarity between the live and dead eggs from a side and the debris from the other side. The non-uniformity of the egg positions, especially when they are located at the dish edges may cause non-detected eggs. The overlapping eggs can be a challenge in this work if the density of eggs is high.

The major objectives of this work are summarised below:

- Design an automatic system for zebrafish inside a tank using computer vision techniques.
- Detect, identify, and count the live eggs inside a petri dish based on scanned images and using image processing methods.
- Introduce cost-effective tools which are affordable and easy to use such as, mobile phone and flatbed scanner.

3.3 Automatic Zebrafish Counting System

During the biologist experiments, the researchers need to count the fish inside the tank. This process is always carried out manually. This work aims to detect and count the number of free-swimming fish.

In the research laboratories, there are many water tanks containing tens or often hundreds of fish. Researchers need to recognise the number of these fish inside the tank. Manual fish counting needs to transfer the fish from tank to another. Handling the fish manually is a harmful procedure, time-consuming, and error-prone [143].

Determining the number of fish in aquarium aims to prevent the overstocking problem in the aquarium, and to determine the number of fish that should be harvested and fed [144].

The published automatic fish counting systems have attempted to find and isolate the fish blobs depending on image processing operations for each frame in a recorded video [145, 146]. These systems have limitation regarding the fish body reflection and overlapping problem. Using underwater videos and depending on the fish shape, a counting system was presented by Fabic et al. in 2013 to count the fish under the sea. They used Zernike moment to analyse the shape of the blobs and to identify the fish in each frame of the video [145].

In 2016, an automatic fish counting based on videos was presented to detect and count the zebrafish inside a tank. The authors used the Gaussian Mixture Model (GMM) for background removing and Hue moment for fish body reflection [143]. Table 3.1 presents a summary for the most related existing systems.

Table 3.1: Fish counting systems

Author	Method	Capturing Tool	Counting Error
Fabic et. al [145]	Canny edge detection with coral blackening background	camera under-sea water	10-20%
Toh et al. [146]	thresholding and averaging	camera on a box covered the fish tank	0-29%
Francisco et al. [143]	thresholding and averaging	camera above the fish tank	5-37%

This work provides a solution for fish counting problems such as noise, the reflection of the fish image on the tank sides. The proposed imaging tool is This system analyses a video using several statistical analysis to detect the fish blobs inside the image using thresholding then count these blobs according to the shape and area feature.

The general methodology of this work is illustrated in Figure 3.1. This system will overcome the reflection and overlapping problems that have not been solved by the existing systems using statistical analysis and depending on the fish shape features behind the image processing techniques.

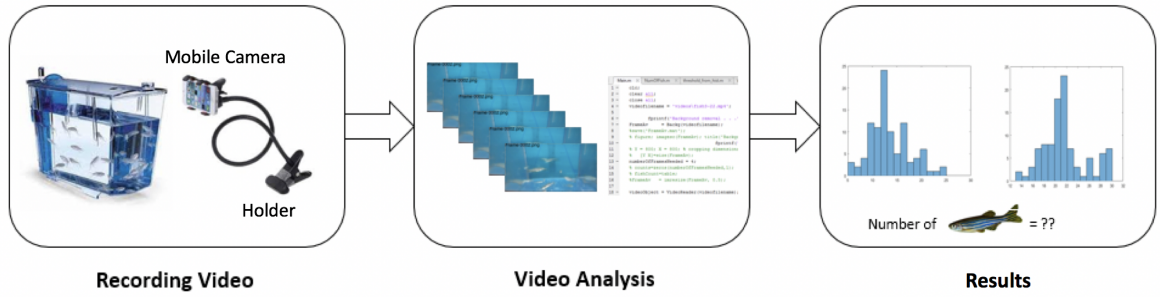


Figure 3.1: Fish counting system

3.3.1 Experiments

Using a mobile phone camera, 54 videos were recorded. The first 29 videos were recorded using a mobile phone camera that was fixed on a holder that was stuck on the table with 30cm between the phone and the water tank, as shown in Figure 3.2. Four tanks were

used in this work; each tank has ten fish with some stones in the bottom of the tank.

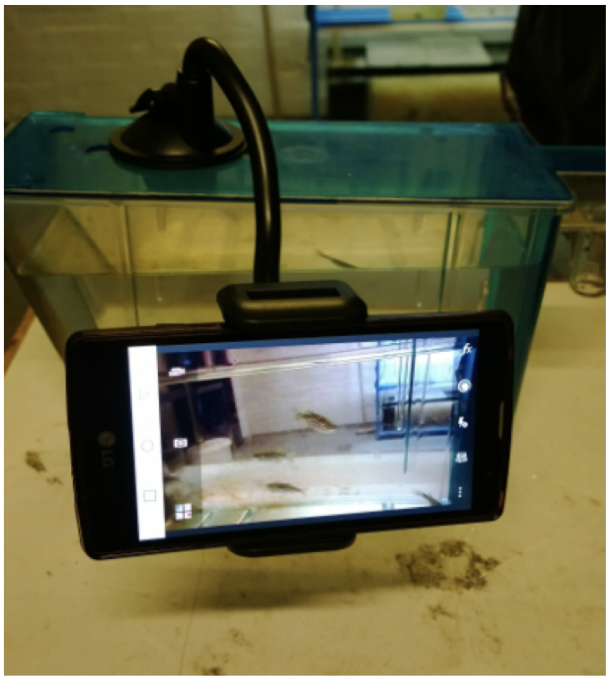


Figure 3.2: First setup

The 25 videos were recorded using a mobile phone camera that was set up using a tripod was fixed 20 cm from the tank, as shown in Figure 3.3. Two tanks were used with five fish in each one, and the length of the videos was around 20 seconds which is about 495 frames. The videos were processed offline to test, analyse and check the proposed work.

The used tanks are standard and have a specific criteria according to the home office recommendations [147]. They are 3.5-litre tanks, 6.5-7 inches deep, 4 inches wide, 10.5 inches in length. The home office recommends no more than five fish per litre, so

the maximum we could put into one tank is 17 fish and its recommended that we don't have groups smaller than 5 per tank.

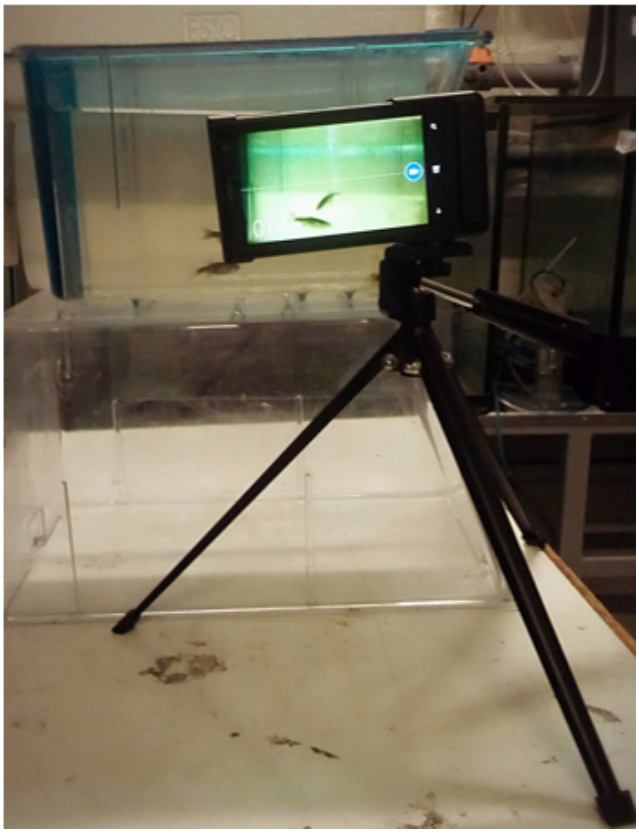


Figure 3.3: Second setup

During the experiments, we noticed that using the tripod was better and steadier than the holder; the holder fell many times in each experiment. Furthermore, the videos using the tripod were more stable than the tripod. In this system, any camera can be used, but we preferred the mobile phone to improve the methodology into a mobile application.

3.3.2 Video Analysis

Some recorded videos were 10 seconds, and others were 20 seconds long. Where the main purpose here is to count the fish, the most important part here is the fish detection. This process is done in several steps that were applied to all frames of each video. The methodology started by finding the background to specify the analysis region. The background of the video frames was obtained by calculating the average of all the video frames. For each frame, a difference frame was found out and binarised, manipulated, and filtered to a specific area, and finally, the number of blobs were recognised. The number of fish in all frames were processed using statistical formulas like mean and mode. Figure 3.4 shows an example of the background subtraction.

- Read the video file
- Find the background of the video
  - Read the video file
  - Read each frame (1 to the total number of frames of the video)
  - find the sum of all of these frames
  - Divide the result matrix on the number of frames
  - Save the result (Background frame)
- For each frame:
  - Find the foreground frame by subtracting the background frame from the frame
  - Threshold the resulting frame to have a binary frame using Otsu's method [148]

  - Specify the area of the blobs to eliminate small blobs using experimental area value
  - Count the number of these blobs
- Find the mean value of the blobs of all frames of the video
- Number of fish = mean (number of blobs of (fr_1 : fr_n))

The binary image was filtered based on an experimental value for the blob area, the major axis, and the minor axis of the blob. The filtered image has the fish blobs, and the unwanted objects were removed. The number of blobs for each frame was found, and the final decision about the number of fish was taken after applying the statistical measurements.

3.3.3 Statistical Analysis

Two statistical values were found to decide the most useful one to get the right number of fish inside the tank. The first value is the mean value $\bar{F_m}$ of the detected blobs in all frames in the video and the second value was the mode value F_d that presents the most frequent number of the blobs in these frames as shown in the following equations.

$$\bar{F_m} = \frac{1}{N} \sum_{n=1}^{N} f_n \tag{3.1}$$

where f_n presents the number of blobs in the frame number n, and N is the total number of the frames.

$$F_d = max_R\{f_1, ..., f_N\} \tag{3.2}$$

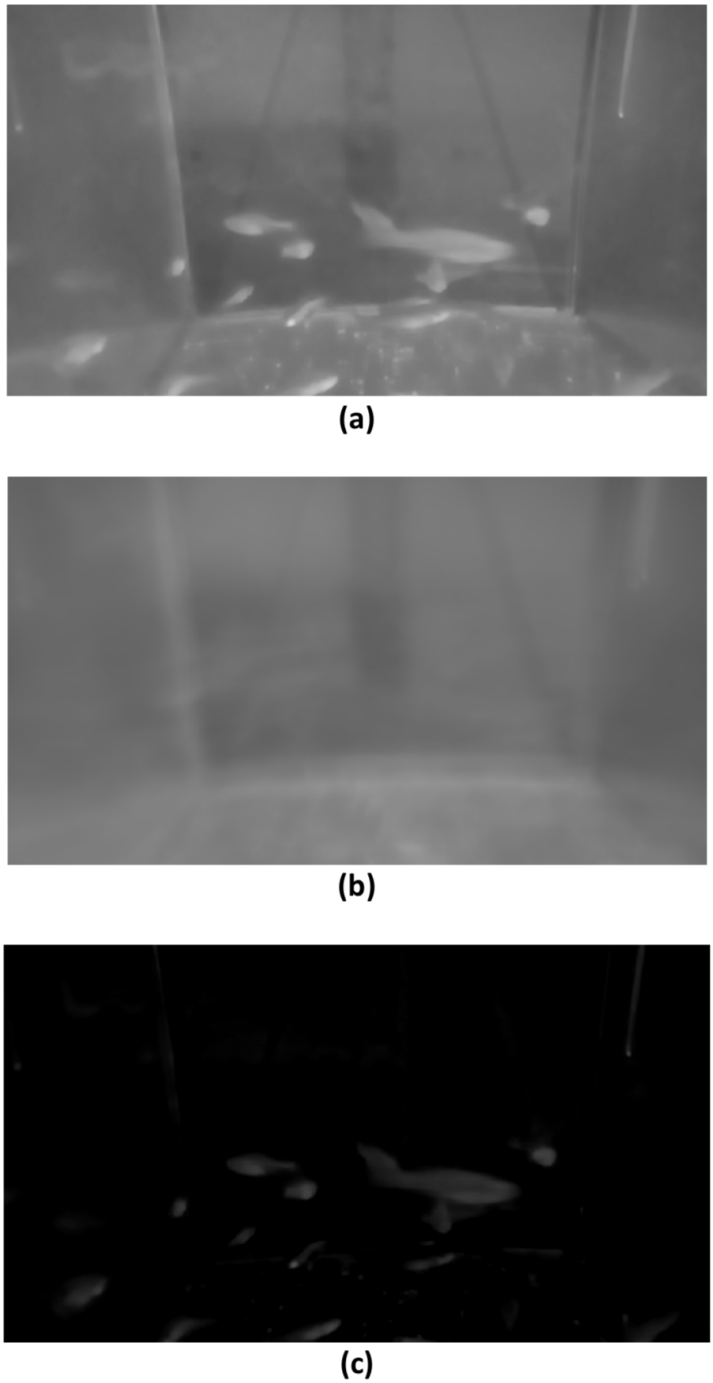


Figure 3.4: Background removing: (a) Frame (b) Background (c) Subtraction result

where max_R is the most frequent value, and f_n is the number of blobs in the frame n.

3.3.4 Results and Discussion

After applying two statistical formulas, mean and mode, the results were analysed and registered. Figures 3.5 and 3.6 show the results of the mean and mode values for 29, 25 videos. As shown in these figures, the mean was closer to the actual number than the mode. This means the counting system is based on the mean value of the number of fish in all blobs.

To find the detection error of the proposed approach, the following equation was applied then the average of these values was

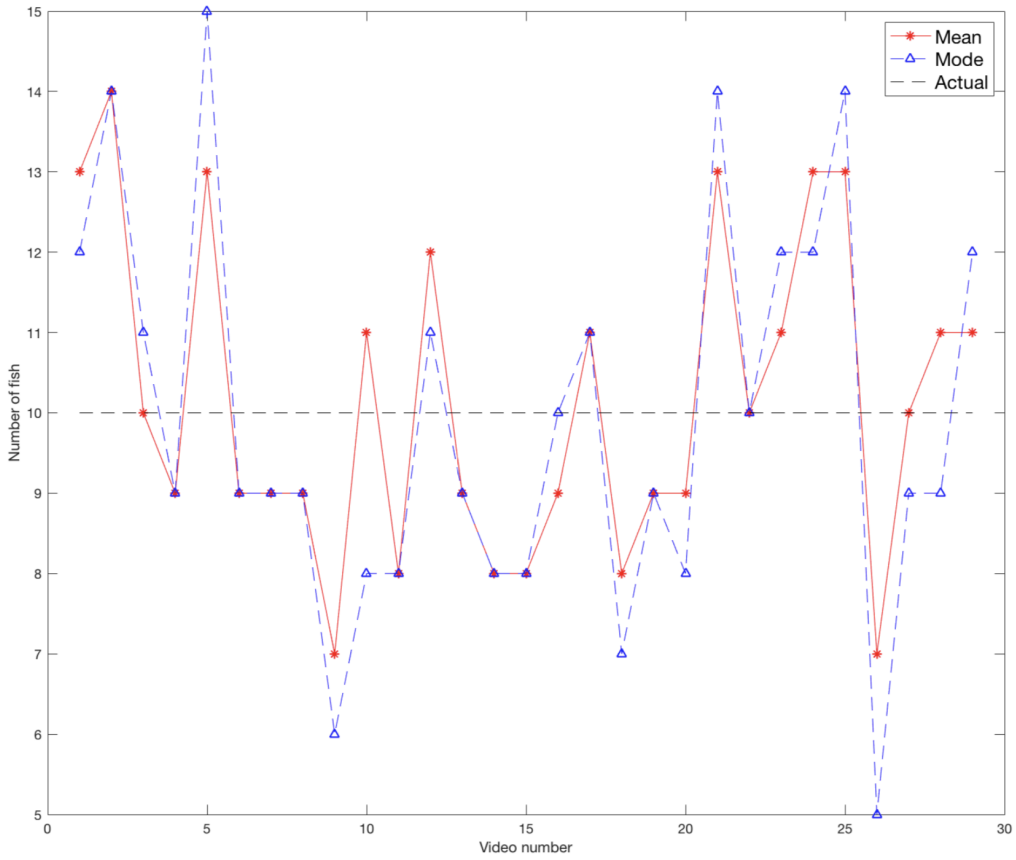


Figure 3.5: Statistical values for each video (10 fish)

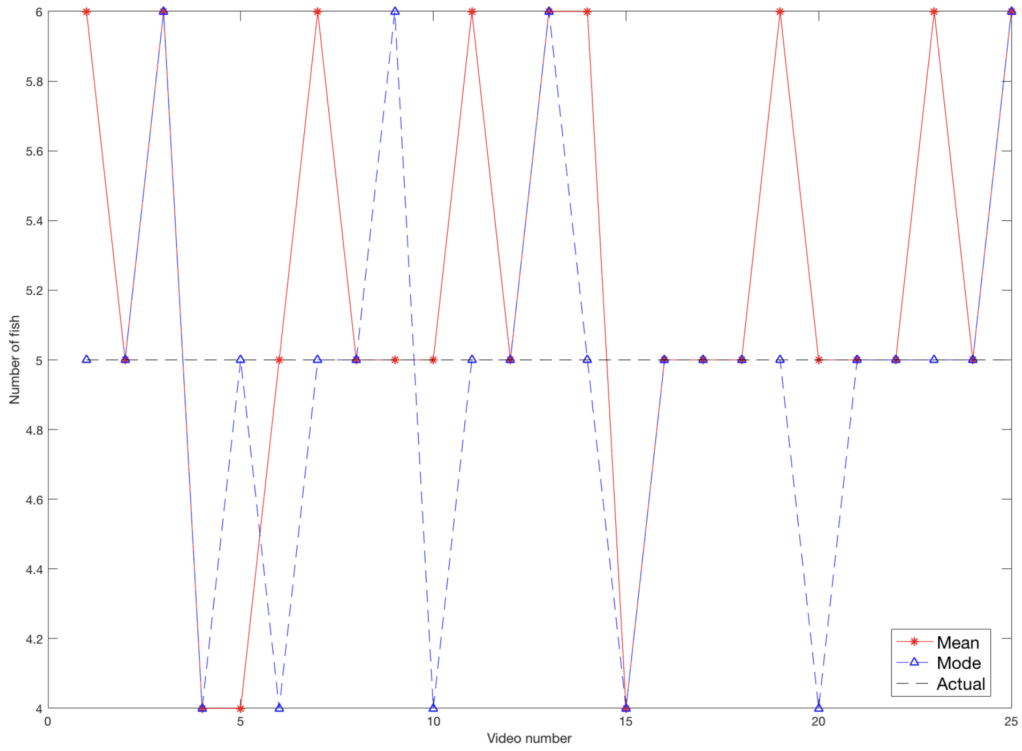


Figure 3.6: Statistical values for each video (5 fish)

calculated to determine the algorithm efficiency. Tables 3.2 and 3.3 show the detecting error for each video that were recorded for 10 and 5 fish. The detecting error was found for the **mean** and the **mode** according to the following equations.

$$e_m = \frac{|\bar{F}_m - F_c|}{F_c} \tag{3.3}$$

$$e_d = \frac{|\bar{F}_d - F_c|}{F_c} \tag{3.4}$$

Where e_m: is the detecting error using the mean value, e_d is the error for the mode value, and F_c is the actual number of the fish.

Table 3.2: Counting error (10 fish)

	Video										
	1	2	3	4	5	6	7	8	9	10	11
e_m	30%	40%	0%	10%	30%	10%	10%	10%	30%	10%	20%
e_d	20%	40%	10%	10%	50%	10%	10%	10%	40%	20%	20%

	Video										
	12	13	14	15	16	17	18	19	20	21	22
e_m	20%	10%	20%	20	%10%	10%	20%	10%	10%	30%	0%
e_d	10%	10%	20%	20%	0%	10%	30%	10%	20%	40%	0%

	Video							
	23	24	25	26	27	28	29	AvgErr
e_m	10%	30%	30%	30%	0%	10%	10%	16.5%
e_d	20%	20%	40%	40%	10%	10%	20%	19.7%

Table 3.3: Counting error (5 fish)

	Video												
	1	2	3	4	5	6	7	8	9	10	11	12	13
e_m	10%	0%	10%	10%	10%	0%	10%	0%	0%	0%	10%	0%	10%
e_d	0%	0%	10%	10%	0%	10%	0%	0%	10%	10%	0%	0%	10%

	Video												
	14	15	16	17	18	19	20	21	22	23	24	25	AvgErr
e_m	10%	10%	0%	0%	0%	10%	0%	0%	0%	10%	0%	10%	4.8%
e_d	0%	10%	0%	0%	0%	0%	10%	0%	0%	0%	0%	10%	3.6%

3.4 Eggs Counting System

Zebrafish eggs are typically transparent and are always sorted one by one manually using a pipette or estimating the number of eggs using 'spawning biomass' depending on the volume of the spawning-stock [149, 150]. The adult female zebrafish can produce hundreds of eggs in every single mating process. Eggs are fertilised and developed in an environment with a temperature range between 25-31^{o}C,

where the optimal degree is 28.5^{o}C [151].

The growth levels and mortality are affected by many factors like the temperature, water quality, and fish size. These factors were presented and analysed in Houde and Zastrow's study [152].

Instead of the traditional laborious ways for eggs counting, automated systems were proposed to find out the number of eggs automatically. In the system proposed by Friedland et al. [153], the petri dish was captured using a scanner after covering the dish by a black cover. The resulting images have the eggs with a black background. The eggs processed and analysed to have the number of the eggs and the size of these eggs. In this study, all eggs were counted as blobs with no need to distinguish between the live and dead eggs.

Counting the live eggs for pelagic fish was proposed in [151] by producing images using a camera, illumination device, and black background. In this study, the live eggs always float, and the dead ones sink, this leads to having images with only live eggs. The research focuses on the capturing method by capturing the surface of the solution only which means the image processing operations for live eggs identification and counting are carried out to clear the image, enhance its illumination, and specify the area of the blob.

The traditional process for counting the number of surviving eggs is a tedious work, time-consuming, and error-prone where the diameter of the pipette is large compared to the egg size, occasionally leading to uncounted eggs during this process. However, the

available automatic counting systems deal with all eggs inside the image, and no effort has been made to distinguish between the live and dead eggs before counting them.

A high-throughput, viable, and cost-effective counting system is proposed to detect and count the live zebrafish eggs inside a petri dish. The proposed system starts with capturing an image for the dish using a low-cost tool, process the images to identify the dead and live eggs, and finally count the live ones. In this work, the counting process is carried out automatically in a short period using a low cost, high-performance imaging tool, which is a flatbed scanner. Figure 3.7 shows the general algorithm of counting the number of zebrafish live eggs.

3.4.1 Data Collection

The collected images of this system have been collected in the Institute of Integrative Biology laboratories in the University of Liverpool in collaboration with our colleagues in this Institute.

For the egg counting purpose, the data were collected for the petri dish with 100mm diameter using a flatbed scanner with a high-speed scanning. The scanner has its source of illumination and a fixed focal length, therefore, no need to consider the exposure or focus on the capturing process.

The region of interest was specified to minimise the scanning

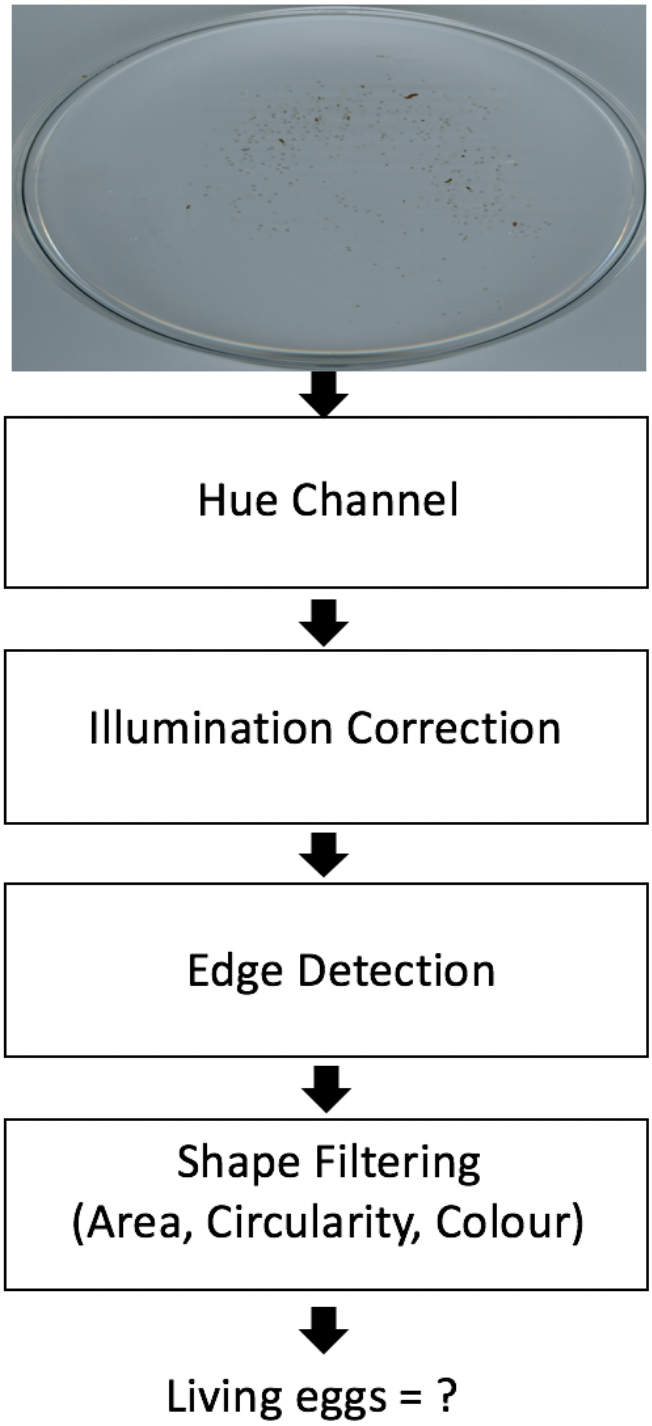


Figure 3.7: Egg counting stages

process time. The petri dish was scanned using 1200 dpi. However, the eggs inside the plate can also be visualised clearly, and the fertilised and unfertilised eggs distinguished using a lower resolution 600 dpi. Figure 3.8 shows an example from the collected images with 6436×6888 pixels in size.

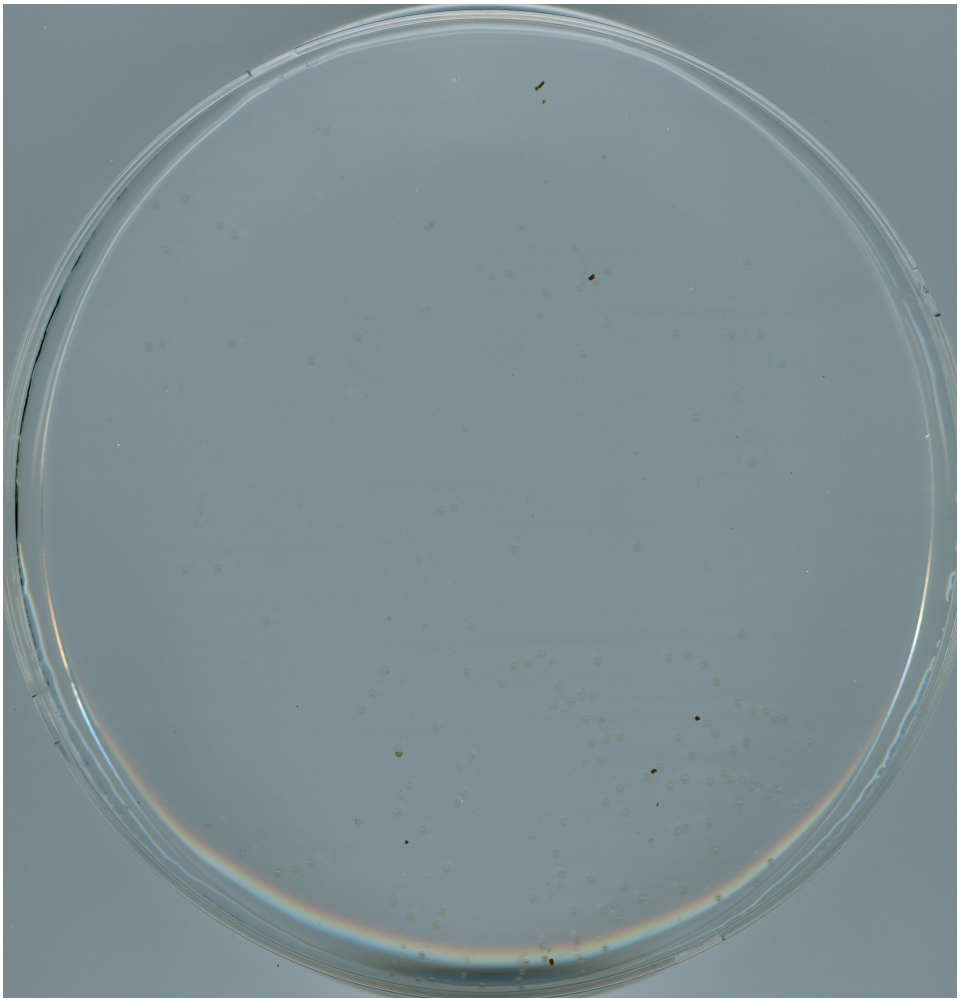


Figure 3.8: Zebrafish eggs inside a petri dish (1200 dpi)

The applied successful experiments were 10, and the capturing process was done two times. The first time was carried out immediately after the breeding process in the morning, which means 4hpf (hours post fertilisation). The second time was carried out

after 24 hours from the first time. Figure 3.9 shows a small part from the dish in Figure 3.8.

