

AI-Based Prediction and Monitoring of Drug Formulation Research and Development

Thesis submitted to the University of Nottingham for the degree of

Doctor of Philosophy

Haoyu Wang

Supervised by

Dr. Chiew Foong Kwong

Dr. David Cho Siu-Yeung

Dr. Boon Giin Lee

Dr. Liang Huang

Date: 12th May 2023

Abstract

When new solid formulations reach a stage of research and development (R&D), the number of experiments to be performed is generally several times higher than expected such that the R&D cost will surge. It is noteworthy that the quality evaluation standard for solid formulations and the dissolution test results take on critical significance to guiding the solid formulation research, whereas the experimental phenomenon of dissolution is difficult to be fully observed by experimenters for the extremely long experiment time of at least 8 h. Thus, a system exhibiting prediction and phenomenon monitoring functions is developed and tested in this study to tackle down the above-described problems.

First, a computer technology-based prediction system for the dynamic solubility of active pharmaceutical ingredient (API) is developed in accordance with the solid formulation R&D process, where a novel equation is employed, instead of polynomial regression. After the experimentally acquired dynamic solubility data are trained, the developed system is capable of predicting the dynamic solubility rate of API following the blade dissolution speed.

Moreover, a novel tablet dissolution rate prediction system is built using artificial neural networks (ANN) and non-linear regression methods. Compare with SVM, ANN can offer a faster prediction speed and suitable for the future big data model. This system can predict dissolution rates with ANN and processing the prediction results based on two novel non-linear regression methods. Accordingly, the demanded system database becomes less than the orthogonal design. Compared with previously developed systems, the built system exhibits higher prediction precision with a less amount of training data. Besides, the system, based on the novel input data design, is enabled to use the data whose formulation composition resembles prediction composition. Furthermore, the input data screening function is introduced into the system based on the prediction function to avert the effect of experimental error.

Lastly, a phenomenon monitoring system covering a camera module and an image recognition module is developed to monitor the dissolution phenomenon of tablets during the dissolution test. The image recognition program based on the region growth, Hue, Saturation, and Value (HSV) is capable of capturing and pixelating tablets of different colors during their movement in the dissolution cup. To be specific, the camera module comprises a visible/infrared light camera and an infrared light source, such that the monitoring system is enabled to recognize tablets in dark and bright environments.

Using ANN, non-linear regression, and image recognition methods, the dynamic solubility of API is predicted, the tablets dissolution rate is predicted based on a smaller database, and the real-time dissolution phenomenon is monitored in a wide variety of environments in this study.

Acknowledgements

First, I would like to extend my sincere gratitude to my supervisors Dr. David Cho Siu-Yeung and Dr. Chiew-Foong Kwong, who have lit up the darkness of my life. Without the invaluable support and kind encouragement from Dr. David Cho Siu-Yeung, I would not have commenced this challenging interdisciplinary research and achieved satisfactory results. Dr. Chiew-Foong Kwong hauled me out of the quagmire in my tough times. His supervision has made my research scientifically and academically organized and his optimistic, positive, and sincere personality has profoundly influenced me in both study and life.

Moreover, I appreciate the Department of Electrical and Electronic Engineering, the University of Nottingham, for granting me the opportunity to complete my doctorate. I would like to thank AIOP and SSNI Research Group for providing with me an excellent study and research environment.

Furthermore, my heartfelt gratitude goes to my family, my mother Liu Zhixin, my father Wang Zhe, my wife Jia Bijun, and my dearest daughter Wang Zitong. With their support, I complete this academic journey.

Table of Contents

Abstract.....	I
Acknowledgements	III
Table of Contents	IV
List of Figures.....	VIII
List of Table	X
List of Acronyms	XII
Chapter 1 Introduction.....	1
1.1 Background and Motivation	1
1.2 Problem Statements	5
1.2.1 Problem 1	5
1.2.2 Problem 2	5
1.2.3 Problem 3	5
1.3 Research Questions.....	6
1.4 Research Objective	6
1.4.1 Research Objective 1	6
1.4.2 Research Objective 2	7
1.4.3 Research Objective 3	7
1.5 Report Organisation	7
Chapter 2 Formulation Research of New Drugs: Trends and Issues	10
2.1 Related Projects	10
2.1.1 DDD Plus	10
2.1.2 ANN in API Dynamic Solubility	10
2.1.3 ANN in Formulation Dissolution Prediction	11
2.1.4 Image Recognition in Dissolution Monitoring	15
2.2 Computer Science and Technology.....	21
2.2.1 Artificial Neural Networks	21
2.2.2 Back Propagation	22
2.2.3 Radial Basis Function Network	23
2.2.4 Hue, Saturation, Value (HSV)	24
2.2.5 Morphological Transformations	24
2.3 Pharmaceutical Technologies.....	25
2.3.1 Tablet	25

2.3.2 Dissolution Test	25
2.3.3 F2 Similarity	27
2.3.4 Dynamic Solubility	28
2.3.5 Biopharmaceutics Classification System (BCS).....	28
2.3.6 Disintegration.....	29
2.3.7 Corrosion.....	29
2.4 Pharmaceutical Instruments	29
2.4.1 High Performance Liquid Chromatography (HPLC)	29
2.4.2 Ultraviolet-Visible	29
2.4.3 Dissolution Tester	31
2.5 Chapter Summary	32
Chapter 3 Artificial Neural Network in Predicting API Dynamic Solubility	34
3.1 Chapter Introduction	34
3.2 Materials and Methods.....	34
3.2.1 Materials	34
3.2.2 Preparation of API Sample	36
3.2.3 Assay for Indometacin API.....	36
3.2.4 Design of Experiment	37
3.2.5 Core Equation	37
3.2.6 Curve Fitting	40
3.3 Result and Discussion	41
3.3.1 BPNN and RBFN Modeling.....	41
3.3.2 Data Prediction.....	43
3.3.3 Data Optimisation	52
3.4 Chapter Summary	55
Chapter 4 Backpropagation Neural Network and Non-linear Regression Approach in Predicting Solid Formulation Dissolution Rate.....	56
4.1 Chapter Introduction	56
4.2 Materials	56
4.3 Pharmaceutical Experimental Methods	59
4.3.1 Assay for Tinidazole	59
4.3.2 Organic Impurities Analysis	61
4.3.3 Impurity Study	62
4.3.4 Dissolution Test Study.....	62

4.3.5 Stability Study.....	65
4.3.6 Tinidazole Tablet Preparation Method	66
4.4 Computer Science Technologies.....	67
4.4.1 Model Test Design.....	67
4.4.2 Problem of Insufficient data leads to prediction errors.....	68
4.4.3 Fully Connected Neural Network Model Design	69
4.4.4 Two novel regression methods	70
4.4.5 Reference Line Regression Method (RLRM).....	73
4.5 Results and Discussion	75
4.5.1 Result of Formulation Experiment.....	75
4.5.2 Prediction Result of EDRM versus RLRM	77
4.5.3 Related Data Reuse	82
4.5.4 The Relationship between Prediction Times and Prediction Stability.....	84
4.5.5 Comparison with others work.....	87
4.5.6 Input Data Screen.....	93
4.6 Chapter Summary	93
Chapter 5 Monitoring and Analyzing Solid Formulation Dissolution Phenomenon with Image Recognition Technologies.....	95
5.1 Chapter Introduction	95
5.2 Materials and Methods.....	95
5.2.1 Materials	95
5.2.2 Sampling	95
5.2.3 Shading	96
5.2.4 Camera and Infrared Light Source.....	96
5.2.5 Instrument Design.....	97
5.2.6 Program Modeling	98
5.3 Result and Discussion	99
5.3.1 Disintegration Tablets Image Recognition Test	100
5.3.2 Large-Size Disintegration Tablets	100
5.3.3 Small Size Disintegration Tablets.....	105
5.3.4 Erosion Tablets	109
5.3.5 Different Color Tablets.....	114
5.3.6 Infrared Condition.....	117
5.4 Chapter Summary	121

Chapter 6 Conclusions and Future Works	123
6.1 Concluding Remarks.....	123
6.2 Future Works.....	124
List of Research Outputs.....	126
Journal Article Publications	126
Invention Patents.....	126
References.....	127

List of Figures

Figure 1.1 Structure of Research	4
Figure 1.2 Structure of Report Organization	8
Figure 2.1 The representative BP network with one hidden layer.	22
Figure 2.2 The RBFN structure.....	23
Figure 2.3 The internal parameter of the dissolution cup, affecting the solvent flow distribution.	26
Figure 2.4 The parameter of the blade, the decisive factor of the solvent flow rate.....	26
Figure 2.5 The location (height) of the blade, affecting the flow rate distribution in the dissolution cup	27
Figure 2.6 A dissolution tester with a sampling function, with a dissolution tester on the left and a sampling instrument on the right.....	31
Figure 3.1 The fitting curve with a limit of Equation 3.11	40
Figure 3.2 The structure of ANN model	41
Figure 3.3 ANN model process.....	42
Figure 3.4 The dissolution test results of Indometacin in water	45
Figure 3.5 The dissolution test results of Indometacin under 25 rpm/water condition	46
Figure 3.6 The dissolution test results of Indometacin under 50 rpm/water condition	47
Figure 3.7 The dissolution test results of Indometacin under 75 rpm/water condition	48
Figure 3.8 The dissolution test results of Indometacin under 100 rpm/water condition	49
Figure 3.9 The dissolution test results of Indometacin under 125 rpm/water condition.	50
Figure 3.10 The dissolution test results of Indometacin under 150 rpm/water condition.	51
Figure 3.11 Data summary of this chapter	52
Figure 3.12 The dissolution test curve of Indometacin under 100 rpm condition.....	53
Figure 4.1 The structure of ANN	69
Figure 4.2 The flow chart of EDRM.....	72
Figure 4.3 The flow chart of RLRM in the program	74
Figure 4.4 The x-axis is the dissolution time while the y-axis is the dissolution rate at different dissolution time points.	77
Figure 4.5 The prediction output of EDRM.....	78
Figure 4.6 The prediction output of RLRM.....	79
Figure 4.7 Comparison of EDRM and RLRM.....	80
Figure 4.8 The contrast curve of the results.....	81
Figure 4.9 The contrast curve of Formulation 15	83

Figure 4.10 The comparison of 10 predictions of Formulation 21 under two trainings	85
Figure 4.11 The comparison of 10 predictions of Formulation 21 under ten trainings	86
Figure 4.12 The comparison of 10 predictions of Formulation 21 under 50 trainings	87
Figure 4.13 Data bulk requirement comparison between Arthur’s model and our model	92
Figure 5.1 The photo of IR-CUT camera module.....	96
Figure 5.2 An IR-CUT camera system structure chart	97
Figure 5.3 The whole instrument system structure chart	97
Figure 5.4 The process of program	99
Figure 5.5 The experimental results of Large Size Disintegration Tablets (a)	102
Figure 5.6 The experimental results of Large Size Disintegration Tablets (b)	103
Figure 5.7 The experimental results of Large Size Disintegration Tablets (c)	104
Figure 5.8 The experimental results of Small Size Disintegration Tablets (a)	106
Figure 5.9 The experimental results of Small Size Disintegration Tablets (b).....	107
Figure 5.10 The experimental results of Small Size Disintegration Tablets (c)	108
Figure 5.11 The experimental results of Erosion Tablets (a)	110
Figure 5.12 The experimental results of Erosion Tablets (b).....	111
Figure 5.13 The experimental results of Erosion Tablets (c)	112
Figure 5.14 The experimental results of Erosion Tablets (d).....	113
Figure 5.15 General Color Wheel from Shanghai Colorcon Coating Technology Co.LTD ..	114
Figure 5.16a White small round plastic plate	115
Figure 5.16b Yellow small round plastic plate.....	115
Figure 5.16c Bright red small round plastic plate.....	115
Figure 5.16d Bright blue small round plastic plate.....	115
Figure 5.16e Deep blue small round plastic plate.....	115
Figure 5.16f Dark red small round plastic plate	115
Figure 5.16g Green small round plastic plate	115
Figure 5.16h Purple small round plastic plate	115
Figure 5.17a System practicability test (a)	118
Figure 5.17b System practicability test (b).....	118
Figure 5.17c System practicability test (c)	118

List of Table

Table 2.1 The comparison of relevant projects	13
Table 2.2 The comparison of related projects.	18
Table 2.3 The contrast of average deviation and factor F2.	27
Table 2.4 The BCS classes	28
Table 3.1 API parameters	35
Table 3.2 The dissolution data of Indometacin in water	44
Table 3.3 The comparison of real data, BPNN prediction data, and RBFN prediction data under 25 rpm/water condition	46
Table 3.4 The comparison of real data, BPNN prediction data, and RBFN prediction data under 50 rpm/water condition	47
Table 3.5 The comparison of real data, BPNN prediction data, and RBFN prediction data under 75 rpm/water condition	48
Table 3.6 The comparison of real data, BPNN prediction data, and RBFN prediction data under 100 rpm/water condition	49
Table 3.7 The comparison of real data, BPNN prediction data, and RBFN prediction data under 125 rpm/water condition.	50
Table 3.8 The comparison of real data, BPNN prediction data, and RBFN prediction data under 150 rpm/water condition	51
Table 3.9 The dissolution data of Indometacin under 100 rpm condition	53
Table 3.10 The relative standard deviation between average values and abnormal data (sample 6)	54
Table 4.1 Tinidazole Parameters	58
Table 4.2 The solubility of tinidazole in pH 1.2, pH 4.0, and pH 6.8 dissolution media and water	59
Table 4.3 The basic information of selected tablet	59
Table 4.4 The assay of Tinidazole.....	60
Table 4.5 Organic impurities analysis of Tinidazole.....	61
Table 4.6 The Impurity names and chemical structural formulas recorded in USP, EP, and CP.	62
Table 4.7 The comparison of dissolution test in Chinese Pharmacopoeia (CP), Japanese Orange Book, and FDA dissolution method database.	63
Table 4.8 The details of dissolution method in Japanese Orange Book	64
Table 4.9 The details of dissolution method in Chinese Pharmacopoeia 2015.....	65
Table 4.10 The stability under different conditions	66

Table 4.11 Composition of 500 mg Tinidazole Tablet	66
Table 4.12 The percentage of seven ingredients	75
Table 4.13 The percentage of drug release in dissolution test at different time Points.....	76
Table 4.14 The prediction results of Formulation 3 with Formulations 13, 16, and 21 as reference line.....	81
Table 4.15 The Formulation 15 prediction results through EDRM and RLRM (second and third column) and the experimental results of Formulation 15 (fourth column).....	83
Table 4.16 10 prediction results of Formulation 21 under two trainings.....	84
Table 4.17 10 prediction results of Formulation 21 under ten trainings.....	85
Table 4.18 10 prediction results of Formulation 21 under 50 trainings.....	86
Table 4.19 Experimental dataset of Arthur Manda’s work	88
Table 4.20a The prediction result using 29 groups of formulation.....	90
Table 4.20b The prediction result using 14 groups of formulation.....	91
Table 4.20c The prediction result using 6 groups of formulation.....	90
Table 5.1 A summarize table of experimental result and discussion	121

List of Acronyms

2D	2 Dimensional
3D	3 Dimensional
5G	Fifth Generation
ANN	Artificial Neural Network
API	Active Pharmaceutical Ingredient
BCS	Biopharmaceutics Classification System
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BP	British Pharmacopoeia
BPNN	Back Propagation Neural Network
CAS number	Chemical Abstracts Service Number
CCNA	Croscarmellose sodium
CFDA	China Food and Drug Administration
ChP/CP	China Pharmacopoeia
DDD PLUS	Dose Disintegration and Dissolution Plus
EDRM	Effective Data Regression Method
EP	European Pharmacopoeia
F2	Similarity Factor
FCNN	Fully Connected Neural Network
FDA	Food and Drug Administration (United States)
FTIR-ATR	Fourier Transform Infrared Spectroscopy Attenuated Total Reflection
HPLC	High Performance Liquid Chromatography
HPMC	Hydroxypropyl methylcellulose
HSV	Hue(H), Saturation(S), Value(V)
IoE	Internet of Everything

IR	Infrared
JP	Japanese Pharmacopoeia
LDR	Light-Dependent Resistor
L-HPC	Low-substituted hydroxypropyl cellulose
MAH	Marketing Authorization Holder
MS	Magnesium stearate
MSE	Mean Square Error
NLT	No Less Than
PC	Personal Computer
PE	Processing Units
pH	Potential of Hydrogen
PH101	Microcrystalline Cellulose PH 101
PRC	People Republic of China
PROC REG	Process Region
PSA	Polar Surface Area
QbD	Quality becomes Design
R&D	Research and Development
RBFN	Radial Basis Function Network
RLRM	Reference Line Regression Method
RMB	Ren Min Bi
RS	Reference Solution
RSD	Relative Standard Deviation
RSM	Response Surface Methodology
RTDPM	Real-Time Dissolution Phenomenon Monitoring
SVM	Support Vector Machine

TLC	Thin-Layer Chromatography
US	United States
US FDA	United States Food and Drug Administration
USP	United States Pharmacopoeia
UV-VIS	Ultraviolet-Visible Spectroscopy
w/w	Weight/Weight

Chapter 1 Introduction

1.1 Background and Motivation

Considerable resources (e.g., time and funding) have been invested in the research and development (R&D) of new drugs over the past few years, with the average cost worldwide (e.g., clinical costs) approaching RMB 15 billion [1]. Furthermore, all operating results fail due to uncertain and unpredictable clinical risks (e.g., whether new drugs can capture a large market share compared with conventional drugs). Thus, given R&D costs, the price of new drugs during the patent protection period tends to be extremely high. In contrast, generic drugs have significantly lower R&D costs since the cost of discovering a new API and conducting clinical trials can be saved (i.e., most generic formulations require bioequivalence tests, instead of clinical trials). The total cost is consistently less than RMB 15 million [2]. Accordingly, numerous generic drugs have been developed after the patent of reference drugs expired.

Generic drugs account for 95% of the drug market in China, 22.8% in Japan, and 18% in Europe. China's quality control system in terms of generic drugs has been lifting several strict restrictions over the past few decades [3]. As a result, a considerable number of generic drugs exhibit poorer quality and efficacy than the reference drugs in the drug market. The China Food and Drug Administration (CFDA) has issued the "Consistency Evaluation of Generic Drug" guidelines for the improvement of the quality of generic drugs [4]. It is imperative for pharmaceutical companies to improve the quality of their products to be equivalent to the reference drugs. Only those formulations that pass the consistency assessment are allowed to be marketed. Accordingly, how to produce generic drugs exhibiting the same qualities and efficacies as the reference formulations has become a formidable challenge facing China's pharmaceutical R&D industry.

Existing research on drug formulation has suggested that dissolution serves as a crucial indicator for quality assessment of new and generic drugs. Formulation technology has been confirmed as a critical step in pharmaceutical production, and the safety and efficacy of drugs are directly dependent on the effects of dynamic solubility on the bioequivalence of formulation products. Most inactive ingredients are naturally occurring substances (e.g., microcrystals and starches). They have complex functional structures where interactions occur (e.g., the inactive ingredients of spherical structures block the pores of the inactive ingredients of vesicular structures). Because of the complexity of factors affecting formulations, researchers tend to rely on their own experience instead of extensive calculations when designing experiments. Accordingly, experiments should be conducted to explore the best prescription amount and formulation process to ensure that the products have exemplary safety, efficacy, and quality control. However, most of the experiments conducted so far are unnecessary.

Dose Disintegration and Dissolution Plus (DDD PLUS) [5], developed by Simulation Plus Company in the United States, is the most widely used software for predicting the dynamic solubility and dissolution rate of solid dosage forms. Major users include the US FDA, the Chinese CFDA and other government agencies, the Academy of Military Sciences, the National Nano Center and other scientific research institutions, as well as Sanofi, GSK and other leading international pharmaceutical companies. First, the types of solute and other experimental conditions are selected. Subsequently, the physical and chemical properties of the API (e.g., its solubility, particle size, crystal form, and moisture) are input. Next, the type and proportion of excipients in its database are selected. The process type should be selected, and the process parameters should be set. Using its database, DDD PLUS will predict dissolution profile for a given formulation, whereas this is not the result. Afterward, one or more sets of experimental data should be used to revise

the results for the final model, used to predict subsequent experiments.

During the prediction of dynamic solubility and dissolution based on a large database, the DDD Plus prediction logic of in the practical project process does not capture all ingredients in the database, especially inactive ingredients. Most inactive ingredients are natural products exhibiting high molecular weight, such that their physicochemical properties are affected by molecular weight. For instance, the molecular weight of commonly used alginic acid ranges from 20,000 to 240,000 [6], such that the physical and chemical properties of products from different raw material manufacturers are different. In addition, the use is affected by the cumbersome processing steps.

The trend line is a curve used by researchers to analyze the trend of the dissolution curve as a function of the dissolution rate at each time point. It aims to evaluate the ratio of the dissolution rate in each period scientifically. However, the long-term lack of relevant studies and lack of a relevant equation to fit the scattered dissolution data scientifically lead to the main method to fit the trend line is polynomial fitting [7].

A simpler forecast method is offered as a substitute for DDD Plus. Some research predicts the solution result by ANN or regressing. Although they have high prediction accuracy, the required input database exceeds the number of orthogonal experiments. Thus, although the prediction of Dissolution Result is theoretically possible, it cannot be used in practice.

In addition to the dissolution curve, the dissolution behavior and disintegration time are also one of the main criteria for evaluating the dissolution. However, the dissolution test usually lasts 8-72 h and even longer for some delayed or controlled release tablets. Accordingly, few people observe the dissolution behavior, but they judge the dissolution behavior like the disintegration time by the dissolution curve. To solve the above-described problem, researchers record the dissolution process with a camera and analyze the images

with software (e.g., dissoGUARO PRO) to generate a volume change curve [8]. However, software (e.g., dissoGUARO PRO) fails to analyze the volume change of moving objects [8], such that the tablet with iron wires should be fixed and analyzed, whereas this method significantly affects the process and results of the dissolution experiment. Besides, the dissolution behavior should be recorded in advance and then analyzed using software, it is unlikely to monitor the dissolution behavior in real time and give a dynamic report. For some photolabile tablets, the monitoring process should be conducted under shading conditions. In existing research, red light has been adopted to reduce the effects of light on tablet stability, whereas several risks remain.

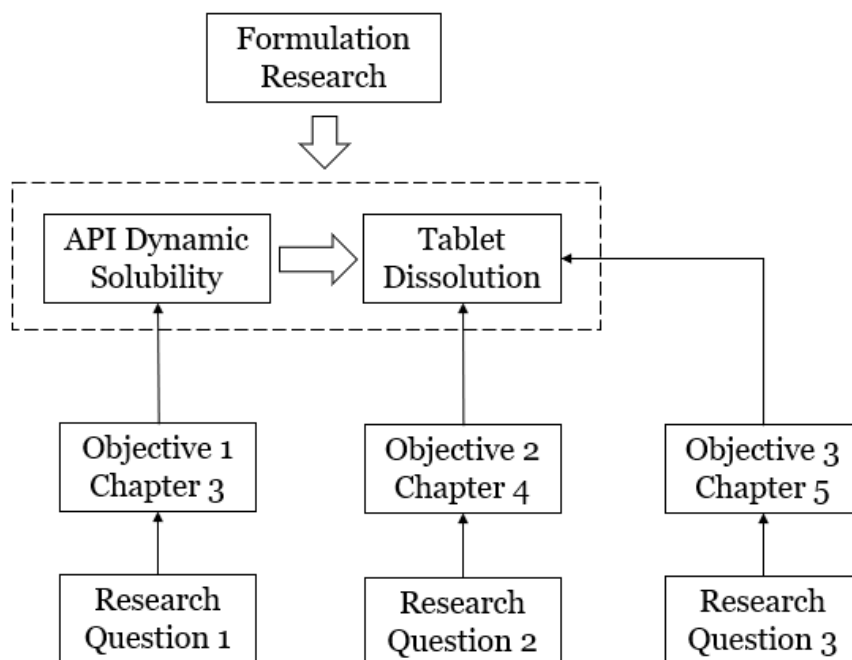


Figure 1.1 Structure of Research

1.2 Problem Statements

1.2.1 Problem 1: API Dynamic Solubility Parameter

DDD PLUS, extensively used in both companies and governments, is the most popular software for dynamic solubility prediction [5]. However, due to algorithm limitations, complex operations and parameter inputs are required for prediction making, which greatly increases the labor cost of the experiment. Besides, the existing trend line formulas are derived from simple polynomial fitting and fail to provide a scientific trend basis.

1.2.2 Problem 2: Prediction of Tablet Dissolution Rate

Despite the constantly conducted prediction research on drug dissolution by neural networks, an applicable neural network prediction method for drug dissolution has not been discovered thus far. The main reason is that previous works using ML, SVM require a large amount of input data to train the model, even larger than the number of orthogonal experiment design.

1.2.3 Problem 3: Image Recognition in Dissolution Phenomenon Monitoring

Image recognition has been extensively adopted to monitor the dissolution phenomenon, whereas it fails to monitor the dissolution phenomenon of the tablet in real-time in practical application. The existing dissolution monitoring system by image recognition cannot monitor the tablets in motion, such that the tablets are only fixed at the bottom of the dissolution cup, such that the dissolution results are affected. In this case, an alternative software is needed for image analysis of the tablets. In addition, for the experimental monitoring of photolabile tablets, there is no special treatment for its light source, and the previous works only use one specific tablet as the test sample, failing to prove its applicability in other conditions.

1.3 Research Questions

- Question 1: How to predict API dynamic solubility in an easy way and derive a scientific equation to fit the trend curve?
- Question 2: How to reduce the ANN model prediction input data less than orthogonal experiment?
- Question 3: How to monitor the moving tablet dissolution phenomenon even generally and automatically in a lightless environment?

1.4 Research Objective

According to the above analysis, this project aims to develop a computer-aided system based on ANN and Image Recognition, as well as supporting hardware equipment, to predict and monitor the entire dissolution process. The whole project mainly includes three main research objectives, and Figure 1.2 illustrates each research objective relationship.

1.4.1 Research Objective 1: To Develop an API Dynamic Solubility Prediction Model Using BPNN and RBFN

The prediction method of the existing application software DDD PLUS is complicated, which requires users to input the parameters of solubility, particle size, crystal form and dissolution environment, and correct the model based on the experimental data before the prediction. Besides, the existing curve fitting method is also a simple polynomial fitting, not a scientific one. Indomethacin will thus be taken as an example for the prediction of dynamic solubility of API under different experimental conditions by ANN, and the input data will be only the experimental data. To ensure the accuracy of the experimental data, the input data will be first filtered by the model. Additionally, a more scientific curve fitting equation will be derived based on the classic equation as a trend line.

1.4.2 Research Objective 2: To Develop a General Formulation Dissolution Prediction Model

ANN's prediction on solid formulation dissolution has not been applied to the practical R&D process. The combination of ANN and non-linear regression will be adopted to reduce the amount of input data for the prediction of the orthogonal experiment on the premise of prediction accuracy, such that it can be applied to the practical production, as well as the research and development. Furthermore, this study will realize the function of reusing related data by leaving a position for the potential factor and set it as 0.

1.4.3 Research Objective 3: To Develop a General Tablet Dissolution Phenomenon Recognition System

The method of region growth and image segmentation will be hereby used for image recognition of the tablet, and the volume change curve of the tablet will be calculated on the pixel area. Besides, a camera system, consisting of a photosensitive switch, an infrared light source and an infrared/transparent light camera, will be designed and installed on the dissolution apparatus. To meet the research and development needs of tablets in different colors, the system will also satisfy the automatic adaptation to the above-mentioned tablets by adjusting the parameter settings.

1.5 Report Organisation

In this study, the risk parameters in the formulation experiment were digitized according to the QbD risk assessment and input into the constructed computer model. Dynamic API solubility, component dissolution rate, and method dissolution behavior are three important quality indices for formulation research and development. Three PC models could be developed for this project to predict API dynamic solubility (Chapter 3), the rate of dissolution of formulations (Chapter 4), and system dissolution behavior (Chapter 5). In general, this model falls divided into three parts:

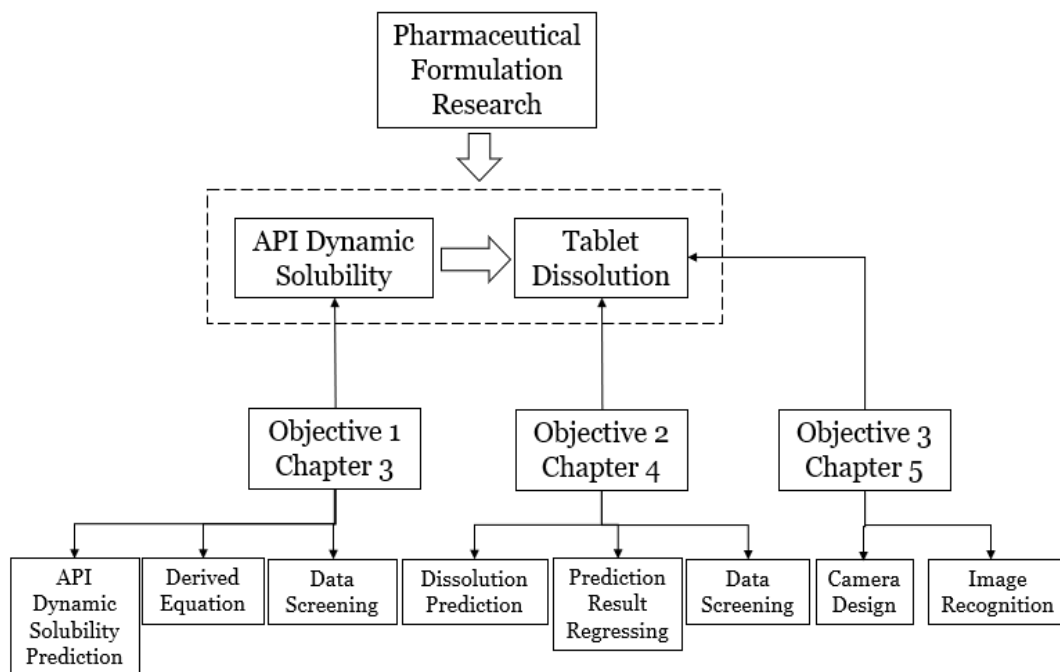


Figure 1.2 Structure of Report Organization

Chapter 3 refers to the prediction of the dynamic solubility of API under different conditions and the facilitation of the API solubility study at the pre-preparation stage.

Chapter 4 is to predict the dissolution profile when the formulation is screened and optimized. Under the insufficient input data, the results can be predicted using algorithms; when the prescription excipients are fine-tuned, the previous similar data can still serve as input data to train the model. Moreover, a data filtering function is substituted into the model to ensure the reliability of input data and avoid the input of abnormal data caused by experimental mis operation. Using the above-described system, the requirement of input data is controlled in orthogonal design number.

Chapter 5 is referring to using image recognition technology to monitor the dissolution behavior in the dissolution process. In general, it takes a long time (over 8 h) to conduct dissolution experiments. During the experiment, most of the time was unsupervised. Thus, it is difficult to monitor the experimental phenomena. The image recognition program and equipment we designed can recognize the moving tablets volume change and analyze the

phenomenon according to the characteristics of various dissolution behaviors in real time. Thus, the volume change curve, the video results, and the analysis results of the system's dissolution behavior can be obtained. Moreover, the image recognition system exhibits a night vision function, such that the dissolution behavior can be observed and analyzed under shading conditions.

Chapter 2 Formulation Research of New Drugs: Trends and Issues

2.1 Related Projects

2.1.1 DDD Plus

DDD Plus refers to an advanced computer program which is adopted by pharmaceutical scientists to simulate the disintegration of drugs in vitro and the dissolution of active pharmaceutical ingredients and excipients under diverse experimental conditions [9]. In general, it has been applied to the research on disintegration, dissolution patterns, and active ingredients of dosage forms. It is noteworthy that a single calibration experiment should be performed in the novel API form. Subsequently, DDD Plus is adopted to predict the effect of formulation changes or experimental parameter changes on the dissolution rate, such that accurate dissolution and disintegration rates can be provided. Moreover, using conventional trial-and-error methods is not long the only way to determine the final formulation [5, 10].

2.1.2 ANN in API Dynamic Solubility

Existing reports have suggested that nearly 40% of the market approved drugs and approximately 90% of compounds failed at the stage of research and development due to the poor dynamic solubility [11, 12]. Extensive studies have suggested that the poor aqueous solubility is a critical limitation against drug bioavailability [13]. Thus, existing research on the dynamic solubility of the drug with a low aqueous solubility is considered a significant step of drug research and development. In 2017, Safa has investigated the effect of different solubilizers on the dynamic solubility of indomethacin using ANN and well predicted its dynamic solubility at different solubilizer concentrations [14]. However, different rotational speeds of the stirring paddle should be explored to obtain the optimal

dissolution profile resolution in the process of developing the dissolution method [15]. Accordingly, an ANN model will be built in this study to predict the indomethacin dynamic solubility at different blade speeds.

2.1.3 ANN in Formulation Dissolution Prediction

Computational method serves as vital alternative to the experimental method. Machine learning algorithm is adopted to predict drug solubility [17] through the analysis and calculation of drug sequence data [16]. In general, commonly used machine learning algorithms comprise the support vector machine [18], neural network algorithm [19], random forest [20], and so forth. CCSOL [21] refers to an SVM-based prediction tool established by FEDERICO et al. in 2012. Besides, it was first proposed to use hydrophobicity, β folding, and α helix as the main features. PaRSnIP [22] refers to a tool established by REDA et al. in 2017. Moreover, existing research has suggested that a high proportion of exposed residues shows a positive correlation with drug solubility, whereas tripeptides comprising multiple histidine and tripeptide fragments are negatively correlated with drug solubility. SOL pro [23] extracted 23 groups of features from the first-level sequence for training the two-stage support vector machine (SVM) architecture. Furthermore, PROSOII [24] refers to a second-level logical classifier with modified Cauchy kernel probability density window model employed by PAWEL et al. However, SVM model has served as a classifier in most existing research, thus considerably limiting the processing ability and slow speed for big data. Deep learning has been confirmed as the core field of artificial intelligence technology thus far [25]. The deep learning model is capable of obtaining more nonlinear relations compared with "shallow learning" (e.g., SVM) [26]. Convolutional neural network is recognized as one of the crucial frameworks of deep learning. In addition, it has been extensively applied to image detection [27], face recognition [28], and audio retrieval [29], and it has achieved favorable results.

Nevertheless, it has been rarely applied to the research field of drug solubility prediction. SAMEERKHURANA et al. [23, 30, 31] of MIT built the DeepSol prediction model in 2018. One-hot was primarily adopted to encode drug sequences, and 21 x 1200 feature matrix was developed. A shallow parallel convolutional neural network model with seven convolution kernels of different sizes was built to predict drug solubility. Arthur Manda [32] predicted the dissolution release at different formulation component ratios using BP and Broyden-Fletcher-Goldfarb-Shanno (BFGS) [57] training algorithm, such that a more effective result was achieved than the RSM method. However, the input data of this method is also over orthogonal design.

The formulation is developed to conform to the needs of treatment or prevention in accordance with specific dosage form requirements. A formulation refers to a medicine that can be finally provided to the subject for use. The formulation has different dosage forms (e.g., tablets, pills, powders, tinctures, patches, injections, aerosols, sprays, ointments, and suppositories). Tablets are employed as the main dosage research of this project [24].

This project places a major focus on the prediction of formulation dissolution result during the changing of formulation composition. Table 2.1 presents the recent research comparison.

Table 2.1 The comparison of relevant projects

Projects	ANN model	Regression Method	Input and output	Amounts of Parameters and Samples
This project	ANN with back propagation net work	EDRM and RLRM	Input: The factors of tablets and the time points of dissolution. Output: Dissolution result in every time points.	21 samples, 7 parameters. 3 types of formulation composition.
Yixin Chen's Project [33]	No algorithm showed in the paper.	The average value of 10 times prediction	Input: The factors of Tablets Output: dissolution result of a single time point	22 samples 4 parameters
Jothi G. Kesavan's Project [34]		PROC REG	Input: hardness, friability, thickness, Output: disintegration time	23 samples , 3 parameters
Uttam MANDAL's project [27]	ANN Using Multilayer Perceptrons	Higuchi equation	Input: The composition of Tablets Output: Dissolution Result of a single time point.	13 samples , 3 parameters
Arthur Manda's Project [32]	ANN using BP and BFGS		Input: ratio of ingredient component Output: Dissolution Result of different time point	29 samples, 7 parameters

As depicted in Table 2.1, most regression methods fail to meet the stability requirement of prediction results due to the prediction times and the regression method. In this project, two methods cover the function of screening abnormal prediction result before regressing. In contrast, this project employs small amount of input data while achieving accurate prediction. Moreover, there are three types of formulation composition in this experiment. Despite the change of the formulation composition in the model, the relevant experimental data remains valid.

In an experiment process, experimental design serves as a vital optimization. In the design and production of experiments, researchers are required to predict the next step of the planned process in the following experiment and then perform experiments to decrease the number of experiments and increase the efficiency of the quantity project based on the experiments. However, human prediction requires a considerable amount of experience, and its results also have partial results prediction, so it is impossible to predict the possibility successfully. Optimization requires 100 or more experiments to complete, so researchers cannot import massive data for analysis [20].

Existing formulation prediction software mainly uses linear regression algorithms or ANN for calculation. For the simulation of the dissolution of the formulation, accurate prediction can be achieved when the data difference is not significant. However, when the data changes significantly, its accuracy will be reduced. Secondly, the existing formulation prediction software needs to input a considerable amount of data to fit the curve before each prediction, perform the curve correction automatically or manually, and then start the prediction. However, because the linear regression algorithm cannot solve the non-linear problem of the dissolution of the preparation, there is always a bottleneck in prediction accuracy. When it reaches a certain level, the accuracy will not increase even though the amount of input data is increased. For some applications of AI used in formulation prediction, they always use a considerable amount of data to train the model. Then they can receive a prediction result from the model. The problem is elucidated as follows: the project is nearly complete when the experiment result reaches that amount [21]. The prediction becomes meaningless.

Recent research [19, 35-48] has estimated tablet dissolution rates using ANNs, whereas none of the above-mentioned ANNs have resolved the issue of excessive input data, such that they become ineffective for pharmaceutical R&D.

In this project, machine learning methods are applied to artificial intelligence, and a neural network is adopted to address nonlinear problems and build a mathematical model and train the model with the experimental data. In this project, all the factors in the experiment, including prescription ratio, preparation process parameters, equipment parameters, and batch size, are digitized. Compared with conventional linear prediction methods, this method requires a considerable amount of known data for training, but the requirements for training data are entirely different. The required training data should only be the preparation development data regarding the prediction (e.g., the formulation data adopting the same or similar excipients as the target formulation, or those employing the same or similar process as the target formulation). Furthermore, the system can become an "omnipotent wise man" under the training of high-throughput data. The trained model is capable of predicting any relevant formulation variety under the premise of a small number of experiments.

This project features the constant effectiveness of predicting the results through algorithms even in the case of insufficient input data. To collect more data during the practical applications, a method is correspondingly designed in this model when the prescription excipients are fine-tuned. The previous similar data can still be used as input data to train the model. In addition, potentially existing problems in future high-throughput data input are also studied. First, since there is a considerable amount of data, abnormal data arising from operator errors and other reasons will be inevitably generated. Accordingly, the automatic data input method of this project is optimized, and artificial intelligence methods are adopted to detect input data and support the manual or automatic screening functions.

2.1.4 Image Recognition in Dissolution Monitoring

Image recognition technology serves as a critical tool to monitor different

experimental phenomena in the absence of experimental personnel. Recording and analysis of dissolution test videos are for the study of tablet behavior during the dissolution test. In Tjaša Felicijan's study [8], an iron wire was adopted to fix the tablet in the bottom of dissolution cup. Subsequently, a 10-second video of the dissolution process was recorded at the respective sampling point. Next, the video was fed into the software named *disso GUARO PRO* to reduce the noise of particles in the dissolution liquid prior to the calculation of the relative volume of the tablet using Microsoft Paint™. The photo instability tablets were recorded using a red-light source, and fixed tablets are capable of providing a stable image to capture but may affect the dissolution result. Besides, the real-time monitoring cannot be achieved by recording the video and then inputting it into other software, such that the monitoring also becomes more complicated. The red-light source is capable of basically solving the recording of photo instability tablets behavior, whereas infrared ray with lower light particle energy is a preferred choice. Wei Li [49] also recorded the dissolution test with a camera. Yutaka Morita [50] determined the trim size of tablets with three tablets fixed in the bottom of dissolution cup by an iron wire. A camera on the top of a dissolution cup was adopted to record changes on the surface of the above-described three tablets, and the space between the above-mentioned three tablets was used to calculate the area changes of tablet surface. Nevertheless, both the tablet fixing and the multi-tablets dissolution tests were limited following the guidance of drug administration in all countries. The overall design of the monitoring dissolution equipment was conducted in accordance with Chinese patents 201721009959 [51]. However, it functions rather simply, only equipped with the functions of real-time monitoring and video recording. In addition, the above-described two designs also fail to monitor the dissolution appearance of photolabile drugs under night vision. Kazarian and van der Weerd [52] explained the physical changes of hydroxypropyl methylcellulose (HPMC) and Buflurdil pyridinal

phosphate tablet during the lysis of the specially designed cells using simultaneous macrophotography and Fourier transform infrared spectroscopy attenuated total reflection (FTIR-ATR) imaging. By utilizing macro photography alone, changes in tablet appearance can be observed, but the substance concentrations cannot be quantified. However, by simultaneously employing FTIR-ATR imaging, both the water immersion and the drug concentration at different locations of the tablet were quantified, thus showing the correlation between the visually observed fronts (i.e., gel formation front and erosion front) and different yield of water. Morita et al. [53] also analyzed the disintegration time of rapidly disintegrating tablets with a camera. Tres et al. [54] adopted photographic images of the dissolution process in specially-designed flow cells to observe the tablets containing varying amounts of felodipine, a drug less soluble in water. To be specific, tablets with a small drug content dissolved completely. Besides, for the tablets with large drug content, they had the same appearance prior to the dissolution, but decreased in the tablet size after the initial swelling. Cao et al. [55] investigated the changes of photographic images with time in the research on the release of HPMC-coated tablets with a camera on the side of the dissolution vessel in a USP paddle apparatus. The initial swelling of the photographic images and the subsequent disintegration of the coating tablet show a correlation with favorable drug release. Tieu et al. [56] provided the side and bottom views of the dissolution vessel with a camera to evaluate different potentials. On that basis, the dissolution process can be monitored via the video. Besides, a conclusion is drawn that the camera is capable of easily evaluating the presence or absence of a tablet and its location in the dissolution vessel. Furthermore, they recommended the application of cameras in the preparation of formulas while studying the camera recordings of the dissolution systems under different vessels, vessel heating types, and camera angles, so as to evaluate the use of video surveillance in several devices.

The drug release process is controlled by two mechanisms of the dissolving drugs, namely disintegration and erosion, which often occur simultaneously [57-59]. Drug solubility affects the gel properties, and drug release and poorly soluble drugs reduce the entanglement of the polymer chains, thus reducing the gel strength [60, 61]. Thus, erosion turns out to be the main release mechanism. For the highly soluble drugs, the release from the matrix is considered primarily diffusion controlled. Furthermore, the highly soluble drugs may serve as a pore former in the gel, such that the perviousness of water can be enhanced. Moreover, they can be considered additional permeability contributors, such that the growth of the gel layer can be boosted [62].

Table 2.2 The comparison of related projects.

Projects	Novel Points
This project	<ol style="list-style-type: none"> 1. Image recognition is adopted to record the pixel to calculate volume change of tablet. The model is capable of automatically tracing the tablet. 2. Based on the volume change of disintegration and corrosion phenomenon, the system can give a result of phenomenon by a decision tree. 3. To address the recording problem of the phenomenon of light unstable drugs, the camera in the system has the recording function under infrared light. Besides, some infrared lights are installed on the camera.
Comparison of Related Articles	
Tjaša Felicijan's Project [8]	<ol style="list-style-type: none"> 1. Use red light to record light unstable tablets. 2. The tablet was fixed in the bottom of dissolution cup by using an iron wire. (It is already against the guidance of dissolution test).

Yutaka
MORITA's
Project [50]

3. Record the 10 seconds video of dissolution at the respective sampling point, then put the video into the dissoGUARO PRO software to reduce the noise of particles in the dissolution liquid. Then use Microsoft Paint to calculate the relative volume.
1. Three tablets were fixed in the bottom of dissolution cup by using an iron wire. (It is already against the guidance of dissolution test).
2. Use a CCD camera to record the 3 drugs surface changing of tablet. (Normally, in the dissolution test, one dissolution cup only allowed one tablets)
3. Use the space between the above-mentioned 3 tablets to calculate the area changing of tablet surface.

WEI LI's Project
[49]

Just use camera to record the phenomenon of dissolution test.

Conclusion

1. None of the methods in the articles can track the dissolution behavior in real-time and give the results in real-time.
2. In addition, only one article considers the dissolution test of photosensitive drugs. However, it still uses visible light. The infrared light source in this project has less influence on light unstable drugs.
3. Both methods of analyzing tablet changes require a fixed tablet. In this project, the image recognition system will track the tablet. It can be tested whether the tablet is stationary or moving in the water.

Comparison of Related Patents

patent number
201721009959
[51]

This utility model is correlated with a dissolution instrument monitoring system that comprises a camera module and a monitoring record host. The camera module covers a bracket, a camera, a universal connector, a switch, as well as a network cable;

the bracket includes an upper beam, a column, and a supporting foot, wherein a universal connector is installed on the lower surface of the upper beam, and the universal connector is then connected to the Camera connection, the network cable connects the camera and the switch. It has a network cable and a power cord from the switch to the monitoring host. The upper beam of the support is selected from a straight shape, a concave shape, or a back shape. The upper beam exhibits a hollow structure, and the network cable is routed from the cavity inside the upper beam. It does not take up the area of the experimental table. Besides, there is no messy cable, and the operation of the experimenter is not affected. Even if the solution of the dissolution medium containing salt is sprinkled on the experimental table, the camera and network cable will not be corroded. It is not accessible to damage and has a long life and good versatility. The system can carry out real-time monitoring while storing and recording videos.

Conclusion

The patent has conducted the overall design of the monitorable dissolution equipment. Nevertheless, its function is relatively single, which merely comprises real-time monitoring and video recording functions. Moreover, the above-described two designs are also unable to monitor the dissolution appearance of photolabile drugs under night vision. The dissolution monitoring system designed in this project can monitor the dissolution appearance under visible light conditions and monitor the dissolution appearance of some photo-labile agents under shading conditions. Furthermore, the system performs image recognition based on the volume and shape changes of the tablet during various dissolutions and records its volume changes. Subsequently, a curve is drawn, and which dissolution behavior it belongs to and which dissolution stage it belongs to are determined at the respective time point.

This study aims to offer a dissolution phenomenon behavior system using image recognition to record video and use the pixel area to calculate the tablet volume change. Besides, the model can automatically trace the tablet. The system, following the volume change of disintegration and corrosion phenomenon, employs decision tree method to give the results of phenomena. The camera in the system exhibits the recording function under infrared light to address the problem of recording the phenomenon of light unstable drugs. Furthermore, some infrared lights are installed on the camera.

2.2 Computer Science and Technology

2.2.1 Artificial Neural Networks

Given the potential of neural networks to replicate the human brain's ability to assimilate instances and make decisions with incomplete information, they have been recognized as one of the most significant existing computing tools by scientists and engineers over the past few years. ANN technology can simulate the pattern recognition ability of the neural network of a human brain and simulating the adaptation of the brain to new situations. Likewise, in terms of a single neuron in the brain, artificial unit neurons receive inputs from numerous external sources, process the information, and make corresponding decisions. The ANN comprises many processing units (PE) and artificial neurons. The strength of the connection among all units is determined by the coefficient or weight. The ANN simulates the learning and generalizing behaviors of the human brain through data modeling and pattern recognition of complex multi-dimensional problems. The significant difference between the ANN model and the statistical model is explained as follows: the former can generalize the correlation between both the independent and the dependent variables without a specific mathematical function. Accordingly, ANN excels in addressing nonlinear problems in multivariable and multi-response systems (e.g., the spatial analysis of quantitative structure-activity relationships in pharmacokinetic research

and structure prediction in drug development) [63-66].

2.2.2 Back Propagation

In back propagation (BP) networks, information only travels forward with layers as a medium, and it is not allowed to flow backwards. Errors obtained as the divergence between the output figure and the target value will be propagated back in the training process. On that basis, the errors can be reduced in the next iteration by changing the weight. Network training refers to find the combination of weights with the minimum number of errors, and the trained network exhibits a function to estimate a given model. In general, the error reducing function of the network is expressed in Eq. 2.1 as follows [67].

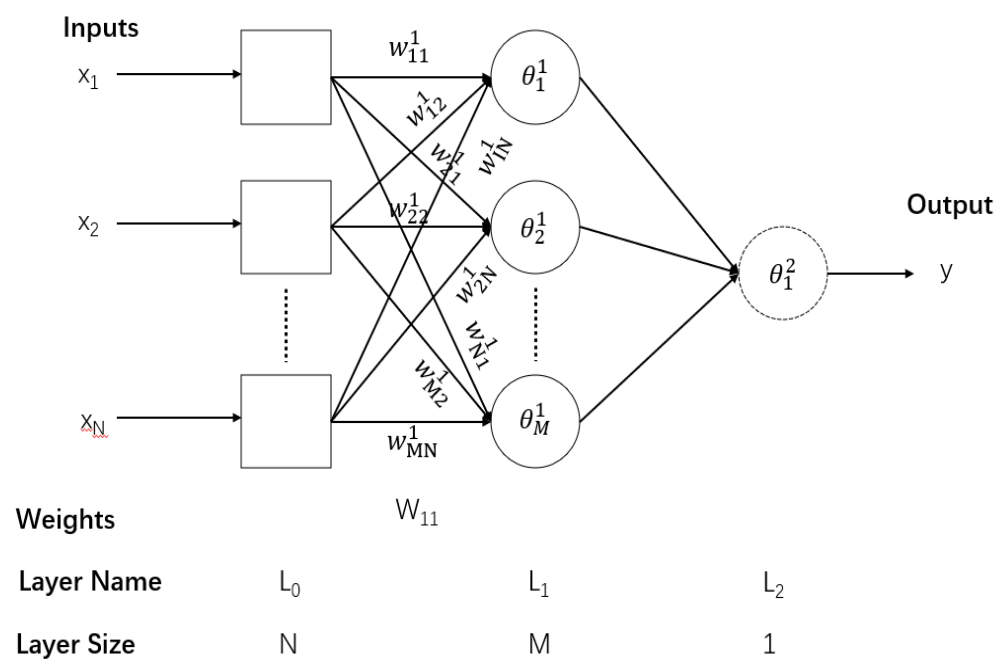


Figure 2.1 The representative BP network with one hidden layer.

$$E = \frac{1}{2} \sum_{i=1}^P \|o_i - t_i\|^2 \quad (2.1)$$

Where O_i , t_i , and p denote the number of output, target, and input modes, respectively.

Learning is conducted through error backpropagation in the BP network, in which the

descent gradient method is adopted. The weight is regulated by Eq. 2.2.

$$w_{ij}(t + 1) = w_{ij}(t) - \eta \frac{\partial E}{\partial w_{ij}} + \mu \Delta w \quad (2.2)$$

Where η represents the learning rate coefficients, μ denotes the momentum terms, and Δw signifies the weight change of the previous layer. The values of η and μ are between 0 and 1. Besides, the BP network is presented in Figure 2.1 [67].