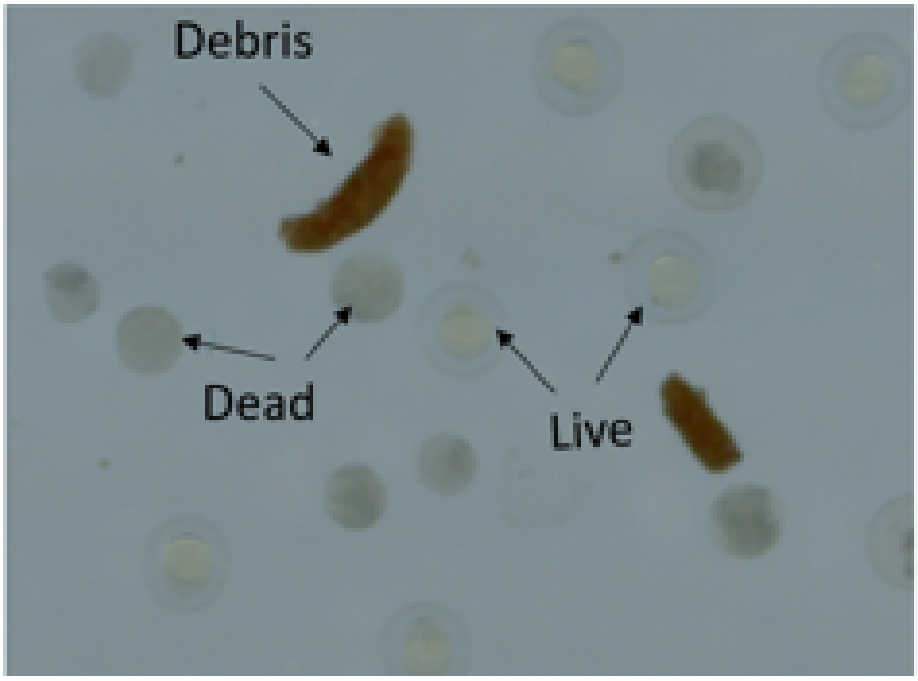


Figure 3.9: Live and dead eggs

The biologist interest in knowing the number of live eggs more than the dead ones, which are so clear and distinguished using the flatbed scanner. The live eggs are transparent by eyes and have a yellow circle inside after having images by the scanner. The dead eggs are dark by eyes and in the collected images using the scanner. The actual number of the eggs inside the petri dish was found and registered by the biologists manually using a pipette and count the eggs one by one.

3.4.2 System Methodology

After several experiments that were carried out in the Integrative Biology laboratories, the collected images have been processed using several operations to get the estimated number of the live eggs where the benefit of using engineering aspects becomes useful for the biologists in their field. To detect and count the eggs, several points should be considered: the similarity in shape between the live and dead eggs, the unwanted objects inside the dish like the food powder and debris, and also the edge of the petri dish that may affect the detection process.

1. HSV converting: the first step is the image converting process where the RGB coloured image from the scanner was converted to the Hue Saturation Value (HSV) colour space. This step was done according to the colour of the live eggs that is extremely yellow, and by turning the image into HSV colour space, the yellow colour becomes dark in the hue channel as shown in Figure 3.11 of the RGB image in Figure 3.10.

2. Illumination correction: In the hue channel, the colours vary from red through yellow, green, cyan, blue, and magenta [86]. The yellow circles were segmented using the hue colour channel. The brightness of the hue image was corrected using the gamma correction process for the image im according to the Eq. 3.5 with γ value =1. This step aims to increase the contrast of the image and improve the difference between the background and the foreground in the image.

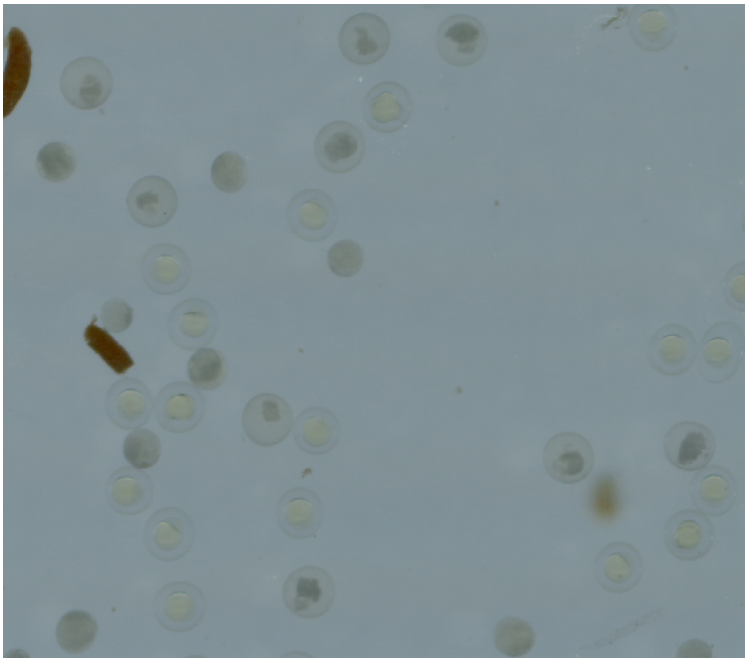


Figure 3.10: Small region of the RGB petri dish image

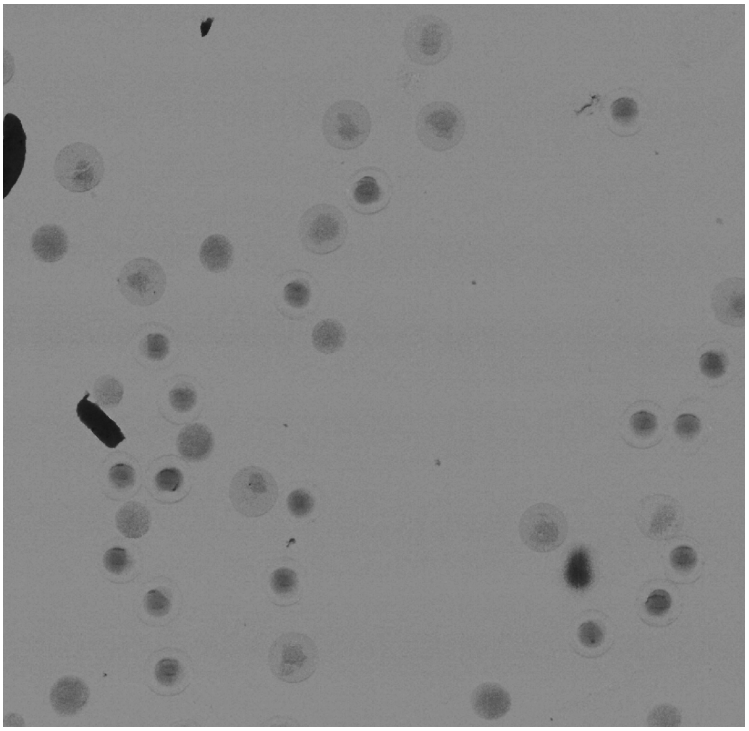


Figure 3.11: Hue channel image of the HSV colour space

$$Correction = 255 \times (im/255)^{\gamma} \tag{3.5}$$

This step makes the system more robust from any an external light that may affect on the image information. Figure 3.12 shows the image after applying the gamma correction.

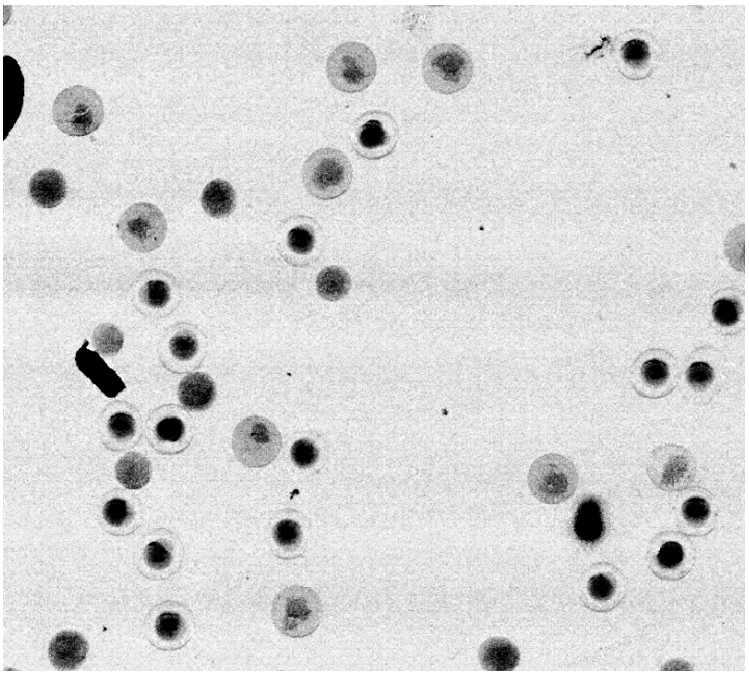


Figure 3.12: Gamma correction result

3. Edge detection: The resulting image was processed using Canny edge detection with a threshold between 0.1 and x; where x was found using the Otsu's thresholding method [148]. The minimum threshold was chosen by trials and error to remove some of the unwanted objects because the debris inside the dish is darker than the eggs. Figure 3.13 shows the resulting image from this step.

4. Circle filtering: The resulting circles are filtered using the Hough Transform [102, 154] and depending on the area, radius, and the

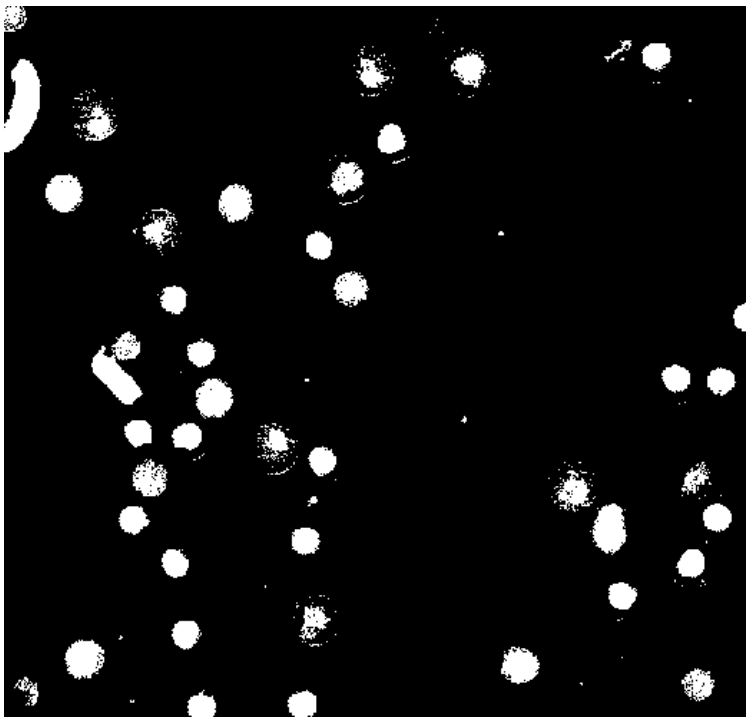


Figure 3.13: Binary image

circularity to remove any other unwanted objects either bigger or smaller than these values. The radius range between 2.6mm and 5.2mm, which equals to 10 and 20 pixels for 1200dpi scanning images. Finally, the filtered circles are counted to have an estimated number of live eggs, as shown in Figure 3.14 below.

3.4.3 Results and Discussion

The manual egg counting process is tedious and time-consuming work and exposed to error, as explained before. The automatic counting system has the main advantage of using engineering solutions to solve other field problems. This new method uses a low cost capturing tool which is a flatbed scanner and the resulted images are processed and analysed to get an estimated number of the living eggs.

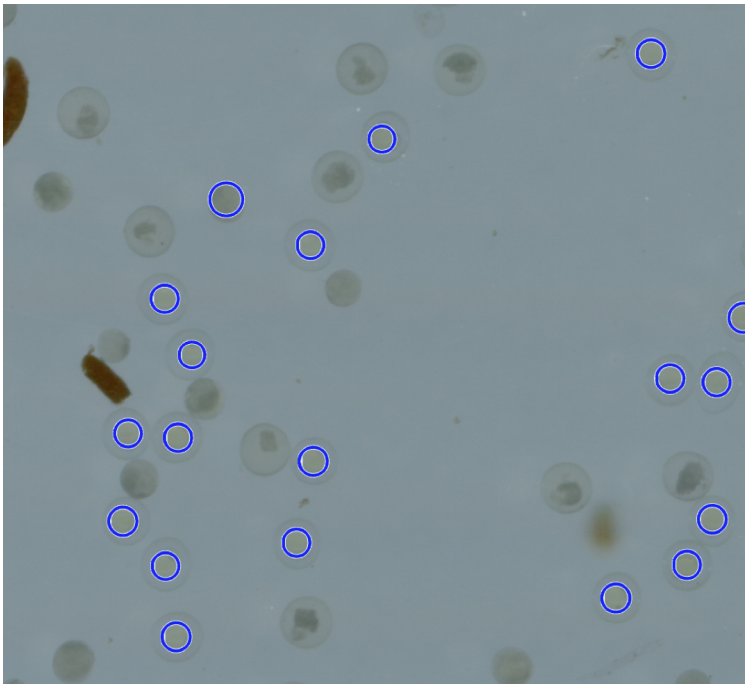


Figure 3.14: Detected circles (live eggs)

Figure 3.15 shows the actual number of eggs versus the predicted number of eggs using this system. The detection error of this system is between 1% and 28% with average detection error 17.5%. The average of the false positive and the false negative were 9%, 8% respectively. The results below show the predicted number compared with the actual number of the zebrafish eggs inside several Petri dishes that were used in the biological laboratories.

As shown in Figure 3.15, the results for 17 images of the zebrafish eggs with 4-24hpf for different dishes with different number of eggs, the figure shows the counted numbers by the system and the actual number of eggs that were counted manually one-by-one by a biologist using a pipette.

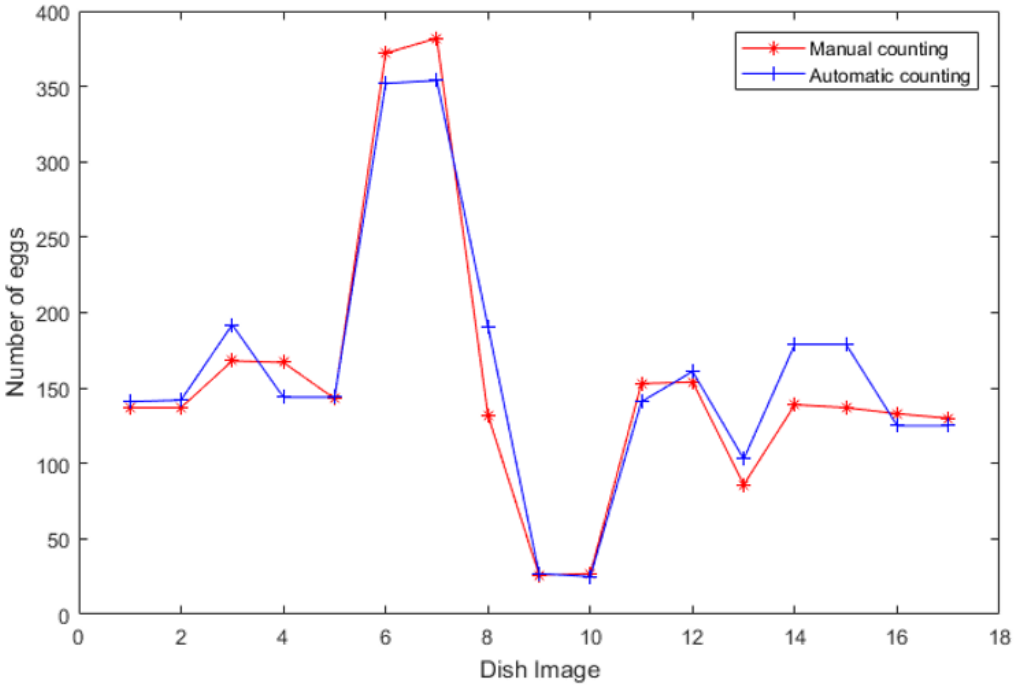


Figure 3.15: Egg counting system results

The major problem of the counting happened because of the false detection and identification processes. Furthermore, the low contrast of some eggs makes it similar to the background, which causes a false detection of the eggs. Only one image provided a 28% counting error, which happened because of using a small petri dish that produced an image with different contrast was not consistent with the system software. If this image is excluded from the system analysis, the average error would be 10%. Some dishes had a lot of debris and food powder that cause noise and false negative detection.

The previous efforts to count the eggs automatically focus on counting all eggs inside the petri dish, and they did not need to distinguish between the live and dead ones. The average error for the proposed system [153] was reduced from 6% to 1% over several

years of work. However, the system counted all eggs and did not need to filter the live eggs from the others. The average error of the proposed system in [151] was 2%. However, this is also counted all eggs inside the dish, and also they used complicated hardware which definitely needs expert intervention.

Using the scanner in this work produced extremely high-resolution images suitable for the system purposes. The proposed system showed a high performance in imaging and counting processes. This system reduced the biologist interventions and present an accurate counting system based on a cost-effective, affordable, and easy to use imaging tool. The system saved the consumed time and reduced the expert efforts where counting thousands of eggs need full concentration and it is error-prone if the biologists forgot the number he reached or if he pulled more than one egg by the pipette. Moreover, the system from producing the image to showing the result, it need 2-3 minutes where the manual counting process needs more than 6 minutes depending on the number of eggs.

3.5 Summary

This chapter presents two automatic counting systems for the zebrafish species. The two systems attempt to avoid manual intervention. Furthermore, these systems give the result in a faster way than manual counting. Considerably, the two proposed systems overcome the manual process which is considered as a time-consuming process and harmful.

Depending on the most important features and using the computer vision aspects, the counting systems are provided as a reliable, accurate, and cost-effective tool to count the zebrafish eggs and the adults without any manual processes and can be done by a non-biological expert.

The proposed systems are economical and can be used by the researchers and the groups to reduce the consuming time and reduce the researcher efforts. The two counting methods can be improved and converted to mobile applications.

Chapter 4

Image Analysis Pipeline For Zebrafish Larvae

4.1 Introduction

In this chapter, an image analysis pipeline is presented. The experiment protocol and the used chemical are briefly explained. The collected images are presented and analysed. The data preparation process is illustrated. The designed classifier is presented and explained. The segmentation, detection, and identification processes are explained and covered. The result of this system regarding the status of all samples is presented and discussed. The research challenges and achievements are introduced.

4.2 Challenges and Objectives

Despite the rapid growth in the use of zebrafish embryos as an experimental model, there is still a lack of automated classification

systems according to several challenges.

- The samples may slide to the edges of the petri dish that affect the detection process and rise the identification error.
- Unwanted objects such as the debris or food particles, where these can interfere with the detection process.
- The fast development of the zebrafish embryos leads to having two classes of the live embryos which have different age and different features, one of them is transparent and the other having a completed growing embryo.
- The unwanted movement of samples inside the wells which rise the detection and the identification errors.

Objectives:

The demand for this system comes from the biological requirements of malformations detection that are happened with hundreds of embryos within a short time relating to ethics and the fast growth of the zebrafish embryos. As an attempt to overcome and address the mentioned problems and limitations, several objectives are drawn and presented:

1. Design an automatic segmentation, classification system for zebrafish eggs using two ways for feature extraction and also a classifier.
2. Using the proposed platform, which is affordable and easy to use, the images are collected automatically, and the biologist only needs to place the dish on the scanning area.
3. Present affordable and time-saving imaging tool with a low cost and friendly use.

4. Present an automatic segmentation, orientation fixing methods where the samples can be located in different positions which were fixed manually according to the existed system.

5. Identify and classify the zebrafish egg/larva if it is healthy or not after collecting images from the Institute of Integrative Biology laboratories by determining numbers, health and presence of abnormalities in zebrafish larvae (under three days post fertilisation) using a high-throughput system.

4.3 Hardware And Data Collection

Due to a large number of samples and the animal protection regulations, it is necessary to use a high-throughput imaging tool. In this work, a successful attempt was carried out using the flatbed scanner. Using this device can produce hundreds of images for four 96-well plates at the same time, where the best camera is not able to do this job. This can raise the system throughput with low cost.

4.3.1 Experiment protocol

The sample images that were used in this work were collected from a dish containing hundreds of zebrafish eggs. The dish images were collected during several biological experiments in the Institute of Integrative Biology laboratories at the University of Liverpool in collaboration with the colleagues in this Institute.

Further images were collected in the Institute of Transla-

tional Medicine at the University of Liverpool with the cooperative with the researchers in this department. The collected images contain healthy embryo and non-hatched ones (chorion) according to their breeding experiments.

Each experiment starts from day 1 to day 5. After that, the samples should be killed according to the animal protection regulations. The biological experiment began by separating the male and female zebrafish adults by standing a plastic wall inside the water tank over the night, then after several hours, this wall was removed and the eggs produced naturally. After that, the eggs were collected and sorted to get only the live eggs. The collected live eggs were put in a particular environment with 28^oc temperature; the last step is adding substances by the biologist.

4.3.2 Chemicals

The collected eggs were subjected to several chemical substances such as Dimethyl sulfoxide (DMSO), Alcohol, the waste nitrogenous compounds Ammonia, (Sodium) Nitrite, (Sodium) Nitrate and metals such as Copper (Sulphate) as well as antimicrobial aquarium treatments.

4.3.3 Hardware

The images were collected using a flatbed document scanner with a high-speed scanning for a petri dish of 100mm diameter. The benefits of using a scanner were manifold. The scanner has its source of

illumination and a fixed focal length. Therefore, there is no need to consider the exposure or focus on the capturing process. The scanner provides the biologist hundreds of sample images in one scan where the traditional way for collecting images in the proposed systems is always carried on using a camera with a microscope which is considered as a time-consuming process.

To reduce the consumed time of the manual imaging process, a flatbed optical scanner was used. The flatbed scanner has been used to capture multi pictures for the well plate(s) which is done in a short period (0.8) second for each well of the plate. Where the manual process requires getting the embryos out of the well plate individually to get images for them using a microscope.

Figure 4.1 shows two examples for the scanned images with two different resolutions; it is clear that the image with higher resolution (1200dpi) is better than the other with the lower one (300dpi). However, the demand here to have a more top quality image to get more apparent features becomes an essential requirement.

According to the swimming bladder that is heavy regarding the embryo body, the embryos take side position, unlike the adult fish [155]. Two main factors were determined after using several scanners in the capturing process, that are illustrated below:

- Resolution of the images: Resolution of the scanner is defined by the number of dots per one inch^2 of the scanning area. De-

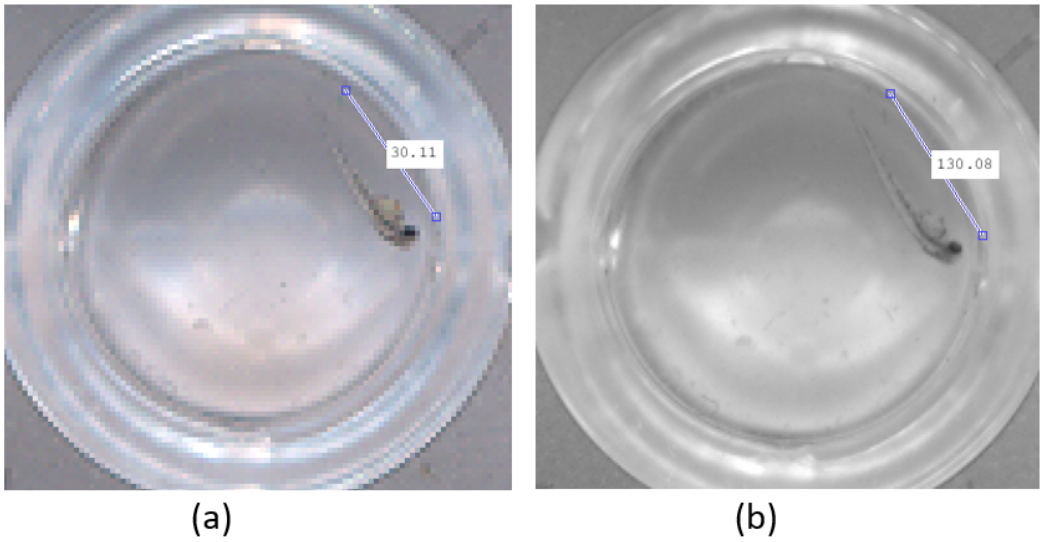


Figure 4.1: Images using scanner with (a) 300dpi (b)1200dpi

pending on the embryo length, which is 3.5mm and after comparing the quality of some images using different resolutions of the scanner, the minimum number of pixels that are required in this work was determined to be 240 pixels for the embryo as shown in Figure 3.5. This means around 69 pixels/mm. Convert this ratio to dot per inch, each millimetre equals to 0.039inch. The minimum resolution should be 1753 dots per inch. In practical, the required device should be a 2400dpi scanner or more. Experimentally, most collected images were gathered using 1200dpi, and it is suitable for the existed four classes and according to the biologist's observation.

- Speed of the scanner: according to the experiments that were done in the Institute of Integrative Biology laboratories in the University of Liverpool, the embryos were slightly moved, and this is discussed in [156] where the authors found that the zebrafish with 3dpf were inactive and allocate at the bottom of

the nursery container. According to their results, the speed of the embryos at 3dpf is 0.38mm/s. By using the scanner, the speed was 6.8mm/s to scan 300mm within 44 seconds.

The Flatbed scanner provides the system by hundreds of images depending on the number of inside the container, the manual interventions are limited, and the major imaging factors like the focal point and the illumination are fixed by the scanner. On the other hand, using the camera with a microscope, which is the traditional imaging tool in most studies needs an expert to fix the illumination and the focal point. Using the camera over the microscope can be considered as a time-consuming process comparing to the proposed tool where the images are taken for each larva separately.

Besides using the camera over the microscope for imaging in Jeanray et al. work [13], they moved the water from the wells, used the side position for all the samples, used a 24-well plate which means a larger size of the wells, and took a picture for each well separately.

4.4 Data Preparation

Due to the system goals, a machine learning model was designed and trained using egg images to classify three classes of eggs after collecting them from the dish images. Each image of the petri dish has two types (classes) of the zebrafish larva according to the sam-

ple age. The live and dead eggs are under 3dpf where the normal (hatched) and the chorion (not hatched) are 4 or 5dpf.

Depending on the egg detection process, the samples were cropped using the egg centres and diameters, which were computed using the Hough transform for circle detection. The egg detection process started by Canny edge detection function using a threshold that was produced from a Sobel edge detection function. This process was followed by several morphological operations. Finally, the circles with a specific range of radius were extracted from the whole image. The resulting images were collected and had been used to classify them into three classes as resulting from the biological experiments were the fourth class (Normal) has been detected and identified depending on the circularity feature.

4.4.1 Embryo Orientation

Before starting the experiments, a tail curvature deformity was expected, an orientation fixing was done to create a common position for all embryo body images, where we deal with the organism that have a unexpected movement. This step is very useful if more experiments will be carried out in the future by adding different toxicities to the samples. The purpose of this work is to allocate the embryo body in a horizontal position, the head on the left side, and the yolk on the bottom. The common position makes the classification process easier and more accurate, specially with the curvature and shortage malformations, this step can be defined as preparing stage where the images will be ready to be processed and analysed.

The general methodology of this work is implemented, as shown in Figure 4.2, which is started with rotation, flipping, and getting a final image with a target position.

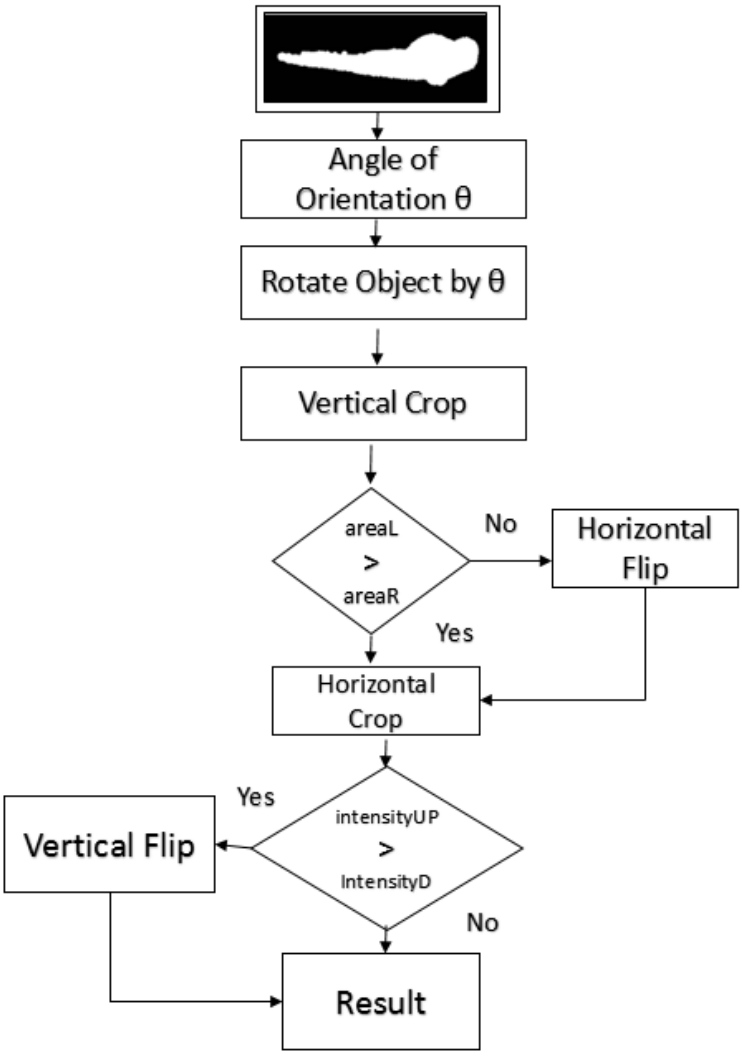


Figure 4.2: General methodology for orientation correction

The steps of this methodology are briefly illustrated below.