2.2.3 Radial Basis Function Network

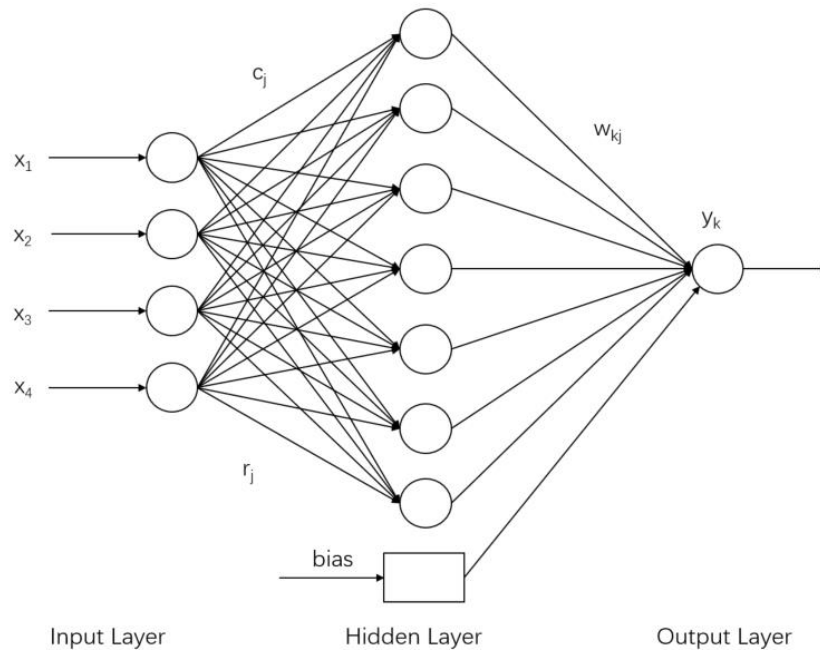


Figure 2.2 The RBFN structure

RBFN Theory has been extensively elucidated in existing research. In this thesis, the basic principle of RBFN is briefly introduced. Figure 2.2 presents the main network structure including the input, hidden, and output layers. Instead of processing information, the input layer diffuses the input vectors to the hidden layer, where there are several RBF units and biases. The respective hidden layer unit is endowed with a radial fundamental function, a relatively central position, as well as a suitable length. Neurons in the hidden

layer apply the radial fundamental function as a non-linear transfer function to determine input information [68].

2.2.4 Hue, Saturation, Value (HSV)

HSV, which was also named Hexcone Model by A.R.Smith in 1978, is a color space created by the properties of colors [69].

(a) Hue (H)

The angle degree is applied to measure the hue. According to the wavelength of the color, the range of value is between 0° to 360° . The value of hue was calculated anticlockwise. For example, red is 0° , green is 120° , and blue is 240° .

(b) Saturation (S)

Saturation is a value of the degree of color close to the spectrum. Colors are the result of mixing spectrum color and white color. The saturation means the percentage of spectrum color.

(c) Value (V)

Value (V) denotes the brightness of colors, which is associated with the radiance of light. The usual range of value is 0% (black) to 100% (white).

2.2.5 Morphological Transformations

Morphological transformation can be interpreted as a simple transformation based on the shape. The objects detected through morphological transformations basically cover binarization images, or color images sometimes. In general, there are two inputs and one output involved in the morphological transformation, i.e., the original image and the core as the inputs in the image after morphological transformation as the output [70].

2.3 Pharmaceutical Technologies

2.3.1 Tablet

In general, most tablets in daily use are regular oral solid tablets, and there are also other types of tablets (e.g., lozenges, sublingual tablets, oral patches, masticatory tablets, dispersible tablets, soluble tablets, effervescent tablets, vaginal tablets, vaginal effervescent tablets, sustained-release tablets, controlled-release tablets, enteric-coated tablets, and oral collapse tablets) [71].

Tablets are usually subject to the technological processes (e.g., wet granulation, dry granulation, and powder direct granulation). The properties of the API and formulation technology process make it inevitable for the dissolution behavior of tablets to differ from each other. There are two dissolution phenomena in standard tablets under normal circumstances (i.e., disintegration and corrosion). In general, the above-mentioned tablets are immediate-release formulations and their inactive pharmaceutical ingredients contain disintegrating agents, such that the disintegration and release of drugs are facilitated. For corrosion, the inactive pharmaceutical ingredients basically comprise weaker or stronger adhesives without disintegrating agents [71].

2.3.2 Dissolution Test

The absorption of drugs with active pharmaceutical ingredients is dependent on their dissolution or release from the formulations, their dissolution under physiological conditions, as well as their infiltration into the gastrointestinal tract. Dissolution takes on critical significance in the absorption of drugs, and the *vitro* dissolution test may predict their *vivo* behaviors. Given the above-mentioned analysis, a *vitro* dissolution test method has been developed for active ingredients. [72]

Two basic *vitro* dissolution measurement techniques (i.e., the stirring beaker method and the flow process) have been developed over the past four decades. Baskets and paddles

have been confirmed as the most applied dissolving devices worldwide. In existing research, the blade method has been employed with the dissolution cup simulating stomach, basket and paddle simulating gastrointestinal peristalsis, and dissolution medium simulating gastric juice. The parameter of the dissolution cup has been considered, and the digestion device is illustrated in the following figures [72].

The parameter of dissolution cup will be considered.

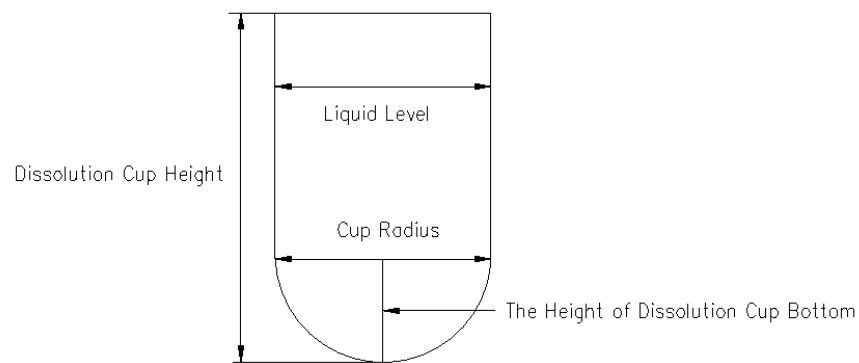


Figure 2.3 The internal parameter of the dissolution cup, affecting the solvent flow distribution.

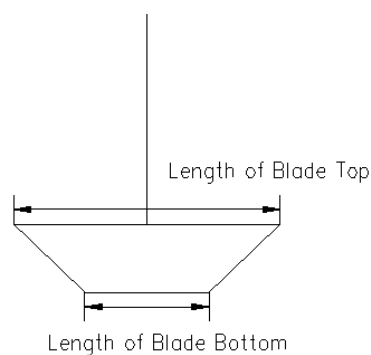


Figure 2.4 The parameter of the blade, the decisive factor of the solvent flow rate.

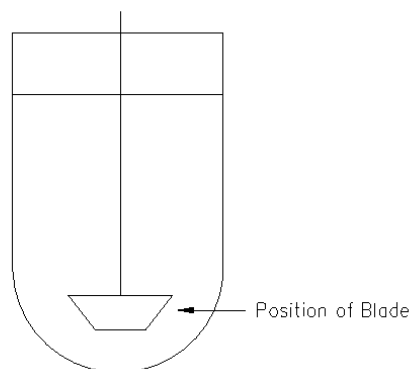


Figure 2.5 The location (height) of the blade, affecting the flow rate distribution in the dissolution cup

2.3.3 F2 Similarity

F2 Similarity Factor [73, 74] is a reference value for evaluating the similarities between two dissolution curves under the same dissolution conditions. Under the computing principle, R_t denotes the dissolution amount of reference formulation; T_t , the dissolution amount of the self-study sample; $(R_t - T_t)^2$, the square of the difference value between the dissolution amount of the reference formulation and that of the self-study sample at the same time point. Notably, in the next step, the sample number (n) is divided directly. Thus, the 10% deviation represented by $F2=50$ refers to the indicated dissolution of 10% of the average difference between the sample points, i.e., the average deviation, rather than the average relative deviation. Table 2.3 lists the supporting data.

Table 2.3 The contrast of average deviation and factor F2.

Average Deviation	2%	5%	10%	15%	20%	30%	50%	70%	100%
F2 Factor	83	65	50	41	35	26	15	8	0

The F2 Similarity Factor is determined through the reciprocal transformation of the logarithmic square root of the total square errors. Similarities between the two drug release curves are elucidated in the following.

Where n represents time points; R_t denotes the percentage of dissolution of drug A (as reference drug); T_t expresses the percentage of the dissolution of drug B (as experimental drug). The value of F_2 should range from 50 to 100 for homologous curves and the equation of F_2 is presented as follows.

The equation of F_2 is illustrated below,

$$F_2 = 50 \log \left\{ \left[1 + \frac{1}{n} \sum_{i=1}^n (R_t - T_t)^2 \right]^{-0.5} \times 100 \right\} \quad (2.3)$$

Where R_t denotes the dissolution rate of the reference formulation; T_t represents the dissolution rate of the experimental sample; $(R_t - T_t)^2$ expresses the square of the difference value between the dissolution amount of the reference formulation and that of the experimental sample at the same time point; n is the number of samples [75].

2.3.4 Dynamic Solubility

Dynamic solubility indicates the mass curve of the solute dissolved in a known quantity of solvent under certain conditions. The dynamic solubility of API serves as a vital parameter affecting the dissolution rates of insoluble drugs [76].

2.3.5 Biopharmaceutics Classification System (BCS)

Biopharmaceutics Classification System (BCS), a guiding framework issued by the Food and Drug Administration (FDA) in the United States, has evaluated the API from three aspects (i.e., permeability, solubility, and immediate release), as demonstrated below [77].

Table 2.4 The BCS classes

Class I	High Solubility and High Permeability
Class II	Low Solubility and High Permeability
Class III	High Solubility and Low Permeability
Class IV	Low Solubility and Low Permeability

2.3.6 Disintegration

Disintegration refers to the physical dissolution of a medicinal product before its absorption while disintegration time denotes the time required to disintegrate pills, tablets, and capsules into granules under certain conditions. The latter is one of the main factors that affect the dissolution of oral drugs [78,79].

2.3.7 Corrosion

Corrosion refers to the gradual disappearance of a non-disintegrating solid formulation during dissolution. In general, it occurs during the dissolving of sustained-release formulations [80].

2.4 Pharmaceutical Instruments

2.4.1 High Performance Liquid Chromatography (HPLC)

High performance liquid chromatography (HPLC), i.e., a crucial branch of chromatography, employs liquids as the mobile phase and adopts the high-pressure infusion system to pump the single solvents with different polarities or mix solvents with different proportions into a stationary phase chromatographic column. The respective component in the column is separated and subsequently enters the detector for testing, such that the sample analysis can be conducted. The above-described method has served as a separation and analysis technique in a wide variety of areas (e.g., chemistry, medicine, industry, agriculture, commodity inspection, and law inspection) [81].

2.4.2 Ultraviolet-Visible

Ultraviolet-Visible (UV-VIS) Spectroscopy, commonly applied for qualitative and quantitative analysis, can measure the absorbance of a substance at a specific wavelength or in a specific wavelength range in the ultraviolet region.

Ultraviolet spectra are the absorption spectra of substances in the near-ultraviolet and

visible regions of 200~400 nm and 400~850 nm, respectively. In general, UV-VIS spectroscopy is adopted to identify, inspect, and determine the drug content in the inspection range of 190~900 nm. Moreover, it applies to the analysis of traces and components with the sensitivity of 10^{-4} ~ 10^{-7} g/ml or lower.

The application of UV-VIS spectroscopy as an analytical method can be justified by two main reasons. The first is light absorption, i.e., the absorption degree of light by substances. From the perspective of electromagnetic waves, all electromagnetic waves are identical except for the only difference in their wavelengths or frequencies. The order of wavelengths from the shortest to the longest is R-rays, X-rays, ultraviolet, infrared, and microwaves. Energy exchange occurs when electromagnetic radiation source interacts with a matter. Spectroscopy can fall into absorption spectroscopy and emission spectroscopy in accordance with the matter conversion direction and radiation energy. When a sample is irradiated by an electromagnetic radiation source, its atoms or molecules will absorb some photons with suitable energy, while the absorption lines or bands will appear at the corresponding wavelength positions, such that the absorption spectrum is formed. Absorption spectroscopy refers to the quantitative, qualitative, and structural analysis employing the absorption spectra of substances. UV-VIS spectroscopy is a type of molecular absorption spectroscopy generated by the transition of outer electrons of atoms in a molecule. Ultraviolet absorption is primarily dependent on the electronic structure of the molecule. Accordingly, the ultraviolet spectrum is known as the electronic spectrum. The light absorbing degree of a substance is examined at different wavelengths. Absorption spectrum refers to a curve drawn with the wavelength as the abscissa and the absorbance as the ordinate. The wavelength range examined in the ultraviolet-visible region is called an ultraviolet-visible spectrum, abbreviated as an ultraviolet spectrum [82].

2.4.3 Dissolution Tester

The essential hardware components of the instrument are equipped with a computer image recognition system, natural light, infrared light, an IR/visible light camera connected to the PC host, a Water Bath Box system, a sampler, a removable window curtain, a dissolution cup, as well as an impeller with the motor control system and the temperature control system of dissolution by the host.

In general, the determination methods of their dissolution and release fall into the basket, paddle, small cup, paddle dish, revolving barrel, flow pool, and simple reciprocating method in accordance with the different properties of solid formulations. To be specific, the basket, paddle, and small cup methods are the most extensively applied. Limited by the observation conditions, only the paddle and small cup methods are applied to the system proposed in this study [83].

A dissolution tester exhibiting an automatic sampling function is presented in Figure 2.6



Figure 2.6 A dissolution tester with a sampling function, with a dissolution tester on the left and a sampling instrument on the right.

2.5 Chapter Summary

The dynamic solubility of API in different solvents serves as a critical parameter affecting the dissolution rate of the formulation. The most applied dynamic solubility prediction software is DDD Plus, which utilizes the linear regression method for calculation. However, the dynamic solubility data is often non-linear, hence this method cannot achieve accurate prediction. Moreover, after predicting the dynamic solubility, the software provides a curve fitting formula that is binomial regression, which is of no guiding significance.

Existing methods of predicting dissolution curves based on computer models fall into two categories (i.e., to predict the dissolution results using neural networks and to predict the model through linear regression or formula calculation). However, all methods require a considerable amount of data to train the model. In contrast, the computer model method is provided with a ratio of factors to a sample size of 1:6.5, whereas that of the second method is 1:5.4 and 1:7.6. In the actual research and development process, the experimental design often follows the orthogonal design. For example, an orthogonal experiment with 4 levels and 4 factors requires 16 experiments, while an orthogonal experiment with 3 levels and 3 factors requires 9 tests, indicating that the number of experiments required by the previous method has exceeded that of conventional orthogonal design.

Moreover, the previous method could only predict a single prescription, and the data were not applicable when the prescription composition was fine-tuned. Furthermore, the input data have not been filtered in existing research, whereas unfavorable operational or experimental environments often result in the generation of abnormal data in the actual application process. Accordingly, the credibility of the data should be screened before they are input into the model.

The existing dissolution behavior monitoring methods are divided into two types. The

first type is to video the dissolution process with a camera; so that the experimenter can analyze the experimental phenomenon by watching it; the second type is to record the dissolution process with a camera, and then process and calculate the images via software. Given the existing patents and articles, the second method has been applied to two projects, where the common problem is that the tablets fixed with iron wire will affect the dissolution results. The dissolution test of photosensitive drugs has been considered by merely one article, whereas it still utilized visible light. Moreover, the two dissolution behavior monitoring methods cannot track the dissolution test of the tablet in real time.

Thus, this study will design a real-time and accurate dissolution appearance analysis system with automatic tracking and infrared recording functions using image recognition technology.

Chapter 3 Artificial Neural Network in Predicting API Dynamic Solubility

3.1 Chapter Introduction

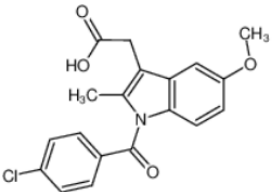
In this chapter, an ANN model with a derived equation, instead of DDD Plus, is built to predict the dynamic solubility. Due to the effect of dynamic solubility on the bioequivalence of the formulations of drug products, formulation technology has been one of the most important steps of pharmaceutical production that directly affect the safety and effectiveness of drugs. Most of the inactive components are natural-occurring substances like microcrystalline and starch with a complex and functional structure, making it possible for them to have interaction effects (e.g., the inactive component with a ball structure plugs the hole of that with a vesicular structure). Given the complexity of factors on formulations, researchers are inclined to rely on their personal experience instead of extensive calculations while planning experiments. Thus, numerous experiments have been performed to investigate the optimal prescription volume and the formulation process to ensure the favorable safety, effectiveness, and quality control of the products, whereas most of which are indeed unnecessary and a waste of resources [84]. Currently, two methods are capable of predicting API dynamic solubility using ANN, i.e., the Back Propagation Neural Network and the Radial Basis Function Network. Furthermore, a newly derived scientific dynamic solubility equation is developed to fit the dissolution curve.

3.2 Materials and Methods

3.2.1 Materials

Indometacin Powder [85]

Table 3.1 API parameters

API Parameters	
API name	Indometacin
Full name	1-(4-Chlorobenzoyl)-5-methoxy-2-methyl-3-indoleacetic acid
CAS number	53-86-1
Molecular formula	C ₁₉ H ₁₆ ClNO ₄
Structure formula	
External properties	The white or yellow crystalline powder
Molecular weight	357.78800
Polar surface area (PSA)	68.53000
Log P	3.92730
Melting point	158~162°C
Crystal form	All the APIs in this research are from the same batch, hence the crystal form is not considered.
BCS classification	II
Solution Parameters	
Solution name	Pure water
pH	7.0
Temperature	37±0.5°C
Ionic strength	N/A
Volume	500 ml

3.2.2 Preparation of API Sample

First, the Indometacin API was dried under reduced pressure at 60°C for 2 h and pulverized in a pulverizer for 30 sec. Then, the pulverized Tinidazole powder passed through an 80-mesh sieve.

3.2.3 Assay for Indometacin API

The standard substance was dried for 4 h at 105°C to constant weight to prepare the standard solution preparation. 28 mg accurately weighted anhydrous substance dissolved into 100 ml methanol in a 100 ml volumetric flask. 5 ml methanol solution was injected into a 50 ml volumetric flask. Next, degassed pure water at pH 7.0 was added till the volume of the solution reached 50 ml. Afterward, the volumetric flask with the prepared solution was put into the ultrasonic mixer for 2 min. The obtained solution was the standard solution.

Next, ultraviolet spectrophotometry was used to analyze the concentration of the sample. A quartz absorption cell filled with blank solvent was put in the ultraviolet-visible (UV-VIS) spectrophotometer to build a baseline, and the reference line was built using the same method. Furthermore, the absorption wavelength was determined at 320 nm. The quartz absorption cell was measured using the UV-VIS spectrophotometry with the reference solution.

The above-described samples should be examined from low concentrations to high concentrations to reduce the experimental errors caused by the rest of the samples. After the respective measurement, the quartz absorption cell should be cleaned with pure water more than three times and then dried by lens paper.

The absorbance of each sample at 320 nm was recorded and then calculated by Equation 3.1.

$$=W_s \times \left[\frac{A_{T(n)}}{A_s} + \sum_{i=1}^{n-1} \left(\frac{A_{T(i)}}{A_s} \times \frac{1}{90} \right) \right] \times \frac{V'}{V} \times \frac{1}{C} \times 90 \quad (3.1)$$

Where W_s denotes the standard weight, C represents the total mass of solute, A_t

expresses the reference solution absorbance at 320 nm wavelength, A_s is the sample absorbance at 320 nm wavelength, V' is the volume of the reference solution, and V represents the volume of the sample solution [86].

3.2.4 Design of Experiment

Step 1: Preparation of Indometacin sample. Six copies of 25 mg Indometacin API were weighed by analytical balance to ensure the samples' consistency. 900 ml degassed pure water at pH 7.0 was added to the respective dissolution cup to create a stable dissolution environment.

Step 2: Dissolution tester parameter setting. The temperature of the system was set as 37.5°C to stabilize the temperature of the dissolution cup at $37 \pm 0.5^\circ\text{C}$. Six different rotation speeds of paddles were set in six parallel experiments (i.e., 25 rpm, 50 rpm, 75 rpm, 100 rpm, 125 rpm, and 150 rpm revolution per minute) to design a gradient experiment. 10 ml of the sample was taken at 5 min, 10 min, 15 min, 30 min, 45 min, 60 min, 120 min, 240 min, and 480 min to be filtered using 0.45 μm filtered hydrophilic film and then injected into the test tube. A 10 ml dissolution medium was introduced to the cup after sampling to keep the constant volume of dissolution medium. A total of six dissolution cups were employed in the respective parallel experiment to repeat this experiment six times to reduce the experimental errors.

Step 3: Six prepared API samples were added from the top of dissolution cups during the dissolution test. After sampling, the test tube was covered by the parafilm to keep the samples from evaporating.

3.2.5 Core Equation

The core equation is derived to solve the academic gap in API dynamic solubility trend equation in curve fitting.

- Core Equation Derivation

The Noyes Whitney Equation: Nernst and Brunner proposed that during the dissolution process, a rapid equilibrium (i.e., saturation) is achieved at the solid-liquid interface and then diffusion occurs through a small diffusion layer of the stagnant solution to enter the bulk solution. In most situations, the diffusion through the diffusion layer is rate-controlled, such that the heterogeneous dissolution process is transformed into a homogeneous one. Thus, this model is also termed the movie model [87]. The dissolution rate can be expressed as follows:

$$\frac{dM}{dt} = \frac{DA}{h} (C_s - C_t) \quad (3.2)$$

Herein, D represents diffusion coefficient while A denotes solid surface and h is diffusion layer thickness.

$$\frac{dQ}{dt} = DA \frac{(C_s - C_b)}{h} \quad (3.3)$$

When the volume of solvent (V) does not change dramatically during the dissolution process, the concentration in the volume solution could be expressed as follows,

$$C_b = \frac{Q}{V} \quad (3.4)$$

Herein, Q represents the amount dissolved while V denotes the solvent volume.

$$C_b = \frac{\frac{4}{3}\pi\rho(a_0^3 - a^3)}{V} \quad (3.5)$$

Hixson-Crowell cubic root law was initially obtained from the hypothesis that the dissolution rate and the surface area of spherical particles are in the direct ratio. [88] Moreover, it can be obtained from a simple diffusion layer model.

$$W^{1/3} = W_0^{1/3} - k_{1/3}t \quad (3.6)$$

$$K_{1/3} = \left(\frac{4\pi\rho N}{3}\right)^{1/3} \frac{DC_s}{\rho h} \quad (3.7)$$

Herein, W denotes the particle weight at time t, W₀ represents the initial particle weight, K_{1/3} signifies composite rate constants, ρ implies the density of the particle, N is the number of particles, D denotes the diffusion coefficient, C_s represents the solubility, h is the diffusion layer thickness, and k' represents a constant.

Wang-Flanagan Equation: By calculating the curvature of the concentration gradient of the diffusion layer around the spherical particle, Wang and Flanagan devised a general method for diffusion-controlled dissolution of a single particle [89].

$$\frac{DC_s}{\rho h} t = a_0 - h \ln \frac{h+a_0}{h+a} \quad (3.8)$$

$$\frac{DC_s}{\rho h} t = \left(\frac{3w_0}{4\pi\rho}\right)^{1/3} - \left(\frac{3w}{4\pi\rho}\right)^{1/3} - h \ln \frac{h+\left(\frac{3w_0}{4\pi\rho}\right)^{1/3}}{h+\left(\frac{3w}{4\pi\rho}\right)^{1/3}} \quad (3.9)$$

Brooke Equation: Brooke obtained the precise equation of powder dissolution with logarithmic normal distribution based on the cubic root single-particle dissolution model. Although the calculation is complex, it can be accomplished by calculators using regular distribution tables [90].

$$\begin{aligned} W_t = r e^{3\left(\mu+\frac{3\sigma^2}{2}\right)} & \left(1 - F\left[\frac{\ln(\tau) - (\mu + 3\sigma^2)}{\sigma}\right]\right) \\ & - 3r\tau e^{2(\mu+\sigma^2)} \left(1 - F\left[\frac{\ln(\tau) - (\mu + 2\sigma^2)}{\sigma}\right]\right) \\ & + 3r\tau^2 e^{\left(\mu+\frac{\sigma^2}{2}\right)} \left(1 - F\left[\frac{\ln(\tau) - (\mu + \sigma^2)}{\sigma}\right]\right) - r\tau^3 \left(1 - F\left[\frac{\ln(\tau) - \mu}{\sigma}\right]\right) \end{aligned}$$

$$F(x) = \frac{\sqrt{2}}{2\pi} \int_{-\infty}^x e^{-x^2/2} dx$$

$$\tau = \frac{2kC_s}{\rho t} \quad (3.10)$$

Herein, $W\tau$ denotes the undissolved weight, μ represents the mean particle diameter, σ is its standard deviation, $r = \pi\rho N/6$, C_s signifies the particle solubility, ρ denotes the particle density, and k represents a constant.

- Derivative Equation

The following equation is the derivative formula of the dynamic solubility of soluble solid elements.

$$y = m + 3k(x - n) - 3k^2(x - n)^2 + k^3(x - n)^3, \quad (3.11)$$

$$\text{when } \frac{dx}{dy} = 0, \quad y=C_s$$

This equation was derived by applying Hixson-Crowell cube root law and Nernst-Brunner expression. Herein, n is the correction coefficient of the experimental environment influencing factor, k represents a constant concerning the physicochemical property and experiment conditions, and m denotes the correction coefficient of sample wetting.

3.2.6 Curve Fitting

Nowadays, people who engage in the research and development of solid formulations generally apply the dissolution curve to analyse the drug dissolution behavior. Under normal conditions, binomials tend to be selected as a trendline to guide the analysis. Nonetheless, with no scientific basis support, binomial formulas do not apply to the description of dissolution behavior. To reduce the experimental errors, the artificial model employs Equation 3.11 as a curve fitting formula, which more closely resembles the practical experiment formula. The fitting curve built by the model is shown as follows.

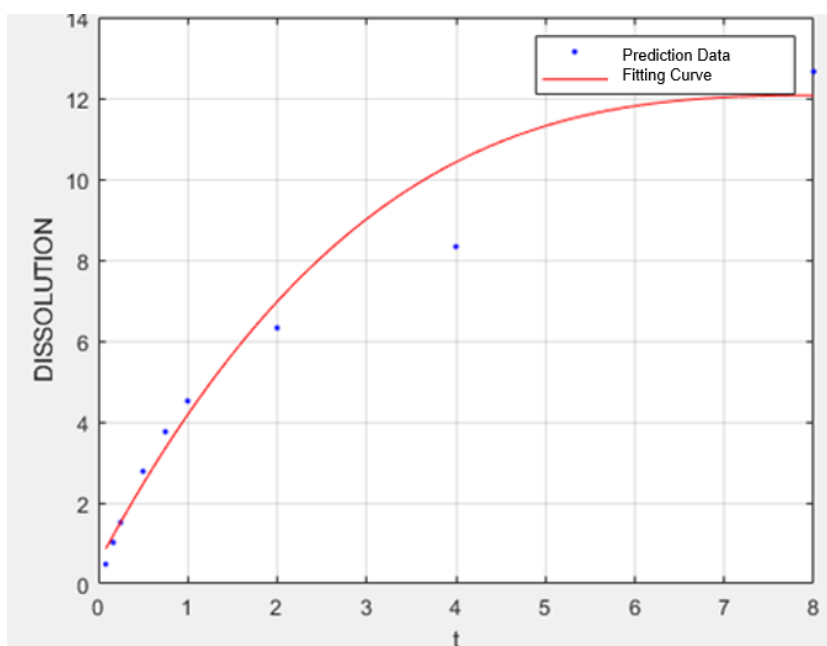


Figure 3.1 The fitting curve with a limit of Equation 3.11

As depicted in Figure 3.1, the fitting curve generated by the model is a regression

curve of scattering that follows Equation 3.11. In figure 3.1, x-axis “t” is dissolution time(hours), y axis “Dissolution” is percentage of API release. Accordingly, this model function can be utilized to address the problem of the dissolution curve trend line.

3.3 Result and Discussion

3.3.1 BPNN and RBFN Modeling

There are typically three layers contains in an ANN model, including an input layer, hidden layers, and output layer. In this model, the input layer includes three elements which are time value of dissolution test, percentage of API concentration and blade speed. In the output layer, there is a neuron that represents the API dynamic solubility value. To avoid the problem of underfitting during the machine learning process, the number of neurons in the hidden layer should be large enough to solve the problem. On the other hand, too many neurons in the hidden layer may result in overfitting. Following the trial-and-error test, the number of neurons in the hidden layer has been determined as three. The ANN model structure is illustrated in Figure 3.2 below.

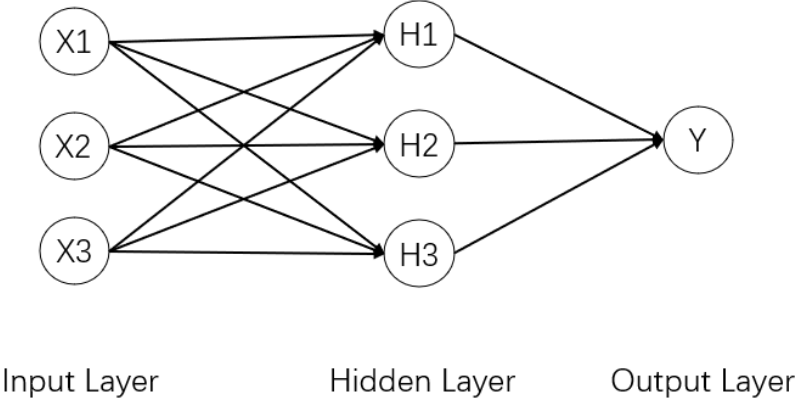


Figure 3.2 The structure of ANN model

In Figure 3.2, the X1-X3 represent: X1, time value (minutes) of dissolution test.; X2, concentration of API, X3, blade speed; and H1-H3, neurons of Hidden Layer. The output is Y, which is the percentage of API released at the time points indicated in X1.

A diagram of the model process is shown in Figure 3.3:

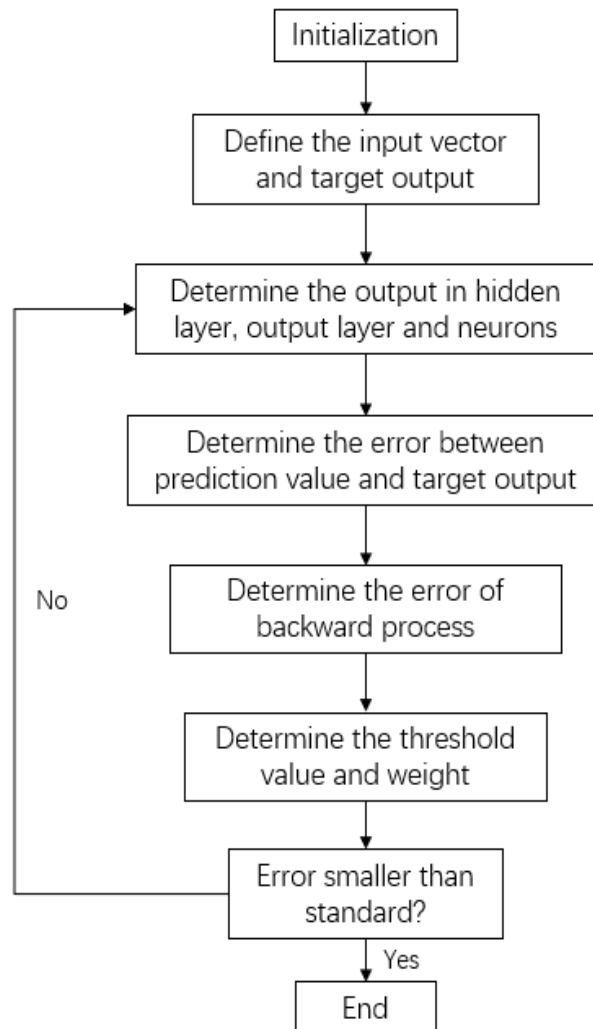


Figure 3.3 ANN model process

The input vector of these three feedforward neural network layers is $X=(x_1,x_2,\dots,x_n)^T$, and if $x_0 = -1$ is added to the input layer, the threshold value of

neurons in the hidden layer can be determined. Assuming that the hidden layer has a vector assume $Y = (y_1, y_2, \dots, y_m)$ as a vector of the hidden layer, and $y_0 = -1$ was added as a vector to the hidden layer, the threshold value for the output layer should be $O = (o_1, o_2, \dots, o_i)^T$. The output vector of the target is $d = (d_1, d_2, \dots, d_k)^T$, the weights between the input layer and hidden layer is $V = (v_1, v_2, \dots, v_m)^T$ and v_j is the corresponding weight vector of j^{th} neuron in the hidden layer. $W = (w_1, w_2, \dots, w_k)^T$ is the weight between the hidden layer and output layer (w_k is the corresponding weight vector of the k^{th} neuron in the output layer.)

Hence, in the input layer of the model the input is $net_j = \sum_{i=0}^n v_{ij} x_i, j = 1, 2, \dots, m$ and the output is $y_j = f(net_j), j = 1, 2, \dots, m$. Besides, the input of the output layer is $net_k = \sum_{j=0}^m w_{jk} y_j, k = 1, 2, \dots, i$, and the output is followed $o_j = f(net_k), j = 1, 2, \dots, i$.

Here, the continuous and derivable function $f(*)$ is the activation function which is the Sigmoid function given by $f(a) = \frac{1}{1 + e^{-a}}$.

Nevertheless, an error will appear between the actual output and the target output. The function of the error $E = \frac{1}{2} \sum_{k=1}^i (d_k - o_k)^2$ is determined by the weight and threshold values. It is possible to improve the performance of the model by adjusting the value of these parameters.

3.3.2 Data Prediction

The same dose of indomethacin was used in this experiment to avoid the errors caused by inter-batch differences. Six rotation speed conditions were measured, i.e., 25 rpm, 50 rpm, 75 rpm, 100 rpm, 125 rpm, and 150 rpm. The temperature condition was determined

as $37.5 \pm 0.5^\circ\text{C}$. Each experiment was repeated six times to reduce experimental errors. The experimental data were calculated by Equation 2.1 to obtain the F2 value. According to the United States of America National Food and Drug Administration (FDA) guidance, $F2 \geq 50$ indicates that the two dissolution curves are similar. The experimental result data is presented as follows:

Table 3.2 The dissolution data of Indometacin in water

Indomedacin Water						
Time (Minute)	25rpm	50rpm	75rpm	100rpm	125rpm	150rpm
0	0	0	0	0	0	0
5	0.25	0.36	5.07	3.50	5.47	4.23
10	2.10	1.45	4.64	4.59	8.61	6.92
15	4.28	3.51	6.29	6.72	11.62	9.30
30	6.25	7.08	10.72	12.26	16.09	14.63
45	7.24	10.48	14.15	16.95	20.28	18.18
60	10.09	11.50	16.44	21.06	27.31	21.80
120	14.44	16.49	22.94	28.67	29.96	29.55
240	20.71	22.64	29.20	33.74	35.09	33.96
480	29.14	33.52	35.59	38.60	40.09	40.97

Figure shows the percentage of Indomedacin dissolved in the water in different time point under changed rotate speed condition.

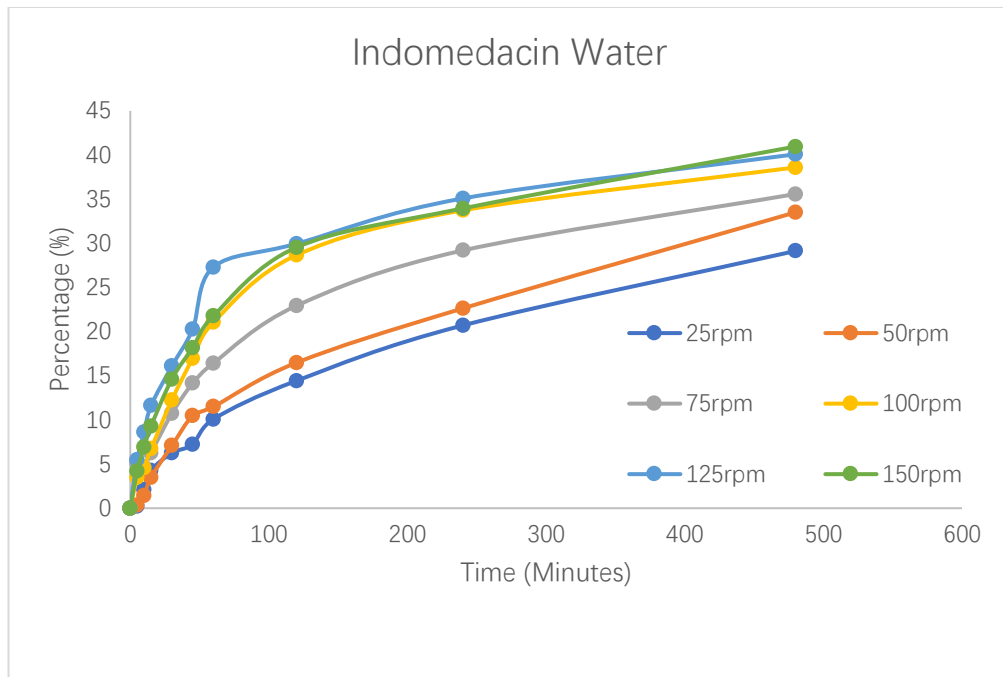


Figure 3.4 The dissolution test results of Indometacin in water

Five sets of the experiment data rather than the average value were randomly selected from six experiments and input into the model. Moreover, the data of group 6 was predicted by the model. We compared the above-mentioned data with the practical value, and the results are presented as follows:

Table 3.3 The comparison of real data, BPNN prediction data, and RBFN prediction data under 25 rpm/water condition

Indometacin 25rpm water			
	Reference	BPNN	RBFN
5min	0.25	0.88	0.57
10min	2.10	2.13	1.69
15min	4.28	3.28	2.77
30min	6.25	6.23	5.73
45min	7.24	8.51	8.31
60min	10.09	10.27	10.52
120min	14.44	14.33	16.24
240min	20.71	18.54	17.88
480min	29.14	29.22	29.92
F2		93.24	88.92

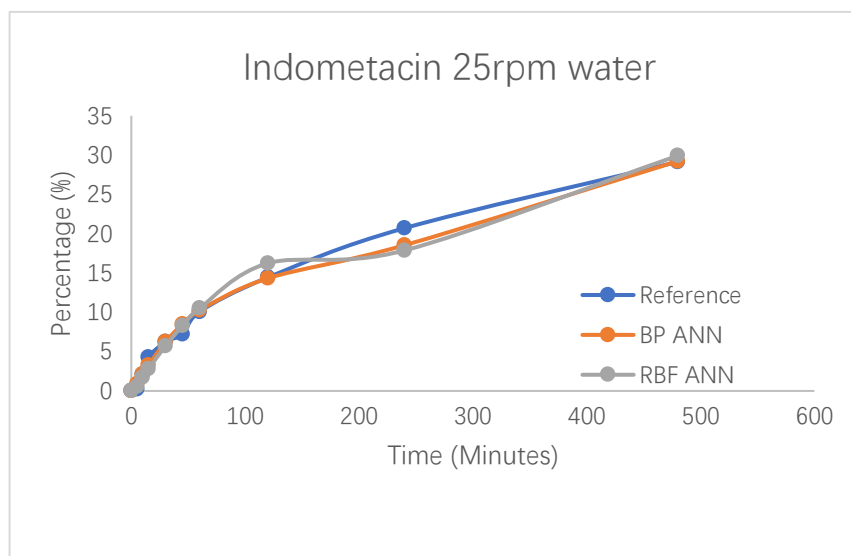


Figure 3.5 The dissolution test results of Indometacin under 25 rpm/water condition

Table 3.4 The comparison of real data, BPNN prediction data, and RBFN prediction data under 50 rpm/water condition

Indometacin 50rpm Water			
	Reference	BPNN	RBFN
0min	0	0	0
5min	0.36	0.76	0.79
10min	1.45	2.38	2.09
15min	3.51	3.89	3.33
30min	7.08	7.76	6.78
45min	10.48	10.73	9.81
60min	11.50	12.98	12.45
120min	16.49	18.05	19.60
240min	22.64	23.64	22.82
480min	33.52	33.28	32.77
F2		93.50	90.62

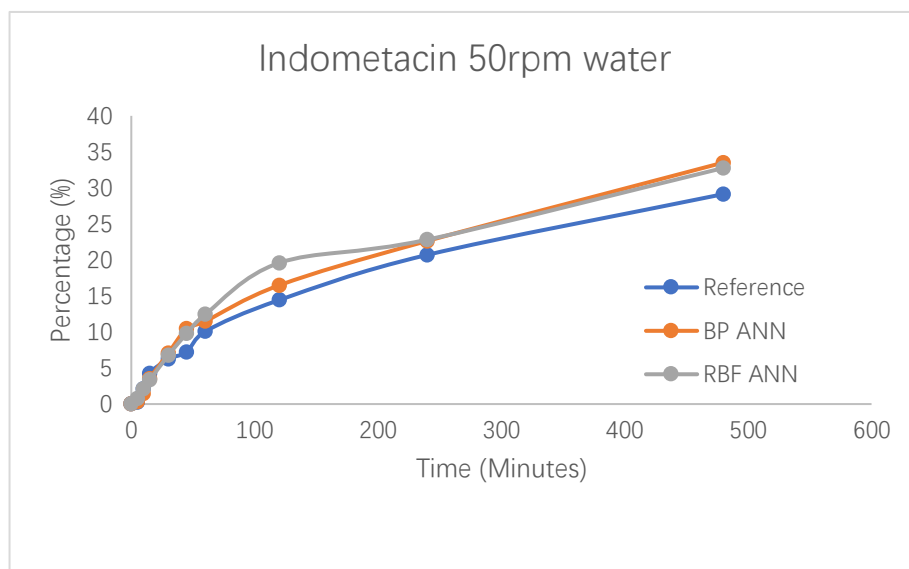


Figure 3.6 The dissolution test results of Indometacin under 50 rpm/water condition

Table 3.5 The comparison of real data, BPNN prediction data, and RBFN prediction data under 75 rpm/water condition

Indometacin 75rpm Water			
	Reference	BPNN	RBFN
0min	0	0	0
5min	5.07	2.38	2.90
10min	4.64	4.12	4.33
15min	6.29	5.77	5.70
30min	10.72	10.12	9.52
45min	14.15	13.60	12.91
60min	16.44	16.33	15.88
120min	22.94	22.49	24.13
240min	29.20	29.40	28.43
480min	35.59	37.24	36.14
F2		91.12	91.46

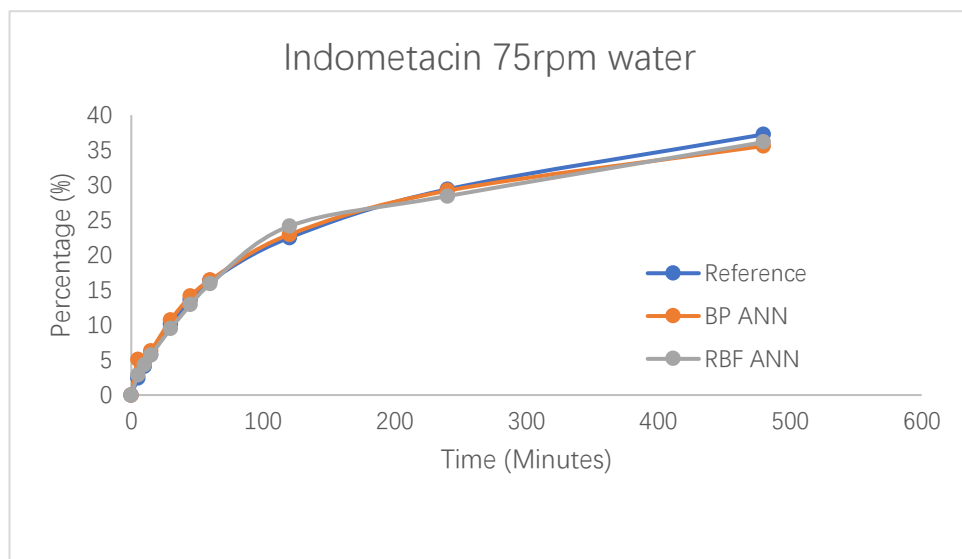


Figure 3.7 The dissolution test results of Indometacin under 75 rpm/water condition

Table 3.6 The comparison of real data, BPNN prediction data, and RBFN prediction data under 100 rpm/water condition

Indometacin 100rpm Water			
	Reference	BPNN	RBFN
0min	0	0	0
5min	3.50	4.77	4.59
10min	4.59	6.83	6.40
15min	6.72	8.74	8.12
30min	12.26	13.72	12.82
45min	16.95	17.66	16.84
60min	21.06	20.75	20.23
120min	28.67	27.72	29.13
240min	33.74	32.40	35.56
480min	38.60	38.37	37.30
F2		88.82	90.46

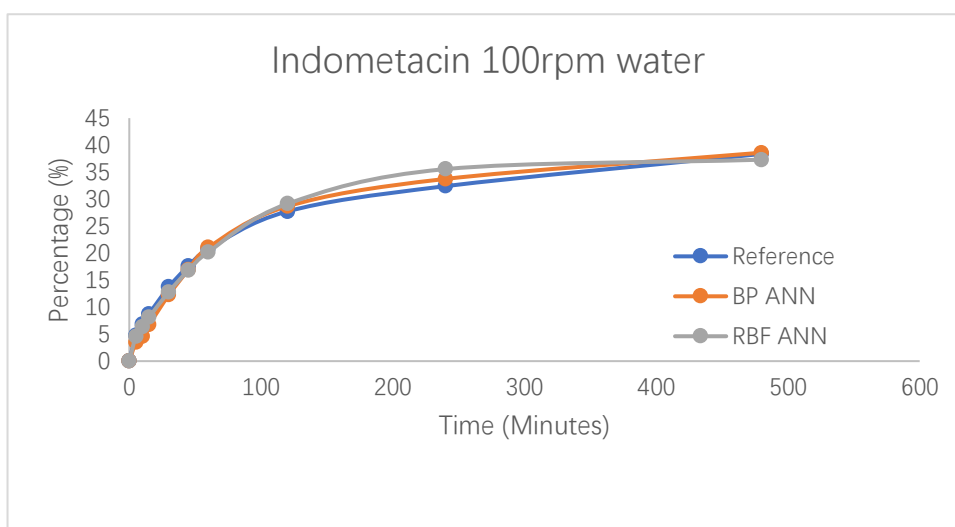


Figure 3.8 The dissolution test results of Indometacin under 100 rpm/water condition

Table 3.7 The comparison of real data, BPNN prediction data, and RBFN prediction data under 125 rpm/water condition.

Indometacin 125rpm Water			
	Reference	BPNN	RBFN
0min	0	0	0
5min	5.47	6.13	7.93
10min	8.61	8.08	9.51
15min	11.62	9.98	11.04
30min	16.09	15.26	15.3
45min	20.28	19.71	19.1
60min	27.31	23.23	22.44
120min	29.96	30.30	31.86
240min	35.09	34.37	36.88
480min	40.09	39.52	39.62
F2		86.55	81.61

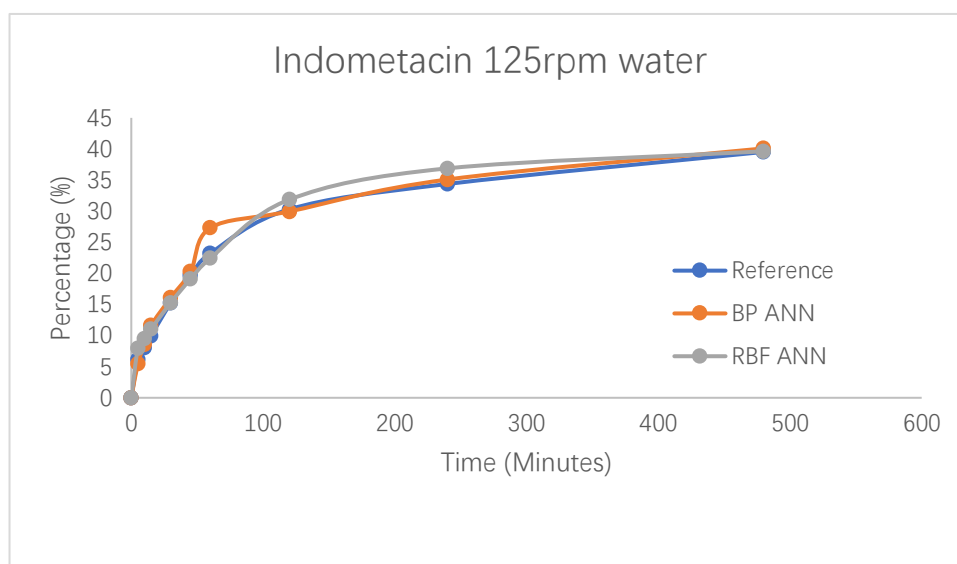


Figure 3.9 The dissolution test results of Indometacin under 125 rpm/water condition.

Table 3.8 The comparison of real data, BPNN prediction data, and RBFN prediction data under 150 rpm/water condition

Indometacin 150rpm Water			
	Reference	BPNN	RBFN
0min	0	0	0
5min	4.23	5.87	10.43
10min	6.92	7.50	12.04
15min	9.30	9.17	13.59
30min	14.63	14.26	17.92
45min	18.18	19.12	21.76
60min	21.80	23.40	25.15
120min	29.55	33.05	34.64
240min	33.96	37.04	39.3
480min	40.97	38.83	39.3
F2		83.29	67.23

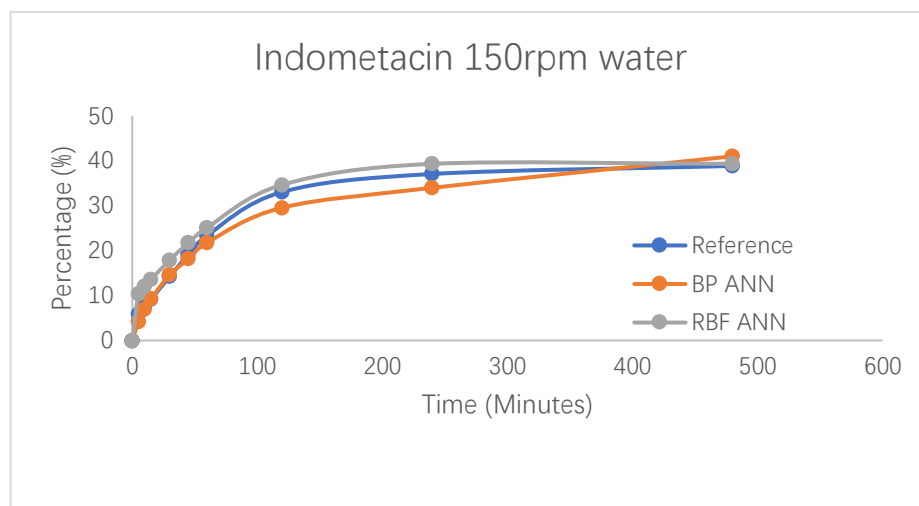


Figure 3.10 The dissolution test results of Indometacin under 150 rpm/water condition.