1. After finding the blob (larva) using many image processing functions that were illustrated in the previous section, the orientation angle value was determined to rotate the embryo by this value. This step produces the horizontal position of the

embryo, as shown in Figure 4.3. This step works successfully with all images with zero error.

Figure 4.3: Embryo rotation: a) before b) after

2. To determine whether the head position on the left or the right side, the embryo blob was divided into two halves, right and left. These two parts were compared with each other according to the sum of non-zero pixels of the two halves, where the bigger one should be the head side, a horizontal flipping will be done if the head on the right side. This step works successfully with all images with zero error. Figure 4.4 shows the result of this step.

Figure 4.4: Horizontal flip: a) before b) after

3. This step aims to check the yolk position to be at the bottom side. To do this, three algorithms were attempted with different results. This step focuses on the left side only that has the head and the yolk.

   The first way was done by comparing the sum of non-zero pix-

els of two parts yield after dividing the blob into two halves (up and down). Where the yolk part is bigger than the head, a vertical flip was applied if needed. The second method was done by calculating the distance between the centre of the eye and the two left corners of the bounding box. If the distance between the upper corner is longer than the distance between the centre of the eye and the lower corner of the bounding box, a vertical flipping of the embryo will be applied. These two methods succeed with the most images, not all of them, 57 images using the first method and 92 images using the second method from 137 images.

But the last attempt achieves the best results; it fixed all images successfully. In this method, the blob (left side) was divided into two halves. These two halves were compared to each other according to the mean intensity values, where the yolk part is always darker than the head. A vertical flipping was done depending on the mean intensity value. Figure 4.5 shows the result of this step.

Figure 4.5: Vertical flip: a) before b) after

Figure 4.6 shows some images of those that were successfully fixed using this procedure.

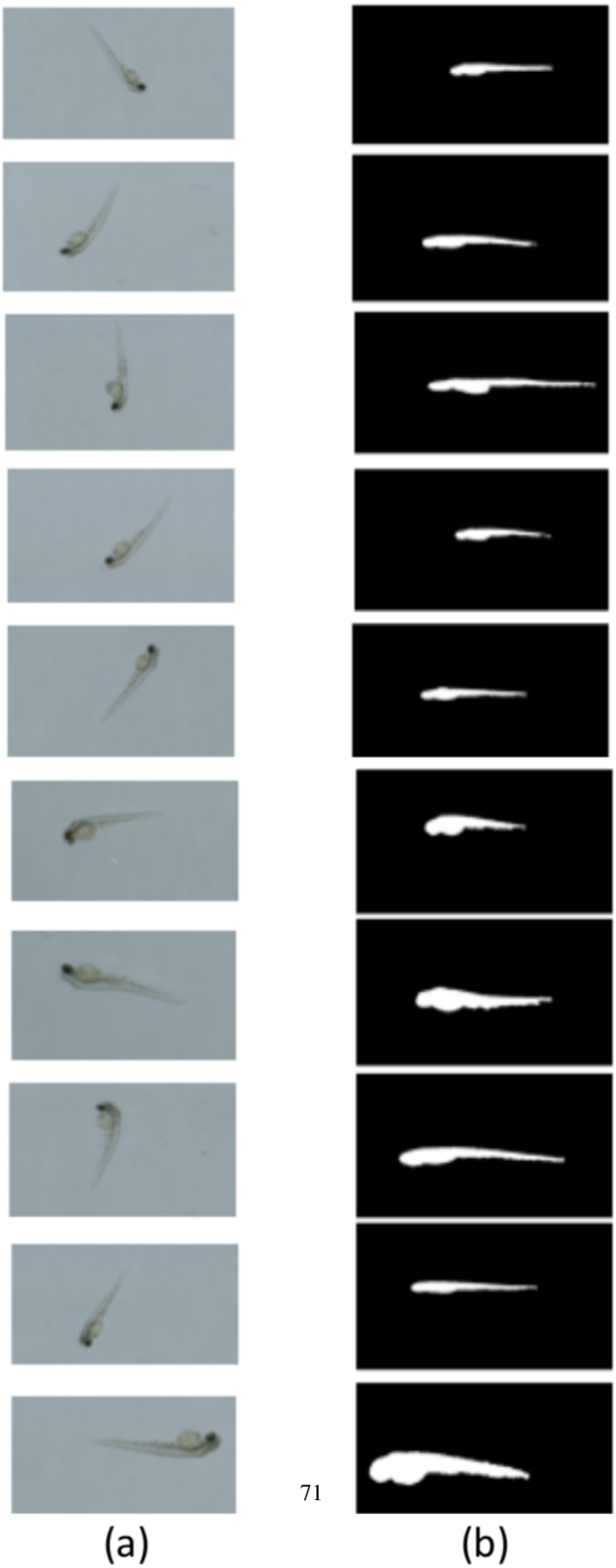


Figure 4.6: Fixing results: a) before b) after

4.5 The System Methodology

The proposed work aims to identify the health and detect the abnormalities of zebrafish embryos from day.1 to day.5 using computer vision algorithms. The proposed system has two automatic parts: one for the data acquisition and the other for data analysis and result finding. The images are comprised of many features which could be extracted automatically or manually. The collected images were gathered using a scanner collecting a large number of high-resolution images (suitable for biological observation) every scan facilitating high-throughput analysis. Figure 4.7 shows the general stages of the proposed system.

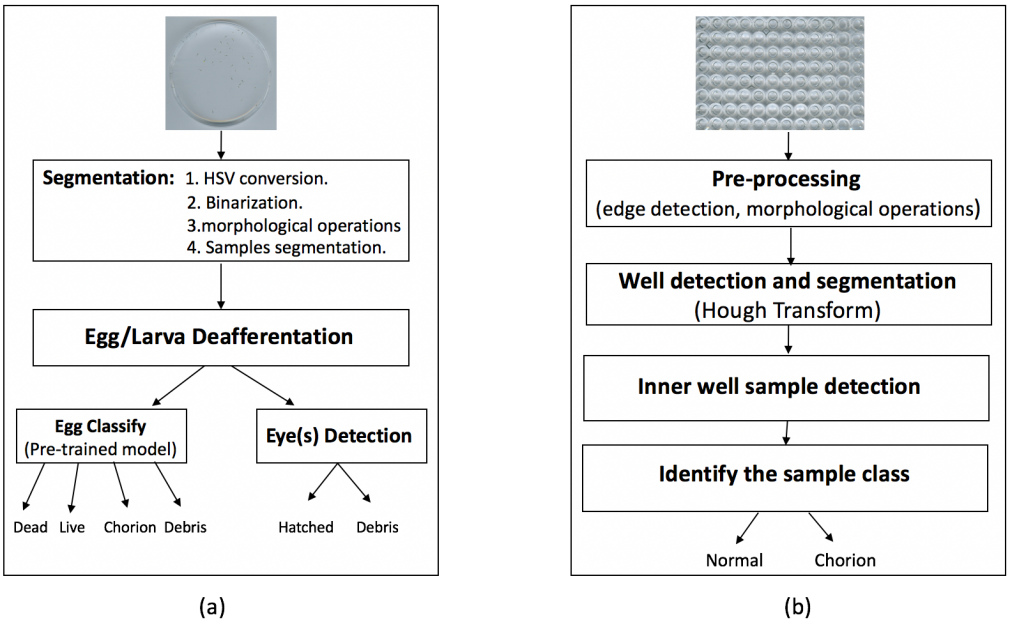


Figure 4.7: General overview of the system (a) dish images (b) n-well plate images

Due to the low number of the biological experiments and the lack of the malformed samples and the low survival rate, the collected images were grouped into: live egg, dead egg, normal

embryo, and chorion. The scanning process is always carried out synchronously with the biological experiments. The experiments ran for up to 5dpf (days post fertilisation), adding chemicals (e.g. application of different chemicals to the holding water) or physical interventions (e.g. temperature) can be applied. The images were prepared and subdivided into four different categories for the software design step. Figure 4.8 shows examples from the four classes.

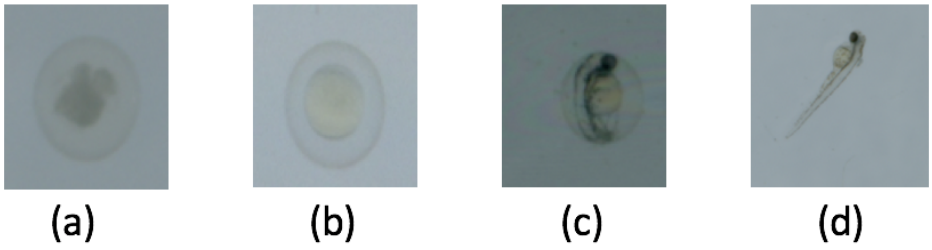


Figure 4.8: Sample classes a) dead egg b) live egg c) chorion d) normal

The images that have been used in the proposed classifier were gathered from the dish and the n-well plate images. The live eggs between 0-40 hpf (hours post fertilisation) are transparent and look like a yellow circle. The dead eggs are white and appear dark. The last two classes which present an older embryo with more than 3 dpf, the chorion is a non-hatched egg, and the normal is the hatched one.

In this work, several stages are combined and integrated to detect and classify the zebrafish larvae according to their age and health. Two different containers were used for data collection during the biological experiments.

4.5.1 Zebrafish Larva Identification and Classification Inside Petri Dish

Using petri dish images and based on computer vision algorithms, an image analysis pipeline to detect and classify the zebrafish larvae.

- **Image pre-processing:**
  This step started with HSV colour space conversion, which experimentally better than the edge detection. Depending on the hue image and using an experimental threshold, the image was binarised. Some morphological operations were applied on the binary image to get an image with the zebrafish larva samples only. Figure 4.9 shows these processes applied to a petri dish image.

- **Egg/Larva Differentiation:**
  According to the ground truth of the collected images, four classes during five days were observed and registered by a biological expert. Three classes of these four classes are non-hatched larvae. Due to the different shape properties between the three classes and the fourth class (hatched), a distinguishing process was done before the classification step.

  The most useful feature to differentiate the egg and larva is the roundness. Depending on the segmented object dimensions (area, perimeter) and using the roundness equation below, the rounding metric was found to decide the next steps.

$$R_m = \frac{4 * \pi * A}{P^2} \tag{4.1}$$

  Where: R_m is the rounding metric, A is the object area, and P is the perimeter of the object.

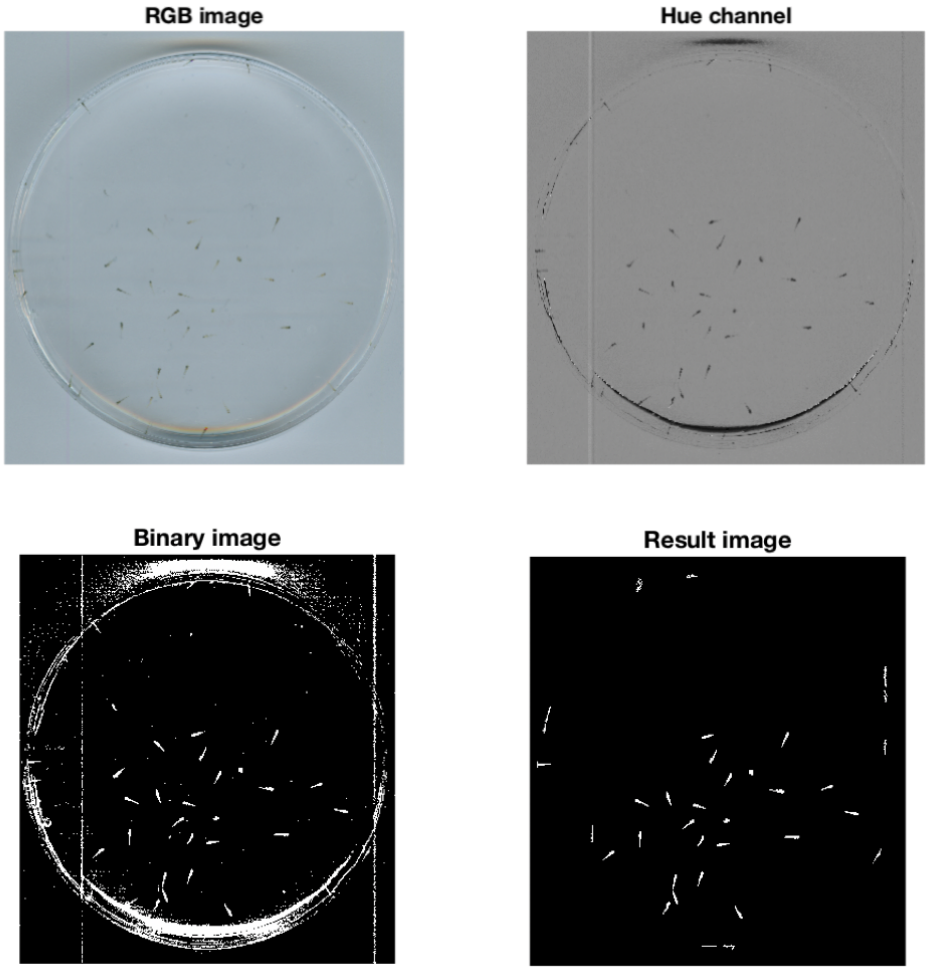


Figure 4.9: Segmentation process (petri dish)

In the early stage of the experiments (1-3 dpf), the objects can be detected and segmented using the Hough Circle Transform (HCT) for circular shape detection. Furthermore, this cannot be efficient with the next stages of the zebrafish larva images (3-5 dpf). Figure 4.10 shows larva and egg samples with the roundness value. The roundness value ranges between 0 to 1, where 1 means the full rounded object.

The decision of this step simplifies the system to deal with egg

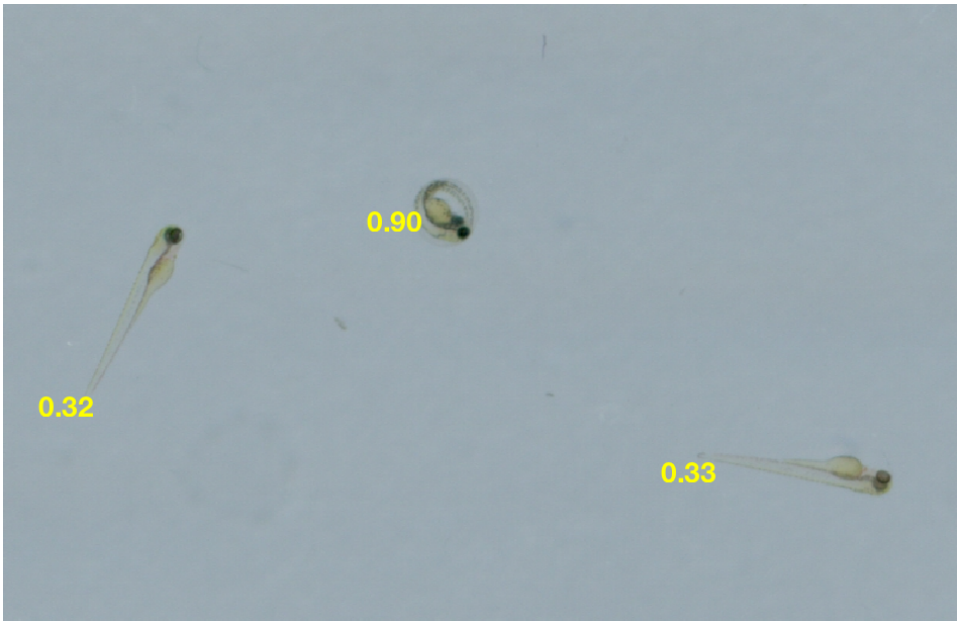


Figure 4.10: Samples roundness feature

or larva. If the sample is identified as an egg, the sample will be tested using a pre-trained model that was designed and trained using a large number of zebrafish egg images covering the three classes and using their texture features as illustrated in the next subsection.

On the other hand, if the sample is detected as a larva, it will be tested again if it is real larva or debris or a part of the dish edges. To check and confirm if the sample is a hatched larva or not, an eye(s) detection process is applied using HCT to detect and find out any dark circle with an experimental radius value which is evaluated based on the larva eye radius. Figure 4.11 shows some results of eye detection to confirm the sample class.

- **Zebrafish larva Classifier:**
  The aim of the classifier designing is to classify the three types

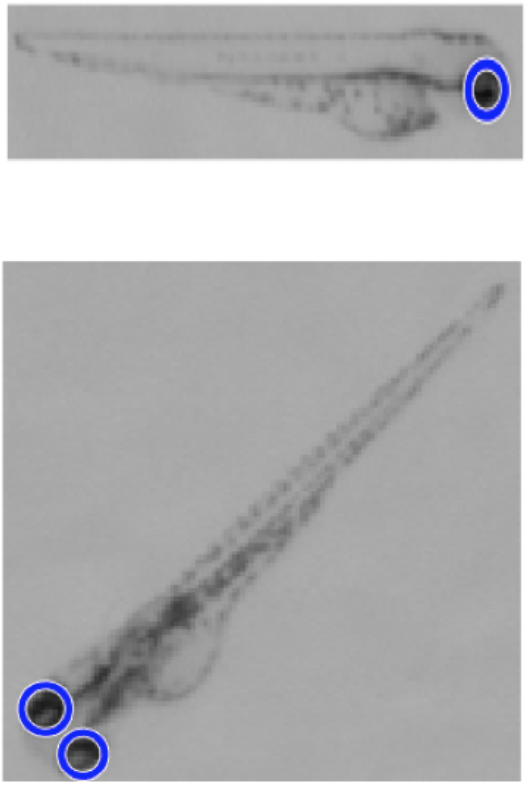


Figure 4.11: Zebrafish larva eye detection

of non-hatched samples. This was done by training CART model using the texture features of the collected data. After cropping the objects from the dish image and divide them into three categories depending on the ground truth.

Several image processing operations are applied on the training set as shown in Figure 4.12 to segment the sample object from the whole image to make sure that the most important and useful features will be extracted where the image may contain debris and unwanted objects. These operations start with the edge detection process using the Canny filter. After detection

of the edges, some dilation using a disk mask with two pixels is done.

To remove any unwanted objects, the largest object is segmented depending on the object areas. The resulting image is produced from the grey image multiplied by the largest object binary image to keep the target object information. The processed images have been used in the next steps to extract the features and using these features to train the classifier on how to predict and classify the three classes.

The next step is the feature extraction were the essential characteristics of the three classes images are extracted and used for the training step. The first attempt was made by focusing on two first order colour features of the egg according to the colour variance between the three classes. According to the colour similarity between the dead and the embryo classes, 22 texture features are extracted from the egg images.

The texture features are useful for the classification process when a wide variation of the grey levels are present in the image. Combining first and second order features aim to have robust features for the classifier training step. The feature vector consists of 24 features for colour and texture image characteristics. For the image I of nxm in size, the mean and the standard deviation values are calculated as follows

$$mean = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij}}{mn} \tag{4.2}$$

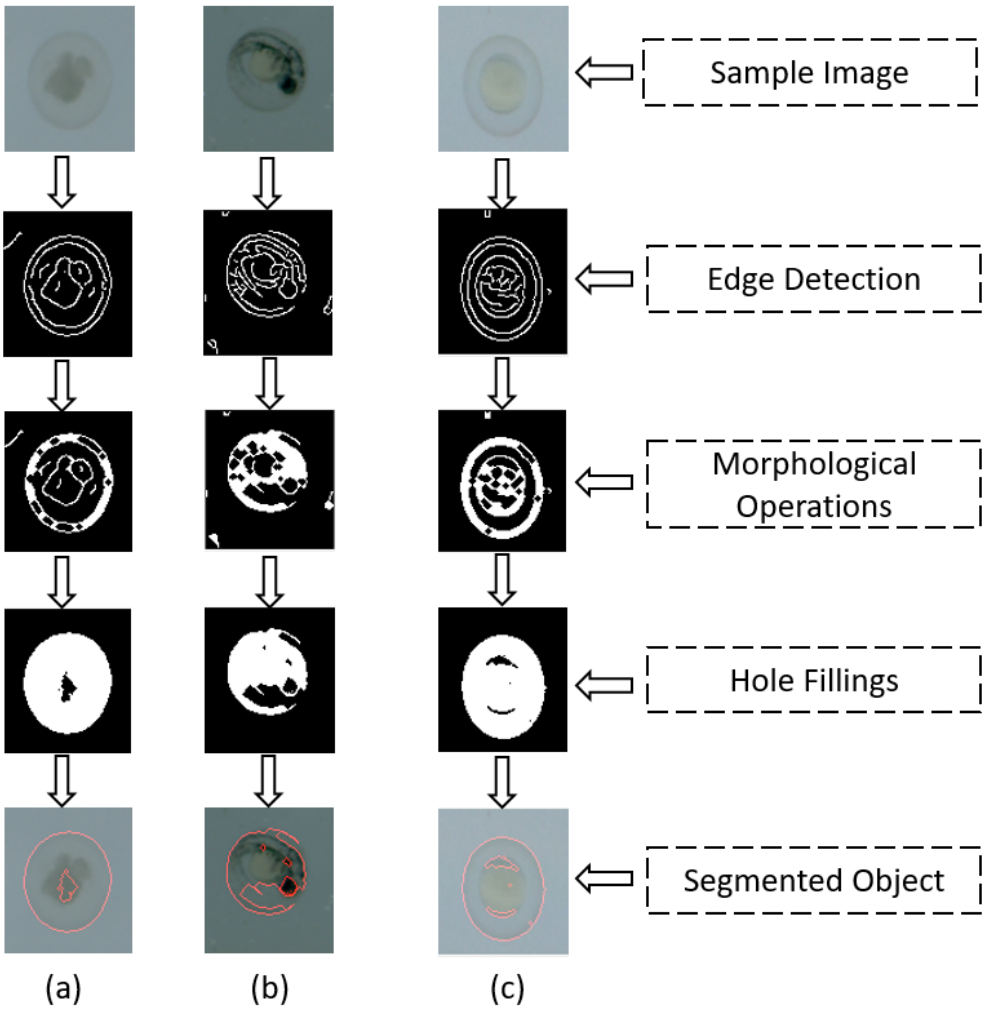


Figure 4.12: Pre-processing operations

$$std = \sqrt{\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}(I_{ij} - mean)^2} \tag{4.3}$$

The first order features provide basic information about the grey level distribution. However, the relative positions of these grey levels have not been provided by the first order features. The second order features describe and analyse if the low grey levels are together or mixed with the high grey levels. These features are calculated as proposed and illustrated in [94, 95, 157].

The second order statistics are calculated depending on a matrix $C_{\theta,d}(I_{p1}, I_{p2)}$ of the relative frequencies that describes how often the two pixels (I_{p1}, I_{p2}) of different or similar gray levels N_g appear as a pair in the image matrix concerning the distance d and the direction θ. The value of this parameter N_g is 8 levels.

Using the co-occurrence matrix, 22 features are extracted where the smoothness, coarseness, and the image texture information are described and quantified. The image contrast, correlation, cluster shade, cluster prominence, energy, homogeneity, entropy, and variance are measured as follows:

1. Auto correlation:

$$f_1 = \sum_i \sum_j (ij)(C(i,j)) \tag{4.4}$$

2. Contrast:

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} C(i,j) \right| |i-j=n| \tag{4.5}$$

3. Correlation:

$$f_3 = \frac{\sum_i \sum_j (ij) C(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{4.6}$$

Where:

$$\mu_x = \sum_i i C_x(i) \tag{4.7}$$

$$\mu_y = \sum_j j C_y(j) \tag{4.8}$$

$$\sigma_x = \sum_i C_x(i) - \mu_x(i) \tag{4.9}$$

$$\sigma_y = \sum_j C_y(j) - \mu_y(j) \tag{4.10}$$

4. Cluster prominence:

$$f_4 = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 C(i,j) \tag{4.11}$$

5. Energy:

$$f_5 = \sum_i \sum_j C(i,j)^2 \tag{4.12}$$

6. Entropy:

$$f_6 = \sum_i \sum_j C(i,j) log(C(i,j)) \tag{4.13}$$

7. Homogeneity:

$$f_7 = \sum_i \sum_j \frac{1}{1+(i-j)^2} C(i,j) \tag{4.14}$$

The rest of the 15 features are correlated using Matlab functions. Cluster shade, dissimilarity, homogeneity using a Matlab function, maximum probability [157]. Sum of squares, difference variance, sum average, sum variance, sum entropy, difference entropy, information measure of correlation1, information measure of correlation2 [94]. Inverse difference, inverse difference normalized, inverse difference moment normalised [95].

The idea of the CART classifier model is presented by conditions. In this model, several questions are answered by the trees sequentially like **If-Then** condition statements. These questions depend on the extracted features from the images. Using CART model related to its efficiency and flexibility.

The tree model is easy to interpret and modify according to the observed internal work. The classification consists of two

main steps, training and testing. The data set is divided for training and testing processes using two quarters-one quarter relation from the whole data set as follows:

Table 4.1: Data Set Division

Class	Training Set	Testing Set	Total
Dead egg	322	161	483
Chorion	14	7	21
Live egg	464	231	695

In the training stage, the feature set (predictors) with class labels (responses) are used to train a CART classifier model. The second step is the testing step in which the classifier performance appears as an essential indication of its capability. To predict the class of a new sample, the designed model follow the decisions in the tree from the root (beginning) node down to a leaf node. The leaf node contains the response. By repeating the first steps for preparing the image and extracting the 24 features, these features (predictors) are provided to the saved classifier model to predict the class (response).

4.5.2 Zebrafish Larva Identification Inside n-well Plate

In some experiments, the biologists used the 48 or the 96-well plate in their experiments to observe the zebrafish larva reaction. Several attempts were set out in the purpose of getting malformed samples. However, the number of experiments was not enough to get defected samples, and most samples died in the early stages. This situation presents a big challenge for the proposed project and ends the work with only four classes.

- **Pre-processing:**

  Several image processing operations were applied to the well plate images to detect the target objects from the whole image. The process started with the canny edge detection algorithm using an experimental threshold value. The result binary image was enhanced using morphological operations, e.g. dilation using a disk mask with 2 pixels size. Figure 4.13 shows the result of the pre-processing step for 48- and 96-well plate.

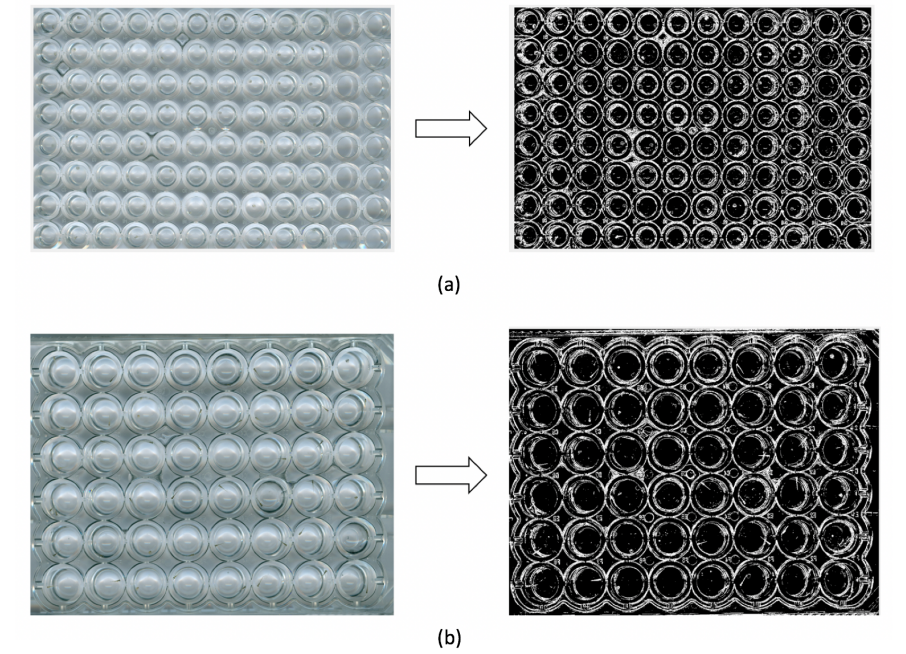


Figure 4.13: Pre-processing operations: a) 96-well plate b) 48-well plate

- **Well detection:**

  To detect and segment the samples inside the wells, the wells should be detected, segmented, and cropped to specify the identification process for each well which is necessary for memory usage and processing time reduction.