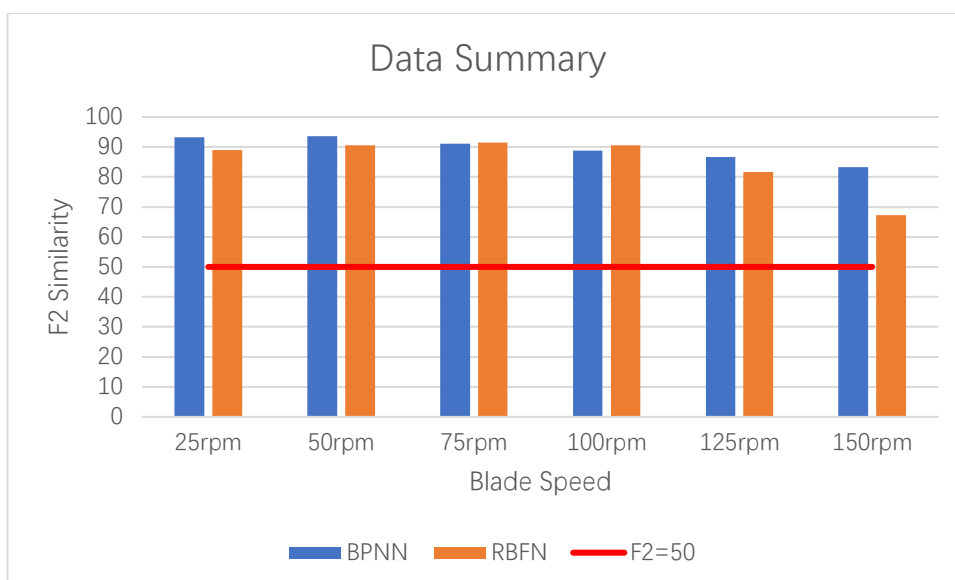


Figure 3.11 Data summary of this chapter

As revealed by the experiment results, all predicted values are highly accurate with the F2 value larger than 80, except for RBFN prediction data under the 150rpm condition. In addition, the F2 value of the RBFN/150 rpm group is 67.23, suggesting the most significant relative standard deviation compared with the practical values. The curve similarity requirement ($F2 \geq 50$) remains significantly higher than that for this standard deviation. In this case, it can be concluded in accordance with the above-described data that the model is endowed with high prediction accuracy that meets the requirements for the dissolution experiment. Furthermore, it takes on prominent guiding significance to the research and development of solid formulations.

3.3.3 Data Optimisation

The respective experiment is completed by the same dissolution tester six times with six dissolution cups to reduce the experimental errors. The average value is adopted to build the dissolution curve.

In this experiment, the acceptable relative standard deviation range is set as 0-20%. As depicted in Table 3.9, the relative standard between the average value and sample 6 substantially exceeds 20%, such that the data of sample 6 is automatically deleted.

Table 3.9 The dissolution data of Indometacin under 100 rpm condition

Indomedacin 100rpm Water From SYPU								
	Sample	Sample	Sample	Sample	Sample	Sample	AVERAGE	RSD
	1	2	3	4	5	6	%	%
5	0.78	0.58	0.39	1.26	0.68	0.87	0.76	39.07
10	1.75	1.85	1.26	1.95	1.46	2.04	1.72	17.42
15	2.44	3.02	2.63	2.83	2.63	3.41	2.83	12.31
30	6.33	6.24	5.55	5.76	6.04	8.57	6.42	17.07
45	9.96	8.99	7.72	8.60	9.37	13.37	9.67	20.31
60	12.05	11.17	9.70	10.98	11.95	17.81	12.28	23.13
120	19.78	15.89	14.22	16.08	17.74	27.61	18.55	25.98
240	26.09	21.99	17.49	22.47	24.23	41.15	25.57	31.89
360	27.97	23.36	20.88	23.55	26.49	44.85	27.85	31.22
480	29.08	24.35	21.47	24.93	27.30	46.83	28.99	31.45

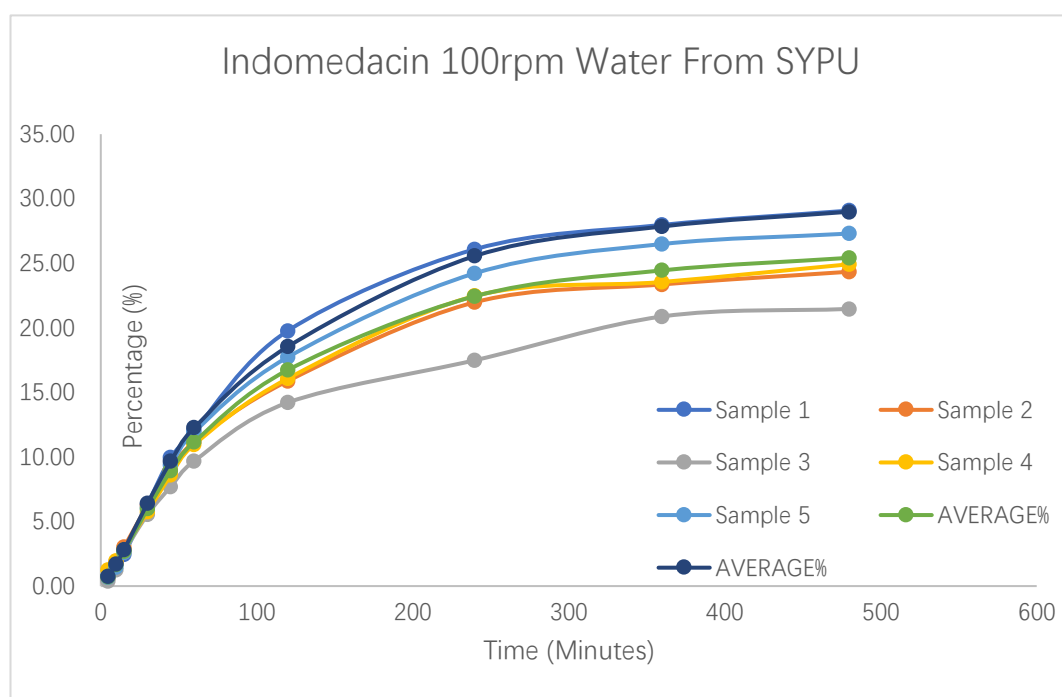


Figure 3.12 The dissolution test curve of Indometacin under 100 rpm condition

All the data in Table 3.9 are input into the model. After screening, the model recognizes sample 6 as a group of abnormal data since the relative standard deviation between sample 6 and the average data takes on more significance than the set value. The value range of relative standard deviation can be set in the main_Fun script. The comparative data between sample 6 and the average value is presented in the following.

Table 3.10 The relative standard deviation between average values and abnormal data (sample 6)

AVERAGE	Sample 6	RSD%
0.76	0.87	10.29789
1.72	2.04	13.26583
2.83	3.41	14.5881
6.42	8.57	23.74494
9.67	13.37	27.07127
12.28	17.81	31.87154
18.55	27.61	34.51896
25.57	41.15	43.07652
27.85	44.85	43.15475
28.99	46.83	43.505

This replication experiment demonstrates that the problem of abnormal data screening has been tackled by keeping the relative standard deviation within the acceptable range. In this way, the model constructed in this study can screen the abnormal data among mass data without manual work.

3.4 Chapter Summary

In this chapter, a novel model for predicting the dynamic solubility for solid active pharmaceutical ingredients is developed based on the ANN model structured by BPNN and RBFN. The domain-driven design DDD Plus application software marketed by Simulations Plus Inc. holds the largest market share of formulation dissolution prediction. However, the algorithm of DDD Plus has limited its development, whose core is based on a formula that calculates the result by inputting it with many parameters, some of which are unnecessary. In addition, no matter how much data are input in this software, its predictive accuracy remains unimproved.

In this model, we only need to consider the required parameters and design a gradient experiment to collect the necessary data for training. With more data being input, the predictive accuracy would be remarkably enhanced.

In addition, we also derived an equation from the Hixson-Crowell cube root law and Nernst-Brunner Equation, which can help set a trend curve that does not link the points by mere binomial.

In future work, the prediction model of formulation dissolution will be established. Meanwhile, the users will save large amounts of resources due to fewer experiments and less experimental expenditure.

Chapter 4 Backpropagation Neural Network and Non-linear Regression Approach in Predicting Solid Formulation Dissolution Rate

4.1 Chapter Introduction

In this chapter, the dissolution rate prediction requirement of input data is reduced using ANN and regression methods. Fully connected neural network (FCNN) has been extensively applied to data screen and prediction of the experimental results of pharmaceutical research. Thus, a novel dissolution result prediction and screening system is modeled using backpropagation networks and regression methods, with 21 sets of dissolution data adopted to train and verify the FCNN model. Based on the design of the input data, the relevant data remains available to train the FCNN model when the formulation composition is changed. In contrast to the orthogonal experiment, Effective Data Regression Method (EDRM) and Reference Line Regression Method (RLRM) facilitate this system to predict the dissolution results more accurately with fewer data. Based on the decision tree, the data screening function is also realized in this system, such that the abnormal data can be filtered during data inputting. This ANN model provides a novel drug prediction system that saves the prediction time and the costs, such that the design of new formulations can be facilitated [91-93].

4.2 Materials

- **Tinidazole**

Tinidazole refers to an anti-anaerobic and antigenic drug for the treatment of bacteria and insect infections, and a nitroimidazole derivative with higher efficacy, a shorter treatment course, better tolerance, and wider distribution in vivo after metronidazole. Pfizer

developed it in the late 1960s and marketed it in 1972 in Switzerland and West Germany under the brand name of Fasigyn. [94] Tinidazole tablets were recorded in the *12th Edition of the Japanese Medical Board* in 1991. In 1993, the domestic company Hubei Institute of Pharmaceutical Industry Co. Ltd. first exclusively developed Tinidazole tablets and obtained the new drug certificate issued by the Ministry of Health of the People's Republic of China (PRC) [29]. In the same year, Lizhou Group Lizhu Pharmaceutical Factory (Lizhu Kuifujing) was granted the production certificate of tinidazole tablets by the Ministry of Health of the PRC. Tinidazole can be observed in Chinese Pharmacopoeia (ChP 2010) [4], Japanese Pharmacopoeia (JP 16) [95], European Pharmacopoeia 7.0 (EP 7.0) [96], British Pharmacopoeia (BP 2013) [97], and the United States Pharmacopoeia (USP 35) [75], whereas the examination methods for relevant substances remain different. High-performance liquid chromatography (HPLC) was performed in ChP 2010, EP 7.0, and BP 2013 for analysis, while thin-layer chromatography (TLC) was employed in USP 35 [75] and JP 16 [95]. In this study, the analysis method refers to the one adopted in EP 7.0.

Physical and chemical properties [96]

Table 4.1 Tinidazole Parameters

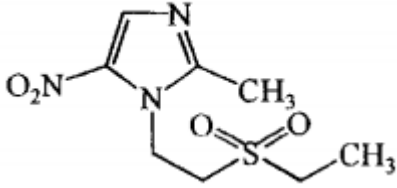
Tinidazole Parameters	
Structure formula	
Molecular formula	C ₈ H ₁₃ N ₃ O ₄ S
Relative molecular mass	247.28
Chemical name	2-methyl-1-[2-(ethylsulfonyl)ethyl]-5-nitro-1 H imidazole
CAS number	19387-91-8
Appearance	White or faint yellow crystal or crystalline powder
Solubility	Soluble in acetone or trichloromethane and slightly soluble in water or ethyl alcohol
Melting point	125~129°C
pKa	1.84

Table 4.2 The solubility of tinidazole in pH 1.2, pH 4.0, and pH 6.8 dissolution media and water [86]

Solubility in different dissolution media (37°C)	pH 1.2: 27.5 mg/ml	pH 4.0: 14.3 mg/ml
	pH 6.8: 13.2 mg/ml	Pure Water: 14.0 mg/ml

- Tinidazole Tablet

Table 4.3 The basic information of selected tablet [77]

Name		Tinidazole tablets		
Chemical name	2-methyl-1-[2-(ethyl sulfonyl) ethyl]-5-nitro-1 H imidazole			
Molecular formula	C ₈ H ₁₃ N ₃ O ₄ S	Expiration date	36 months	
Dosage form	Tablet	Dose	500 mg	
Reference formulation approval number	US: NDC 0178-8500-20	Marketing Authorization Holder (MAH)	Mission Company/US	Pharmaceutical

4.3 Pharmaceutical Experimental Methods

4.3.1 Assay for Tinidazole

The whole assay method including mobile phase preparation, standard solution preparation, sample solution preparation, chromatographic system setting, system stability validation, analysis method, and acceptance criteria.[96]

Table 4.4 The assay of Tinidazole

Mobile phase preparation	Acetonitrile, methanol, and water (10:20:70)
Standard solution preparation	0.1mg/ml of USP Tinidazole RS was prepared as follows. A proper amount of USP Tinidazole RS was transferred to a suitable volumetric flask, and methanol was added to 10% of the flask volume. Then, dilute with Mobile Phase to volume.
Sample solution preparation	0.1mg/ml of Tinidazole was prepared as follows. A proper amount of Tinidazole was transferred to a suitable volumetric flask, and methanol was added to 10% of the flask volume. Then, dilute with Mobile Phase to volume.
Chromatographic system	<p>Mode: LC;</p> <p>Detector: UV 320 nm;</p> <p>Column: 3.0-mm X 25-cm and 5-μm packing L7;</p> <p>Flow rate: 0.5 ml/min;</p> <p>Injection Volume: 20 μL;</p> <p>Run Time: 1.5 times the retention time of tinidazole.</p>
System stability	<p>Sample: Standard solution;</p> <p>Suitability Requirements:</p> <p>Tailing Factor: No more than 1.5;</p> <p>Relative Standard Deviation: No more than 0.73%.</p>
Analysis	<p>Samples: Standard solution and sample solution;</p> <p>The percentage of tinidazole (C₈H₁₃N₃O₄S) in the portion of Tinidazole taken is calculated as follows:</p> $\text{Result} = (r_u/r_s) \times (C_s/C_u) \times 100$ <p>r_u implies peak response of tinidazole from the sample solution;</p> <p>r_s represents peak response of tinidazole from the standard solution;</p> <p>C_s denotes the concentration of USP Tinidazole RS in the standard solution (mg/ml);</p> <p>C_u indicates the concentration of Tinidazole in the sample solution (mg/ml).</p>
acceptance criteria	100% \pm 2% on the dried base

4.3.2 Organic Impurities Analysis

The method of organic impurities analysis including mobile phase, sample solution, chromatographic system, standard stock solution, standard solution, system stability, analysis formula, and acceptance criteria. [86]

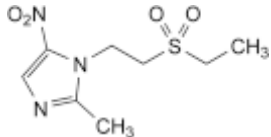
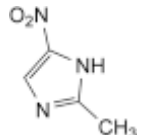
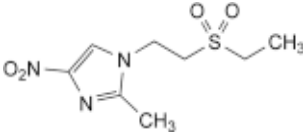
Table 4.5 Organic impurities analysis of Tinidazole

Mobile phase, sample solution, and chromatographic system	Refer to Assay Analysis
Standard stock solution 1	Use the standard solution from the Assay Analysis
Standard stock solution 2	USP Tinidazole Related Compound A RS (0.05 mg/ml) and USP Tinidazole Related Compound B RS (0.05 mg/ml) were prepared as follows. A proper amount of USP Tinidazole Related Compound A RS and USP Tinidazole Related Compound B RS was transferred to a suitable volumetric flask, and methanol was added to 10% of the flask volume. Then, dilute with Mobile Phase to volume.
Standard solution	USP Tinidazole RS (0.1 µg/ml) and USP Tinidazole Related Compound B RS (0.2µg/ml) in mobile phase from standard stock solution 1 and standard stock solution 2.
System stability	Sample: Standard Solution; Suitability Requirements: Resolution: NLT 2.0 between Tinidazole Related Compound A and Tinidazole Related Compound B; Relative standard deviation: No more than 5.0% for each peak;
Analysis	Samples: Sample solution and standard solution. The percentages of tinidazole related compound A and tinidazole related compound B in the portion of Tinidazole taken is calculated as follows: Result = $(r_u/r_s) \times (C_s/C_u) \times 100$

	Name	Relative retention time	Acceptance criteria (no more than)
Acceptance criteria	Tinidazole Related Compound A	0.6	0.2%
	Tinidazole Related Compound B	0.7	0.2%
	Tinidazole	1.0	--
	Any Unspecified Impurity	--	0.10%
	Total Impurities	--	0.4%

4.3.3 Impurity Study

Table 4.6 The Impurity names and chemical structural formulas recorded in USP, EP, and CP. [4, 77, 96]

USP Name	EP Name	CP Name	Structural Formula
Tinidazole	Tinidazole	Tinidazole	
Impurity A	Impurity A	Impurity I	
Impurity B	Impurity B	N/A	

4.3.4 Dissolution Test Study

- Dissolution Test Method Study

Tinidazole refers to a BCS II drug with poor water solubility but good permeability, whose absolute bioavailability is $\geq 100\%$. Among the pharmacopeias of various countries,

only the Chinese Pharmacopoeia 2015 and Indian Pharmacopoeia 2010 incorporate Tinidazole tablets, whereas the former does not contain the “dissolution” item. The dissolution methods, limits, and detection methods of Tinidazole tablets given in the respective pharmacopoeia are presented in Table 4.7:

Table 4.7 The comparison of dissolution test in Chinese Pharmacopoeia (CP), Japanese Orange Book, and FDA dissolution method database. [4, 77, 86]

Source	Dissolution Medium	Dissolution Method	Assay	Limit
Chinese Pharmacopoeia 2015	HCl solution (9~1000ml) or Water (900ml)	Basket Method 100 rpm	UV 317nm	30 min, $\geq 80\%$
Japanese Orange Book	900 ml Water	Blade Method 50 rpm	UV 317nm	45 min, $\geq 75\%$
US FDA Dissolution Method Database	900 mL Water	Basket Method 100 rpm	N/A	N/A

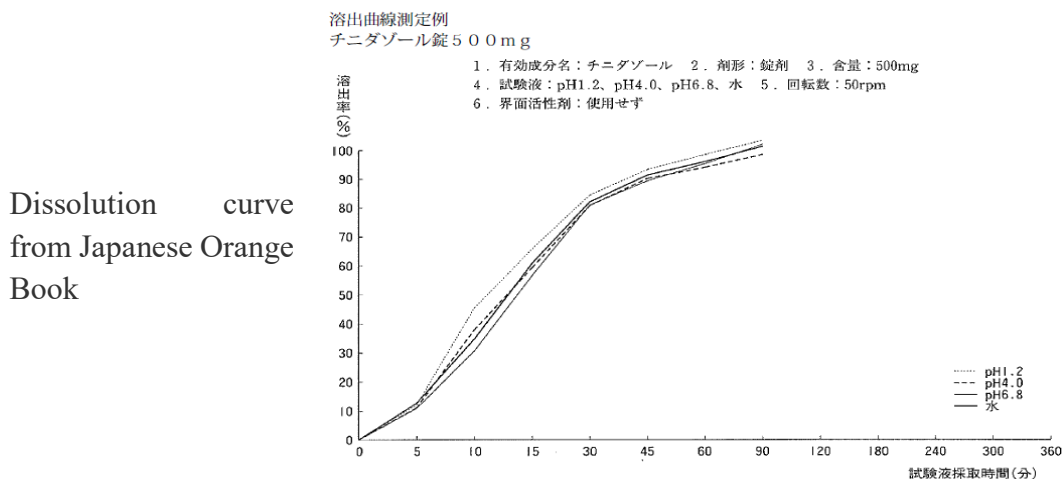
● Japanese Orange Book Method [86]

Table 4.8 The details of dissolution method in Japanese Orange Book

Dissolution method	Dissolution Method-2 (Blade Method)
Dissolution medium	900 mL water
Rotation speed	50 rpm
Surfactant	Not used

According to the dissolution determination method (paddle method), the sample was put in 900ml of water and the rotation speed was set as 50 rpm. Following the method, after a specified time, take an appropriate amount of solution, filter it, and discard at least 10ml of primary filtrate. Afterward, accurately measure an appropriate amount of continuous filtrate and dilute it with water to obtain the test solution containing 13 μ g solution per 1ml. In addition, accurately weigh 0.027g of tinidazole reference substance pre-dried at 105°C for 2 hours, put it into a 100 ml volumetric flask, add water to dilute to the scale, and shake it evenly. Then, accurately measure 5ml of the solution, put it into a 100ml volumetric flask, add water to dilute to the scale, and shake it. The resultant solution is the reference solution. According to UV-VIS spectrophotometry, measure the absorbance of these two solutions at 317 nm wavelength and calculate the dissolution rate of each tablet, which shall meet the requirements.

Limit 500mg Dose/Dissolution Rate 45min \geq 75%



- Chinese Pharmacopoeia 2015 Method [4]

Table 4.9 The details of dissolution method in Chinese Pharmacopoeia 2015

Test method	Dissolution Method-1 (Basket Method)
Dissolution medium	HCl Solution (9~1000ml) or water (900ml)
Rotation speed	100 rpm
Dissolution method	Take the product and determine its dissolution and release (method 1 of general rule 0931) with a hydrochloric acid solution (9~1000) or 900 ml water (plain tablets). The dissolution medium is 900 ml and the rotation speed is 100 rpm. Operate according to the law. After 30 min, take 10ml of the solution, filter it, accurately measure 2ml of the continuous filtrate, place it in a 100ml volumetric flask, dilute it with water to the scale, and shake it well. According to the UV-VIS spectrophotometry (general rule 0401), determine the absorbance at the wavelength of 317 nm. Take another appropriate amount of tinidazole reference substance, weigh it accurately, and add 0.002 mol/L hydrochloric acid solution or water (plain tablets) to quantitatively dilute it to obtain a solution containing about 12% sample. Determine the dissolution rate of each tablet through the same method.
Limit	≥80%

Given the analysis of previous experimental data, tinidazole is a rapidly disintegrating tablet, whereas the disintegration speed of 100 rpm is too fast for tinidazole dissolution. Thus, the identification of product quality is insufficient. Consequently, the method of Japanese Pharmacopoeia is adopted for the dissolution experiment in this study.

4.3.5 Stability Study

The stability in different dissolution media is presented in Table 4.10. [4]

Table 4.10 The stability under different conditions

Factors	Details
Water	Indoor temperature: Stable in 24 hours.
pH1.2, pH4.0, pH 6.8 dissolution media	Indoor temperature: Stable in 24 hours.
Light	Under the light intensity of 240,000 lm·h, the solid-state changed from light yellow to light yellow brown, but the content remained unchanged; the content of the aqueous solution decreased; in the dissolution media and aqueous solutions of pH 1.2, pH 4.0, and pH 6.8, the content decreased by 1.3%, 15.3%, 7.1%, and 6.1% respectively.

4.3.6 Tinidazole Tablet Preparation Method

Table 4.11 Composition of 500 mg Tinidazole Tablet [98]

Ingredients	Amount
Tinidazole (TNZ)	500 mg
Microcrystalline Cellulose PH 101 (PH101)	5.0%-20.0% w/w
Starch	1%-20% w/w
Croscarmellose sodium (CCNA)	0.5%-7% w/w
Hydroxypropyl methyl cellulose (HPMC)	0.5%-2% w/w
Magnesium stearate (MS)	0.5%-1.0% w/w
Low-substituted hydroxypropyl cellulose (L-HPC)	2.5%-5% w/w

The procedure of Tinidazole tablet preparation can be divided into 7 steps.

Step 1: API treatment. Dry the tinidazole powder in a vacuum drying oven under reduced pressure of 1kpa at 60°C for 2 h. After natural cooling in a low-pressure environment of 1kpa, the API was taken out and put into a grinder for grinding and

screening.

Step 2: Preparation of the adhesive. Weight the prescription amount of HPMC, add some boiling water, and use a glass rod to continuously stir the liquid till the HPMC completely disperses. Then, use ambient temperature water to make up the remaining prescription water, continue to stir it for 30 min, and then leave it to stand for 12 h.

Step 3: Total mixing. The tinidazole API is mixed with inactive ingredients except for the binder and magnesium stearate and put into the V-type mixer for mixing for 30 min.

Step 4: Granulation. Put the mixed sample into the granulator, turn on the shearing knife, and slowly add the binder.

Step 5: Drying. After the granulation, the granules were put into the tray, which was then put into the oven to dry till the moisture content was qualified.

Step 6: Granules. Put the dried granules into the finishing machine for granulation.

Step 7: Compressing tablets. Add magnesium stearate to the granules and put them in a V-type mixer for mixing for 5 min. After adjusting the tablet press speed, pre-compression parameters, and tablet thickness parameters, start tablet compression.

4.4 Computer Science Technologies

4.4.1 Model Test Design

First, a formulation experiment with different prescriptions and prescription volumes is designed to examine the model's adaptability. The predicted results vary significantly due to the small amount of training data, thus affecting the prediction accuracy. Accordingly, the regression value was calculated to process the predicted value after multiple predictions. In addition, a gradient experiment is designed to verify the improvement of prediction accuracy when the amount of input data increased. A total of 21 sets of prescriptions and dissolution data with three different excipients are used in this test due to the limited experimental conditions.

Thus, the test is divided into three parts, i.e., testing the impact of training frequency on the stability of the results, testing the effect of the training dataset on the prediction accuracy, and making predictions based on 21 sets of data to test their experimental adaptability.

4.4.2 Problem of Insufficient data leads to prediction errors

As some external factors, such as the position of the tablet in the dissolution cup (the tablets always not drop in the middle of the cup then put in), the tablet weight (the allowable tablet weight error is $\pm 10\%$), the hardness of the tablet (the hardness of the same thickness is always more significant when the tablet with higher weight) and the test results of tablets with the same prescription are always different. Accordingly, there will be an error between our test value and the expected value. When the input data is the test value, the experimental error will lead to the deviation of the practical prediction. The more significant the amount of input data, the smaller the prediction error caused by experimental error. It is why the model's prediction accuracy is unstable when the data is insufficient. Thus, most ANN applications [19, 35-48] to predict formulations will use large amounts of data to train the model and make predictions. For instance, in Saman Sarraf's study [99], he used an ANN model to predict the Betamethasone release rate, and it prepared over 80 samples.

However, the input layer neurons are only 5. The prediction effect can be proved academically, but 80 experiments are enough to complete a complete orthogonal experiment in practical applications. According to the orthogonal experiment design of 5 factors and five levels, 25 experiments can result. Accordingly, using a large amount of data to train the model has no practical application significance. In this project, the model should be helpful to assist formulation research and development from the beginning of the formulation project. Thus, 21 samples with 7 influence factors were prepared.

4.4.3 Fully Connected Neural Network Model Design

There are few information resources available for drug development, and the acquisition of information is expensive. Consequently, the primary issue this thesis seeks to address is how to use the version to achieve convergence and draw accurate conclusions with a limited dataset.

As the range of model layers and internal parameters are significantly reduced, model convergence is facilitated. Accordingly, this chapter estimates the dissolution charge using a layer of an absolutely related neural network.

The critical elements of this study are the prescription ratio, the dissolution time, and the dissolution rate (i.e., percentage of drug dissolution) at different time points.

To establish connections among time points, the formulation ratio and the dissolution time were set as the input dataset and the dissolution rate at each time point was set as the output data.

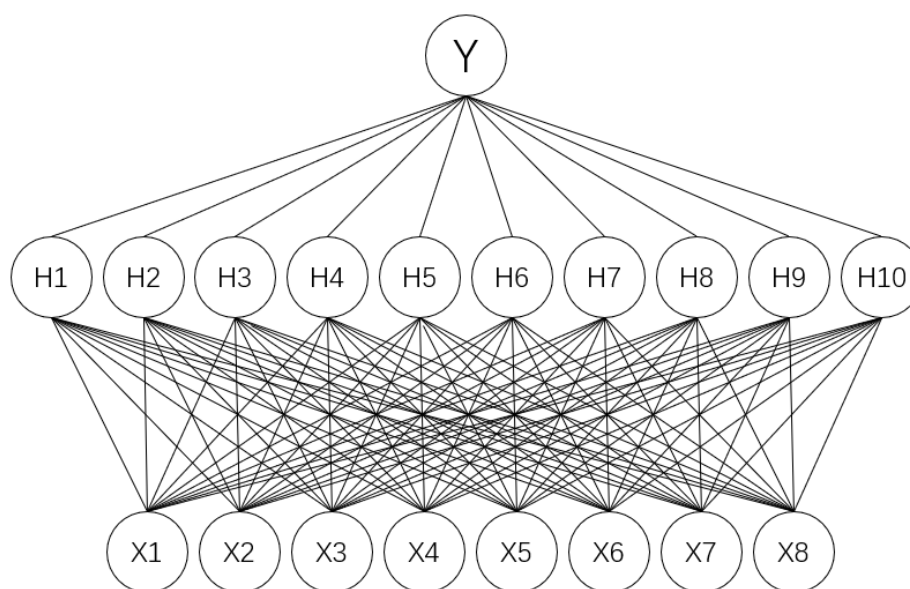


Figure 4.1 The structure of ANN

In Figure 4.1, the X1-X8 means: X1, time value (minutes) of dissolution test.; X2,

percentage of Tinidazole; X3, percentage of microcrystalline cellulose PH 101, X4, percentage of starch; X5, percentage of croscarmellose sodium; X6, percentage of hydroxypropyl methyl cellulose; X7, percentage of magnesium stearate; X8, percentage of low-substituted hydroxypropyl cellulose. H1-H9, neurons of Hidden Layer. Y is the output, which is the percentage of drug release in the time points of X1.

As depicted in Figure 4.1, a three-layer fully connected neural network combines eight independent variables in the input layer and one response variable in the output layer. The output of the node to the next layer is calculated by summing the input of the previous layer and calculating the activation function. The activation function employed in this project is the sigmoid function described as:

$$S(y_j) = \frac{1}{1+e^{-y_j}} \quad (4.1)$$

Where $S(y_j)$ denotes the output from the j-th node, in which y_j is defined as follows:

$$y_j = \sum_{i=1}^n W_{ij}x_i + B_j \quad (4.2)$$

Herein, x_i denotes the input of the i-th node in the previous layer, n represents the total number of nodes, W_{ij} signifies the corresponding weight, and b_j implies the bias. The ANN was iteratively trained to minimise the mean square error (MSE). The gradient of the MSE performance function was used to adjust the network weights and biases till the MSE reached 10^{-5} . This paper used MATLAB R2020b version (Mathworks Inc., Natick, MA, USA) to develop and train ANN, which could automatically generate the initial weights and deviations of the network.

4.4.4 Two novel regression methods

- Effective Data Regression Method (EDRM)

The model is trained using multiple sets of logical data, such that the predicted data primarily fluctuates around the practical value as long as the model is valid. First, multiple

predicted results of the same input dataset are obtained through multiple training and prediction of the probability and statistics method. Subsequently, the prediction result is obtained after the respective training serves as a decision tree. If it conforms to the requirements, it will be retained; otherwise, it will be deleted. Using statistical methods, the model calculates the standard deviation and the average of all predicted data. If the absolute difference between the predicted value and the average value exceeds the standard deviation, the data will be automatically deleted. On that basis, the model is capable of eliminating the abnormal data generated when the predicted data converges toward the optimal local solution. Lastly, an average curve is generated by determining the average of all the data. The stability of the predicted value is generally determined by the training frequency and prediction time. Figure 4.2 shows the flow chart of EDRM in program.

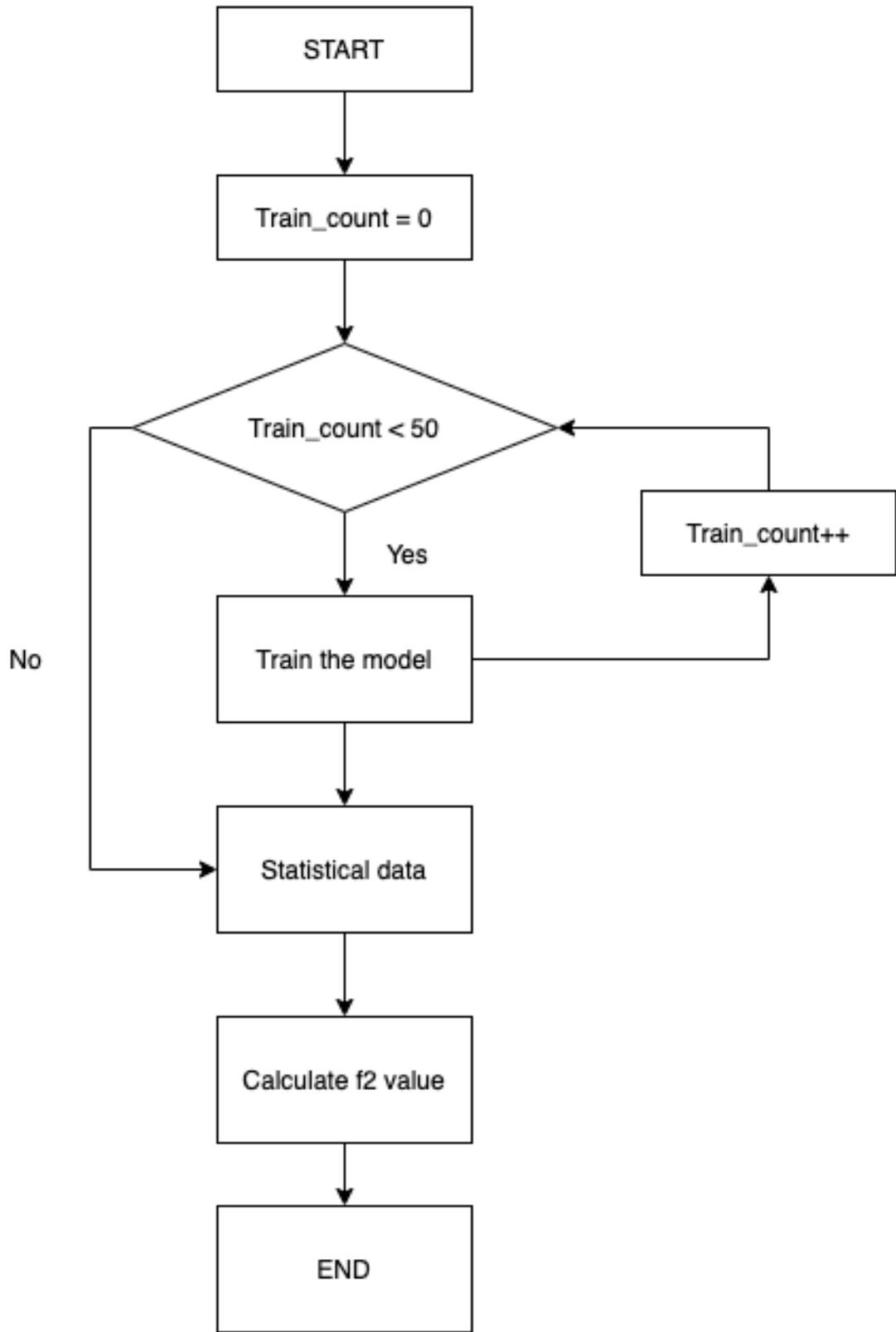


Figure 4.2 The flow chart of EDRM

4.4.5 Reference Line Regression Method (RLRM)

Based on the first method, the second method changes the screening method for abnormal data generated when the predicted data converges toward the optimal local solution. First, one or more sets of experimental data from the training are set as the reference data. According to the amount of experimental data, only one dataset is selected for testing. The reference data should also be tested by the program data screening system (see the data screening model). After each model training, the model creates predictions based on the dataset, compares the predicted results with the data, and calculates the F2 value. When F2 is below the set value, which is initially 65 and can be customised according to the project, the model is considered to converge toward the optimal local solution. Then, the program automatically recognises it as an abnormal model and deletes it. When the number of models that meet the requirements reaches the set value, which is initially 50 and customisable, the filtered data is averaged to obtain the regression line. Similar to EDRM, the value of the curve tends to stabilise when the training frequency predictions increases.

The progress and prediction results of Formulation 3 of RLRM in the program are presented as follows:

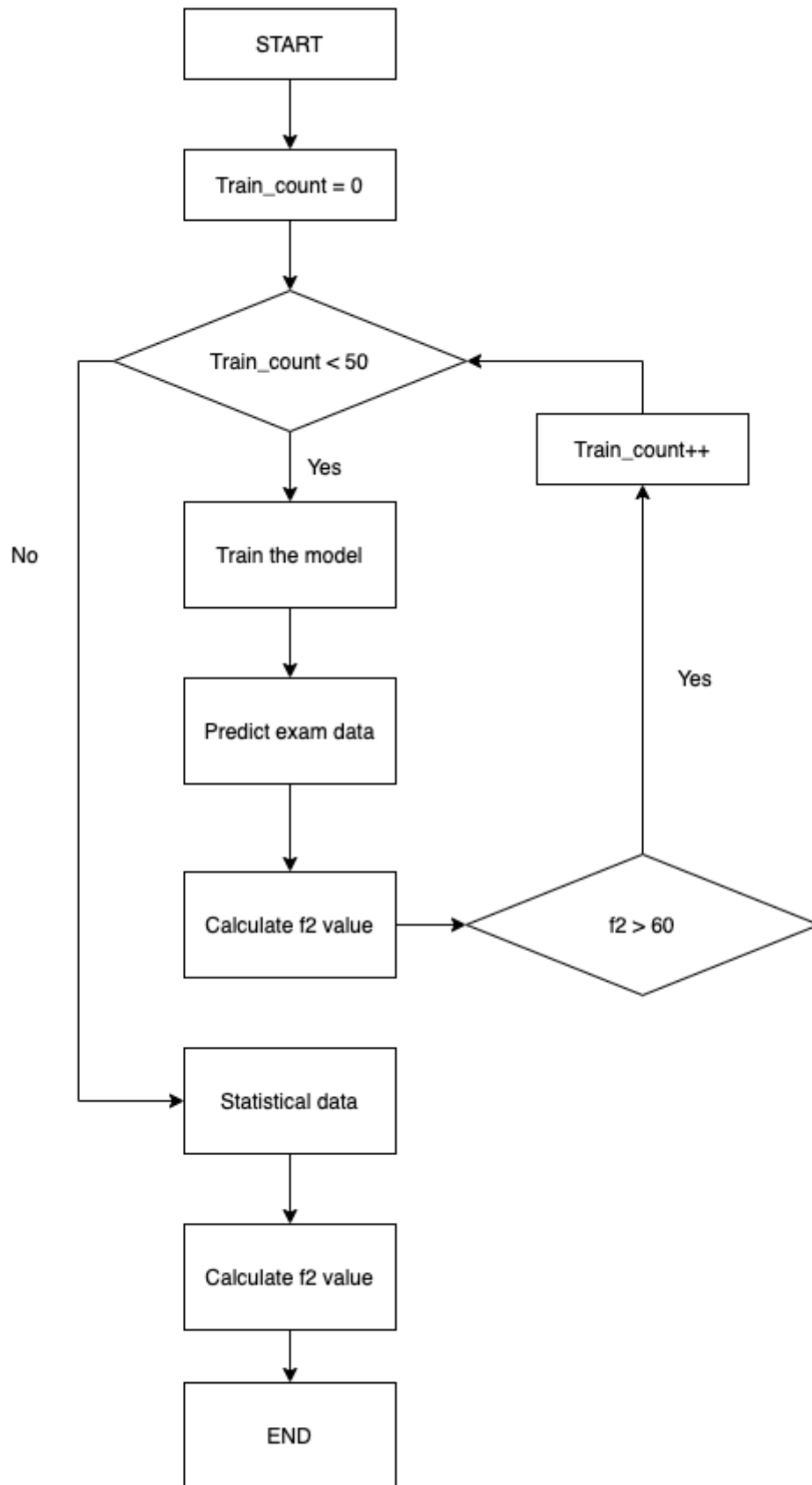


Figure 4.3 The flow chart of RLRLM in the program

4.5 Results and Discussion

4.5.1 Result of Formulation Experiment

The percentage of ingredients in each formulation experiment is presented in Table 4.12. The experimental method is described in the section of Materials and Methods.

Table 4.12 The percentage of seven ingredients

Sample Number	Tinidazole%	PH101 %	Starch %	CCNA %	HPMC %	MS %	L-HPC %
1	76.92	9.54	9.54	3.00	0.71	1.00	0.00
2	76.92	13.39	6.69	1.00	1.00	1.00	0.00
3	76.92	10.04	10.04	1.00	1.00	1.00	0.00
4	76.92	6.86	13.72	0.50	1.00	1.00	0.00
5	76.92	6.96	13.92	0.50	0.80	1.00	0.00
6	69.44	14.03	14.03	0.50	1.70	1.00	0.00
7	69.44	9.12	18.24	0.50	1.70	1.00	0.00
8	71.43	19.64	4.91	3.00	0.55	1.00	0.00
9	71.43	19.10	4.77	3.00	0.70	1.00	0.00
10	71.43	19.89	3.98	3.00	0.70	1.00	0.00
11	71.43	19.10	4.77	3.00	0.77	1.00	0.00
12	71.43	18.30	4.57	4.00	0.77	1.00	0.00
13	71.43	19.83	3.97	3.00	0.77	1.00	0.00
14	88.62	5.32	2.66	0.00	0.53	0.30	2.66
15	88.62	3.70	1.85	0.00	0.53	0.30	5.00
16	88.62	6.31	1.58	0.00	0.53	0.30	2.66
17	71.50	17.40	5.10	5.0	0.00	1.00	0.00

18	71.50	16.50	5.0	6.0	0.00	1.00	0.00
19	71.50	16.40	5.10	6.0	0.00	1.00	0.00
20	71.50	15.40	5.10	7.0	0.00	1.00	0.00
21	71.50	17.40	5.10	5.0	0.00	1.00	0.00

Note: The first column is the number of formulation compositions

The dissolution test results of these formulations at 37°C in water are presented in Table 4.13 as follows.

Table 4.13 The percentage of drug release in dissolution test at different time Points

Sample number	0 min	5 min	10 min	15 min	30 min	45 min	60 min
1	0%	33.54%	82.84%	91.43%	94.22%	94.84%	92.59%
2	0%	31.32%	78.00%	95.27%	99.27%	100.04%	98.91%
3	0%	19.26%	48.30%	78.00%	98.55%	99.11%	98.29%
4	0%	7.31%	20.22%	34.25%	70.32%	90.82%	95.41%
5	0%	22.02%	54.07%	78.45%	94.96%	96.34%	95.49%
6	0%	13.64%	30.97%	45.14%	78.20%	95.56%	95.47%
7	0%	9.74%	21.53%	32.88%	62.05%	83.21%	93.92%
8	0%	68.55%	83.08%	87.39%	90.57%	92.31%	92.71%
9	0%	64.77%	90.53%	93.61%	95.86%	96.00%	95.21%
10	0%	70.54%	85.85%	89.07%	92.00%	91.18%	90.56%
11	0%	57.90%	85.99%	91.64%	93.81%	94.09%	96.60%
12	0%	53.23%	83.70%	88.98%	95.80%	95.66%	96.61%
13	0%	62.77%	74.11%	81.47%	86.66%	88.77%	88.26%

14	0%	68.63%	78.65%	85.52%	92.67%	92.54%	93.73%
15	0%	69.19%	80.81%	86.37%	89.55%	90.81%	92.10%
16	0%	47.79%	66.95%	77.02%	86.66%	90.64%	92.93%
17	0%	48.40%	69.66%	81.59%	90.64%	96.05%	99.03%
18	0%	55.27%	68.13%	76.18%	88.44%	95.54%	99.41%
19	0%	45.91%	58.46%	69.19%	80.30%	86.73%	92.12%
20	0%	26.98%	49.31%	67.57%	81.26%	88.85%	92.40%
21	0%	62.75%	78.55%	84.81%	91.29%	94.03%	95.35%

Note: The first column is the number of formulation compositions.

4.5.2 Prediction Result of EDRM versus RLRM

First, the first 20 sets of data were selected among the 21 sets of data to train the model. Then, the predicted value of the 21st dataset was obtained. The model was trained 50 times and the predicted values were compared. The test results are presented as follows.

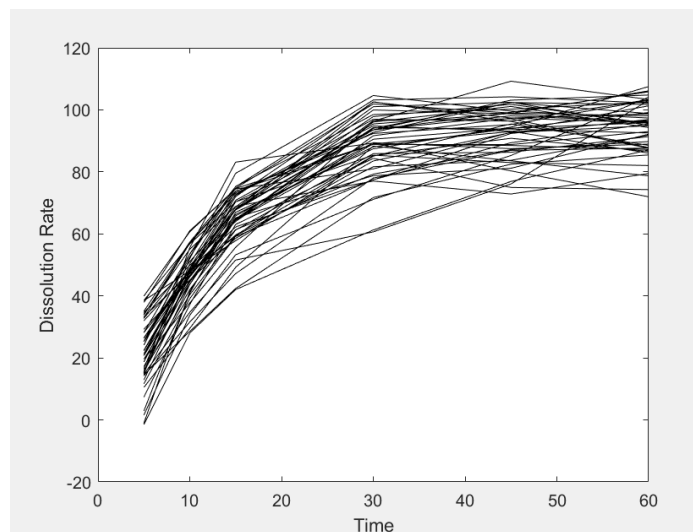


Figure 4.4 The x-axis is the dissolution time while the y-axis is the dissolution rate at different dissolution time points.

After 50 predictions of Formulation 21, it is found that the results of each test vary significantly. According to theoretical analysis, this can be attributed to the following two reasons:

(1) After the neural network model was trained, the model did not fully converge due to insufficient training data. Thus, the value predicted by the model would be larger or smaller than the practical value, but it remains within the usual error range.

(2) The prediction data for the model would occasionally converge toward a locally optimal solution. In this case, the model's predicted value would significantly diverge from the practical value, thus resulting in abnormal data.

To address the above-described two problems, this study combines statistics and random forest methods to propose two methods, i.e., EDRM and RLRM.

(a) EDRM

The progress and the prediction results of Formulation 3 through EDRM in the project are presented in Figure 4.5.

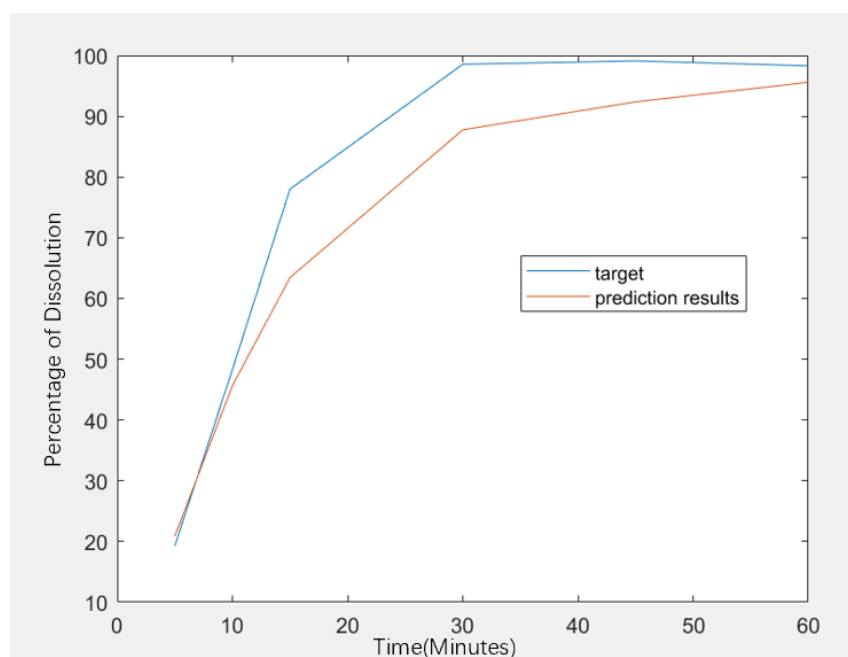


Figure 4.5 The prediction output of EDRM

In Figure 4.5, the blue line denotes the practical value obtained from the experiment while the red line represents the predicted value. After the two values were substituted into the F2 calculation formula, the F2 was 54.48, which indicates that the two curves are similar. This demonstrates that this method can address the problem.

(b) RLRM

After substituting the RLRM for EDRM in the last test, the model generates the following results,

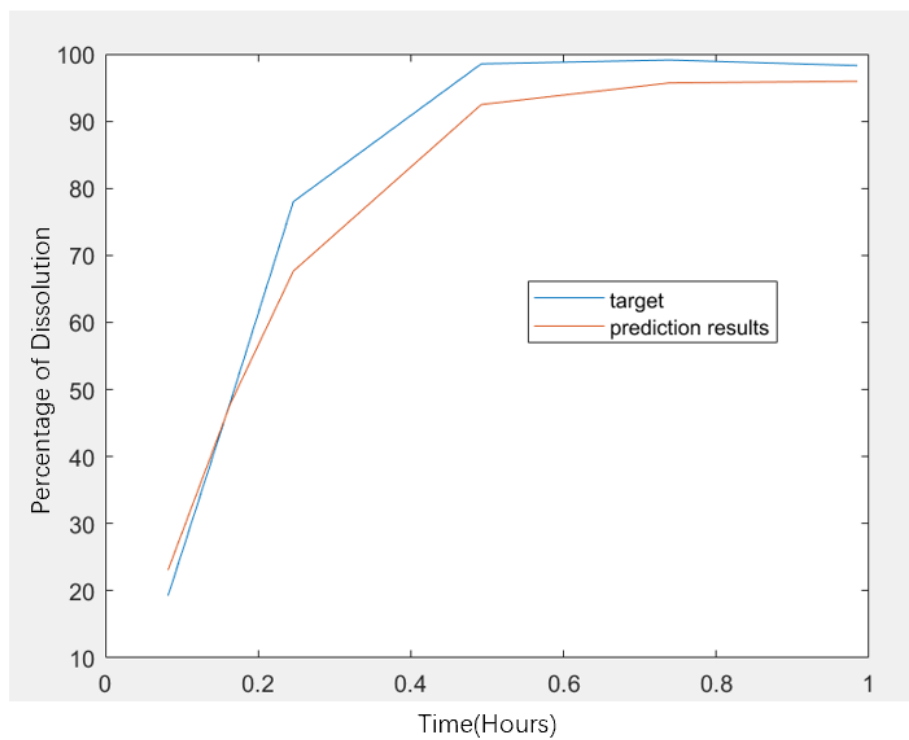


Figure 4.6 The prediction output of RLRM

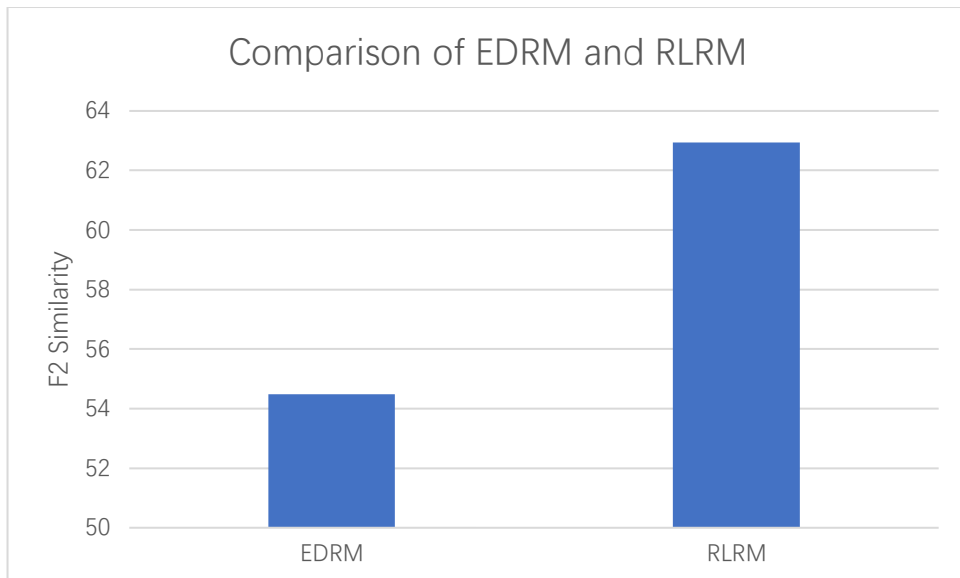


Figure 4.7 Comparison of EDRM and RLRM.

As depicted in Figure 4.6, the blue line represents the practical value obtained in the experiment, while the red line denotes the predicted value of the model. After the two values are substituted into the F2 calculation formula, F2 is 62.94, suggesting that the RLRM generates better results compared with EDRM.

In RLRM, the choice of the reference line takes on critical significance. The predicted value changes with the change of the reference line, such that choosing an appropriate reference line will increase the prediction accuracy. The prediction accuracy of the model is tested on the reference line, i.e., Formulation 13, Formulation 16, and Formulation 21, to predict Formulation 2. The 19 datasets other than prediction data and reference line are regarded as input data. The results are listed in Table 4.14 and Figure 4.8.

Table 4.14 The prediction results of Formulation 3 with Formulations 13, 16, and 21 as reference line.