  Well detection is an important step for the individual identifi-

cation process for zebrafish larvae. Well detection process was done using Hough Circle Transform (HCT) algorithm due to the well shape and depending on the well radius, which is the same for all wells. As shown in Figure 4.14, the well radius of each image with any size using 1200 dpi imaging resolution is 300±100.

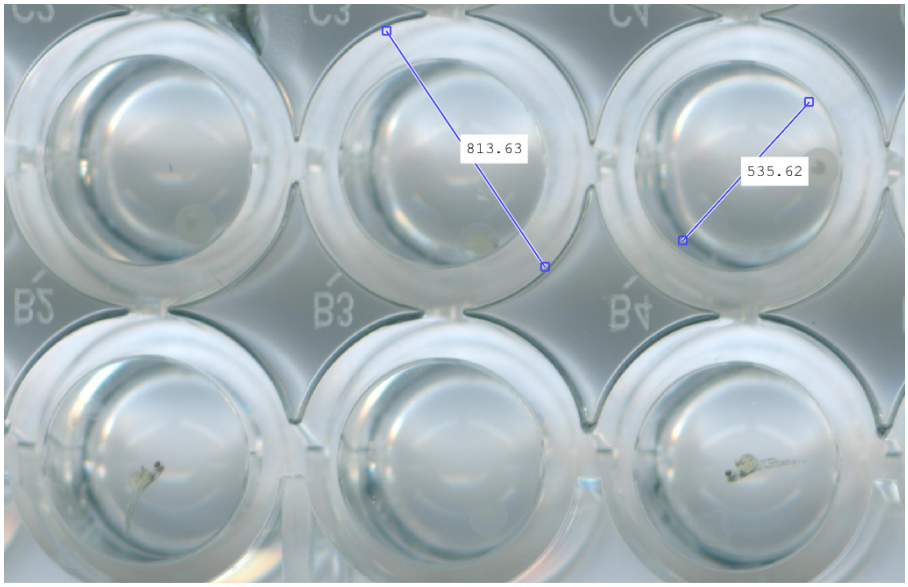


Figure 4.14: Inner and outer well diameter

Using the HCT for well detection present a robust detection and segmentation technique, it is working efficiently with any size of images and any size of wells. Furthermore, the image orientation, scaling, and the number of wells is neglected where the well segmentation using HCT algorithm find out the required circles with specific properties and robust to these significant problems in image processing. Comparing with the edge detection and plate dimensions segmentation process that was proposed in [158], the HCT was higher performance than this method and more robust for the system consideration.

- **Sample detection and identification:**
  At this stage and after specify all wells as shown in Figure 4.15 and Figure 4.16, the sample in each well will be detected, segmented, and identified its class.

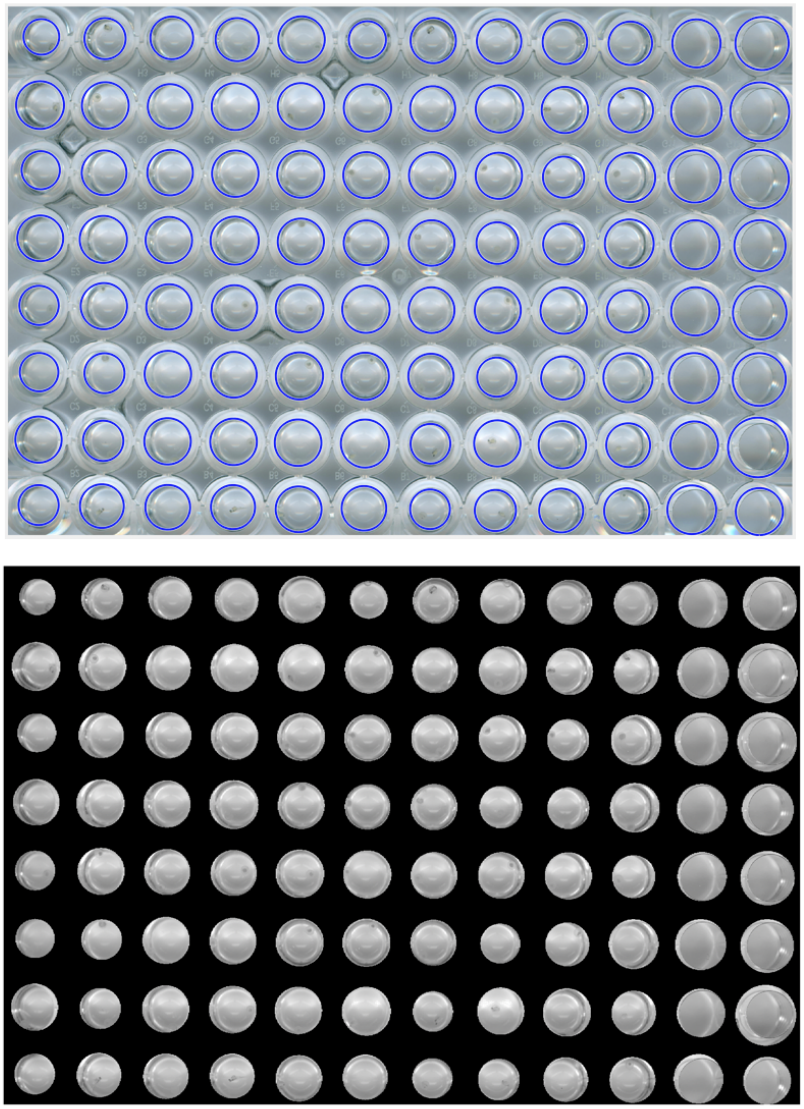


Figure 4.15: Wells segmentation

The sample segmentation started with canny edge detection

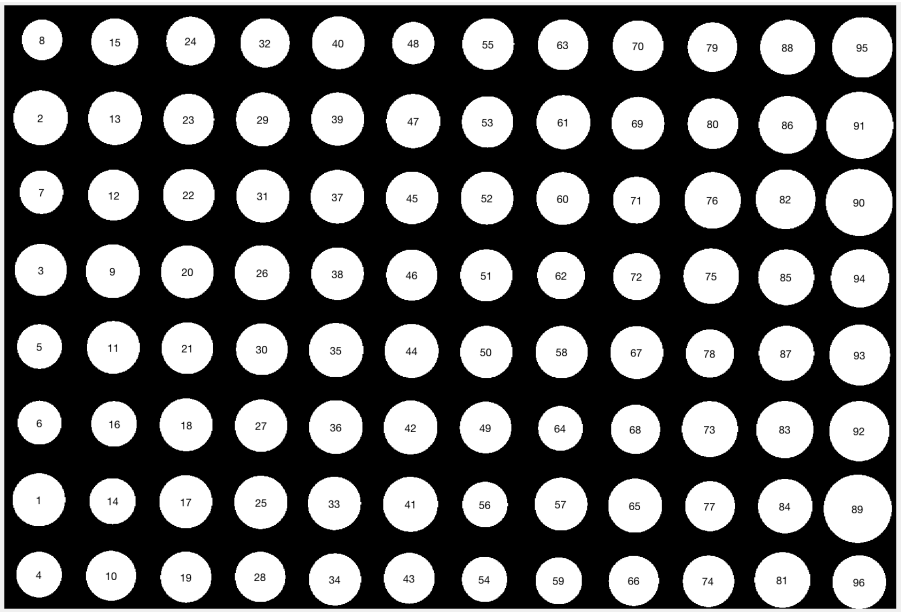


Figure 4.16: Wells labeling

with an experimental threshold value. The image was processed using two linear dilation masks with size 3 pixels and two different angles. The result processed image has many unwanted objects. The outer ring was removed depending on the object perimeters to reduce the number of unwanted objects and deal with inner parts of the well only. After removing the outer circle, the holes are filled, and the image was filtered again depending on the object areas to segment the experimental sample. Figure 4.17 shows the inner well processing and sample segmentation.

The identification process, as mentioned before for the petri dish analysis. Using the properties of the sample and depending on the object shape, the identification process was carried out, and the sample class was shown for the system user. The n-well plate differs from the petri dish that has two types of samples (hatched and non-hatched zebrafish larva). Moreover, the wells

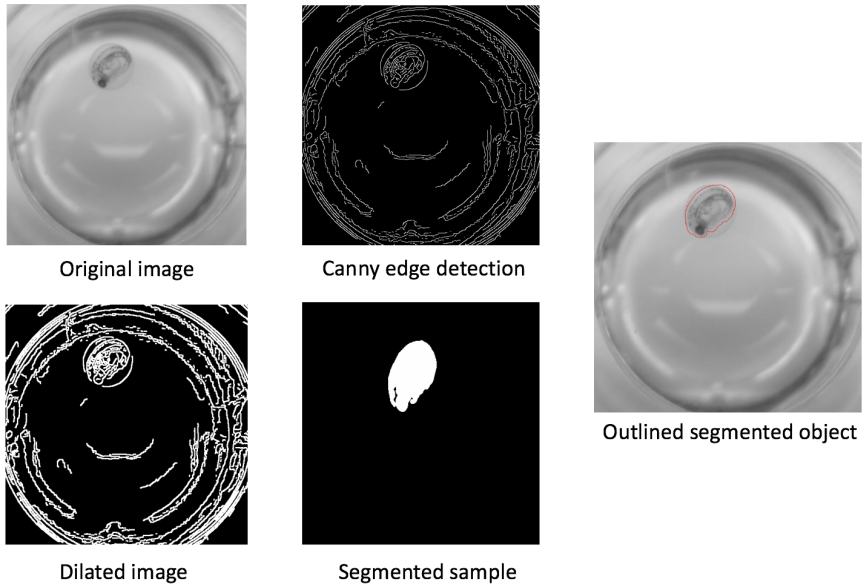


Figure 4.17: Zebrafish larva segmentation

are narrow, and the chance of getting merged samples with the well edges is higher than the petri dish container.

4.6 Results and Discussion

Using the n-well plate for zebrafish larva analysis and observation provide high detection and identification results. However, the segmentation of well contents is crucial especially when the samples are close to the edges which cause a wrong object detection.

By applying all the mentioned steps and testing them on the plate images, the detection process succeed with all inner samples without considering those close with the edge, and it was segmented with the edge as one object as shown in Figure 4.18.

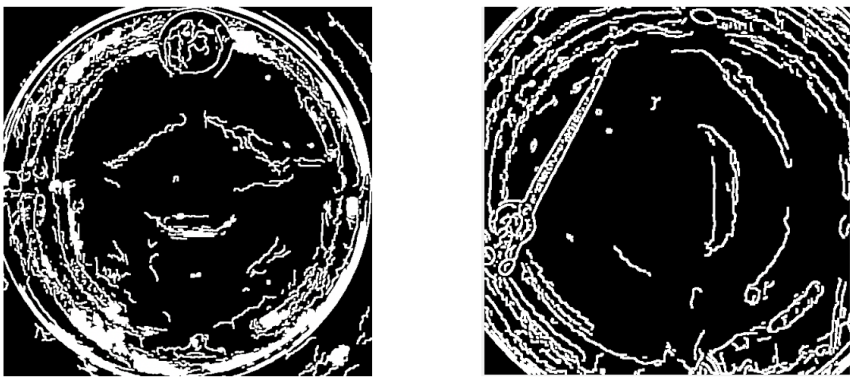


Figure 4.18: Zebrafish larva close to the well edge

The detection and classification error for the n-well plate samples was zero after discarding all empty wells and those with sliding samples to the edges. Once the sample is located in the well centre, the system able to detect, segment, and classify the larva. Figure 4.19 shows examples of the zebrafish larva inside 48-well and 96-well plates that were correctly detected and classified using the proposed system.

The system can detect and classify the samples inside the plate wells. However, many samples (about 60%) are close to the well edges and cause detection and classification error. On the other hand, using the petri dish for the same purposes showed a higher performance and overcoming the sample sliding to the problem of the edges.

The proposed image analysis pipeline for samples of a petri dish shows a higher performance compared with the n-well plate; the complexity time is extremely the same. This system is used to detect, segment, and classify the eggs/larvae within the whole dish

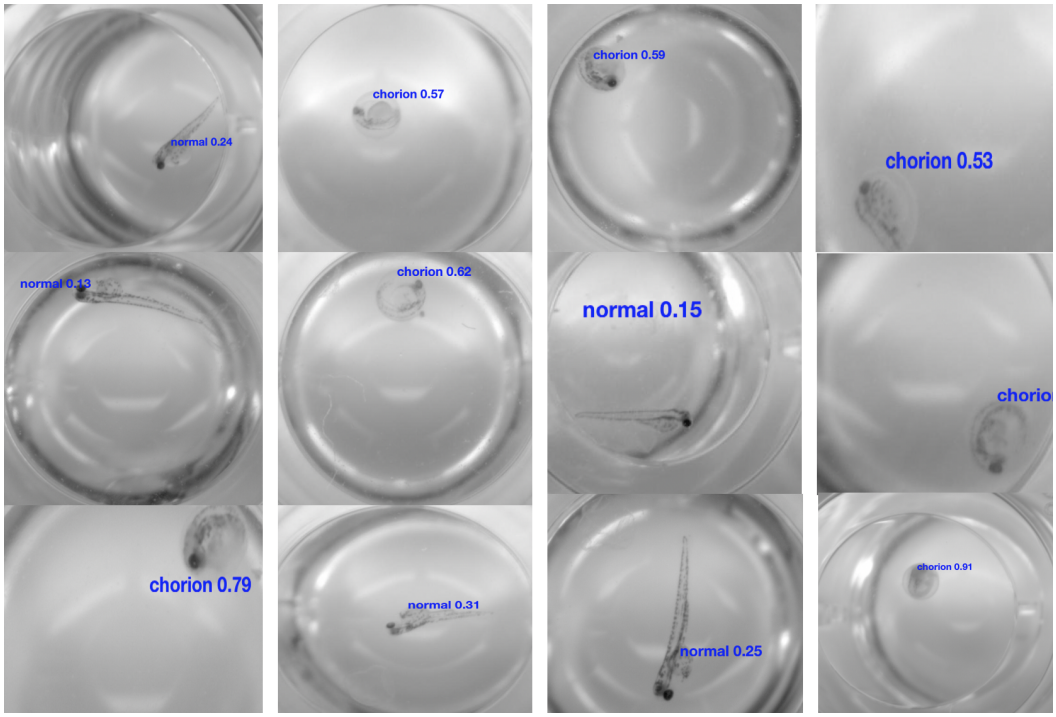


Figure 4.19: Examples of well-plate classification

image. Figures 4.20 and 4.21 show examples of small parts of different dishes where the egg/larva samples are detected successfully and classified correctly. The label **L** for the live egg and **D** for the dead one.

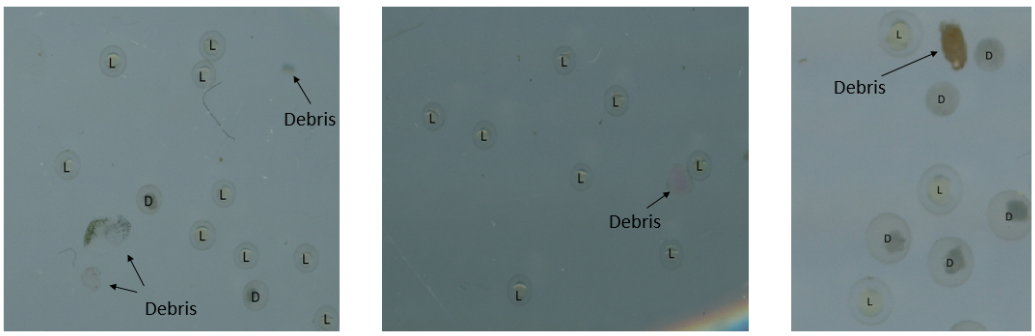
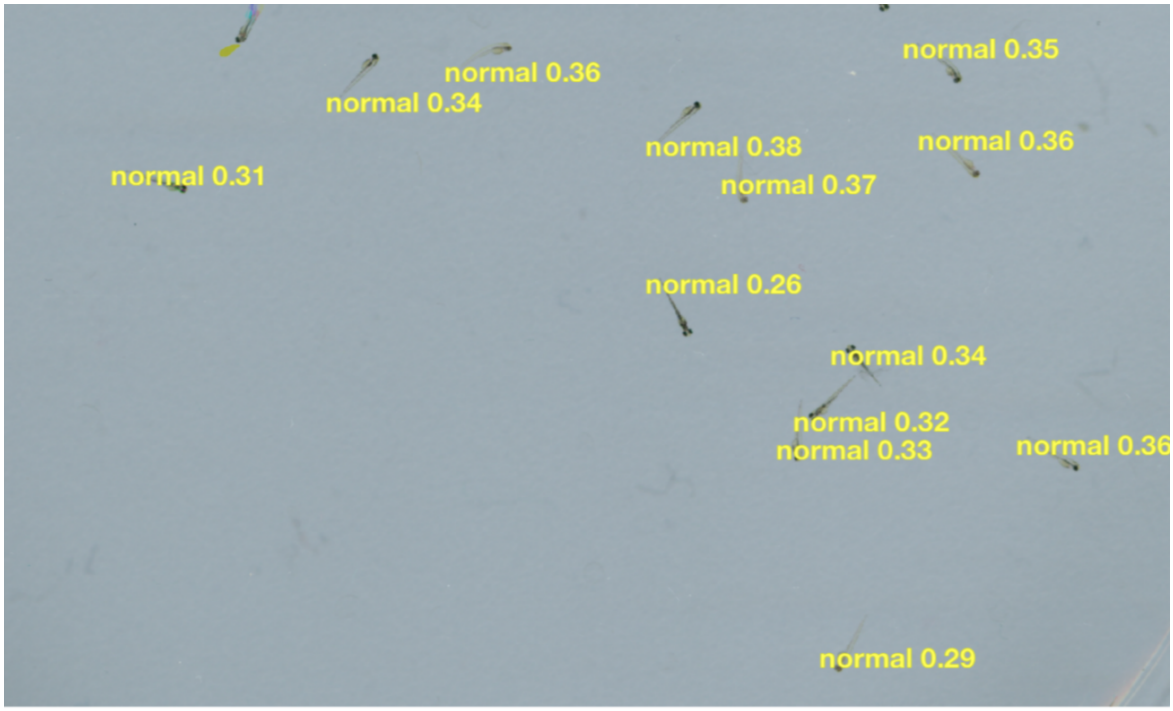
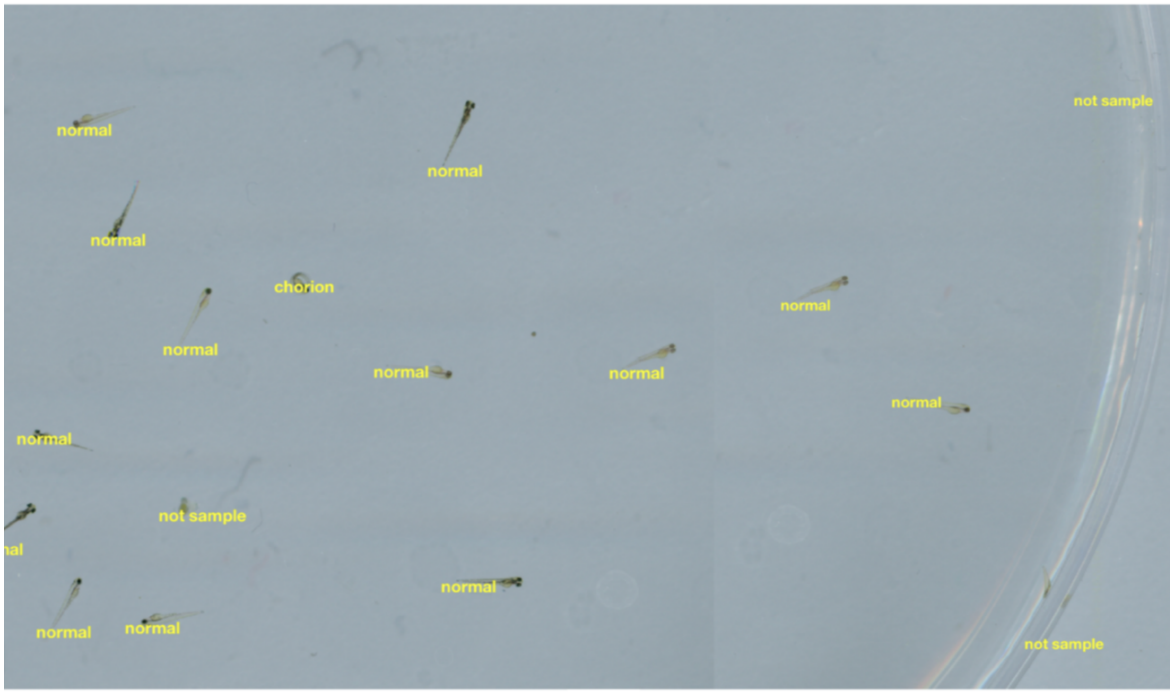


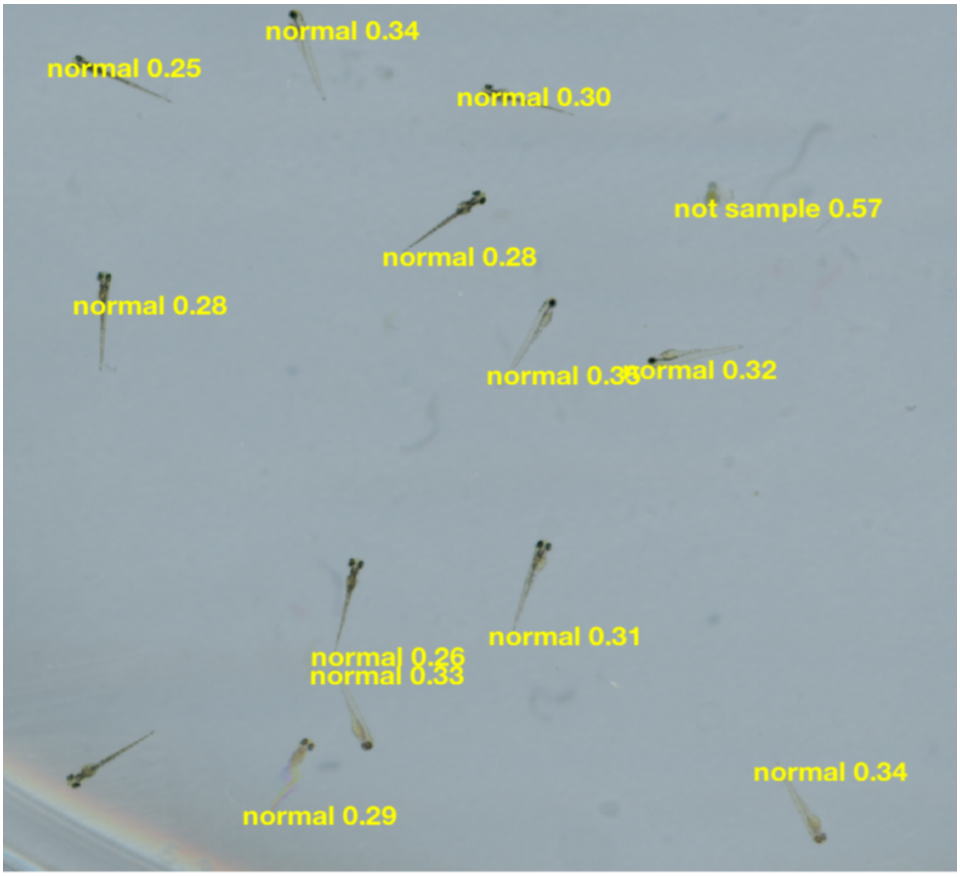
Figure 4.20: Classification results for the whole dish image


normal 0.35
normal 0.36
normal 0.34
normal 0.38
normal 0.36
normal 0.31
normal 0.37
normal 0.26
normal 0.34
normal 0.32
normal 0.33
normal 0.36
normal 0.29

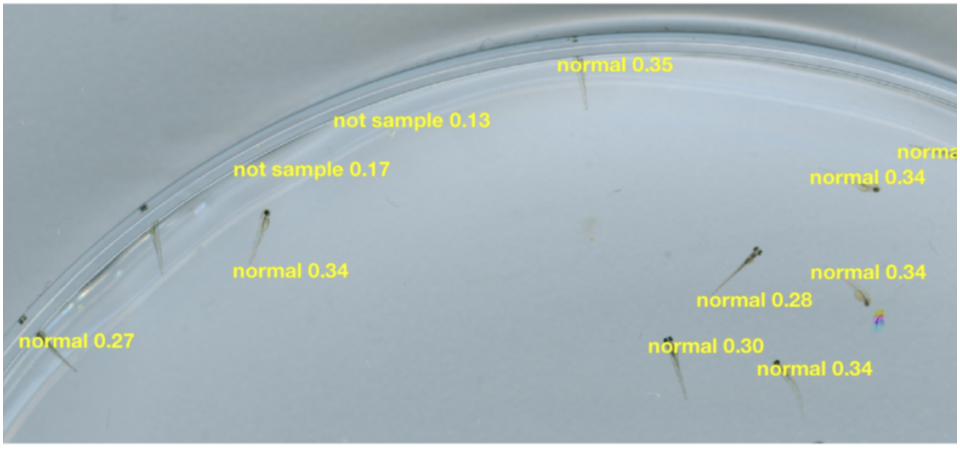
(a)


not sample
normal
normal
normal
chorion
normal
normal
normal
normal
normal
normal
not sample
normal
normal
normal
not sample

(b)



(c)



(d)

Figure 4.21: Identification results for the whole dish image: a) Example 1 b) Example 2 c) Example 3 d) Example 4

As shown in Figure 4.20 and Figure 4.21, the debris which is considered as an unwanted object is discarded, the edge parts are identified as not organism, and the target objects are detected, segmented, and classified efficiently. However, in some cases, the proposed system failed in detection and classification processes, as shown in Figure 4.22. Some samples are misclassified, which can be manipulated either by cross-validation for the classifier or reducing the number of extracted features by using a feature selection algorithm.

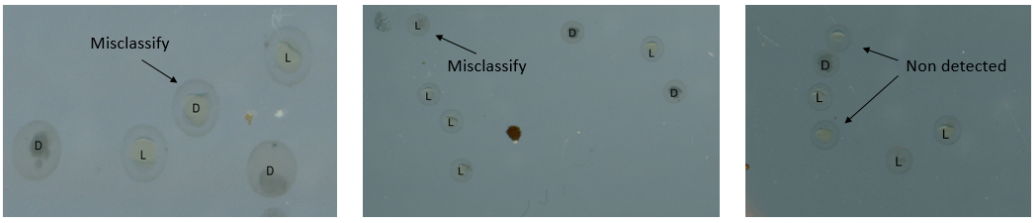


Figure 4.22: Drawbacks of the classifier

Using the scanner images provides us with a large number of samples of outstanding quality. However, these images are taken from a bigger image that contains hundreds of eggs. Partitioning process to get every sample and deal with it as a single image distort the sample images in many cases which are considered as a big challenge in this work.

4.7 Summary

In this chapter, a novel zebrafish larva image analysis pipeline is starting from detection to segmentation and classification based on two container image types. Two types of container with different ages were tested, analysed, and compared to the ground truth. By

extracting the most important features for both colour and texture characteristics of the image, a high-performance classification model was presented and evaluated with 97% testing accuracy to classify the sample image of the zebrafish egg into three classes (live egg, dead egg, chorion) depending on its status.

Using the flatbed scanner is presented as a low-cost, effective imaging tool that saves the consuming time where the one-shot provides the system by hundreds of sample images. Furthermore, this tool affordable and easy to use by the biologists with the least imaging problems.

This system covers the most common malformation types considering the nature of these embryos (orientation, position) that is fixed using an automatic fixing method.

Besides the benefits that are provided by using the flatbed scanner for data collection, the proposed system is assessed by the biologists as an effective and time-saving process for their experiments. The traditional way of capturing images for the samples is always carried out individually for hundreds of samples, which is a time-consuming process. Also, the biologist needs to analyse each sample to decide its status, which is also considered as a time-consuming process.

By using the proposed system, the biologist intervention is limited, and the experiment time should be shorter than usual. This system presents an image analysis pipeline that automatically

detects, classifies, and identifies the zebrafish embryo abnormalities using a high-throughput model for both the petri dish and the n-well plates.

Chapter 5

Zebrafish Embryo Malformation Classification

5.1 Introduction

In this chapter, multi-class and multi-label classification systems are proposed based on two machine learning algorithms to classify the zebrafish larva malformations. The first proposed algorithm is the decision tree, and the other algorithm is the pretrained CNNs. Nine classes are identified by the proposed machine learning algorithms for multi-class purposes. Most images have more than one label where a multi-label based on the biologist observations. The result of these two concepts is presented and discussed. A comparison between the presented systems and the ground truth is presented. The most challenges of this research are mentioned.

The proposed work aims to detect and classify the zebrafish embryo malformations using different types of features, which were

analysed and optimised with the classifier parameters to classify nine classes of larvae (5dpf). The proposed work presents a binary relevance multi-label classification using Decision And Regression classification (CART) model integrating with shape and colour features. Furthermore, this system able to identify all the deformity classes that may appear on a one sample body at the same time. On the other hand, a classification system was proposed based on the pretrained CNNs for multi-class and multi-label problems.