Time	Formulation 13	Formulation 16	Formulation 21	Experimental data
0	0	0	0	0
5	44.39	42.20	44.42	31.32
10	63.76	68.55	69.26	78.00
15	79.06	86.00	84.74	95.27
30	97.26	103.17	100.99	99.27
45	98.62	100.62	100.04	100.04
60	97.28	97.92	98.09	98.91
F2	49.11	56.96	55.30	N/A

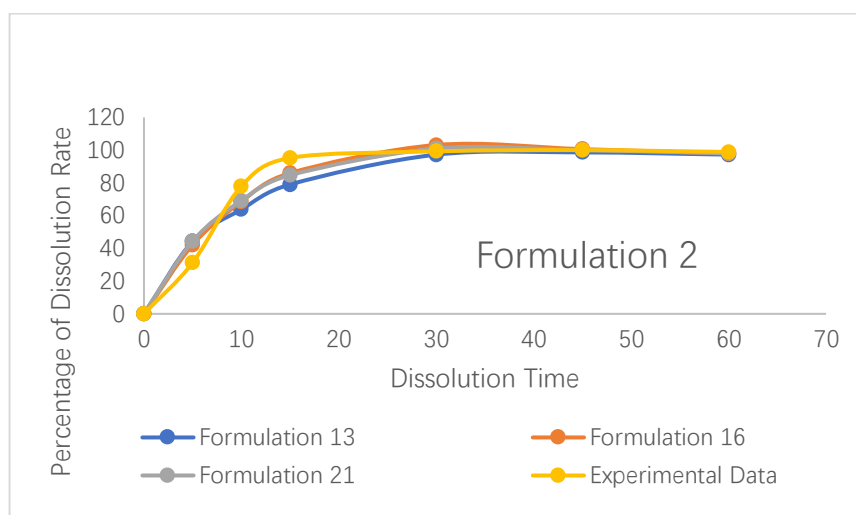


Figure 4.8 The contrast curve of the results

Given the above analysis of the two methods, the algorithm of EDRM is more stable and can conform to the model requirements, whereas that of RLRM requires higher accuracy of reference data. When there is sufficient standard reference data, RLRM will outperform EDRM.

4.5.3 Related Data Reuse

In the practical formulation experiment, the types of excipients in the prescription tend to be fine-tuned. The existing articles that predict the results of formulation experiments through neural networks only predict a single prescription by screening the prescription ratio. If the types of excipients in the prescription are changed during its optimization process, new data and training will be required for the model, such that the previous experimental data will be wasted. The preliminary data remains available for a method designed to conform to the requirements for changing the types of excipients. First, the auxiliary materials in all samples are set as input parameters, the amount of unadded auxiliary materials is set as 0, and the above data are input into the ANN model for training with other input data. In the experiment, a total of three prescriptions are designed. To test the feasibility of the proposed method, each of them has one excipient that differs from the excipients of the other two prescriptions.

The second formulation composition has only three groups of samples, but it has five variables. Mathematically, no solution for this problem. Nonetheless, it has been perfectly addressed through the method of the project.

In the test, Formulation 15 is selected as under test data, which is the second formulation composition. In EDRM, the data of the other 20 formulations is set as input data, and the training times reach 100. In RLRM, Formulation 21 is set as the reference line, while the data of the other 19 formulations is set as input data to train the model. EDRM is adopted to predict Formulation 15 and determine the F2 value.

Table 4.15 The Formulation 15 prediction results through EDRM and RLRM (second and third column) and the experimental results of Formulation 15 (fourth column)

Time	EDRM	RLRM	Experimental Data
0	0	0	0
5	61.53	57.98	69.19
10	72.01	64.26	80.81
15	78.34	73.65	86.37
30	90.90	86.58	89.55
45	99.97	91.52	90.81
60	98.64	93.89	92.10
F2	56.33	50.37	N/A

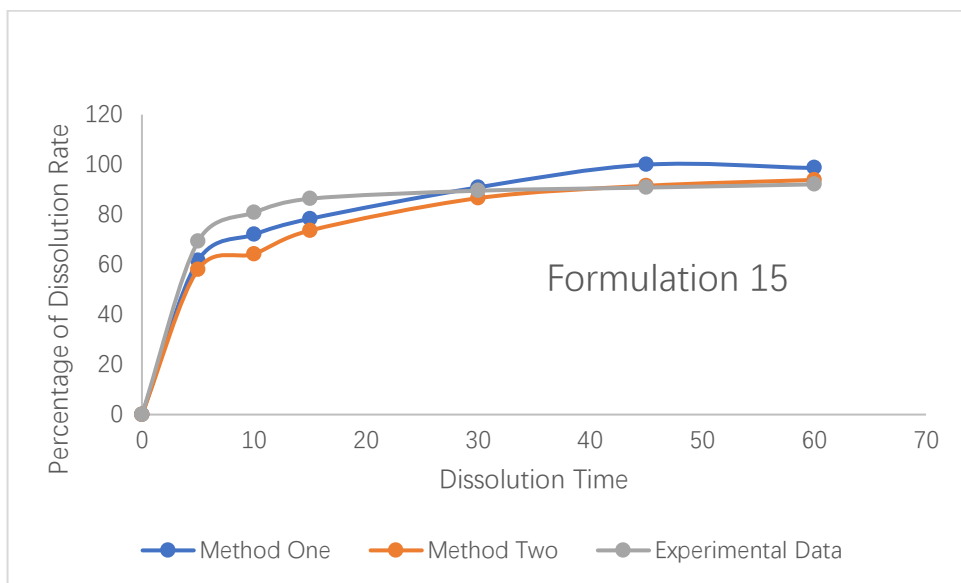


Figure 4.9 The contrast curve of Formulation 15

In summary, observed from the prediction results, the F2 value exceeds 50, implying that the similarity rate between the predicted curve and the practical experimental value

exceeds 90%. This demonstrates that our method effectively solved the problem that data could not continue the training after the types of prescription excipients were fine-tuned.

4.5.4 The Relationship between Prediction Times and Prediction Stability

Observed from the above two methods, the number of predictions is particularly important. The gradient experiment was designed to investigate the relationship between the frequency of predictions and the accuracy of the model. A dataset was selected for each of the three prescriptions for testing with Formulation 13, Formulation 16, and Formulation 21 as the target prediction data values. Then, the model was trained and predicted twice, 10 times, and 50 times for each dataset values. The respective training is repeated 10 times, in which the predicted results are analyzed to determine RSD. The experimental results are presented as follows.

Table 4.16 10 prediction results of Formulation 21 under two trainings

	1	2	3	4	5	6	7	8	9	10	RSD
5	52.89	51.25	55.98	52.50	56.41	60.51	49.60	58.88	54.38	56.27	3.40
10	68.65	67.20	71.00	69.38	69.71	70.70	62.31	71.16	71.46	68.28	2.73
15	79.40	76.43	78.58	79.62	77.35	77.15	71.18	78.09	81.12	76.80	2.67
30	92.34	89.08	87.62	88.25	85.69	88.81	83.21	85.57	90.19	89.60	2.63
45	95.64	93.76	91.90	92.84	91.04	98.40	89.05	92.13	93.82	93.11	2.55
60	101.2	95.59	95.15	96.94	96.91	99.15	93.09	96.51	96.46	93.76	2.39
F2	58.75	56.10	63.98	60.17	60.76	63.63	47.96	64.46	65.62	60.70	5.19
Sum of RSD							16.37				

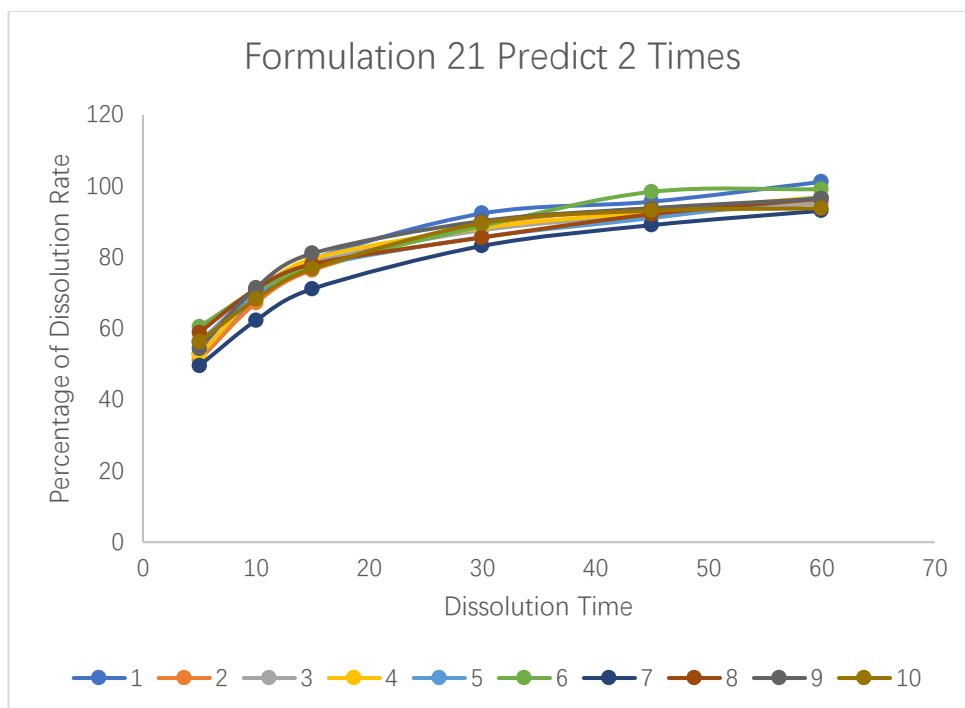


Figure 4.10 The comparison of 10 predictions of Formulation 21 under two trainings

Note: Formulation 21 was predicted 10 times.

Table 4.17 10 prediction results of Formulation 21 under ten trainings

	1	2	3	4	5	6	7	8	9	10	RSD
5	55.03	56.48	57.92	57.22	53.71	55.86	56.25	55.08	55.38	54.05	1.32
10	70.00	69.32	71.82	71.26	70.22	70.50	69.72	68.63	70.94	70.25	0.93
15	78.20	78.29	79.46	80.15	79.08	78.57	79.16	77.67	79.51	78.66	0.74
30	89.92	88.78	90.57	88.80	89.85	89.69	90.04	90.44	90.17	88.39	0.75
45	95.08	93.63	94.84	93.22	95.59	93.86	93.46	94.69	92.70	93.05	0.97
60	98.24	96.05	98.22	96.90	98.23	97.14	97.00	98.68	96.19	98.44	0.97
F2	62.27	63.07	68.16	67.41	61.78	64.15	64.17	60.58	65.05	61.54	2.50
Sum of RSD											5.69

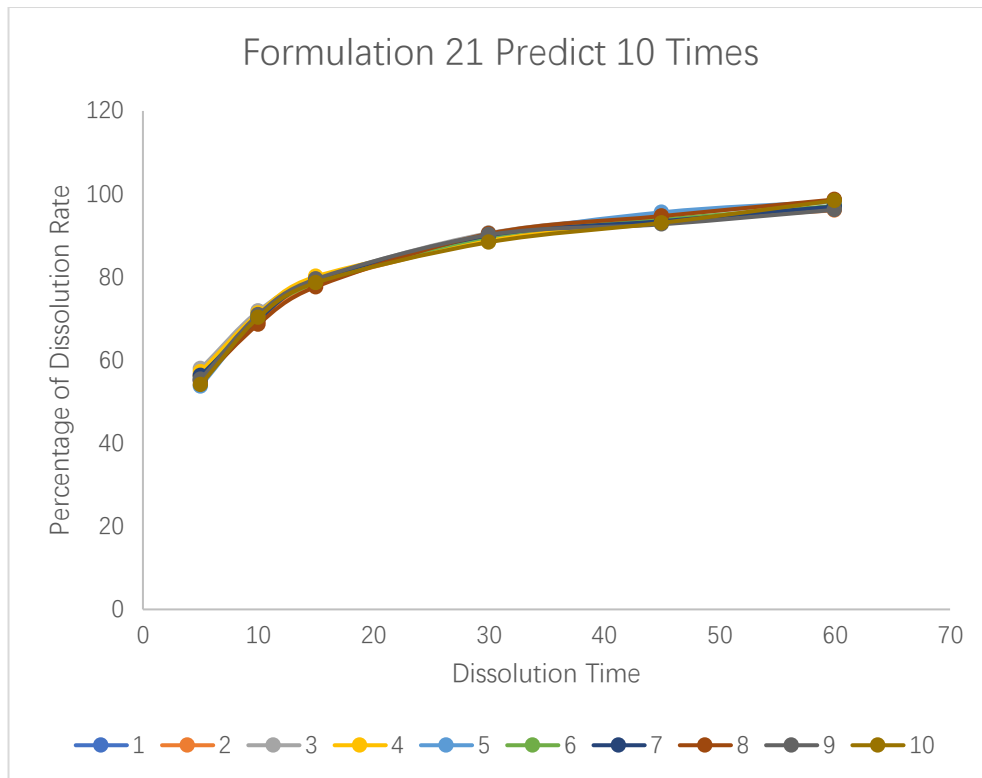


Figure 4.11 The comparison of 10 predictions of Formulation 21 under ten trainings

Note: Formulation 21 was predicted 50 times.

Table 4.18 10 prediction results of Formulation 21 under 50 trainings

	1	2	3	4	5	6	7	8	9	10	RSD
5	57.70	55.69	56.26	55.94	56.54	56.32	56.55	56.13	56.19	55.33	0.63
10	70.97	69.54	70.46	70.24	70.97	70.94	70.89	70.27	70.68	69.56	0.55
15	79.30	78.53	79.02	78.72	79.87	79.96	80.13	79.26	79.79	78.16	0.67
30	89.83	89.54	89.41	88.98	89.88	89.62	90.28	89.61	89.11	89.07	0.41
45	94.50	93.54	94.38	93.93	93.95	93.80	94.33	94.08	94.51	93.86	0.33
60	97.29	96.65	96.78	97.45	97.11	97.20	97.93	97.00	98.04	97.01	0.45
F2	66.99	62.99	64.85	63.81	66.48	66.18	66.36	64.71	65.17	62.27	1.58
Sum of RSD							3.03				

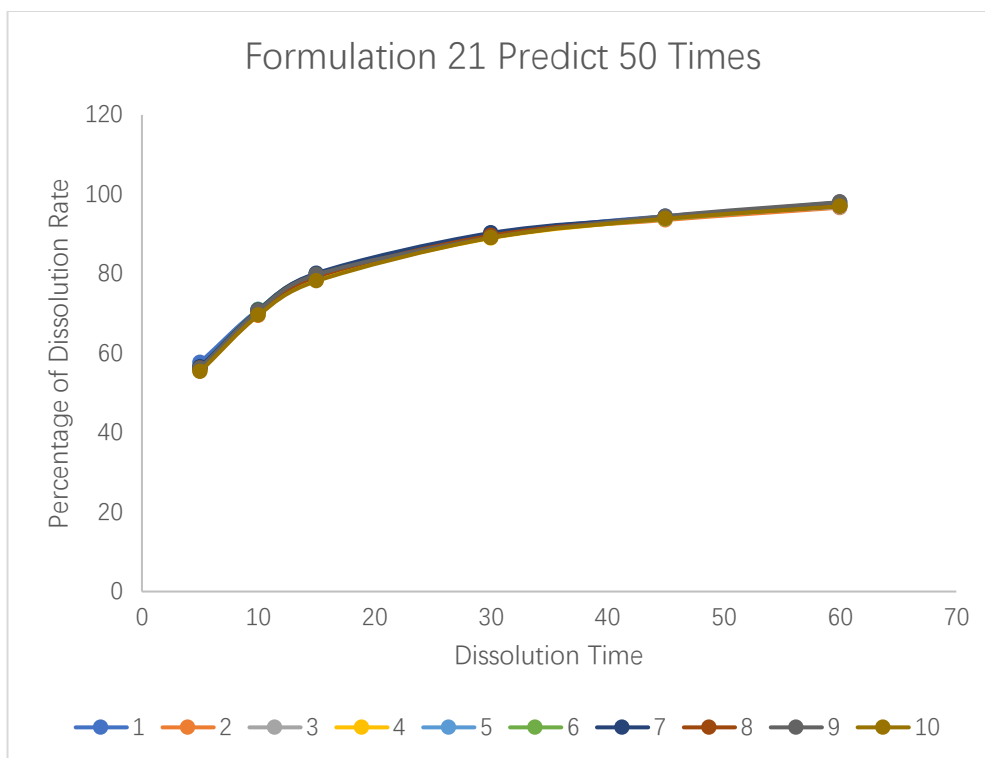


Figure 4.12 The comparison of 10 predictions of Formulation 21 under 50 trainings

In brief, when the number of predictions is 2, 10, and 50, the RSD values of the 10 tests data reach 16.37, 5.69, and 3.03, respectively. Moreover, the RSD of F2 reaches 5.19, 2.50, and 1.58. When the number of predictions only reaches 2, there is a case where F2 is only 47.96. The above result confirms that at least one data of this test converges toward the optimal local solution. As revealed by the test results, more prediction times can reduce the amount of abnormal data and prediction errors of the results.

4.5.5 Comparison with others work

An ANN model is designed by Arthur Manda in 2019 to predict the dissolution release of formulation. This model is tested through the experiment of prednisone pellets with seven factors. Table 4.19 lists the experimental data.

Table 4.19 Experimental dataset of Arthur Manda's work[32]

No.	X1	X2	X3	X4	X5	X6	X7	Y1	Y2	Y3	Y4
1	60	2	1	30	12.8	6.4	12.8	75.3	88.2	100.1	100.3
2	60	1	2	25	13.2	6.6	13.2	31.1	53.3	66.8	84.6
3	60	1	1	30	13.2	6.6	13.2	37.1	56	65	75.2
4	70	1	2	30	9.2	4.6	9.2	24.6	39.8	51.3	59.4
5	70	1.5	2	25	9	4.5	9	32.8	51.8	66.5	74.6
6	50	1.5	1	30	17	8.5	17	80.4	89.7	90.7	90.2
7	60	2	3	30	12.8	6.4	12.8	71.5	82.8	86.2	90.5
8	60	2	2	25	12.8	6.4	12.8	77.5	88.2	95.1	99.1
9	60	1.5	2	30	13	6.5	13	63.4	78.8	85.2	87.9
10	60	1.5	2	30	13	6.5	13	47.2	62.3	71.9	77.6
11	70	1.5	3	30	9	4.5	9	18.5	32.8	43.3	51.1
12	50	2	2	30	16.8	8.4	16.8	80.2	82	91.3	80.9
13	60	1.5	2	30	13	6.5	13	64	76.7	82.8	84.8
14	60	1.5	1	35	13	6.5	13	68.7	83.5	86.2	86.7
15	60	1.5	2	30	13	6.5	13	72.6	82.6	86.6	86.5
16	60	1.5	3	35	13	6.5	13	50.8	67.1	73.9	80.8
17	60	2	2	35	12.8	6.4	12.8	64.1	80	85.3	90.5
18	60	1	3	30	13.2	6.6	13.2	25.1	46.4	61.4	71.7
19	50	1	2	30	17.2	8.6	17.2	63.7	79.9	85.3	84
20	50	1.5	2	25	17	8.5	17	58.9	77.1	80.9	81.5
21	50	1.5	2	35	17	8.5	17	67.3	72.3	77.1	77.2
22	50	1.5	3	30	17	8.5	17	55.6	72.1	76.6	75.4
23	50	1.5	2	35	9	4.5	9	17.8	33	44.4	53.9

24	70	1.5	1	30	9	4.5	9	24.7	43.2	55.8	65.2
25	60	1.5	3	25	13	6.5	13	70.7	82.3	85.4	87
26	70	2	2	30	8.8	4.4	8.8	46.4	75.7	91.5	99
27	60	1.5	1	25	13	6.5	13	57.9	74.8	81	83.2
28	60	1	2	35	13.2	6.6	13.2	43.2	68.7	91.2	99.3
29	60	1.5	2	30	13	6.5	13	61.6	74.6	77.7	85.9

In Table 4.19, No. represents the mark of formulation component. The percentage of microcrystalline cellulose of the formulation(X1), the percentage of sodium starch glycolate(X2), the minute of spherization time(X3), the rap per minute of extrusion speed(X4), the percentage of tween 80(X5), the percentage of PEG 400(X6) and the percentage of Eudragit RL 30 D(X7) of the above-mentioned 29 formulations are demonstrated. Furthermore, the prednisone release at 15min(Y1), 30min(Y2), 45min(Y3) and 60min(Y6) of the above-described 29 formulations are presented.

This study can be designed as a 7-factor and 3-level orthogonal experiment based on the design method of orthogonal experiments, which can be completed using 18 groups of experiments. However, 29 groups of experiments were adopted in Arthur Manda's project to train the model, with nearly 60% more experimental groups than that of the orthogonal design. In contrast, RLRM and EDRM methods are promising in reducing the requirement for input data.

The experimental data are screened under the condition that for the respective factor based on the distribution of the dataset, and there are at least three horizontal values. In accordance with the gradient experimental design, the datasets with 29 groups of formulation, 14 groups of formulation, and six groups of formulation are designed, RLRM and EDRM are adopted to predict the results, and then F2 values are determined. The ANN

model prediction time is set as 100 to reduce the error of the prediction value. The prediction results are listed in Table 4.20.

Table 4.20a The prediction result using 29 groups of formulation

No.	R15	R30	R45	R60	R-F2	E15	E30	E45	E60	E-F2
1	76.38	91.36	97.48	97.18	77.53	75.82	91.45	98.41	97.31	79.42
2	35.93	50.83	64.40	78.96	68.77	36.32	50.12	63.35	80.34	68.71
3	38.59	63.08	72.78	80.01	61.40	39.73	64.48	70.62	78.43	62.64
4	20.06	33.80	52.30	60.01	70.24	21.88	34.36	51.66	60.19	74.54
5	28.14	48.72	54.23	72.31	58.03	28.10	50.24	54.38	73.24	58.84
6	74.21	80.35	95.21	96.25	58.27	74.97	81.40	94.76	98.61	58.11
7	71.52	78.21	88.65	90.51	77.76	69.59	79.47	90.69	90.00	75.24
8	68.97	83.26	92.35	95.21	62.73	66.96	83.63	94.63	96.52	61.18
9	61.24	77.22	80.12	85.88	74.73	60.82	76.58	80.64	84.15	72.50
10	31.44	58.13	78.11	78.21	52.82	30.16	58.39	79.22	78.07	51.04
11	33.18	41.25	45.35	53.54	53.09	34.68	38.77	43.65	53.82	52.80
12	74.13	75.13	80.11	83.12	56.58	72.80	75.58	78.18	82.23	54.11
13	61.22	73.30	89.12	95.16	59.26	60.60	74.28	87.13	94.33	62.12
14	63.75	79.88	83.14	91.25	68.67	63.03	82.33	80.95	90.54	67.49
15	68.91	77.13	83.14	89.26	69.38	67.09	76.80	81.94	86.78	66.21
16	52.33	67.11	79.13	81.82	76.54	51.00	65.62	79.60	82.45	74.61
17	61.24	77.12	88.99	95.51	70.75	62.79	76.64	87.13	97.95	68.06
18	33.20	45.11	61.28	77.11	64.99	33.84	46.27	60.63	77.22	63.87
19	65.12	81.18	85.32	84.12	92.91	63.21	83.36	84.37	83.63	84.16
20	51.92	72.14	84.94	81.52	65.77	53.05	72.53	82.48	79.44	69.59
21	63.12	73.31	74.12	74.51	75.40	63.47	73.01	76.16	75.95	81.67

22	66.12	75.11	80.61	87.41	53.71	67.72	72.94	78.38	88.27	52.41
23	14.01	31.90	43.75	53.31	82.33	11.92	31.71	45.37	51.14	72.84
24	25.16	44.75	63.76	64.65	68.88	27.55	44.16	65.94	65.83	63.41
25	75.27	79.42	82.91	89.76	73.25	73.40	80.45	83.13	91.35	75.34
26	45.33	76.97	93.52	97.53	87.22	46.28	75.66	91.19	97.91	96.97
27	53.80	70.30	84.46	89.62	65.69	51.45	72.23	86.81	87.54	64.54
28	40.71	67.68	90.32	101.01	85.69	38.82	65.38	91.47	99.22	76.67
29	60.60	67.21	75.25	85.17	69.54	58.26	69.24	73.25	85.53	69.93

Table 4.20b The prediction result using 14 groups of formulation

No.	R15	R30	R45	R60	R-F2	E15	E30	E45	E60	E-F2
1	70.12	75.53	97.92	98.84	57.62	70.22	75.90	96.62	98.27	57.67
2	36.02	51.08	64.54	77.74	66.79	36.80	51.37	65.29	77.74	66.25
3	38.73	62.02	73.26	78.55	62.86	38.55	60.85	74.39	79.59	61.63
4	20.18	33.17	51.65	60.59	69.07	18.77	32.16	52.29	60.24	65.26
5	29.19	47.78	53.24	71.01	56.39	29.77	48.96	54.47	71.60	58.97
6	75.71	79.48	95.66	95.90	58.22	75.84	78.20	95.25	96.60	56.56
11	32.83	42.18	44.08	53.26	53.03	31.75	41.20	42.77	52.09	55.05
12	67.70	82.78	93.59	94.74	51.21	66.92	82.70	94.33	94.55	50.67
14	62.21	78.43	81.09	85.76	65.18	61.93	78.14	81.91	85.81	65.30
16	52.42	66.63	78.31	80.69	79.53	52.33	65.78	79.58	82.18	74.41
17	62.09	77.10	89.35	94.79	72.31	62.24	78.35	88.99	93.76	76.61
20	75.25	73.78	78.97	84.58	53.29	73.97	72.47	77.88	84.64	54.21
21	61.75	72.68	74.55	75.32	73.72	62.57	73.58	75.90	76.44	78.09
25	62.45	80.49	82.02	91.31	64.50	62.21	81.31	82.73	89.81	65.96

Table 4.20c The prediction result using 6 groups of formulation

No.	R15	R30	R45	R60	R-F2	E15	E30	E45	E60	E-F2
1	67.43	77.39	99.46	95.54	57.22	68.55	77.99	99.67	93.77	57.71
2	34.49	56.14	54.96	74.75	54.65	32.80	57.15	56.77	71.72	53.56
4	25.74	46.45	64.06	67.28	54.09	26.31	45.78	65.47	64.41	54.33
5	31.98	55.17	49.32	68.19	51.38	32.62	58.55	48.42	70.64	50.21
16	51.60	56.84	63.80	75.55	55.57	48.26	59.83	63.99	73.65	56.86
21	65.71	79.48	85.66	90.90	52.45	63.40	78.55	84.09	88.32	55.97

In Table 4.20, No. represents the mark of formulation. The prediction results of prednisone release at 15min(R15), 30min(R30), 45min(R45), 60min(R60) by RLRM and 15min(E15), 30min(E30), 45min(E45), 60min(E60) by EDRM are set. Furthermore, the F2 similarity of RLRM(R-F2) and the F2 similarity of EDRM(E-F2) are presented.