5.2 Challenges and Objectives

Recently, zebrafish embryos are widely used in the pharmacological and toxicological experiments and studies. The proposed system aims to detect and classify the malformations that may cause for the embryos after adding chemical substances.

The most challenge faced the proposed work is presented by the high similarity between the different classes. Some classes are hardly identified and classified by an expert that reflected on the proposed system performance. Furthermore, the high number of types compared with the small size of the data set.

To reach the best reliable system for zebrafish embryo screening, several objectives were drawn using the proposed system:

Objectives:

- Using efficient and more accurate machine intelligence algorithms for detection and classification of malformations for zebrafish larvae of 5dpf sample ages.
- Using different types of features using the ROI which integrated with a high-quality classifier.
- Optimise all the extracted features and tuning them with the classifier parameters to classify nine classes of larva malformations (5dpf).
- Identify all the deformity classes that may appear on a one sample body at the same time.

5.3 Data Set

All the used images of this work are collected from public resources with unrestricted use and reproduction. The data set is available from [13, 42]. This data set was captured and collected by Jeanray et al., in which the images are taken in different sessions. These images were captured using an Olympus SZX10 microscope joined with an Olympus XC50 camera with consideration of the light illumination that comes from a condenser passing into the embryo. By using this capturing tools, the generated images are 2575×1932 pixels of resolution in 8-bit TIFF format. The number of images that were used in this system was 944 images for training and 230 images for testing process.

The data set consists of different types of malformation to different body parts such as the embryo tail, the vessels, and the whole body shape on the fifth day after fertilisation (5dpf). Adding

several chemical substances into the surrounded liquid of the larvae causes changes in the tail shape, so it goes to be up or down, also the larvae may not be hatched (i.e. chronic type) or having edema. Figure 5.1 shows examples from this data set [42].

Figure 5.1: Malformation Examples: a) Chorion b) Dead c) Down-Curved-Tail d) Edema e) Hemostasis f) Necrosed-yolk-sac g) Short-tail h) Up-curved-tail/fish i) Normal

To integrate the position of the embryo in all images with the proposed algorithm, the embryo should take a fixed scene. This position should be aligned horizontally with the head aligned at the

left side.

To detect any gradient position of the embryo, each image was identified if it has any an unusual position by calculating the angle of the orientation then rotates the image counterclockwise by the value of the oriented angle as shown in Figure 5.2 The angle was found from the following equation:

$$\theta = \tan^{-1}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \tag{5.1}$$

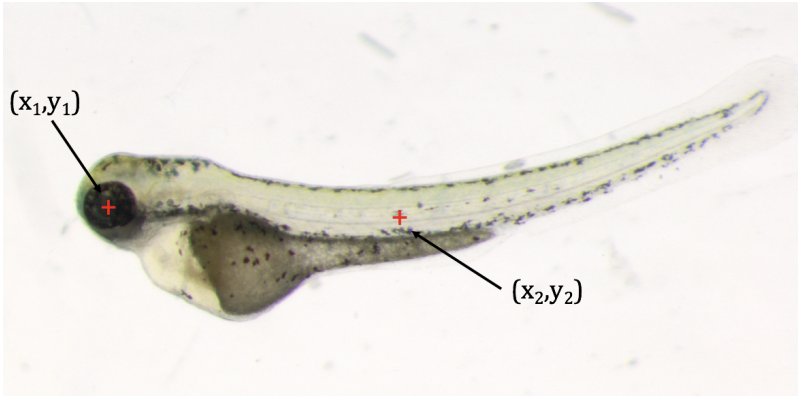


Figure 5.2: Orientation fixing points

5.4 Binary Relevance And Decision Tree

The proposed system aims to design and develop a novel automated detection and classification system for zebrafish embryo malformations with high performance. In this system, nine classes of zebrafish larva (eight types of deformities and the normal class) have been analysed and classified. Figure 5.3 shows a general view of the system.

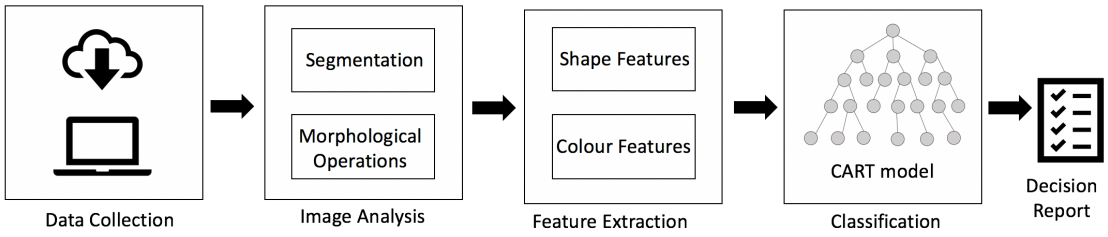


Figure 5.3: General structure of the system

5.4.1 Image Analysis

Having the most important information of the images will improve the classifier performance because it reduces any noise or distortion that may affect the desired detection. To generate high informative images, several image processing functions were applied to distinguish the larva body from the whole image.

The image resolution was down-sampled to quarter the original resolution (483×644) to upgrade the system work and minimise the load on the CPU and get the result faster. The embryo region was automatically extracted from the surrounding background using an automatic thresholding process. The thresholding value was found using the minimum numeric value after applying two algorithms; the first one that was developed depending on the cumulative histogram of the image. The other one that depends on the colour level distribution is adopted from [159].

The biggest blob was identified after labelling all blobs inside the image. Then the dilation and the erosion processes were applied to process the border of the object (larva) to be more noticeable and recognised. As illustrated in Figure 5.4, the raw image has many

small objects around the larva body that should be eliminated to have apparent features of the image using filtering and thresholding and other morphological operators such as dilation, erosion, and closing, which yield the segmented image with the target object (larva). Finally, the segmented object is cropped to get only the embryo body, as shown in Figure 5.4.

Figure 5.4: Image pre-processing

5.4.2 Feature Extraction

To distinguish between the different types of malformations and to identify if the larva is normal or not, 14 features were chosen and extracted to get the most valuable characteristics and information for the nine classes. The extraction process was carried out depending on the region of interest (ROI) and using the shape and the colour features according to each class characteristics. Table 5.1 shows a summary of these features.

The most important feature that indicates the chorion deformity is the roundness of the object. To detect the non-hatching abnormality, the rounding metric was found for the sample images (hatched and non-hatched). This metric was found depending on

Table 5.1: Extracted features

Class	**Features**
Chorion	Rounding metric
Curved tail/fish	Quadratic coefficient, curvature angle
Short tail	The length of the larva body, distance from eye centre to the end point of the body
Dead	Mean intensity value
Necrosed yolk	Mean intensity value, standard deviation
Edema	Area of yolk, height of the larva body
Hemostasis	Area of red spots

the object perimeter and the object area and using the following Equation 5.2.

$$R_m = 4\pi.\frac{area}{perimeter^2} \tag{5.2}$$

The value of the rounding metric for the chorion class identified between 0.85-1.00, this metric evaluate the rounding degree of the object where 1.00 means the object is entirely circular, and for any other values, the object classified as hatched.

The other classes that were studied are related to the curvature abnormality of the zebrafish larvae tail (up or down) is detected using geometric features. The study in [160] described tail curvature of the zebrafish embryos as a "surrogate marker" that is used to identify the kidney cyst disease effectively. This work demonstrated how the ValProic Acid (VPA) could reduce the disease effects on the kidney in the earlier stages and before being worse. As a result of high-throughput screening, the responsible gene of the tail deformity can be identified and depending on this the suitable chemical substances can be added by the biologist to

stop the disease progression.

Using the Region Of Interest (ROI) only which is the tail side and finding the gradient to get the tail edges, which is used to find the quadratic equation coefficient, which indicates the tail curvature. After specifying the horizontal and the vertical directions of the tail edge which are computed using the following equation.

$$\nabla F = \frac{\partial F}{\partial x}\hat{i} + \frac{\partial F}{\partial y}\hat{j} \tag{5.3}$$

This step was done to restrict the next steps to be applied in the desired direction which is the positive y-direction as shown in Figure 5.5.

The curve fitting was used to calculate the degree of the curvature. For the normal shaped embryos, the body should be straight where any curving appears with their bodies classified as a malformation. Then to decide if the curve is up or down, it depends on the curvature direction is it up or down. After applying the

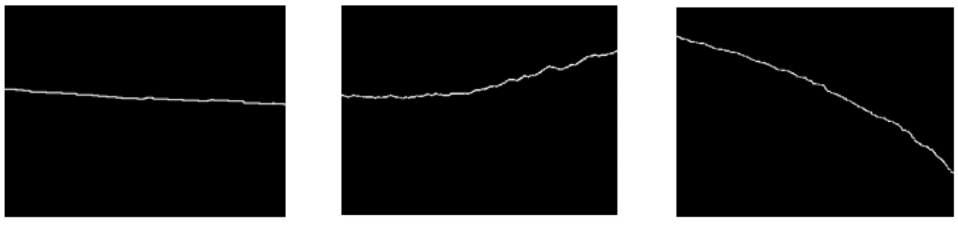


Figure 5.5: Positive y-axis gradient of the cropped tail (a) normal tail (b) up-curved tail (c)down-curved tail

polynomial fitting function the three coefficient values (a, b, c) of

the quadratic are obtained from

$$f(x) = ax^2 + bx + c \tag{5.4}$$

Examples of the obtained polynomial curve fitting for the y-axis gradient of normal, curved-up and curved-down tails are shown in Figure 5.6. These coefficients indicate embryo's abnormality (i.e. curved-up or curved-down). The value of a is varied between -8×10^{-4} and 10×10^{-4}. The negative value of a indicates the up-curved, the positive over 1 indicates the down-curved, and the value around zero indicates the un-curved sample body.

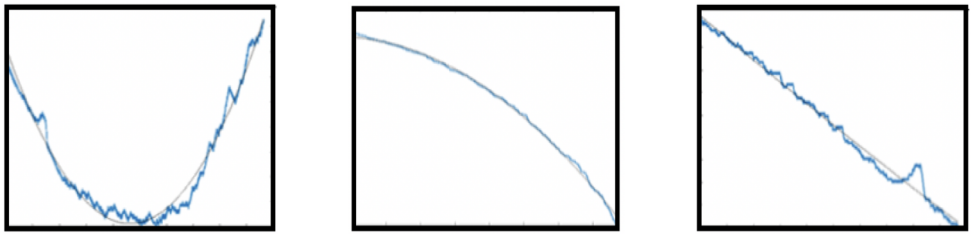


Figure 5.6: Polynomial curve fitting for the y-axis gradient (a) up-curved tail (b) down-curved tail (c) normal tail

To get more robust features and rise the system performance, the curvature angle of the body was found depending on the eye centre, the body centre, and the end point of the object body as shown in Figure 5.7.

Finding the first point, which is the centre of the eye (a_x, a_y) is found after thresholding the eye area, which is always black. After segmenting the eye blob, the centre of this blob was found depending on the blob area properties. The second point is the centre of mass of the embryo body (b_x, b_y) which is found using the

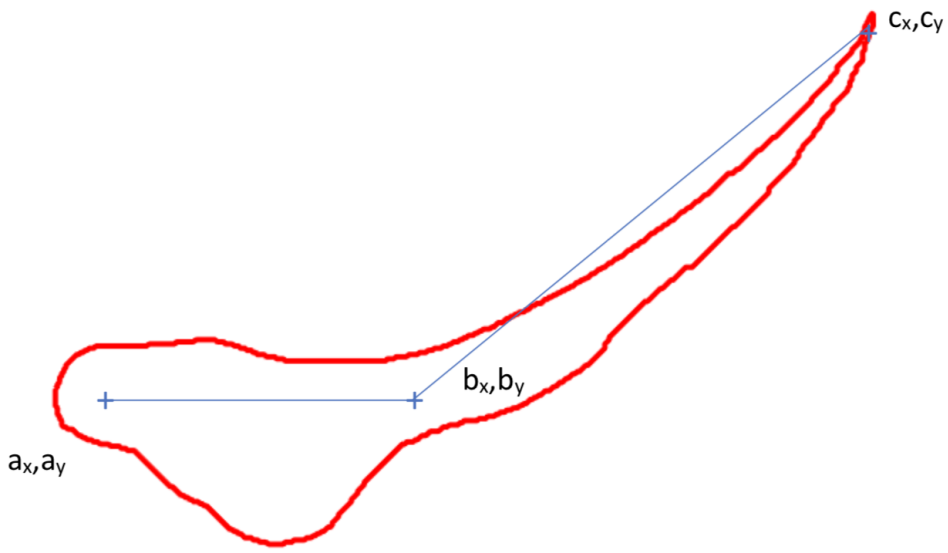


Figure 5.7: Curvature angle points

weighted position vectors for x and y direction for a binary image $I(x,y)$ with MxN size as shown in the following equations.

$$b_x = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} x.I(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N} I(x,y)} \tag{5.5}$$

$$b_y = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} y.I(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N} I(x,y)} \tag{5.6}$$

The third point presents the last point of the tail (c_x, c_y), which varies for the three classes (up-curved tail, down-curved tail, and non-curved). After finding the three points, the angle between them is found using the cosine rule with the line lengths between the three points, as shown in Figure 5.8.

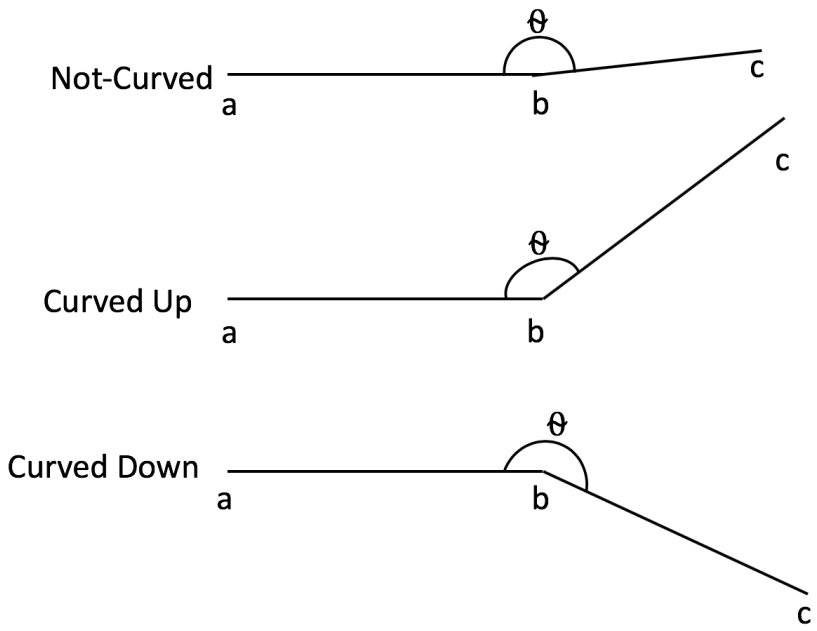


Figure 5.8: Curvature angle for different classes

To have features for all the malformation classes, the yolk should be detected, segmented, and identified if it is normal or deformed. The edema class is determined when the yolk becomes inflated, and the under eye region becomes bigger than the normal. The darkness of the yolk cells indicates the death of these cells, which is identified by necrosed class. The dead category happens when the whole body cells are dead with a dark colour appearance which is biologically identified as a non-survivor sample.

To segment the yolk region only, several mathematical and image processing operations are applied to detect and segment this region. The object corner points are recognised by finding a matrix of the object peaks depending on the extreme points of the object convex hull, as shown in Figure 5.9 and Figure 5.10. This property is one of the region properties that describe many shape

features of the desired object. The interesting point as shown in Figure 5.10 is A or B, and in this work A was used as a start point for yolk segmentation, which is the first point in the corners matrix.

Figure 5.9: The convex hull of the larva body

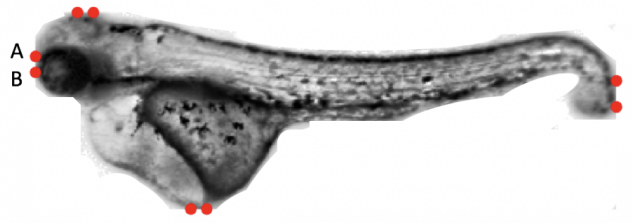


Figure 5.10: The extreme points of the body

Starting from A to the end of the image, the segmentation was carried out by segmenting the lower part of the body from point A to the end. In some cases, the eye is segmented with the yolk, which may affect the feature extraction process performance. To avoid any effect of the eye existing, the eye region was removed as shown in Figure 5.11 using the points in Figure 5.12 based on the vertical projection of the image which presents the number of pixels in each column of the image.

By noticing the point A is the minimum here, a vertical projection was found and plotted to find out the most suitable way

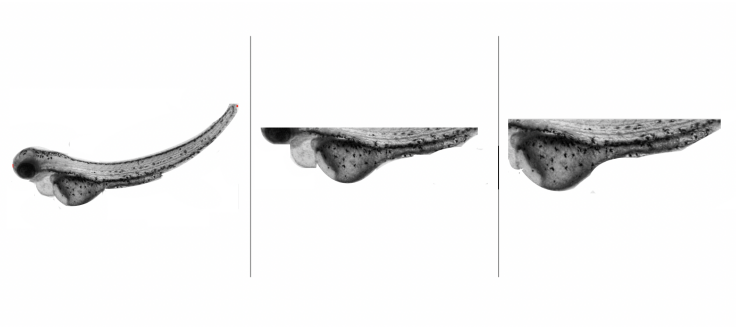


Figure 5.11: Example of the eye segmenting with the yolk

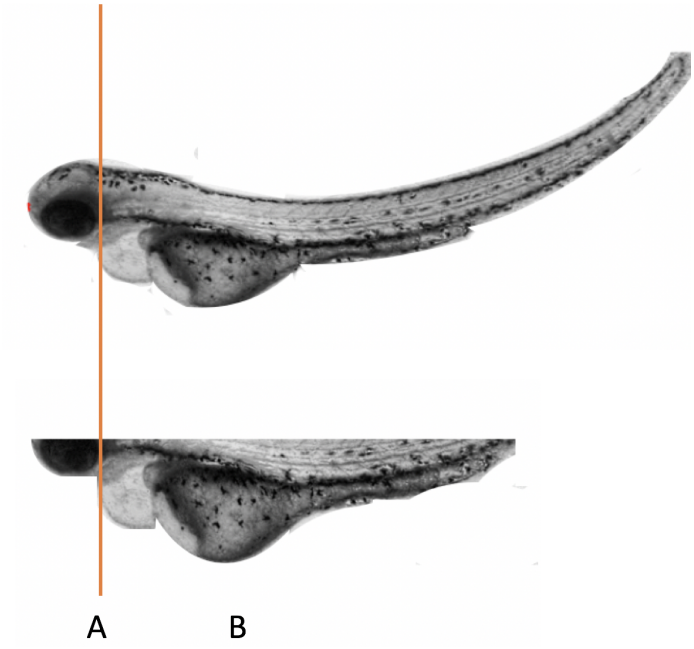


Figure 5.12: The segmented yolk

for head or eye removing. Figure 5.13 shows the vertical projection vector, which describes the image matrix rows. The signal of the projection was smoothed, then the minimum value (A) was found starting from the maximum point (B). By using point A, the head was removed, and the yolk sac region only was resulted. The area of the yolk region is one of the extracted features to identify the edema class.

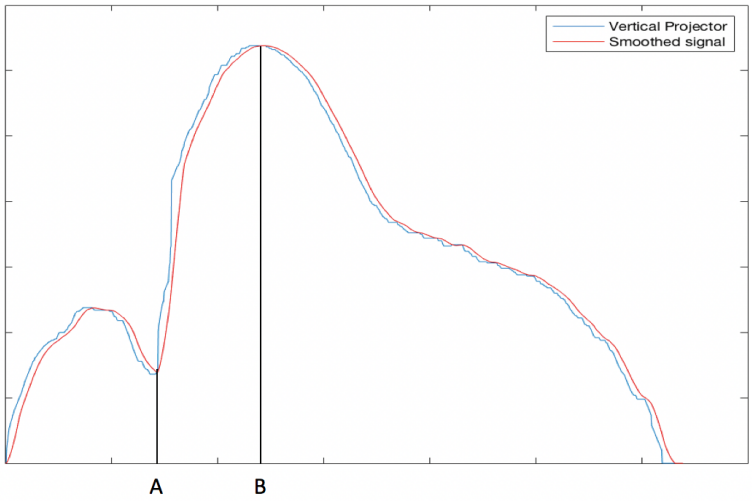


Figure 5.13: The vertical projection

To figure out if the sample has a necrosed yolk or not, the colour features for the segmented yolk were found. The necrosed yolk is identified by dead yolk cells when all yolk cells are dark. Two colour features were chosen and extracted, which are the mean intensity value for the non-zero double value and the standard deviation for the same target (yolk).

The hemostasis deformity happens when the head or the yolk has a red pigment on which is biologically identified as a defence process to prevent any loss of the blood in the injury status where the circulatory system prevents this from happening [161]. To detect this region, the CIELab colour space was used then the area for the segmented region was found and analysed which differentiate between the hemostasis and the other classes.

The CIELab colour space was utilised to identify the colours that are visualised by the human eyes. The first channel (*L) presents the illumination where the channel (*a) identified the val-

ues of red, magenta, and green. By computing the filled relative area of the (*a) channel and (*b) channel as shown in Figure 5.14, this feature will be able to judge if there is a red spot or not. Figure 5.15 shows images from hemostasis class with the detected red spots that were marked by the green colour.

To determine the survivor samples, which means ascertaining if the larva is alive or dead, the colour features such as the mean intensity value and the standard deviation value for the whole object region were calculated. The colour features are useful to identify the darkness of the embryo body. Furthermore, the number of pixels of the lowest grey levels showed differentiation between the dead and live samples.

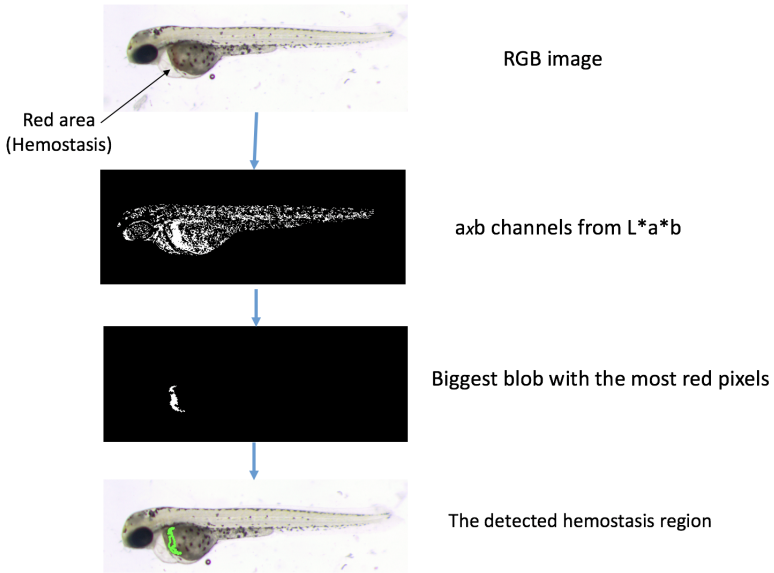


Figure 5.14: Red spots detection

Figure 5.16 shows the essential features for dead sample detection and depending on these values; three features were used for classification.

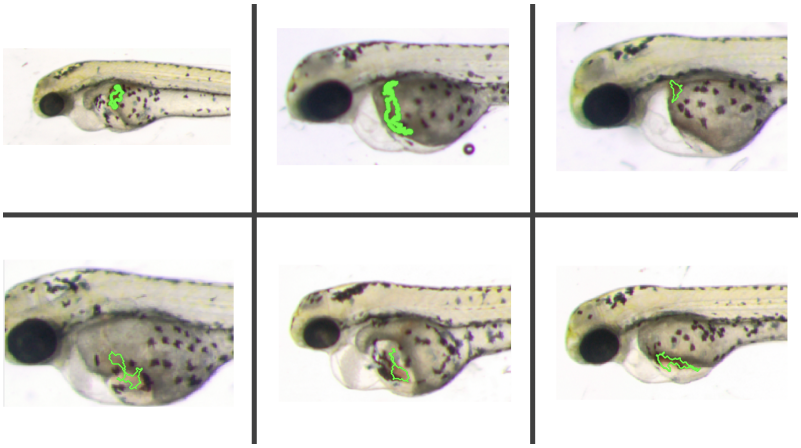


Figure 5.15: Hemostasis red spot detection

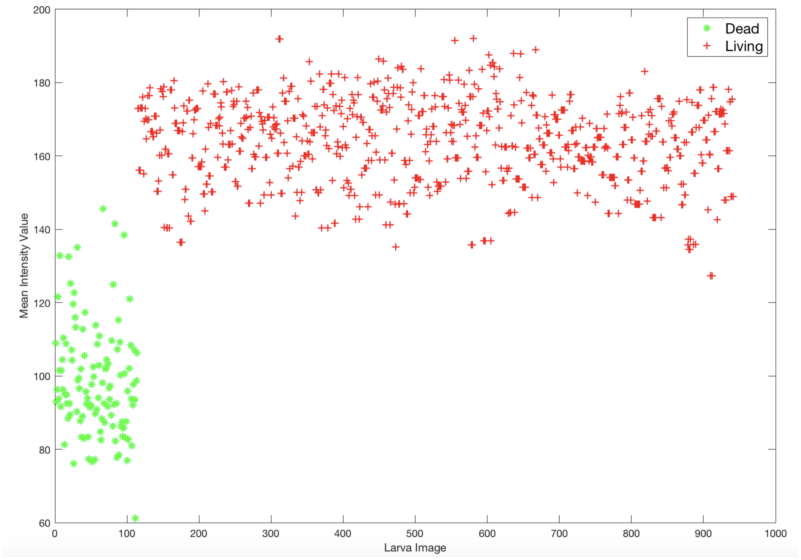


Figure 5.16: The most effective features for dead class

To recognise the shortage deformity which is presented by the loss in the posterior vertebrae of the zebrafish larva's spinal column which is caused by some stimulus [162]. To determine this deformation, two features were extracted. The first one is the length of the whole body for the horizontal view where the second one is the distance from the centre of eye to the endpoint of the tail using the Euclidean distance between two points (centre of the eye and

the last point of the tail). Figure 5.17 shows these features for short embryo and not-short ones.

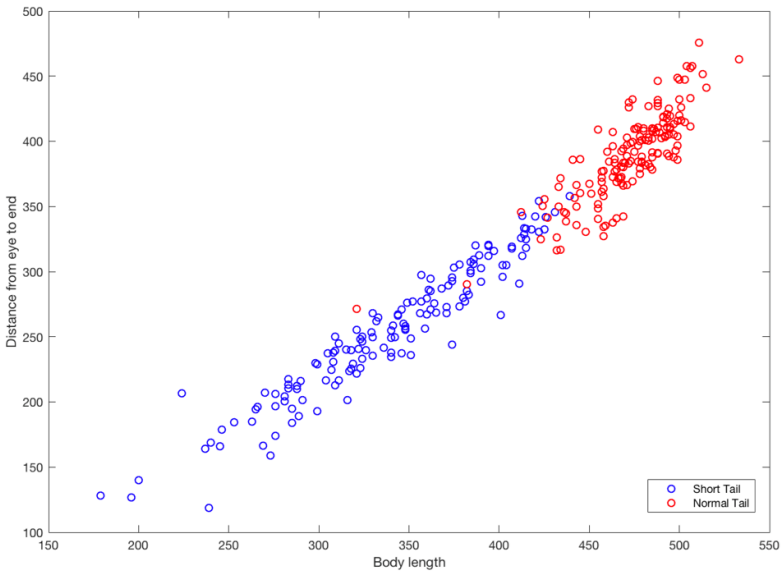


Figure 5.17: The extracted features for short class

5.4.3 Training classification models

Based on the extracted features, 8 CART models were built and trained using the 14 feature values. Each model presents a binary model to distinguish between the normal sample with the defective one. The goal is to detect and classify the zebrafish embryo sample status if it has normal or abnormal growth. Furthermore, the classification process here is a multi-label classification system; some of the added chemical substances cause deformities for different parts of the larva's body. The CART model is used according to its characteristics such as the compatibility with the small size of data set and the flexibility. The CART model was developed by Breiman in 1984 [163]. The main concept of the CART is the condition, several questions with type ***If-Then*** questions have been answered by the

model depending on the feature values.

The classification consists of two main steps, training and testing. The training stage depends on the binary relevance concept, which means training eight binary CART models using the extracted features. Each binary model has its feature vector (predictors) depending on each model label (response) in the testing step, where the classifier performance and capability will be determined. To predict the class of none labelled sample, the trained model follows the decisions in the tree from the root node down to the leaves. The last node (leaf) contains the response (label). The testing step for the new image starts with extracting the features which are provided to the saved classifier model to predict the class.

The CART models were trained using the k-fold cross-validation with ten folds, as shown in Figure 5.18. The cross-validation process is used to test the model prediction ability during the training stage to languish the overfitting problem and to get an indication about the model generalisation for the unknown data set. Cross-validation try to find the most accurate prediction performance depending on the fitness measures [164].

5.4.4 Testing the trained models

Once the tree models were built, the testing stage will start to find out the classification system accuracy and performance. The validation process was carried out during the training stage, and after validated the eight models, a set of images were provided for

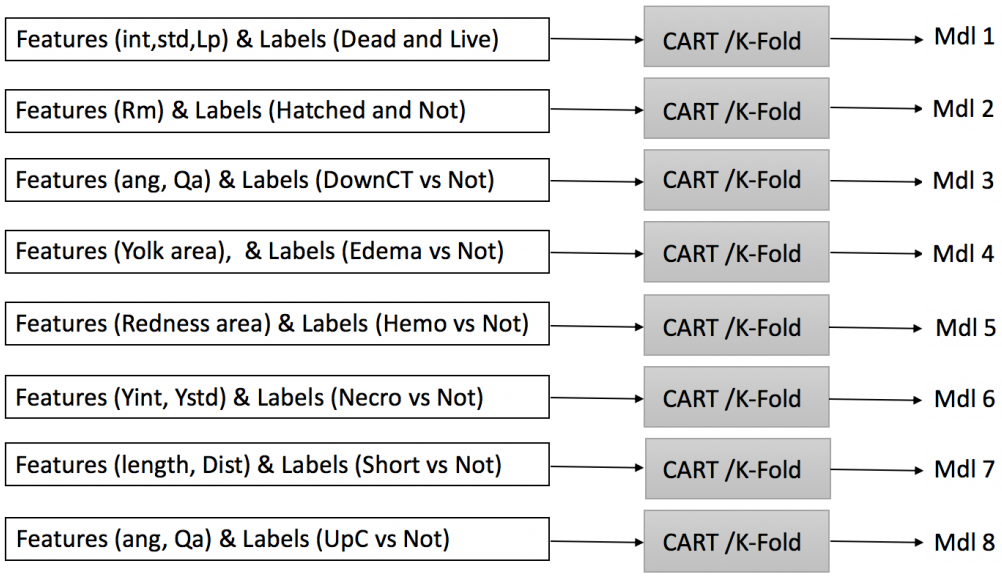


Figure 5.18: Training stage (Binary relevance)

multi-label testing. The testing process was carried out, as shown in Figure 5.19. The testing stage starts by providing the system by a new image. Using the CART model to test if the sample is live or dead. If the sample is live, then we need to identify if it is hatched or not. The live hatched sample will be examined using the seven models in a parallel way to detect and specify if it has any malformation(s).

5.5 Deep Learning

Deep learning algorithm represents one of the most recent algorithms of the machine learning approaches based on a set of learning layers. The significant difference between the deep learning algorithm and the other machine learning approaches is the feature extraction process which is the main step in the supervised and the

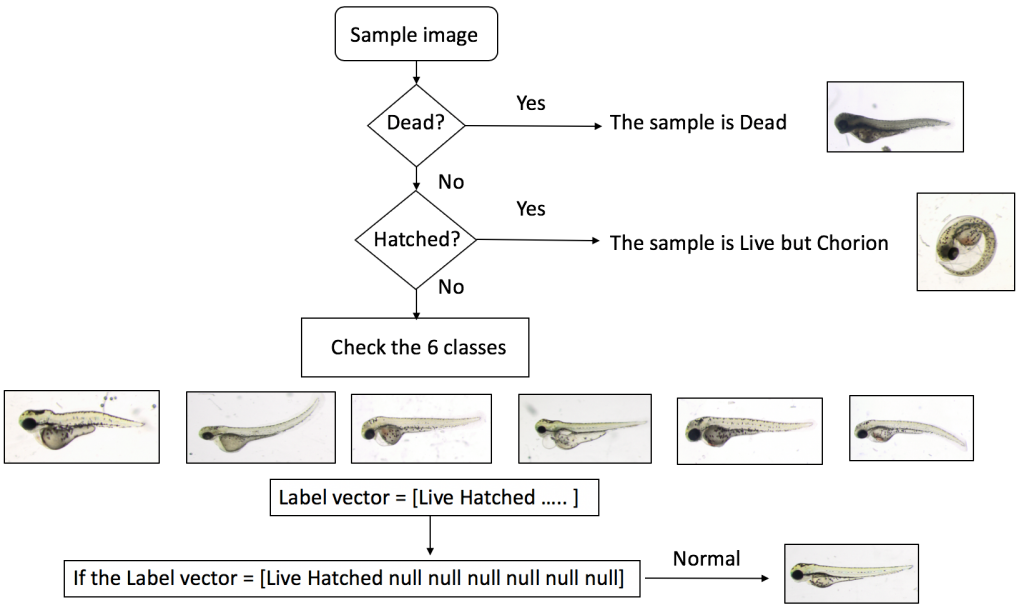


Figure 5.19: The testing stages

unsupervised classification.

The feature extraction process is a significant challenge in the machine learning algorithms where it might represent a particular data and not work with others. The feature extraction process is traditionally done manually and based on the human vision, and this is considered as a time-consuming process for most machine learning algorithms.

Recently, deep learning algorithm represents a high-active research field which becomes a high-performance tool for images classification. Deep learning has complex layers for the feature extraction process of the input images effectively. However, the ineffectual extracted features comparing with the manual feature extraction process based on the machine learning approaches where only particular information of the input data.

The most advanced machine learning algorithm represented by Convolutional Neural Networks (CNNs) has been widely used by researchers for image classification [138, 165, 166] . The proposed work aims to classify the nine classes of zebrafish larvae using two types of pretrained CNN with different characteristics and parameters.

Recently in 2018, one study was proposed for zebrafish larva classification based on deep learning [167]. This study used Jeanray et al. data set [13, 42] and classify them using a fine-tuned VGG 16 model. The result showed lower accuracy when the number of classes was increased due to the high similarity between the larva classes and the small size of the data set. The overlapping classes were not mentioned by the author where this data set has a multi-label problem and to deal with it as a multi-class problem the overlapped classes should be concerned.

5.5.1 CNN architecture

Two kinds of the pretrained convolutional neural networks are used, tested, and analysed using the data set of zebrafish larvae [13, 42]. Figures 5.20 and 5.21 show the general structure, and the main layers of the two used CNN. The first one is the GoogleNet with 144 layers and the second CNN is ResNet50 with 177 layers.

According to the small size of the data set, pretrained networks were used for image classification where training a CNN from

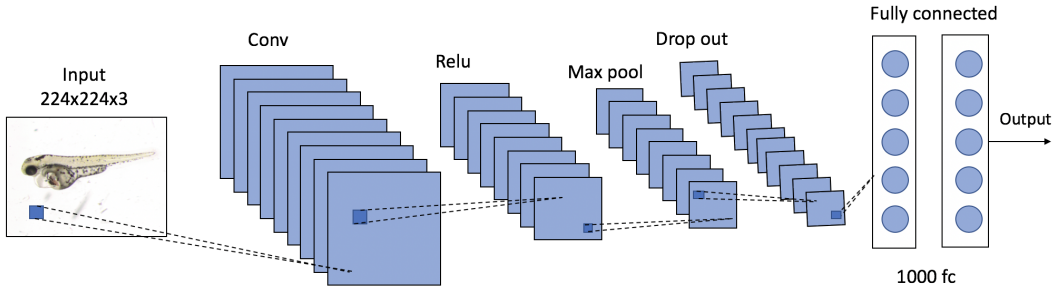


Figure 5.20: GoogleNet architecture

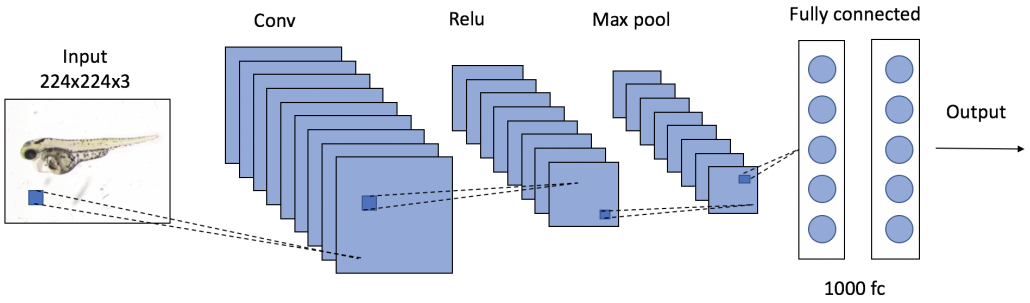


Figure 5.21: ResNet50 architecture

scratch successfully needs a huge data set to be trained efficiently and produce sufficient value of the network weights.

Using a pretrained CNN and reuse the network weights is defined as Transfer Learning (TL) process, and this helps the network to train faster than training from scratch. However, the pretrained networks extract low-level features such as edges, curves, and many other [168].

The deep networks generate a high level of features. However, the deeper network has a high training error and worse per-

formance, as shown in Figure 5.22. The main idea of the residual networks where the DesNet50 is one of this type of deep network is to use layers and skip other layers which is defined as identity shortcut connection [169].

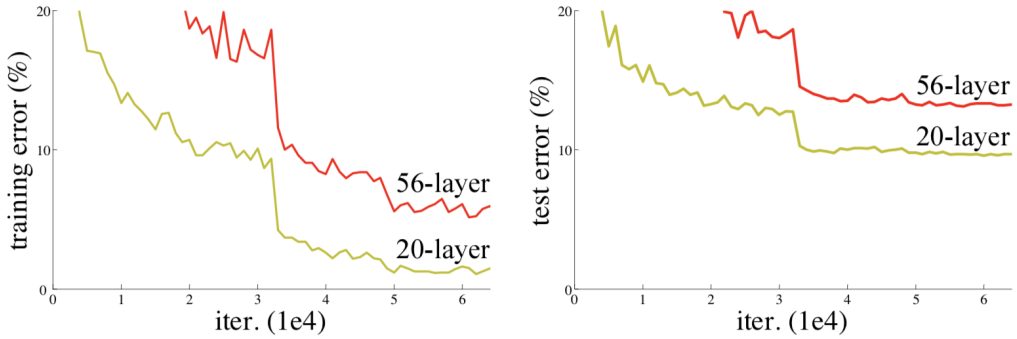


Figure 5.22: Training error and testing error for two network depths

5.5.2 Training and Validation

The used data set her is zebrafish larvae images with different types of malformations [13, 42]. The input images are re-sized to 224x224x3 pixels according to the pretrained CNNs parameters. And the data set was divided into a training set 70% and 30% for system evaluation.

First, a multi-class problem of the data set was addressed and implemented using the two types of CNN (ResNet50, GoogleNet). Where the images have overlapped labels, a multi-label classification system is proposed using the same CNN.

By using the pretrained CNNs, the same weights and the

same stack of layers were used according to the small size of data set and the weights reusing advantages. The most advantage of using the pretrained CNNs is the training time saving with lower complexity.

For multi-class classification, eight classifier models were trained and validated. Each model is a binary classifier model similar to the proposed CART model. For example, the dead larva images were used versus all the live larva images and the same with the rest of malformation types.

On the other hand, a multi label classifier is proposed to classify the samples that have more than one label. The categories were managed as shown in Table 5.2 besides the dead and chorion classes.

First, the pretrained Resnet50 was used, which comprises of 50 layers, convolution layers, max pool layers, fully connected layers, and output layer for feature extraction and classification [169]. The second attempt was carried out using the pretrained GoogleNet which comprises 22 layers, convolution layers, max pool layers, fully connected layers, dropout layers, and output layer [170] which are used for feature extraction and classification.

5.6 Results and Discussion

The multi-label system aims to detect and classify zebrafish embryo malformations and to name all the deformation types that appeared on each testing sample that has more than one malformation class.

Table 5.2: Multi label classes

Class Abbreviation	Classes
DENS	Down curved, Edema, Necrosed, Short tail
DNE	Down curved, Edema, Necrosed
DNEH	Down curved, Edema, Necrosed, Hemostasis
DNSEH	Down curved, Edema, Necrosed, Hemostasis, short tail
NE	Edema, Necrosed
NEH	Edema, Necrosed, Hemostasis
SNE	Short tail, Edema, Necrosed
SNEH	Short tail, Edema, Necrosed, Hemostasis
UNE	Up curved, Edema, Necrosed
UNES	Up curved, Edema, Necrosed, short tail
UN	Up curved, Necrosed
UNEH	Up curved, Edema, Necrosed, Hemostasis
UNH	Up curved, Necrosed, Hemostasis
UNHSE	Up curved, Edema, Necrosed, Hemostasis, short tail

The data set was publicly available for use and reproduction [42].

The images in the collected data set have a large size (2575×1932 pixels), which are down-sampled to quarter this size to be (644×483) pixels. This step has many advantages like accelerating the calculations and feature extraction process. This can be carried out using a typical CPU without need to get a new processor. Furthermore, the capturing process cannot always be with a high resolution; this has been considered as a challenge to get a high classification performance with low-resolution images.

5.6.1 The Performance Equations

To verify the system performance, a set of unknown images were provided to the system, and the system performance was analysed and found out. By using a set of unknown images for multi-label processing, the evaluation metrics should be calculated and registered using the true set of labels (T) and the predicted set of labels (P). The first metric is the Hamming loss which is identified as the fraction between the wrong labels and the total number of labels [171] for N of L labels instances which is found depending on the following equation

$$Hamming - Loss = \frac{1}{|N|.|L|} \sum_{i=1}^{|N|} \sum_{j=1}^{|L|} xor(y_{i,j}, z_{i,j}) \tag{5.7}$$

The second important metric is the exact match which find the percentage of samples or instances that have been correctly detected all the labels in the ground truth, which is done according to the following equation

$$Exact - Match = \frac{1}{L} \sum_{i=1}^{L} P_i = T_i \tag{5.8}$$

The micro averaging was used for the B binary problems of c labels (trees). The computation for TP, TN, FP, FN will be calculated depending on the micro averaging criteria as following:

$$m(B) = B\left(\sum_{i=1}^{c} TP(Y_i), \sum_{i=1}^{c} FP(Y_i), \sum_{i=1}^{c} TN(Y_i), \sum_{i=1}^{c} FN(Y_i) \right) \tag{5.9}$$

In the reported evaluation metrics and depending on the micro averaging formula, the recall, precision, the F score, and the

exact match were defined as follows [172]:

$$precision = \frac{\sum_{i=1}^{c} TP(Y_i)}{\sum_{i=1}^{c} TP(Y_i) + \sum_{i=1}^{c} FP(Y_i)} \tag{5.10}$$

$$Recall = \frac{\sum_{i=1}^{c} TP(Y_i)}{\sum_{i=1}^{c} TP(Y_i) + \sum_{i=1}^{c} FN(Y_i)} \tag{5.11}$$

$$Specificity = \frac{\sum_{i=1}^{c} TN(Y_i)}{\sum_{i=1}^{c} TN(Y_i) + \sum_{i=1}^{c} FP(Y_i)} \tag{5.12}$$

$$F - score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{5.13}$$

5.6.2 Results and discussion

From Table 5.3, it is noted that the system performance is favourable and promising. After several stages from segmentation to feature extraction and finally the multi-label classification, the result is satisfactory able to get the results as group of labels which can be considered as deep analysis of the samples. Furthermore, the system can be improved and modified by extracting more features. The following results are found for the multi-label processing and after merging the dead and chorion classes and after re-labelling the images by experts where the overlapped images with multiple classes were not mentioned in the collected data set.

The art of the feature extraction process is to extract the most critical features with the most intelligent technique which is presented here by specifying the ROI to avoid the time consuming during the training and testing processes. Furthermore, the ROI

Table 5.3: Multi-label evaluation

Metric	Value
Accuracy	93%
Precision	88%
Recall	85%
Specificity	82%
F-score	87%
Hamming Loss	20%
Exact Match	48%

segmentation offers a compelling and vital step according to the high similarity of the images of different classes.

The system reads the sample image, extracts the features, tests the image using the classification models and depending on the feature values. Finally, the result will be provided as a vector to describe the sample characteristics. Figure 5.23 shows an example of the system results.

Figure 5.23: Zebrafish larva with detected malformations

By relating to the testing accuracy in Table 5.4, we notice that the chorion class has 80% accuracy. In fact, the system succeeds to detect four images from five correctly, and the false classified image was identified as dead according to its colour features.

The low number of images in this class affected the accuracy value of this class. Also, some of the dead class images have a high similarity with the chorion, as shown in Fig. 5.24. To avoid classifying the dead as chorion, the input image has been tested if it is live or dead if it is dead no need for further testing. The chorionic class will be tested only when the sample is live.

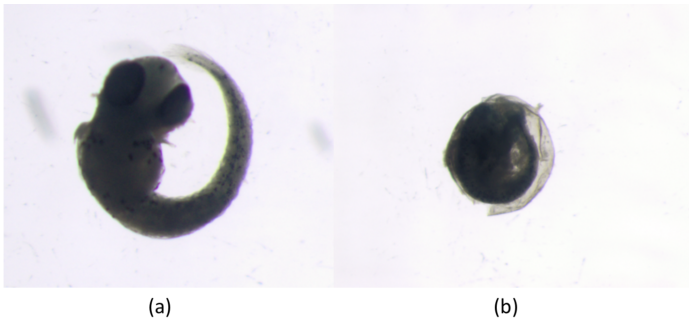


Figure 5.24: Misclassified: a) Dead larva b) Chorion larva

The difficulty of hemostasis detection is occurred because of the heart existing, and this may conflict the heart with the red spots, which are the abnormal blood cells. Figure 5.25 shows an example of larva classified wrongly as hemostasis that happened because of the heart existing.

Figure 5.25: Misclassified hemostasis sample

By comparing the proposed system accuracy with the existing one [13] we conclude that the efficiency was higher for most difficult classes such as Hemostasis class, Edema class, down-curved tail, and Up-curved fish class despite the distortion that occurred of the images because of the down-sampling process. The existing system used the ensembles of the decision tree algorithm that was presented in [173] and based on the pixel values. The method started with extracting 1000 subwindows, describing these subwindows by normalised raw pixel values for the three channels (RGB). These values for each subwindow were provided to the extremely randomised decision tree, and it was ensembled for all the extracted subwindows.

Table 5.4: Evaluation comparison of the proposed system with the existing one

Class	**Existing System (Binary-models) [13]**	**Proposed System (Binary-models)**	**Proposed System (Multi-label)**
Dead	99%	100%	87%
Chorion	90%	80%	80%
Short tail	89.9%	92%	92%
Down curved tail	82.68%	94%	98%
Necrosed yolk	95%	92%	84%
Edema	73.85%	96%	97%
Hemostasis	54.57%	94%	78%
Up curved Tail/Fish	72-80%	100%	100%
Normal	90%	100%	98%

The most challenging class to be detected and recognised is the hemostasis class then the edema class. In the proposed work, these two classes provide us with satisfactory results and better than the existing classification systems in spite of the detection dif-

ficulty and the high similarity between types.

The proposed system showed a high success rate in segmentation, region specifying, feature extraction, and the multi-label classification. The number of malformations that were analysed and studied was eight, and the number of classes was nine, including the non-defective embryos (Normal). These classes represent different types of deformity that effect on the body shape and the internal parts of the zebrafish embryos. Seeing the internal deformations related to larva body transparency. According to this feature, other classes of malformation can be added to these classes if the other parts of the zebrafish larva body are affected by adding chemical substances.

Two types of pretrained CNN were used to classify the zebrafish deformity types. The accuracy of the binary models and the multi-label classifier for both networks are shown in Table 5.5.

The multi-label performance is shown in Figure 5.26. As we noticed, this accuracy was affected by the small number of images in each category and the high number of classes. This may be improved by adding more images or using any augmentation algorithm to increase the data set size.

The obtained results from this system are promising where the proposed system used the pretrained CNN and reuse their weights without any pre-processing or augmentation of the data set images. Furthermore, this work does not require plenty of knowl-

edge about malformation characteristics and features.

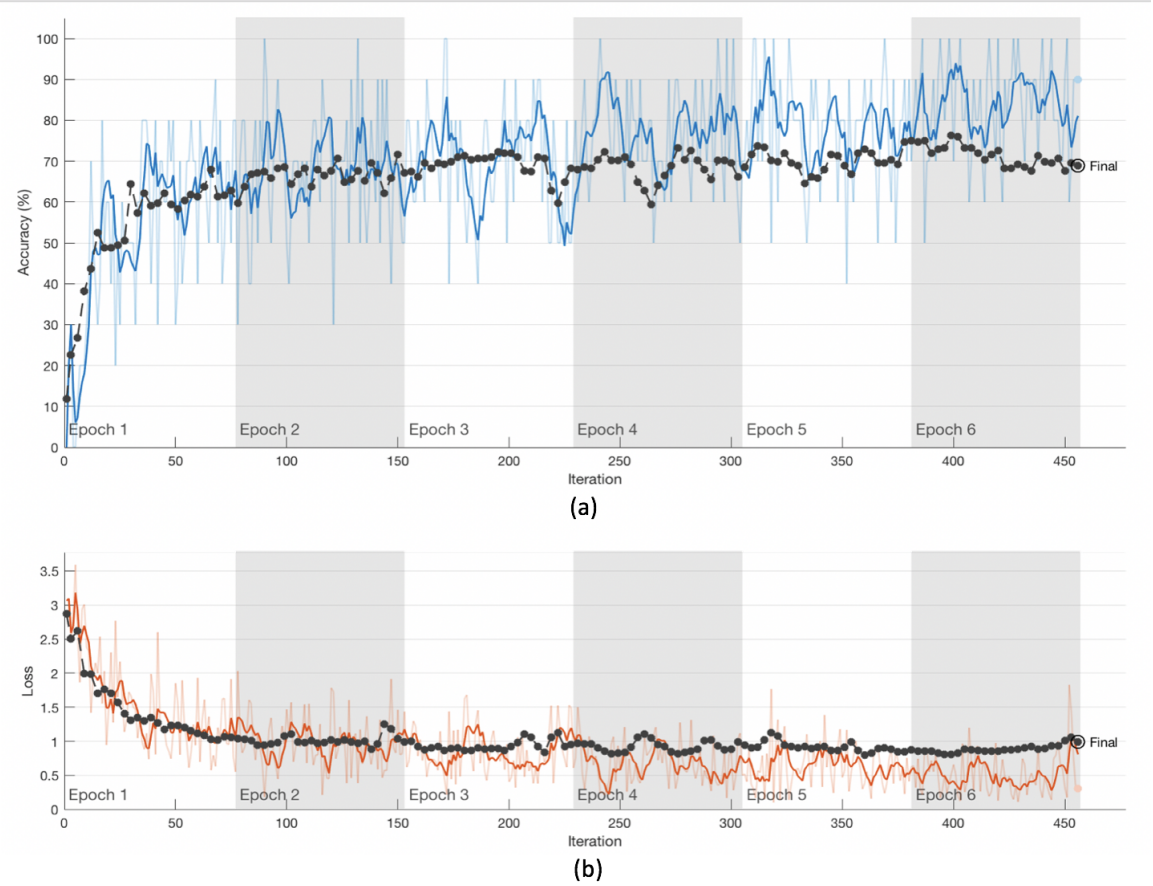


Figure 5.26: Multi-label classification accuracy using GoogleNet

Table 5.5: Testing accuracy of DL models

Classification model	**Accuracy (Resnet50)**	**Accuracy (GoogleNet)**
Dead vs All Live	97%	100%
Chorion vs All Hatched	100%	100%
Down vs Not Down	85%	99%
Up vs Not Up	98%	100%%
Edema vs Not Edema	90%	80%
Necrosed vs Not Necrosed	88%	97%
Short tail vs normal tail	91%	94%
Hemostasis vs Not Hemostasis	87%	89%
Multi Label classifier	59%	69%

5.7 Summary

The proposed systems set out to design and improve the automatic intelligent system to classify the zebrafish larva malformations using a multi-label classification system. One of the most important reasons for the proposed work is the lack of automatic detection and classification systems in this field. However, the attempt of one research group to have similar system has many challenges and limitations. Two methods are presented in this research, using binary relevance CART model and using pretrained CNNs.

The first method is CART models based on binary relevance. Extracting features using the ROI only instead of the whole image can affect the accuracy of the system. Obtaining the essential features (shape, colour intensity) make the proposed classification system more robust and increase the system's efficiency. This multi-label system shows a sufficient ability to identify the sample status and to mention all detected types when it has more than one de-

formation which can be more beneficial to the biologist to find out all the added chemicals effects.

The second method was done using two types of pretrained CNNs. The most significant challenge with this system is the small size of the data set and the high similarity between classes. Based on the CNNs, both ResNet50 and GoogleNet models show high performance with the multi-class problem, but due to the mentioned challenges the accuracy of the multi-label is inadequate.

By comparing the proposed two methods with each other and with the existed system, we noticed that the multi-class classification system accuracy using pretrained CNNs was higher than the CART models and the previously existed system [13]. However, the multi-label classification based on binary relevance CART has a considerable accuracy comparing with the pretrained CNNs performance. The existed system work was a multi-class system, and the multi-label has not been mentioned in their work. The data set images were re-labelled by biological experts.

Any field can get benefits from the engineering aspects, and biology science is one of these fields that success to employ machine learning to help them and solve their experiment problems. Furthermore, this system was widely accepted by the biologists to help them in their experiments.

The importance of this work is presented by the cooperative between two fields, using computer vision aspects to automate the animal screening experiments, to solve significant problems, and the

impact of this work has an accurate system to help the biologists in the screening processes. This work can be extended to cover more types of malformation.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this thesis, a set of successful classification, identification, and counting systems based on computer vision techniques are provided to help the biologists in their work and overcome the manual observation problems. The importance of this work comes from the biological experiments requirement of zebrafish larvae monitoring for hundreds of samples in a short time relating to the ethics and the fast growing of the embryos. The main generated conclusions from this work are summarised as follows.

1. The developed image analysis pipeline provides the biologists by a fully automatic monitoring system without any manual intervention using a high-performance detection, segmentation, and classification processes. However, the system faced a problem with the slide objects to the well sides. To overcome this limitation, a wider container has been used (petri dish) where

most samples sit on the plate centre. This system opens the door for better and bigger image analysis system if the number of classes can be increased. The advantage of this system over all proposed systems that all processes starting from capturing till having a status report of the embryos is going automatically with a low-cost tool and high-throughput. Also, this system covers the most common malformation types considering the nature of these embryos (orientation, position) that is fixed using an automatic fixing method by using the most important features for all classes depending on body shape, texture, and colour. The proposed system has two automatic parts: one for the data acquisition and the other for data analysis and result finding.

By using the proposed system, the biologist intervention is limited, and the experiment time should be shorter than usual. The proposed method is also set out to demonstrate a cost-effective way to carry out clinical relevant research so that it can be carried out in any lab with basic equipment (the scanner and computer are very affordable).

2. Where many drugs are tested on zebrafish embryos that are produced in large number within few hours, the manual detection process of these malformations is a tedious operation, and many embryos will be discarded without being analysed according to the European regulations.

   One of the most important reasons for the proposed project is the lack of automatic detection and classification systems in the

field. However, one research group proposed a similar system with many challenges and limitations. Those limitations were overcome in the multi-label classification system by extracting more essential features, focusing on all types of features not only intensity, using a high performance classifier, integrate the system of all malformation types where the embryo may have one, two, or three classes at the same time according to the kind of the chemical substance.

The multi-label CART system shows a higher performance than CNN. This might happen because of the extracted feature types. The CART model test specific features for each label where the deep learning extract the features for the whole image where the different classes have high similarity and the size of the data set is small.

3. Using the flatbed scanner is presented as a cost-effective imaging tool that saves the consumed time where the one-shot provides the system by hundreds of sample images. Furthermore, this tool affordable and easy to use by the biologists with the least imaging problems. Besides the benefits that are provided by using the flatbed scanner for data collection, the proposed capturing tool is assessed by the biologists as an effective and time-saving process for their experiments.

4. The proposed two counting systems provide the biologists by accurate, automatic, and cost-effective counting systems that minimise the human intervention and the animals harmful and pains that is happened during the manual counting process.

These counting systems received wide acceptance by the biologists to help them in their experiments.

5. The cooperation between two fields will provide a novel solution to have an automatic system that overcomes the biological experiment difficulties and provide the community with the benefits of the engineering techniques. The impact of the proposed analysis systems is presented by solving significant problems face the biologists in their work.

6.2 Future Work

Many directions are recommended to enhance the zebrafish larva analysis systems based on computer vision, which are presented in this thesis. The following points summaries some of these recommendations:

1. The image analysis pipeline for zebrafish egg/larva classification can be extended for more deformity types based on a low-cost tool (flatbed scanner).

2. The two counting systems can be developed into a smart phone application using phone camera settings. The proposed two systems can be used with other types of fish.

3. The proposed multi-label CNN for zebrafish larva classification can be developed using a different type of the deep learning

network or by tuning the network parameters.

4. The fish counting system can be improved using a different way for background subtraction such as the Low-Pass-Filter (LPF) technique for smoothing and noise reduction purposes.
5. Detect the most important features using another computer vision algorithm like scale-invariant feature transform (SIFT) which is a popular algorithm for local feature extraction.
6. Despite the cost-effective of the used imaging tool, but it can be replaced by a different type of imaging tools such as a high-resolution camera or a higher resolution flatbed scanner or 3D scanner. Furthermore, the well plate that has been used in the experiments can be replaced by a lower number of wells like the 24-well plate to minimise the side stick problem.
7. A binary relevance based on pre-trained CNN can be used to rise the multi-label classification performance based on data availability.

Bibliography

[1] E. Linney, L. Upchurch, and S. Donerly, "Zebrafish as a neurotoxicological model," Neurotoxicology and Teratology, vol. 26, no. 6, pp. 709–718, 2004.

[2] B. Canada, G. Thomas, K. Cheng, and J. Wang, "Automated segmentation and classification of zebrafish histology images for high-throughput phenotyping," in 2007 IEEE/NIH Life Science Systems and Applications Workshop. IEEE, 2007, pp. 245–248.

[3] G. Lieschke and P. Currie, "Animal models of human disease: zebrafish swim into view," Nature Reviews Genetics, vol. 8, no. 5, p. 353, 2007.

[4] R. Carvalho, J. de Sonneville, O. Stockhammer, N. Savage, W. Veneman, T. Ottenhoff, R. Dirks, A. Meijer, and H. Spaink, "A high-throughput screen for tuberculosis progression," PloS One, vol. 6, no. 2, p. e16779, 2011.

[5] J. Kanungo, S. Lantz, and M. Paule, "In vivo imaging and quantitative analysis of changes in axon length using transgenic zebrafish embryos," Neurotoxicology and teratology, vol. 33, no. 6, pp. 618–623, 2011.

[6] T. Liu, G. Li, J. Nie, A. Tarokh, X. Zhou, L. Guo, J. Malicki, W. Xia, and S. Wong, "An automated method for cell detec-

tion in zebrafish," Neuroinformatics, vol. 6, no. 1, pp. 5–21, 2008.

[7] I. Torjesen, "Number of animals used in science increased slightly in 2013, home office reports," 2014.

[8] E. Directive, "Directive 2010/63/eu of the european parliament and of the council of 22 september 2010 on the protection of animals used for scientific purposes," Official Journal of the European Union, vol. 20, p. L276, 2010.

[9] R. Alshut, J. Legradi, U. Liebel, L. Yang, van WJ, U. Strähle, R. Mikut, and M. Reischl, "Methods for automated high-throughput toxicity testing using zebrafish embryos," in Annual Conference on Artificial Intelligence. Springer, Karlsruhe, 2010, pp. 219–226.

[10] D. Arslanova, T. Yang, X. Xu, S. Wong, C. Augelli-Szafran, and W. Xia, "Phenotypic analysis of images of zebrafish treated with alzheimer's γ-secretase inhibitors," BMC biotechnology, vol. 10, no. 1, p. 24, 2010.

[11] A. Vogt, A. Cholewinski, X. Shen, S. Nelson, J. Lazo, M. Tsang, and N. Hukriede, "Automated image-based phenotypic analysis in zebrafish embryos," Developmental dynamics: an official publication of the American Association of Anatomists, vol. 238, no. 3, pp. 656–663, 2009.

[12] N. Jeanray, R. Marée, B. Pruvot, O. Stern, P. Geurts, L. Wehenkel, and M. Muller, "Phenotype classification of zebrafish embryos by supervised learning," Toxicology Letters, no. 211, pp. 18–23, 2012.

[13] ——, "Phenotype classification of zebrafish embryos by supervised learning," PloS One, vol. 10, no. 1, p. e0116989, 2015.

[14] A. Tharwat, T. Gaber, M. Fouad, V. Snasel, and A. Hassanien, "Towards an automated zebrafish-based toxicity test model using machine learning," Procedia Computer Science, vol. 65, pp. 643–651, 2015.

[15] R. Alshut, J. Legradi, L. Yang, U. Strähle, R. Mikut, and M. Reischl, "Robust identification of coagulated zebrafish eggs using image processing and classification techniques," in Proc. of the 19th GMA-FA 5.14 Computational Intelligence Workshop, 2009, pp. 2–4.

[16] F. Mosconi, T. Julou, N. Desprat, D. Sinha, J. Allemand, V. Croquette, and D. Bensimon, "Some nonlinear challenges in biology," Nonlinearity, vol. 21, no. 8, p. T131, 2008.

[17] P. Mazzarello, "A unifying concept: the history of cell theory," Nature Cell Biology, vol. 1, no. 1, p. E13, 1999.

[18] L. N. Magner, A history of the life sciences, revised and expanded. CRC Press, 2002.

[19] P. Vallance and T. Smart, "The future of pharmacology," British journal of pharmacology, vol. 147, no. S1, pp. S304–S307, 2006.

[20] S. TF, "What is toxicology.(cambridge toxicology group publication). boston: Cambridge toxicology group," 2008.

[21] W. Detrich III, M. Westerfield, and L. Zon, The zebrafish: disease models and chemical screens. Academic Press, 2011, vol. 3.

[22] L. Yang, N. Ho, R. Alshut, J. Legradi, C. Weiss, M. Reischl, R. Mikut, U. Liebel, F. Müller, and U. Strähle, "Zebrafish embryos as models for embryotoxic and teratological effects

of chemicals," Reproductive Toxicology, vol. 28, no. 2, pp. 245–253, 2009.

[23] O. Bhusnure, J. Mane, S. Gholve, S. Thonte, P. Giram, and J. Sangshetti, "Drug target screening and its validation by zebrafish as a novel tool," Pharm. Anal. Acta, vol. 6, no. 426, pp. 10–4172, 2015.

[24] C. Chakraborty, C. Hsu, Z. Wen, C. Lin, and G. Agoramoorthy, "Zebrafish: a complete animal model for in vivo drug discovery and development," Current Drug Metabolism, vol. 10, no. 2, pp. 116–124, 2009.

[25] k. Hisaoka and C. Firlit, "Ovarian cycle and egg production in the zebrafish, brachydanio rerio," Copeia, vol. 1962, no. 4, pp. 788–792, 1962.

[26] R. Froese and D. Pauly, "Danio rerio," FISHBASE. Paris: WorldFish Center, pp. 376–379, 2007.

[27] C. Kimmel, "Genetics and early development of zebrafish," Trends in Genetics, vol. 5, pp. 283–288, 1989.

[28] A. Hill, H. Teraoka, W. Heideman, and R. Peterson, "Zebrafish as a model vertebrate for investigating chemical toxicity," Toxicological sciences, vol. 86, no. 1, pp. 6–19, 2005.

[29] Q. Al-Jubouri, "Automatic computer vision systems for aquatic research," Ph.D. dissertation, University of Liverpool, 2017.

[30] P. McGrath and C. Li, "Zebrafish: a predictive model for assessing drug-induced toxicity," Drug Discovery Today, vol. 13, no. 9-10, pp. 394–401, 2008.

[31] E. Loucks and S. Ahlgren, "Assessing teratogenic changes in a zebrafish model of fetal alcohol exposure," Journal of Visualized Experiments: JoVE, no. 61, 2012.

[32] M. Carvan III, E. Loucks, D. Weber, and F. Williams, "Ethanol effects on the developing zebrafish: neurobehavior and skeletal morphogenesis," Neurotoxicology and Teratology, vol. 26, no. 6, pp. 757–768, 2004.

[33] A. Amsterdam, K. Sadler, K. Lai, S. Farrington, R. Bronson, J. Lees, and N. Hopkins, "Many ribosomal protein genes are cancer genes in zebrafish," PLoS Biology, vol. 2, no. 5, p. e139, 2004.

[34] M. Newman, G. Verdile, R. Martins, and M. Lardelli, "Zebrafish as a tool in alzheimer's disease research," Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, vol. 1812, no. 3, pp. 346–352, 2011.

[35] Z. Sun, A. Amsterdam, G. Pazour, D. Cole, M. Miller, and N. Hopkins, "A genetic screen in zebrafish identifies cilia genes as a principal cause of cystic kidney," Development, vol. 131, no. 16, pp. 4085–4093, 2004.

[36] P. Henion, D. Raible, C. Beattie, K. Stoesser, J. Weston, and J. Eisen, "Screen for mutations affecting development of zebrafish neural crest," Developmental genetics, vol. 18, no. 1, pp. 11–17, 1996.

[37] S. Neuhauss, L. Solnica-Krezel, A. Schier, F. Zwartkruis, D. Stemple, J. Malicki, S. Abdelilah, D. Stainier, and W. Driever, "Mutations affecting craniofacial development in zebrafish," Development, vol. 123, no. 1, pp. 357–367, 1996.

[38] S. Wilson, L. Ross, T. Parrett, and S. Easter, "The development of a simple scaffold of axon tracts in the brain of the embryonic zebrafish, brachydanio rerio," Development, vol. 108, no. 1, pp. 121–145, 1990.

[39] I. Masai, Z. Lele, M. Yamaguchi, K. Komori, A. Nakata, Y. Nishiwaki, H. Wada, H. Tanaka, Y. Nojima, M. Hammerschmidt, S. Wilson, and H. Okamoto, "N-cadherin mediates retinal lamination, maintenance of forebrain compartments and patterning of retinal neurites," Development, vol. 130, no. 11, pp. 2479–2494, 2003.

[40] C. Nusslein-Volhard and R. Dahm, Zebrafish. Oxford University Press, 2002.

[41] H. Teraoka, W. Dong, and T. Hiraga, "Zebrafish as a novel experimental model for developmental toxicology," Congenital Anomalies, vol. 43, no. 2, pp. 123–132, 2003.

[42] N. Jeanray, R. Marée, B. Pruvot, O. Stern, P. Geurts, L. Wehenkel, and M. Muller, "Data from: Phenotype classification of zebrafish embryos by supervised learning," 2015. [Online]. Available: https://doi.org/10.5061/dryad.23d30

[43] T. Braunbeck, M. Böttcher, H. Hollert, T. Kosmehl, E. Lammer, E. Leist, M. Rudolf, and N. Seitz, "Towards an alternative for the acute fish lc50 test in chemical assessment: the fish embryo toxicity test goes multi-species-an update," ALTEX-Alternatives to animal experimentation, vol. 22, no. 2, pp. 87–102, 2005.

[44] U. Strähle, S. Scholz, R. Geisler, P. Greiner, H. Hollert, S. Rastegar, A. Schumacher, I. Selderslaghs, C. Weiss, H. Wit-

ters et al., "Zebrafish embryos as an alternative to animal experiments—a commentary on the definition of the onset of protected life stages in animal welfare regulations," Reproductive Toxicology, vol. 33, no. 2, pp. 128–132, 2012.

[45] S. Weigt, N. Huebler, T. Braunbeck, F. von Landenberg, and T. Broschard, "Zebrafish teratogenicity test with metabolic activation (mdart): effects of phase i activation of acetaminophen on zebrafish danio rerio embryos," Toxicology, vol. 275, no. 1-3, pp. 36–49, 2010.

[46] N. Sipes, S. Padilla, and T. Knudsen, "Zebrafish—as an integrative model for twenty-first century toxicity testing," Birth Defects Research Part C: Embryo Today: Reviews, vol. 93, no. 3, pp. 256–267, 2011.

[47] R. Creton, "Automated analysis of behavior in zebrafish larvae," Behavioural Brain Research, vol. 203, no. 1, pp. 127–136, 2009.

[48] C. Dlugos, S. Brown, and R. Rabin, "Gender differences in ethanol-induced behavioral sensitivity in zebrafish," Alcohol, vol. 45, no. 1, pp. 11–18, 2011.

[49] N. Abaid, C. Spinello, J. Laut, and M. Porfiri, "Zebrafish (danio rerio) responds to images animated by mathematical models of animal grouping," Behavioural Brain Research, vol. 232, no. 2, pp. 406–410, 2012.

[50] S. Chen, Y. Zhu, W. Xia, S. Xia, and X. Xu, "Automated analysis of zebrafish images for phenotypic changes in drug discovery," Journal of Neuroscience Methods, vol. 200, no. 2, pp. 229–236, 2011.

[51] M. Schutera, T. Dickmeis, M. Mione, R. Peravali, D. Marcato, M. Reischl, R. Mikut, and C. Pylatiuk, "Automated phenotype pattern recognition of zebrafish for high-throughput screening," Bioengineered, vol. 7, no. 4, pp. 261–265, 2016.

[52] B. Zion, N. Doitch, V. Ostrovsky, V. Alchanatis, R. Segev, A. Barki, and I. Karplus, "Ornamental fish fry counting by image processing," Agricultural Research Organization, Bet Dagan, 2006.

[53] R. Labuguen, E. Volante, A. Causo, R. Bayot, G. Peren, R. Macaraig, N. Libatique, and G. Tangonan, "Automated fish fry counting and schooling behaviour analysis using computer vision," in 2012 IEEE 8th International Colloquium on Signal Processing and its Applications. Melaka, 2012, pp. 255–260.

[54] W. Spomer, A. Pfriem, R. Alshut, S. Just, and C. Pylatiuk, "High-throughput screening of zebrafish embryos using automated heart detection and imaging," Journal of laboratory automation, vol. 17, no. 6, pp. 435–442, 2012.

[55] J. Taylor, C. Saunter, B. Chaudhry, D. Henderson, G. Love, and J. Girkin, "Heart synchronization for spim microscopy of living zebra fish," in Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XVIII, vol. 7904. International Society for Optics and Photonics, 2011, p. 79040B.

[56] C. Pylatiuk, D. Sanchez, R. Mikut, R. Alshut, M. Reischl, S. Hirth, W. Rottbauer, and S. Just, "Automatic zebrafish heartbeat detection and analysis for zebrafish embryos," Zebrafish, vol. 11, no. 4, pp. 379–383, 2014.

[57] N. Strachan, "Length measurement of fish by computer vision," Computers and electronics in agriculture, vol. 8, no. 2, pp. 93–104, 1993.

[58] D. White, C. Svellingen, and N. Strachan, "Automated measurement of species and length of fish by computer vision," Fisheries Research, vol. 80, no. 2-3, pp. 203–210, 2006.

[59] N. R. C. U. C. to Review Risk Management in the DOE's Environmental Remediation Program and N. R. C. U. C. to Review Risk Management in the DOE's Environmental Remediation Program, Building consensus through risk assessment and management of the Department of Energy's environmental remediation program. National Academies, 1994.

[60] A. Avdesh, M. Chen, M. Martin-Iverson, A. Mondal, D. Ong, S. Rainey-Smith, K. Taddei, M. Lardelli, D. Groth, G. Verdile et al., "Regular care and maintenance of a zebrafish (danio rerio) laboratory: an introduction," JoVE (Journal of Visualized Experiments), no. 69, p. e4196, 2012.

[61] S. Kalasekar, E. Zacharia, N. Kessler, N. Ducharme, J. Gustafsson, I. Kakadiaris, and M. Bondesson, "Identification of environmental chemicals that induce yolk malabsorption in zebrafish using automated image segmentation," Reproductive Toxicology, vol. 55, pp. 20–29, 2015.

[62] R. Cipolla, S. Battiato, and G. Farinella, Computer Vision: Detection, recognition and reconstruction. Springer, 2010, vol. 285.

[63] D. Ballard and C. Brown, "Computer vision. 1982," Prenice-Hall, Englewood Cliffs, NJ, 1982.

[64] M. Sonka, V. Hlavac, and R. Boyle, Image processing, analysis, and machine vision. Cengage Learning, 2014.

[65] É. Puybareau, D. Genest, E. Barbeau, M. Léonard, and H. Talbot, "An automated assay for the assessment of cardiac arrest in fish embryo," Computers in Biology and Medicine, vol. 81, pp. 32–44, 2017.

[66] S. Berghmans, J. Hunt, A. Roach, and P. Goldsmith, "Zebrafish offer the potential for a primary screen to identify a wide variety of potential anticonvulsants," Epilepsy research, vol. 75, no. 1, pp. 18–28, 2007.

[67] S. Bhat and M. Liebling, "Cardiac tissue and erythrocyte separation in bright-field microscopy images of the embryonic zebrafish heart for motion estimation," in 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, 2009, pp. 746–749.

[68] P. Chan, C. Lin, and S. Cheng, "Noninvasive technique for measurement of heartbeat regularity in zebrafish (danio rerio) embryos," BMC biotechnology, vol. 9, no. 1, p. 11, 2009.

[69] J. Cachat, A. Stewart, E. Utterback, P. Hart, S. Gaikwad, K. Wong, E. Kyzar, N. Wu, and A. Kalueff, "Three-dimensional neurophenotyping of adult zebrafish behavior," PloS one, vol. 6, no. 3, p. e17597, 2011.

[70] C. Cario, T. Farrell, C. Milanese, and E. Burton, "Automated measurement of zebrafish larval movement," The Journal of physiology, vol. 589, no. 15, pp. 3703–3708, 2011.

[71] S. Kato, T. Nakagawa, M. Ohkawa, K. Muramoto, O. Oyama, A. Watanabe, H. Nakashima, T. Nemoto, and K. Sugitani, "A

computer image processing system for quantification of zebrafish behavior," Journal of neuroscience methods, vol. 134, no. 1, pp. 1–7, 2004.

[72] T. Liu, J. Lu, Y. Wang, W. Campbell, L. Huang, J. Zhu, W. Xia, and S. Wong, "Computerized image analysis for quantitative neuronal phenotyping in zebrafish," Journal of neuroscience methods, vol. 153, no. 2, pp. 190–202, 2006.

[73] R. Liu, S. Lin, R. Rallo, Y. Zhao, R. Damoiseaux, T. Xia, S. Lin, A. Nel, and Y. Cohen, "Automated phenotype recognition for zebrafish embryo based in vivo high throughput toxicity screening of engineered nano-materials," PloS one, vol. 7, no. 4, p. e35014, 2012.

[74] D. Mandrell, L. Truong, C. Jephson, M. Sarker, A. Moore, C. Lang, M. Simonich, and R. Tanguay, "Automated zebrafish chorion removal and single embryo placement: optimizing throughput of zebrafish developmental toxicity screens," Journal of laboratory automation, vol. 17, no. 1, pp. 66–74, 2012.

[75] R. Peravali, J. Gehrig, S. Giselbrecht, D. Lütjohann, Y. Hadzhiev, F. Müller, and U. Liebel, "Automated feature detection and imaging for high-resolution screening of zebrafish embryos," Biotechniques, vol. 50, no. 5, pp. 319–324, 2011.

[76] M. Saydmohammed, L. Vollmer, E. Onuoha, A. Vogt, and M. Tsang, "A high-content screening assay in transgenic zebrafish identifies two novel activators of fgf signaling," Birth Defects Research Part C: Embryo Today: Reviews, vol. 93, no. 3, pp. 281–287, 2011.

[77] X. Xu, X. Xu, X. Huang, W. Xia, and S. Xia, "A high-

throughput analysis method to detect regions of interest and quantify zebrafish embryo images," Journal of biomolecular screening, vol. 15, no. 9, pp. 1152–1159, 2010.

[78] C. Zanella, M. Campana, B. Rizzi, C. Melani, G. Sanguinetti, P. Bourgine, K. Mikula, N. Peyriéras, and A. Sarti, "Cells segmentation from 3-d confocal images of early zebrafish embryogenesis," IEEE transactions on Image Processing, vol. 19, no. 3, pp. 770–781, 2009.

[79] B. Al-Saaidah, W. Al-Nuaimy, M. Al-Hadidi, and I. Young, "Zebrafish larvae classification based on decision tree model: A comparative analysis," Advances in Science, Technology and Engineering Systems Journal, vol. 3, no. 4, pp. 347–353, 2018.

[80] B. Al-Saaidah, W. Al-Nuaimy, M. Al-Taee, I. Young, and Q. Al-Jubouri, "Identification of tail curvature malformation in zebrafish embryos," in 2017 8th International Conference on Information Technology (ICIT). IEEE, 2017, pp. 588–593.

[81] T. Carron and P. Lambert, "Color edge detector using jointly hue, saturation and intensity," in Proceedings of 1st International Conference on Image Processing, vol. 3. IEEE, Austin, 1994, pp. 977–981.

[82] A. Hanbury, "The taming of the hue, saturation and brightness colour space," in Proceedings of the Seventh Computer Vision Winter Workshop, Bad Aussee, Austria. Citeseer, 2002.

[83] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the hsv color space for image re-

trieval," in Proceedings. International Conference on Image Processing, vol. 2. IEEE, 2002, pp. II–II.

[84] R. Gonzalez and P. Wintz, "Digital image processing(book)," Reading, Mass., Addison-Wesley Publishing Co., Inc. Applied Mathematics and Computation, no. 13, p. 451, 1977.

[85] F. Woelk, I. Schiller, and R. Koch, "An airborne bayesian color tracking system," in IEEE Proceedings. Intelligent Vehicles Symposium, 2005. IEEE, 2005, pp. 67–72.

[86] S. James, "Face image retrieval with hsv colour space using clustering techniques," The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), vol. 1, no. 1, pp. 17–20, 2013.

[87] A. Robertson, "The cie 1976 color-difference formulae," Color Research & Application, vol. 2, no. 1, pp. 7–11, 1977.

[88] J. Helmke, M. Schulz, and H. Bauch, "Sediment-color record from the northeast atlantic reveals patterns of millennial-scale climate variability during the past 500,000 years," Quaternary Research, vol. 57, no. 1, pp. 49–57, 2002.

[89] T. Cavanaugh, "Applications of spectrophotometry for paleoclimate interpretations from lacustrine sediment records," Ph.D. dissertation, Middlebury College, 2017.

[90] O. Liew, P. Chong, B. Li, and A. Asundi, "Signature optical cues: emerging technologies for monitoring plant health," Sensors, vol. 8, no. 5, pp. 3205–3239, 2008.

[91] E. Alpaydin, Introduction to Machine Learning. MIT press, 2014.

[92] L. Shapiro, "Stockman., gc (2001)," Computer Vision, pp. 279–325.

[93] K. Anding, G. Polte, D. Garten, and V. Musalimov, "Out-of-focus region segmentation of 2d surface images with the use of texture features," Naučno-tehničeskij Vestnik Informacionnyh Tehnologij, vol. 15, no. 5, pp. 796–802, 2015.

[94] R. Haralick, K. Shanmugam et al., "Textural features for image classification," IEEE Transactions on systems, man, and cybernetics, no. 6, pp. 610–621, 1973.

[95] D. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," Canadian Journal of Remote Sensing, vol. 28, no. 1, pp. 45–62, 2002.

[96] L. Shapiro and G. Stockman, "5, 7, 10," Computer Vision. Upper Saddle River, New Jersey: Prentice-Hall, Inc, pp. 157–158, 2001.

[97] V. Leavers, "Use of the radon transform as a method of extracting information about shape in two dimensions," Image and vision computing, vol. 10, no. 2, pp. 99–107, 1992.

[98] P. Hough, "Method and means for recognizing complex patterns," Dec. 18 1962, uS Patent 3,069,654.

[99] R. Duda and P. Hart, "Use of the hough transformation to detect lines and curves in pictures," SRI international Menlo park ca artificial intelligence centre, Tech. Rep., 1971.

[100] E. Achtert, C. Böhm, J. David, P. Kröger, and A. Zimek, "Global correlation clustering based on the hough transform," Statistical Analysis and Data Mining: The ASA Data Science Journal, vol. 1, no. 3, pp. 111–127, 2008.

[101] H. Liu, Y. Qian, and S. Lin, "Detecting persons using hough circle transform in surveillance video." in International Conference on Computer Vision Theory and Applications (2), 2010, pp. 267–270.

[102] H. Yuen, J. Princen, J. Illingworth, and J. Kittler, "Comparative study of hough transform methods for circle finding," Image and Vision Computing, vol. 8, no. 1, pp. 71–77, 1990.

[103] P. Assheton and A. Hunter, "A shape-based voting algorithm for pedestrian detection and tracking," Pattern Recognition, vol. 44, no. 5, pp. 1106–1120, 2011.

[104] R. Satzoda, S. Sathyanarayana, T. Srikanthan, and S. Sathyanarayana, "Hierarchical additive hough transform for lane detection," IEEE Embedded Systems Letters, vol. 2, no. 2, pp. 23–26, 2010.

[105] A. Bajwa and R. Kaur, "Fast lane detection using improved hough transform," Journal of Computing Technologies, vol. 2, no. 5, pp. 10–13, 2013.

[106] L. De Marchi, E. Baravelli, G. Cera, N. Speciale, and A. Marzani, "Warped wigner-hough transform for defect reflection enhancement in ultrasonic guided wave monitoring," Mathematical Problems in Engineering, vol. 2012, 2012.

[107] M. Mirmehdi, G. West, and G. Dowling, "Label inspection using the hough transform on transputer networks," Microprocessors and Microsystems, vol. 15, no. 3, pp. 167–173, 1991.

[108] Y. Sun and B. Nelson, "Biological cell injection using an autonomous microrobotic system," The International Journal of Robotics Research, vol. 21, no. 10-11, pp. 861–868, 2002.

[109] C. Melani, M. Campana, B. Lombardot, B. Rizzi, F. Veronesi, C. Zanella, P. Bourgine, K. Mikula, N. Peyriéras, and A. Sarti, "Cells tracking in a live zebrafish embryo," in 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2007, pp. 1631–1634.

[110] L. Rokach and O. Maimon, Data mining with decision trees: theory and applications. World Scientific, 2008, vol. 69.

[111] A. Painsky and S. Rosset, "Cross-validated variable selection in tree-based methods improves predictive performance," IEEE Transactions on Pattern Analysis and machine intelligence, vol. 39, no. 11, pp. 2142–2153, 2017.

[112] A. Khan, B. Baharudin, L. Lee, and K. Khan, "A review of machine learning algorithms for text-documents classification," Journal of Advances in Information Technology, vol. 1, no. 1, pp. 4–20, 2010.

[113] R. Greiner and J. Schaffer, "Aixploratorium-decision trees," University of Alberta ,Edmonton, 2001.

[114] G. James, D. Witten, T. Hastie, and R. Tibshirani, An introduction to statistical learning. Springer, 2013, vol. 112.

[115] L. Kou, G. Markowsky, and L. Berman, "A fast algorithm for steiner trees," Acta Informatica, vol. 15, no. 2, pp. 141–145, 1981.

[116] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees. wadsworth int," Group, vol. 37, no. 15, pp. 237–251, 1984.

[117] J. Powell, S. Krotosky, B. Ochoa, D. Checkley, and P. Cosman, "Detection and identification of sardine eggs at sea

using a machine vision system," in Oceans 2003. Celebrating the Past Teaming Toward the Future (IEEE Cat. No. 03CH37492), vol. 1. IEEE, 2003, p. 175.

[118] J. Amrhein, C. Stow, and C. Wible, "Whole-fish versus filet polychlorinated-biphenyl concentrations: An analysis using classification and regression tree models," Environmental Toxicology and Chemistry, vol. 18, no. 8, pp. 1817–1823, 1999.

[119] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, p. 436, 2015.

[120] L. Deng, D. Yu et al., "Deep learning: methods and applications," Foundations and Trends in Signal Processing, vol. 7, no. 3–4, pp. 197–387, 2014.

[121] A. Marblestone, G. Wayne, and K. Kording, "Toward an integration of deep learning and neuroscience," Frontiers in Computational Neuroscience, vol. 10, p. 94, 2016.

[122] A. Ivakhnenko and V. Lapa, Cybernetic predicting devices. CCM Information Corporation, 1965.

[123] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, 2015, published online 2014; based on TR arXiv:1404.7828 [cs.NE].

[124] A. Ivakhnenko, "Polynomial theory of complex systems," IEEE transactions on Systems, Man, and Cybernetics, no. 4, pp. 364–378, 1971.

[125] R. Dechter, Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems, 1986.

[126] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner et al., "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[127] W. Long, Z. Lu, and L. Cui, "Deep learning-based feature engineering for stock price movement prediction," Knowledge-Based Systems, vol. 164, pp. 163–173, 2019.

[128] P. Domingos, "A few useful things to know about machine learning." Commun. ACM, vol. 55, no. 10, pp. 78–87, 2012.

[129] S. Zha, F. Luisier, W. Andrews, N. Srivastava, and R. Salakhutdinov, "Exploiting image-trained cnn architectures for unconstrained video classification," arXiv preprint arXiv:1503.04144, 2015.

[130] Wikipedia contributors:By Aphex34 - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=45679374, "Convolutional neural network — Wikipedia, the free encyclopedia," 2019, [Online; accessed 15-April-2019]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Convolutional_neural_network&oldid=891636189

[131] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner et al., "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[132] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in International conference on artificial neural networks. Springer, 2010, pp. 92–101.

[133] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in Proceedings of the fourteenth interna-

tional conference on artificial intelligence and statistics, 2011, pp. 315–323.

[134] B. Al-Bander, W. Al-Nuaimy, B. Williams, and Y. Zheng, "Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc," Biomedical Signal Processing and Control, vol. 40, pp. 91–101, 2018.

[135] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in International Conference on Medical Image Computing and Computer-assisted Intervention. Springer, Nice, 2013, pp. 246–253.

[136] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," Medical Image Analysis, vol. 35, pp. 18–31, 2017.

[137] S. Lawrence, C. Giles, A. Tsoi, and A. Back, "Face recognition: A convolutional neural-network approach," IEEE transactions on neural networks, vol. 8, no. 1, pp. 98–113, 1997.

[138] A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. Swetter, H. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, p. 115, 2017.

[139] Y. Pan, W. Huang, Z. Lin, W. Zhu, J. Zhou, J. Wong, and Z. Ding, "Brain tumor grading based on neural networks and convolutional neural networks," in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015, pp. 699–702.

[140] I. Wallach, M. Dzamba, and A. Heifets, "Atomnet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery," arXiv preprint arXiv:1510.02855, 2015.

[141] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke, "The rise of deep learning in drug discovery," Drug Discovery Today, vol. 23, no. 6, pp. 1241–1250, 2018.

[142] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognition," in International Workshop on Human Behavior Understanding. Springer, Amsterdam, 2011, pp. 29–39.

[143] F. Silvério, A. Certal, C. de Ferro, J. Monteiro, J. Cruz, R. Ribeiro, and b. p. y. o. Silva, J, "Automatic system for zebrafish counting in fish facility tanks."

[144] D. Alderton, Freshwater Aquariums: Basic Aquarium Setup and Maintenance. i5 Publishing, 2012.

[145] J. Fabic, I. Turla, J. Capacillo, L. David, and P. Naval, "Fish population estimation and species classification from underwater video sequences using blob counting and shape analysis," in Underwater Technology Symposium (UT), 2013 IEEE International. IEEE, 2013, pp. 1–6.

[146] T. Y, N. T, and L. B, "Automated fish counting using image processing," in Computational Intelligence and Software Engineering, 2009. CISE 2009. International Conference on. IEEE, Wuhan, China, 2009, pp. 1–5.

[147] B. Reed, M. Jennings et al., "Guidance on the housing and care of zebrafish danio rerio." Guidance on the housing and care of zebrafish Danio rerio., 2011.

[148] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[149] A. Davis and G. Paulik, "The design, operation, and testing of a photoelectric fish egg counter," The Progressive Fish-Culturist, vol. 27, no. 4, pp. 185–192, 1965.

[150] J. Alheit, "Use of the daily egg production method for estimating biomass of clupeoid fishes: a review and evaluation," Bulletin of Marine Science, vol. 53, no. 2, pp. 750–767, 1993.

[151] Y. Duan, L. Stien, A. Thorsen, Ø. Karlsen, N. Sandlund, D. Li, Z. Fu, and S. Meier, "An automatic counting system for transparent pelagic fish eggs based on computer vision," Aquacultural Engineering, vol. 67, pp. 8–13, 2015.

[152] E. Houde and C. Zastrow, "Ecosystem-and taxon-specific dynamic and energetics properties of larval fish assemblages," Bulletin of marine science, vol. 53, no. 2, pp. 290–335, 1993.

[153] K. Friedland, D. Ama-Abasi, M. Manning, L. Clarke, G. Kligys, and R. Chambers, "Automated egg counting and sizing from scanned images: rapid sample processing and large data volumes for fecundity estimates," Journal of Sea Research, vol. 54, no. 4, pp. 307–316, 2005.

[154] T. Atherton and D. Kerbyson, "Size invariant circle detection," Image and Vision Computing, vol. 17, no. 11, pp. 795–803, 1999.

[155] D. Parichy, M. Elizondo, M. Mills, T. Gordon, and R. Engeszer, "Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living

fish," Developmental dynamics, vol. 238, no. 12, pp. 2975–3015, 2009.

[156] S. Lindsay and R. Vogt, "Behavioral responses of newly hatched zebrafish (danio rerio) to amino acid chemostimulants," Chemical Senses, vol. 29, no. 2, pp. 93–100, 2004.

[157] L.-K. Soh and C. Tsatsoulis, "Texture analysis of sar sea ice imagery using gray level co-occurrence matrices," IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 2, pp. 780–795, 1999.

[158] Q. Al-Jubouri, W. Al-Nuaimy, M. Al-Taee, J. Luna, and L. Sneddon, "Automated electrical stimulation and physical activity monitoring of zebrafish larvae," in 2015 IEEE Conference on Applied Electrical Engineering and Computing Technologies (AEECT). Amman, 2015, pp. 1–6.

[159] J. Kittler and J. Illingworth, "Minimum error thresholding," Pattern Recognition, vol. 19, no. 1, pp. 41–47, 1986.

[160] Y. Cao, N. Semanchik, S. Lee, S. Somlo, P. Barbano, R. Coifman, and Z. Sun, "Chemical modifier screen identifies hdac inhibitors as suppressors of pkd models," Proceedings of the National Academy of Sciences, pp. 21 819–21 824, 2009.

[161] P. Jagadeeswaran, M. Gregory, K. Day, M. Cykowski, and B. Thattaliyath, "Zebrafish: a genetic model for hemostasis and thrombosis," Journal of Thrombosis and Haemostasis, vol. 3, no. 1, pp. 46–53, 2005.

[162] K. Griffin, R. Patient, and N. Holder, "Analysis of fgf function in normal and no tail zebrafish embryos reveals separate mechanisms for formation of the trunk and the tail," Development, vol. 121, no. 9, pp. 2983–2994, 1995.

[163] B. Leo, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees," Wadsworth International Group, 1984.

[164] G. Seni and J. Elder, "Ensemble methods in data mining: improving accuracy through combining predictions," Synthesis Lectures on Data Mining and Knowledge Discovery, vol. 2, no. 1, pp. 1–126, 2010.

[165] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[166] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," Computational intelligence and neuroscience, 2016.

[167] G. Tyagi, N. Patel, and I. Sethi, "A fine-tuned convolution neural network based approach for phenotype classification of zebrafish embryo," Procedia Computer Science, vol. 126, pp. 1138–1144, 2018.

[168] S. Liawatimena, Y. Heryadi, A. Trisetyarso, A. Wibowo, B. Abbas, E. Barlian et al., "A fish classification on images using transfer learning and matlab," in 2018 Indonesian Association for Pattern Recognition International Conference (INAPR). IEEE, 2018, pp. 108–112.

[169] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[170] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[171] H. Blockeel, L. Schietgat, J. Struyf, A. Clare, and S. Dzeroski, "Hierarchical multilabel classification trees for gene function prediction," in Probabilistic Modeling and Machine Learning in Structural and Systems Biology, 2006, pp. 9–14.

[172] E. Tanaka, S. Nozawa, A. Macedo, and J. Baranauskas, "A multi-label approach using binary relevance and decision trees applied to functional genomics," Journal of Biomedical Informatics, vol. 54, pp. 85–95, 2015.

[173] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine learning, vol. 63, no. 1, pp. 3–42, 2006.

Appendix A: HP Scanjet G4050 Photo Scanner

HP Scanjet G4050 Photo Scanner

The HP Scanjet G4050 Photo Scanner is designed for families and photo enthusiasts with large numbers of photos, slides, negatives or medium/large format transparencies which they want to convert into professional-quality, colour-accurate digital files.

HP-exclusive 6-colour high-definition scanning for photos, slides, film and negatives.

- Convert packs of photos, slides, film and negatives into vivid, colour accurate digital images and get the highest quality scans[1] with 6-colour, 96-bit scanning – exclusive to HP. Get high definition results at 4800 x 9600-dpi resolution[2].

Scan stacks of photos, slides, films or negatives, and automatically remove dust and scratches.

- Scan multiple originals – sixteen 35 mm slides, 30 negative frames, two medium format (120 mm) or one large format film frame or four 10 x 15 cm prints – and save as separate files. Remove dust and scratches using built-in infrared technology.

Preserve your memories at the touch of a button – plus easy colour restoration and photo fixing.

- Digitise photos, slides, film and negatives using one-touch buttons. Bring old originals back to life with a range of intuitive tools. Restore faded colour, remove red-eye and bring detail out of dark photos, all using HP Real Life technologies.

[1] Based on independent tests of in-class scanners, Dec 06. http://digitalkamera.image-engineering.de/index.php/Publications
[2] Not available when scanning negatives in 6-colour; maximum resolution may be affected by PC system factors

HP Scanjet G4050 Photo Scanner

TECHNICAL SPECIFICATIONS

Scan Technology	Charge Coupled Device (CCD)
Scan Type	Flatbed; **Colour Scanning**:
Scan Resolution	**Hardware**: Up to 4800 x 9600 dpi **Optical**: Up to 4800 dpi **Enhanced**: Unlimited
Scan File Format	Windows®: BMP, JPEG, TIFF, TIFF compressed, PNG, PCX, Flashpix (FPX), PDF, PDF searchable, RTF, HTM, TXT; Macintosh: TIFF, PICT, JPEG, GIF, FlashPix, Plain Text, PDF, HTML, Rich Text
Scan Input Modes	Front-panel: Scan (reflective scans from the glass), Scan film, Copy, Scan to PDF; HP Photosmart software; user application via TWAIN; Transparent Materials Aapter (transmissive scans of film materials)
Image Scaling Or Enlargement Range	10 to 2400% in 1% increments
Grayscale Levels/Bit Depth	256; **Bit Depth**: 96-bit
Multi Feed Detection	No
Auto Document Feeder Capacity	None
External I/O ports	1 Hi-Speed USB 2.0
Standard Connectivity	Hi-Speed USB 2.0
Network Ready	None
Control Panel	4 front-panel buttons (Scan, Scan film, Copy, Scan to PDF)
Media Types Supported	Paper (plain, inkjet, photo, banner), envelopes, labels, cards (index, greeting), 3-D objects, 35 mm slides and negatives (using transparent materials adapter), iron-on transfers
Twain Version	Version 1.9
Software Included	HP Photosmart software includes integrated OCR, HP Photosmart Share, faded colour restoration, dust and scratch removal, HP Adaptive Lighting, HP Image Editor (HP Red-eye Removal capabilities)
Compatible Operating Systems	Windows 10, Windows 8, Windows 7, Windows Vista, Windows XP (32-bit and 64-bit), Windows XP, Windows 2000, Mac OS X v10.6, Lion and Mountain Lion
Minimum System Requirements	**PC**: Windows 10, Windows 8, Windows 7, Windows Vista: 1.3 GHz processor, 1 GB RAM (2 GB for 64-bit); Windows XP: 1.3 GHz processor, 512 MB RAM (1 GB for 64-bit); Windows 2000: 800 MHz processor, 256 MB RAM, for all systems: 450 MB available hard disk space, USB port, CD-ROM drive, 800 x 600 SVGA monitor, 16-bit color; **Mac**: Mac OS X v10.3.9, v10.4 and higher; PowerPC (G3, G4 and G5 processor), Intel® Core processor; 128 MB RAM, [250 MB recommended (required for v10.4 and higher)], 250 MB available hard disk space

Operating Environment	**Operating Temperature Range:** 10 to 35° C **Storage Temperature Range:** -40 to 60° C **Recommended Operating Temperature Range:** 10 to 35° C **Recommended Humidity Operating Range:** 15 to 80% RH **Non-Operating Humidity Range:** 0 to 90% RH
Power	**Power Consumption:** 25 watts maximum **Power Requirements:** Input voltage 100 to 240 VAC (+/- 10%), 50/60 Hz (+/- 3 Hz)
Energy Star	Yes
Regulatory compliance information/safety	IEC 60950-1:2001, national derivatives, associated voluntary and mandatory certifications: Canada (cUL), China (CCC), Russia (GOST), European Union (UL GS), Taiwan (BSMI), USA (UL)
Electromagnetic compatibility	EU (CE Declaration of Conformity), US (FCC Declaration), Australia (ACA), New Zealand (ACA), Russia (GOST), Korea (MIC), Taiwan (BSMI), China (CCC)
Product Dimensions	303 x 508 x 108 mm; **Packaged:** 596 x 184 x 439 mm
Product Weight	5.3 kg; **Packaged:** 6.89 kg
What's In The Box	**L1957A:** HP Scanjet G4050 Photo Scanner; power supply adapter/power cord; USB cable; CD-ROMs with software and User's Guide; Setup and Support Guide
Warranty	One-year commercial hardware warranty, Web support included. Warranty may vary by country as required by law.
Service And Support Options	**UH253E** HP Care Pack, Next Business Day Exchange Service, 3 years **UH248E** HP Care Pack, Next Business Day Exchange Service, 2 years (except Turkey, EEM and Russia) **UJ997E** HP Care Pack, Standard Exchange Service, 3 years (EEM and Russia only) **UH247E** HP Care Pack, Return To Depot, 3 years (Turkey only)

Count on dependable support to create the best image possible. We can help you enhance your printing and imaging environment, protect your IT investment, and grow your business—with expert support that's affordable, personal, and tailored to your needs through HP Care Pack Services.

For more information, visit our website at http://www.hp.com

HP Scanjet G4050 Photo Scanner L1957A

Published in EMEA November 2015 4AA0-4447EEE

Appendix B: 96-well Plate Technical Sheet

eppendorf

Technical Data Sheet
Eppendorf Cell Culture Plate, 96-Well
English (EN)

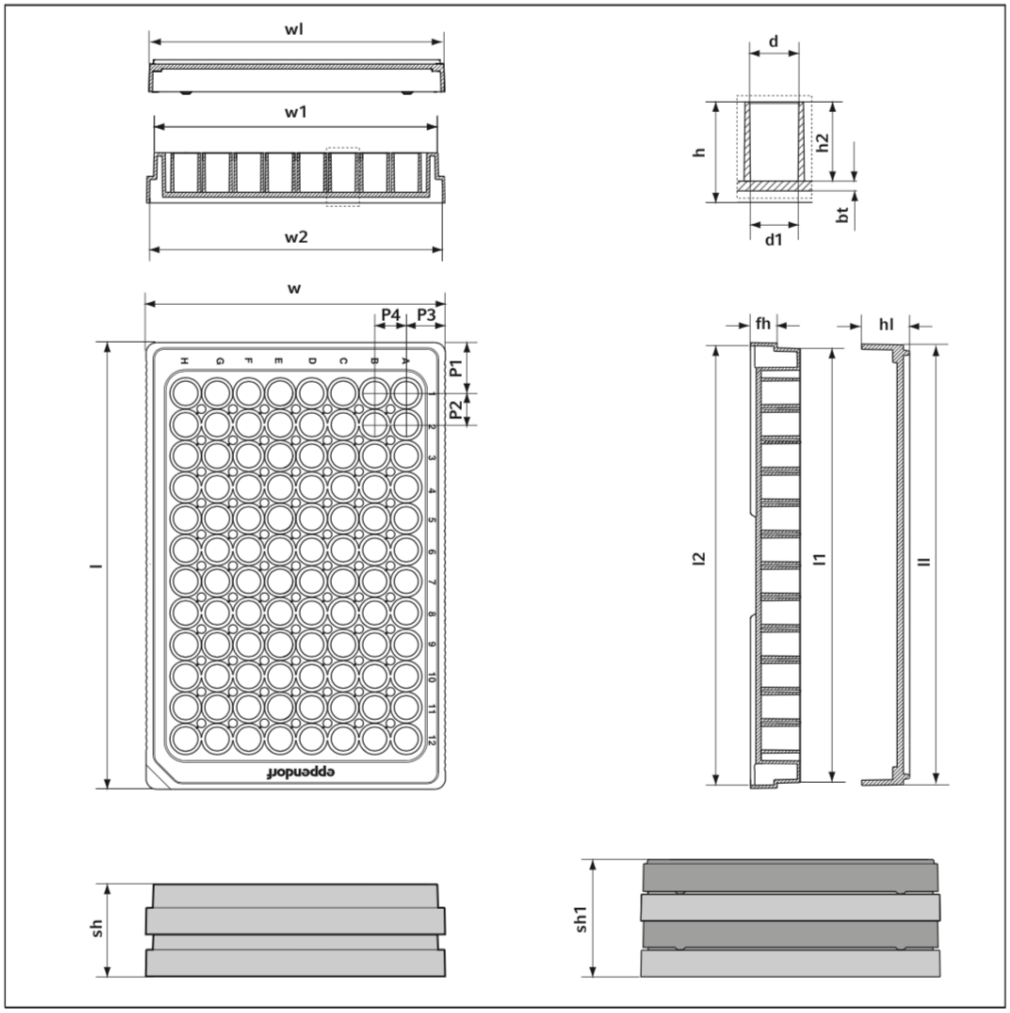
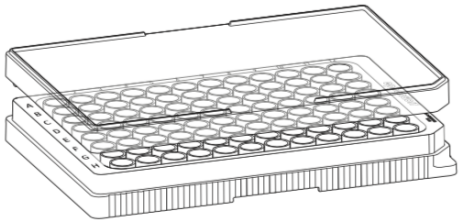


Plate dimensions [mm]							
Length			Width			Height	
l	l1	l2	w	w1	w2	h	fh
127.8	124.3	125.3	85.5	81.1	81.9	14.4	7.6
Well dimensions [mm]				**Positions of well center [mm]**			
Diameter		Depth	Thickness	Position			
d	d1	h2	bt	P1	P2	P3	P4
7.0	6.8	11.3	1.4	14.4	9.0	11.2	9.0
Stack height [mm]				**Lid dimensions**			
without lid (sh)		with lid (sh1)		Height (hl)	Width (wl)	Length (ll)	
21.5		33.7		9.1	84.3	127.6	

Technical Data Sheet
Eppendorf Cell Culture Plate, 96-Well
English (EN)



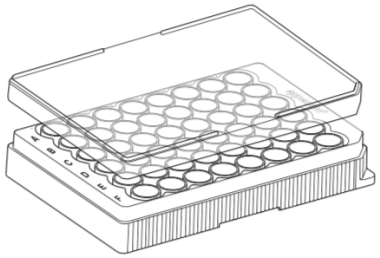
Descripition	Eppendorf Cell Culture Plate, 96-well 0030 730.119: with tissue culture treated surface 0030 730.127: with tissue culture treated surface 0030 730.011: with non-treated surface	
Dimensions	Plate with lid: 127.8 x 85.5 x 17.6 mm (l x w x h)	
Weight	Plate 43 g, lid 20 g	
Filling volume	Working volume/well: 0.1 mL/well to 0.2 mL/well Working volume moat: 5 mL	
Growth area	37.0 mm^2/well	
Material	Polystyrol, meets USP Class VI specifications	
Color	Transparent	
Operating temperature	-86 °C to +60 °C	
Centrifugability*	Can be centrifuged up to 2500 x *g*. When arranged in a stack (5 plates) can be centrifuged up to 300 x *g**.	
Autoclavable	—	
Shelf life	5 years	
Packing unit	Pcs/bag: 0030 730.119 and 0030 730.011: 1 plates/bag 0030 730.127: 10 plates/bag	Pcs/case: 0030 730.119 and 0030 730.011: 80 plates/box 0030 730.127: 200 plates/box
Lot-specific certificates	Sterility: Sal 10^{-6} (according to ISO 11137-2:2007) Pyrogens: <0,001 EU/mL (according to Ph. Eur. 2.6.14) RNase: $<10^{-9}$ Kunitz units DNase: $<10^{-6}$ Kunitz units Human DNA: <2 pg Bacterial DNA: <50 fg Tissue culture surfaces tested by cell growth.	
Certificates of quality	Quality Assurance of surface properties. Products are non-cytotoxic according to DIN EN ISO 10993-5. Products are produced in a class 7 clean room environment according to ISO 14644-1 and EG-GMP Class C. Products conform to ANSI/SLAS 1-2004 to ANSI/SLAS 4-2004. Material suppliers of Eppendorf do not use or intentionally incorporate slip agents, biocides, plasticizers, bisphenol A, latex, antistatic agents, metallic dyes or mineral oil in the materials Eppendorf uses for the production of Eppendorf Cell Culture Consumables.	
Eppendorf Quality Management	ISO 9001/13485/14001	

*The centrifugation stability of each plate type generally depends on the centrifuge and its accessories, the ambient conditions and the liquid used.

Your local distributor: www.eppendorf.com/contact
Eppendorf AG · 22331 Hamburg · Germany
eppendorf@eppendorf.com · www.eppendorf.com

Appendix C: 48-well Plate Technical Sheet

Technical Data Sheet
Eppendorf Cell Culture Plate, 48-Well
English (EN)



Description	Eppendorf Cell Culture Plate, 48-well 0030 723.112: with tissue culture treated surface 0030 723.015: with non-treated surface	
Dimensions	Plate with lid: 127.8 x 85.5 x 20.0 mm (l x w x h)	
Weight	Plate 55 g, lid 20 g	
Filling volume	Working volume/well: 0.3 mL/well to 0.5 mL/well	
Growth area	85.6 mm^2/Well	
Material	Polystyrol, meets USP Class VI specifications	
Color	Transparent	
Operating temperature	-86 °C to +60 °	
Centrifugability*	Can be centrifuged up to 2500 x *g*. When arranged in a stack (4 plates) can be centrifuged up to 300 x *g**.	
Autoclavable	—	
Shelf life	5 years	
Packing unit	1 Plate/bag	60 plate/case
Lot-specific certificates	Sterilität: Sal 10^{-6} (according to ISO 11137-2:2007) Pyrogene: <0.001 EU/mL (according to Ph. Eur. 2.6.14) RNase: $<10^{-9}$ Kunitz units DNase: $<10^{-6}$ Kunitz units Human DNA: <2 pg Bacterial DNA: <50 fg Tissue culture surfaces tested by cell growth.	
Certificates of quality	Quality Assurance of surface properties. Products are non-cytotoxic according to DIN EN ISO 10993-5. Products are produced in a class 7 clean room environment according to ISO 14644-1 and EG-GMP Class C. Products conform to ANSI/SLAS 1-2004 to ANSI/SLAS 4-2004. Material suppliers of Eppendorf do not use or intentionally incorporate slip agents, biocides, plasticizers, bisphenol A, latex, antistatic agents, metallic dyes or mineral oil in the materials Eppendorf uses for the production of Eppendorf Cell Culture Consumables.	
Eppendorf Quality Management	ISO 9001/13485/14001	

*The centrifugation stability of each plate type generally depends on the centrifuge and its accessories, the ambient conditions and the liquid used.

Your local distributor: www.eppendorf.com/contact
Eppendorf AG · 22331 Hamburg · Germany
eppendorf@eppendorf.com · www.eppendorf.com

eppendorf

Technical Data Sheet
Eppendorf Cell Culture Plate, 48-Well
English (EN)

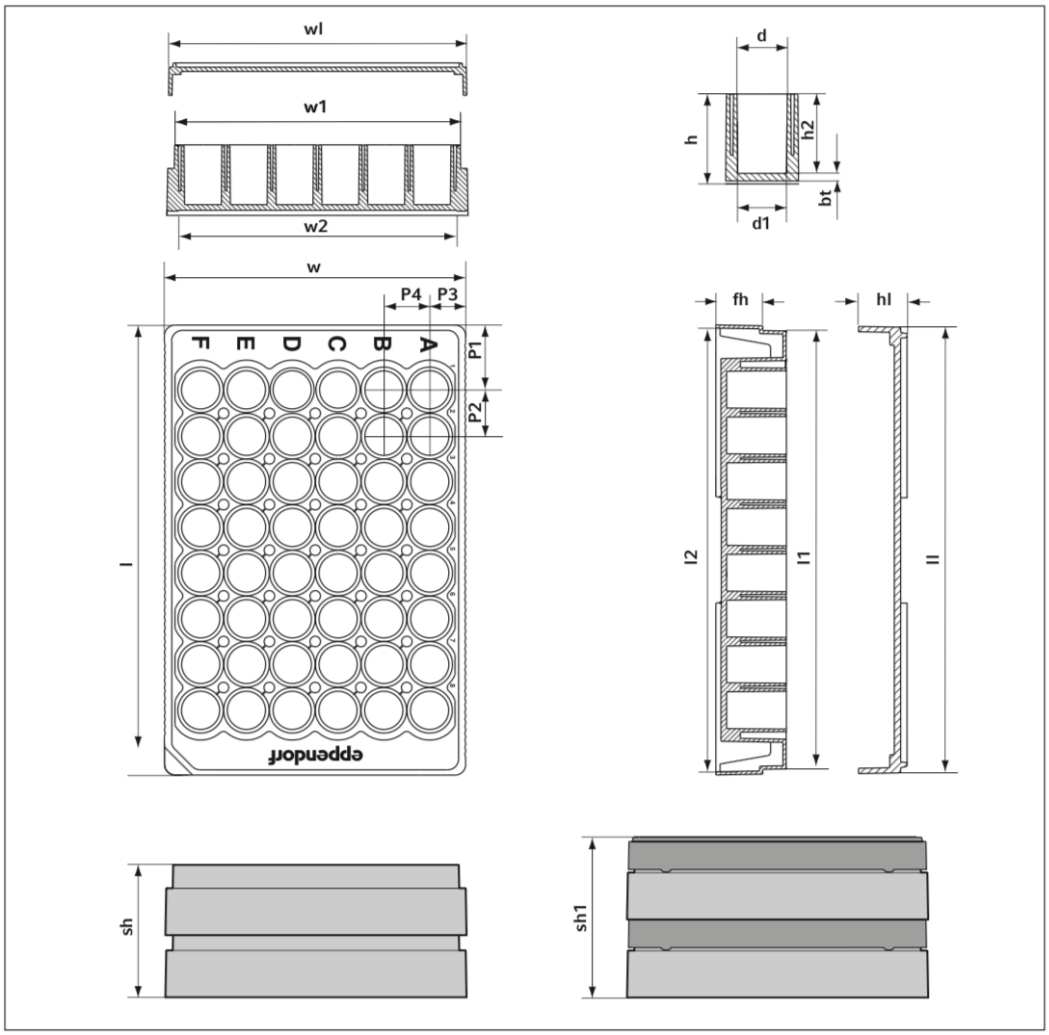
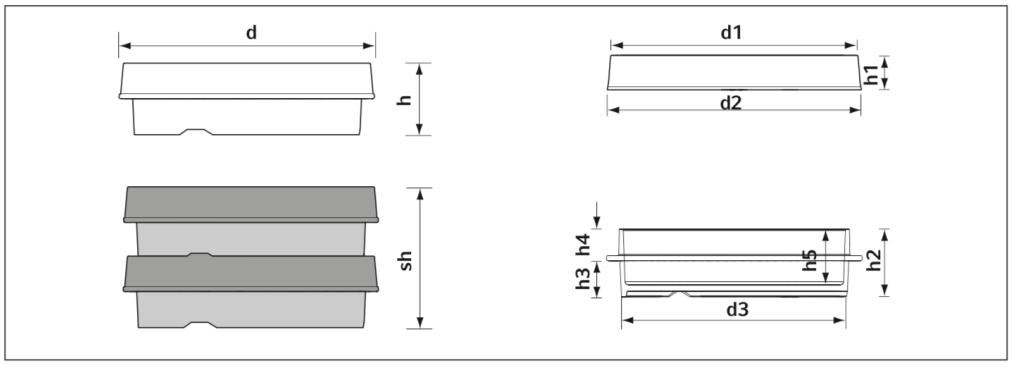


Plate dimensions [mm]							
Length			Width			Height	
l	l1	l2	w	w1	w2	h	fh
127.8	124.4	125.3	85.5	81.0	81.9	20.0	13.3
Well dimensions [mm]				**Positions of well center [mm]**			
Diameter		Depth	Thickness	Positions			
d	d1	h2	bt	P1	P2	P3	P4
10.7	10.4	17.0	1.4	18.4	13.0	10.2	13.0
Stack height [mm]				**Lid dimensions [mm]**			
without lid (sh)		with lid (sh1)		Height (hl)	Width (wl)	Length (ll)	
38.7		45.0		9.1	84.3	127.6	

Appendix D: Petri Dish 100mm Technical Sheet

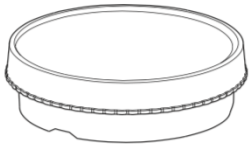
eppendorf

Technical Data Sheet
Eppendorf Cell Culture Dish 100 mm
English (EN)



Dimensions dish and lid [mm]			
Diameter dish with lid (d)	Inner diameter lid (d1)	Outer diameter lid (d2)	Inner diameter dish (d3)
93.8	88.3	92.9	85.0
Lid height (h1)	Dish height (h2)	Height dish with lid (h)	
11.2	19.4	21.2	
Dish height below handling ring (h3)	Dish height above handling ring (h4)	Inner height dish (h5)	
8.9	10.5	17.5	
Stack height [mm]			
with lid (sh)			
41.8			

Technical Data Sheet
Eppendorf Cell Culture Dish 100 mm
English (EN)



Descripition	Eppendorf Cell Culture Dish, 100 mm 0030 702.115: with tissue culture treated surface 0030 702.018: with non-treated surface	
Dimensions	Dish with lid: 93.8 x 19.4 mm (d x h)	
Weight	Dish 11.9 g, lid 8.4 g	
Filling volume	8.0 to 10.0 mL	
Growth area	56.8 cm^2	
Material	Polystyrol, meets USP Class VI specifications	
Color	Transparent	
Operating temperature	-86 °C to +60 °C	
Centrifugability	—	
Autoclavable	—	
Shelf life	5 years	
Packing unit	10 dishes/bag	300 dishes/case
Lot-specific certificates	Sterility: Sal 10^{-6} (according to ISO 11137-2:2007) Pyrogene: <0,001 EU/mL (according to Ph. Eur. 2.6.14) RNase: $<10^{-9}$ Kunitz units DNase: $<10^{-6}$ Kunitz units Human DNA: <2 pg Bacterial DNA: <50 fg Tissue culture surfaces tested by cell growth.	
Certificates of quality	Quality Assurance of surface properties. Products are non-cytotoxic according to DIN EN ISO 10993-5. Products are produced in a class 7 clean room environment according to ISO 14644-1 and EG-GMP Class C. Material suppliers of Eppendorf do not use or intentionally incorporate slip agents, biocides, plasticizers, bisphenol A, latex, antistatic agents, metallic dyes or mineral oil in the materials Eppendorf uses for the production of Eppendorf Cell Culture Consumables.	
Eppendorf Quality Management	ISO 9001/13485/14001	

Your local distributor: www.eppendorf.com/contact
Eppendorf AG · 22331 Hamburg · Germany
eppendorf@eppendorf.com · www.eppendorf.com