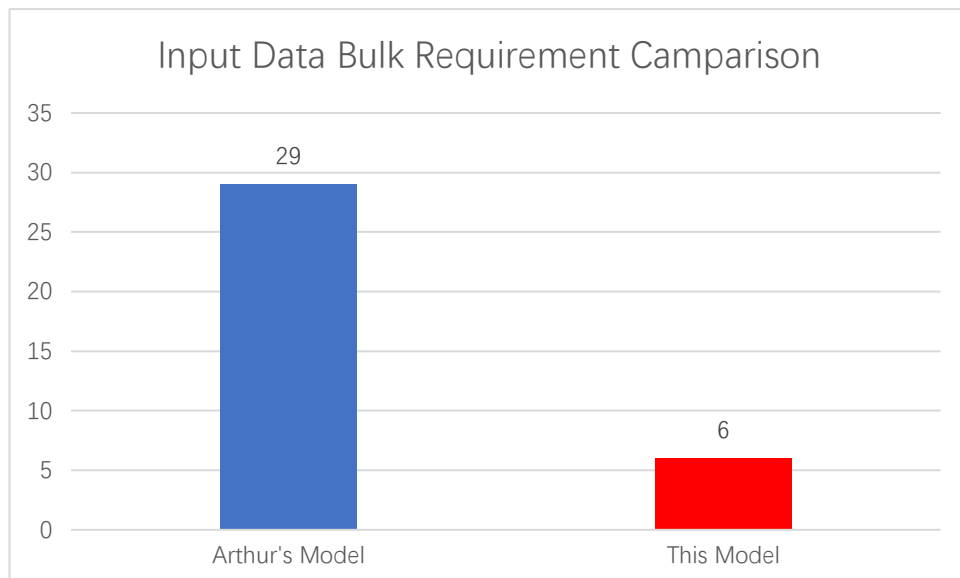


Figure 4.13 Data bulk requirement comparison between Arthur's model and our model

As depicted in Table 4.20 and Figure 4.13, F2 similarities in RLRM and EDRM

prediction results exceed 50. Compared with Arthur Manda's model, RLRM and EDRM can use a dataset with only six groups of formulation to predict the formulation dissolution curve with high accuracy instead 29 groups. In brief, the model of this study can reduce the database requirement less than orthogonal design data volume.

4.5.6 Input Data Screen

In practical application, large deviations of data often occur due to improper or incorrect operation of formulations by analysts during the experiment. In this study, data screening is of great importance. In this model, an initial predicted by the model was set as a training number. The program will predict the input data after the initial input data, compare it with the practical input data, and calculate the F2 value. The F2 value can be calculated by setting an appropriate F2 interval as the allowable error range. When F2 is beyond this interval, a prompt will pop up, hence researchers can manually delete the data. In addition, when the amount of data is large, it may take more time to conduct manual screening. At this time, it can be automatically filtered based on the F2 value. The program will automatically delete the suspected error data without asking whether to keep it.

4.6 Chapter Summary

In conclusion, the designed model is a data prediction model that can be applied in the early stage of formulation experiments or when the amount of data is small. When the data is abundant, the effect of noise is reduced due to the correction of a large amount of data, hence the prediction accuracy of the model will be improved. Moreover, when the prescription is changed, the data of the previous similar prescription can still be used to train the new prescription model. This method also tackles the problem that the preliminary research results cannot be referred to after the composition of the prescription excipients is fine-tuned. Furthermore, the designed model has the input data screening function that

immediately affects the input of abnormal data under the condition of a small amount of data, which lays a solid foundation for future high-throughput data training.

The uncontrollable noise will affect the prediction data during the experiment. The cause of the error may lie in the poor loading of tablets, whereas the prediction accuracy remains $F2 > 50$. The algorithm of EDRM is more stable and efficient than RLRM. The prediction accuracy of RLRM is affected by the reference line. Thus, an optimal curve should be selected first, and then it should be employed as a reference line when RLRM is adopted, thus increasing the prediction accuracy.

In contrast to previous research using the same database, RLRM and EDRM reduce the requirement of input data from 29 to 6 at high accuracy. With high prediction accuracy and low input data requirement, these two methods can be applied in the real pharmaceutical research and development project.

Chapter 5 Monitoring and Analyzing Solid Formulation Dissolution Phenomenon with Image Recognition Technologies

5.1 Chapter Introduction

Dissolution test is considered the critical quality index in the research and development of solid formulation, especially the evaluation of drug bioequivalence. However, it exhibits a low-level operability, and it is tedious, making it always overlooked. Existing research required fixed tablet and analyzed the recorded video using disso GUARO PRO and Microsoft Paint™. To this end, a novel image recognition system is developed to automatically track the moving tablet and simultaneously analyze the volume change. Besides, image recognition technology has been generally adopted to monitor the dissolution process, and the camera system with visible light and infrared camera functions are placed on the dissolution tester. The system is capable of collecting the plate image for binary processing, and then recording and calculating its pixel area, which can automatically record the volume change of tablets in the dissolution test, either the disintegration or the corrosion[100, 101].

5.2 Materials and Methods

5.2.1 Materials

Albendazole Tablets 200mg/Tablet (Sino-American Tianjin Skincare Pharmaceutical Co., LTD.), Famotidine Tablets 20mg/Tablet (Guangdong Bidi Pharmaceutical Co., LTD.), Lemon mints.

5.2.2 Sampling

There are two kinds of sampling, namely automatic sampling and manual sampling.

The automatic sampling system can automatically pump the sample within the required period. In contrast, manual sampling requires an experimenter to draw the sample using a special syringe [102].

5.2.3 Shading

There are two methods to shading, i.e., making the bath box with LED glass and applying shade cloth or masks when a shading experiment is required. When necessary, the mask is placed on the dissolution meter to isolate the external light source.

5.2.4 Camera and Infrared Light Source

An IR-CUT camera with the function of both infrared camera and visible light camera by automatic switch was used to record the image in both day and night. Six IR lights were installed on that camera to achieve a dark environment light source. Furthermore, a light-dependent resistor (LDR) was installed on the camera to switch the IR and visible light modes. Figure 5.1 presents the photo of the IR-CUT camera and the structure of the IR-CUT camera system.



Figure 5.1 The photo of IR-CUT camera module

As shown in Figure 5.1, 6 infrared light sources are posed on both left and right of the camera. A blue element besides of the camera is the LDR.

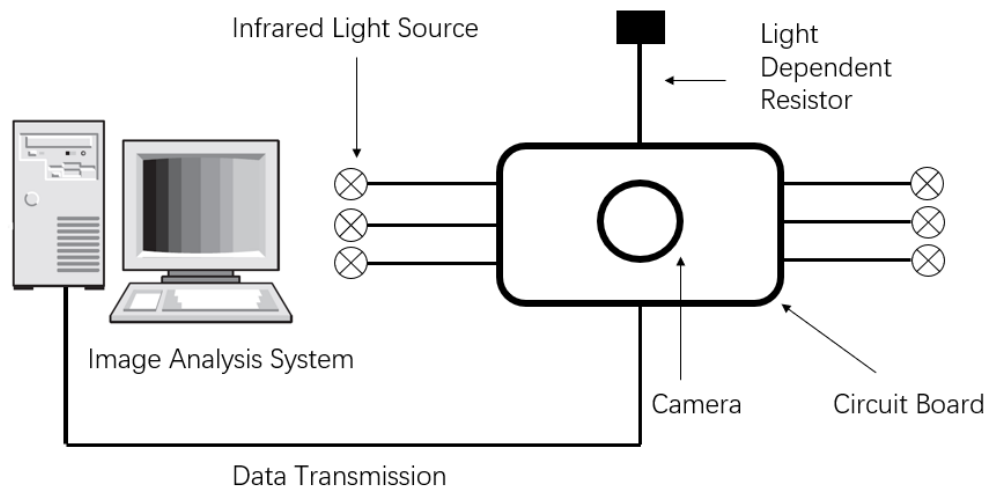


Figure 5.2 An IR-CUT camera system structure chart

As shown in Figure 5.2, the system is with image analysis system (include hardware and software), infrared light source, camera, circuit board and LDR.

5.2.5 Instrument Design

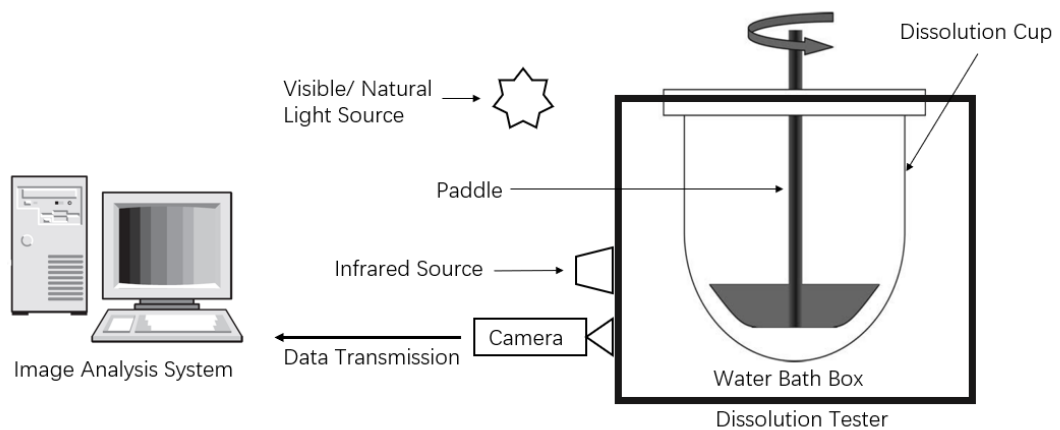


Figure 5.3 The whole instrument system structure chart

Figure 5.3 is the whole instrument system structure chart with both software and hardware for the dissolution test. The instrument system contains an image analysis program, a computer, a dissolution tester, visible/infrared light sources and an IR-CUT camera/normal camera. The image would be transferred into the image analysis program from the camera and the scatter diagram would be output as a result after analysis.

5.2.6 Program Modeling

Four steps showed below are used to achieve the image recognition in the image analysis program.

Step 1: Get the original image

The first step is starting the program to call the camera and then get a frame of image.

Step 2: Image preprocessing

The original image is resized to increase the speed of image processing. Afterward, box blur is adopted to remove the noise to achieve the condition of image contour connection.

Step 3: Image segmentation

Image binarization based on the threshold method sets tablet as white while setting the background as black. Subsequently, the white hole is removed using morphological transformation, and the white area contour is found to calculate the relative area. In accordance with the size of the white area, whether it is a tablet can be determined. A “central point” of this contour is determined using the average method (the average method refers to setting the average x-axis of top point and bottom point as “central point” x-axis and setting the average y-axis of left point and right point as “central point” y-axis) when the contour is determined as target tablet. Lastly, the tablet’s position is confirmed after the region grows based on the “central point.”

Step 4: Data visualization

The relative area of the tablet is determined in accordance with the pixel of region growth image. A scatter diagram with trend line is plotted in real time.

The diagram below presents the whole process of this program.

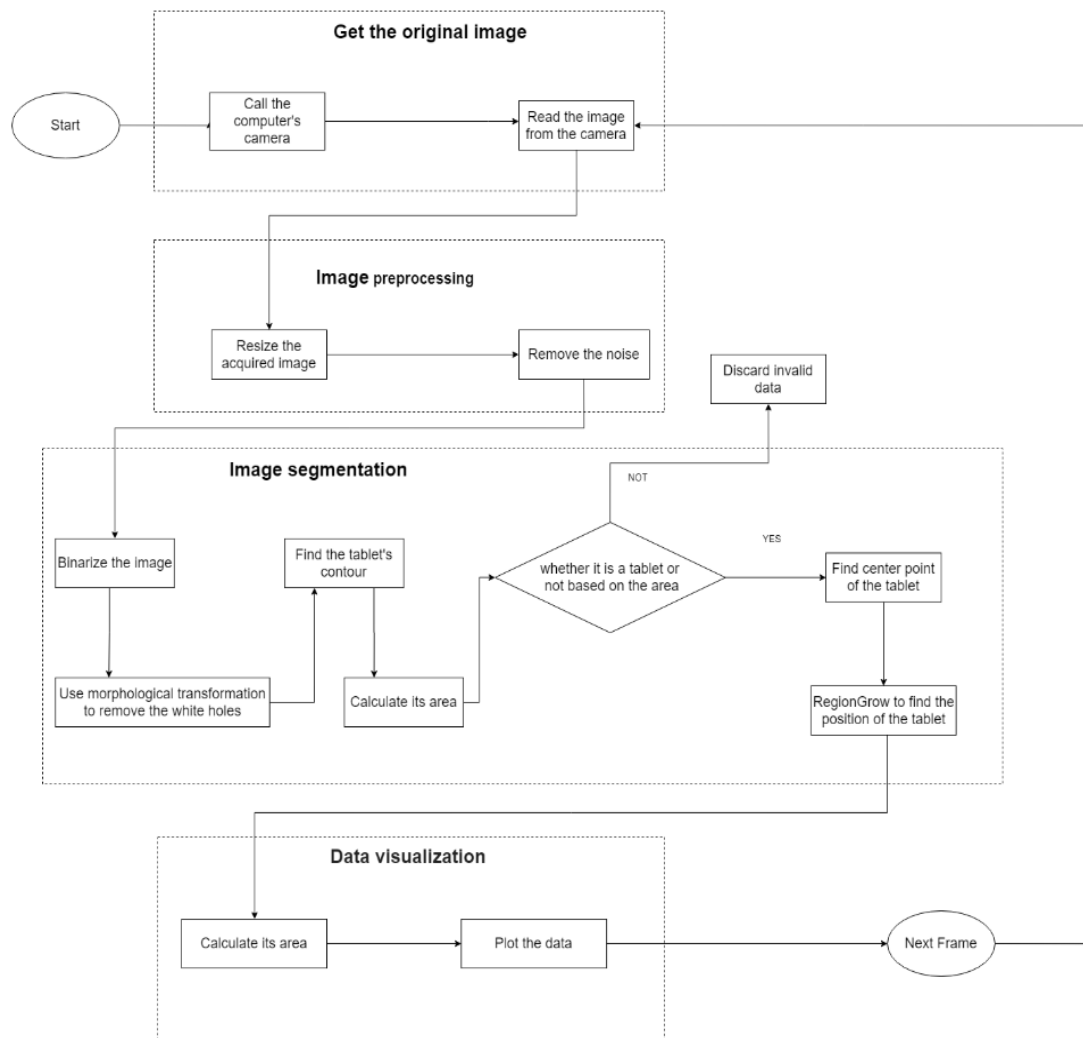


Figure 5.4 The process of program

5.3 Result and Discussion

In this project, the problem of capturing the moving tablet is settled. Moreover, this model solves the problem of the real-time monitoring of the moving tablet while completing the application testing. Previous projects only listed several limited examples

of specific tablets, whereas they failed to prove their capability of performing image recognition on tablets in all cases. Thus, real-time monitoring of tablets of different volumes, different colors, and different dissolution behaviors are performed to demonstrate the broad adaptability and application potential of the model.

5.3.1 Disintegration Tablets Image Recognition Test

The program suitability tests for disintegrating tablets fall into large-size tablets and small volume tablets.

Large-Size tablets: Since the inactive ingredients in large-size tablets are insoluble in water, they will suspend in the liquid under the action of the paddle. The dissolution medium will become cloudy till it is invisible to the naked eye in the disintegration process due to the overmuch insoluble ingredients of large-size tablets. Consequently, the resolution ability of the program in the case of noise interference needs to be tested with large-size tablets as the test standard. If the high-dose tablets are successful, the system could meet most of the experimental requirements.

Small-volume tablets: Small-volume tablets are the limit test for the resolution of the system. Since the recognition of tablets in the system is based on area, the system may ignore small volume tablets. This experiment uses small-volume tablets for limit testing. If there is a good result of the test, it means that the system's adaptability to tablet volume can meet the needs of routine experiments.

5.3.2 Large-Size Disintegration Tablets

Sample: Albendazole Tablets 200mg/Tablet (Sino-American Tianjin Skincare Pharmaceutical Co., LTD.)

Appearance: White coated round medicine, white or similar white after removing the coating

Prescription composition:

(a)Active ingredient: Abendazole

(b)Inactive ingredients include lactose, starch, polyvinylpyrrolidone, sodium carboxymethyl starch, sodium saccharin, magnesium stearate, microcrystalline cellulose, sodium dodecyl sulfate, hydroxypropyl methylcellulose, low substituted hydroxypropyl cellulose, and cassava starch. Lactose and cassava starch were used as the fillers, hydroxypropyl cellulose and sodium carboxymethyl starch were disintegrators, and magnesium stearate was the glidant.

Lactose acts as the filling agent of the tablet, starch acts as the binder and disintegrating agent of the tablet, magnesium stearate and sodium dodecyl sulfate serve as the lubricant of the tablet, hydroxypropyl methylcellulose serves as the binder of the tablet, and sodium saccharin serves as the flavouring agent to make the tablet taste better. Polyvinylpyrrolidone acts as the solubiliser to strengthen the solubility after the dissolution of benznidazole. Sodium carboxymethyl starch is the disintegrating agent. The disintegration promotion function of many disintegrants is affected by hydrophobic excipients (e.g. magnesium stearate and sodium dodecyl sulfate in this prescription), but sodium carboxymethyl starch is less affected by them. The excipients, which play a functional disintegration role, achieve rapid and significant swelling through rapid water absorption. Disintegration rate and phenomenon are the key factors of tablet dissolution rate.

The analysis of its prescription reveals that this formulation conforms to the characteristics of the samples required in this experiment. The experimental results are presented in Figure 5.5, 5.6 and 5.7. The respective group of figures comprises three figures. The first image is a scatter diagram of volume changes over time and a real-time trend line. The unit of the horizontal axis is second, and the vertical axis is in pixels. A screenshot of

the practical tablet is illustrated in the second figure, and the image after binarization is presented in the third figure.

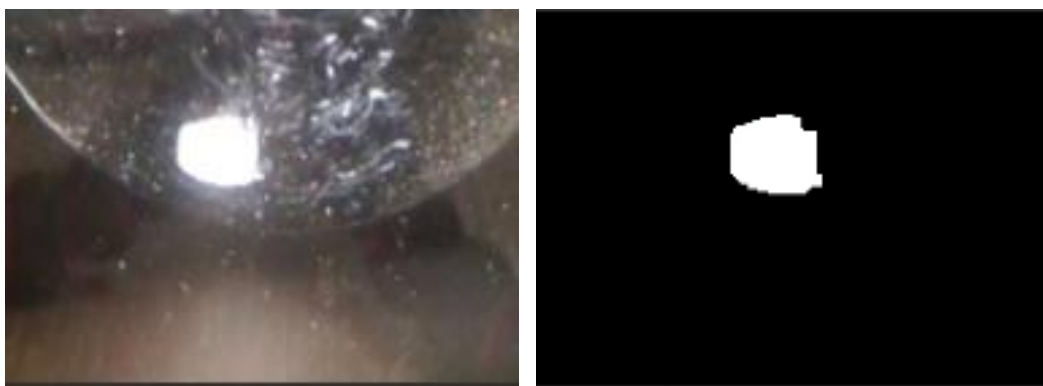
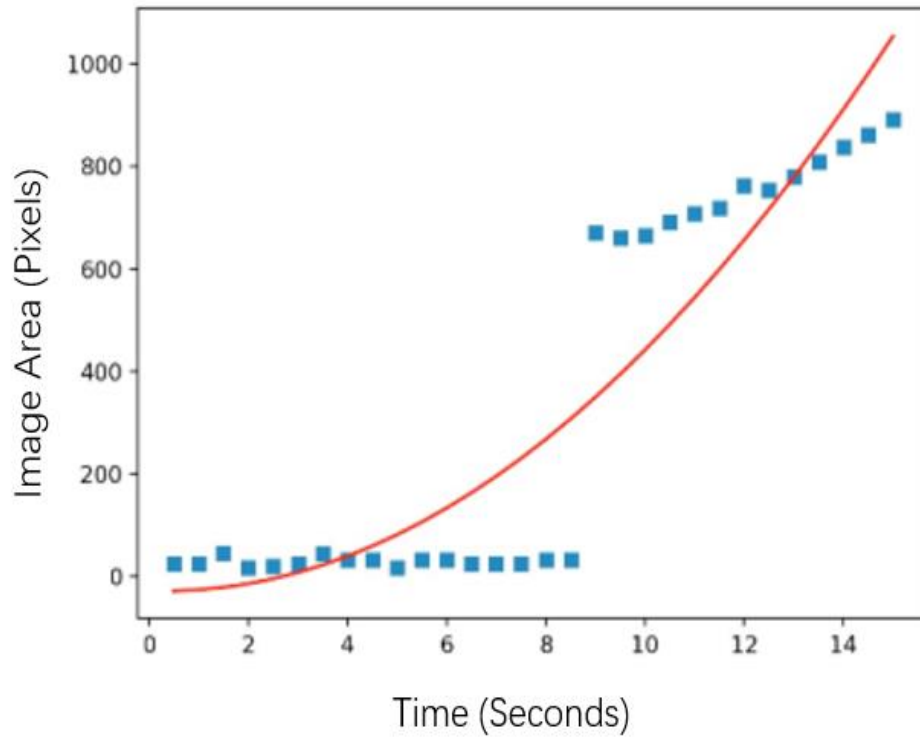


Figure 5.5 The experimental results of Large Size Disintegration Tablets (a)

As depicted in Figure 5.5, after the tablets are put in, the system starts to identify the tablets, draw a scatter diagram of the area of the respective period in accordance with the pixel area, and dynamically adds a trend line to guide the observation based on the trend

scatter.

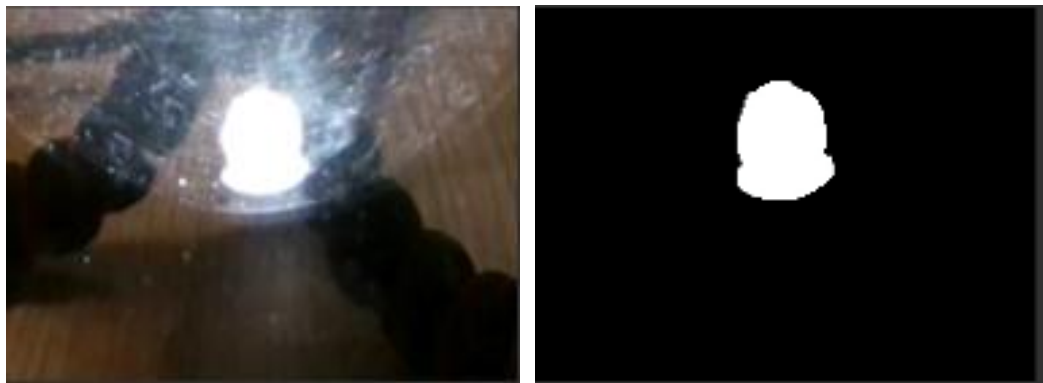
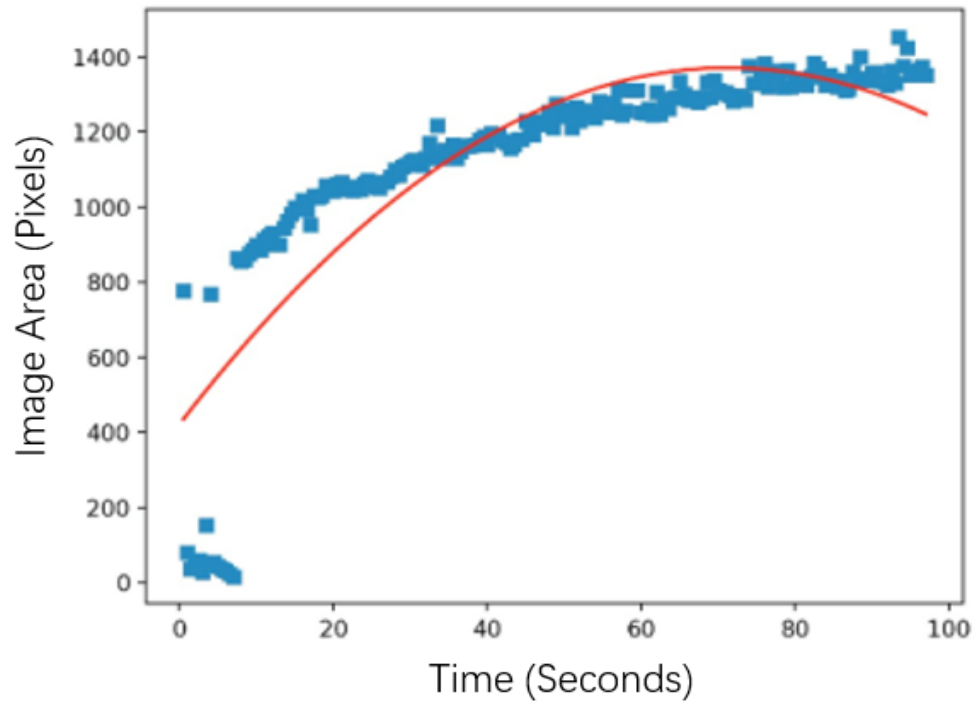


Figure 5.6 The experimental results of Large Size Disintegration Tablets (b)

As depicted in Figure 5.6, the tablets start to absorb water and get imbibition, appearing like a cake shape. Furthermore, the disintegration begins. The trend of the volume can be judged following the curve and scatter diagram.

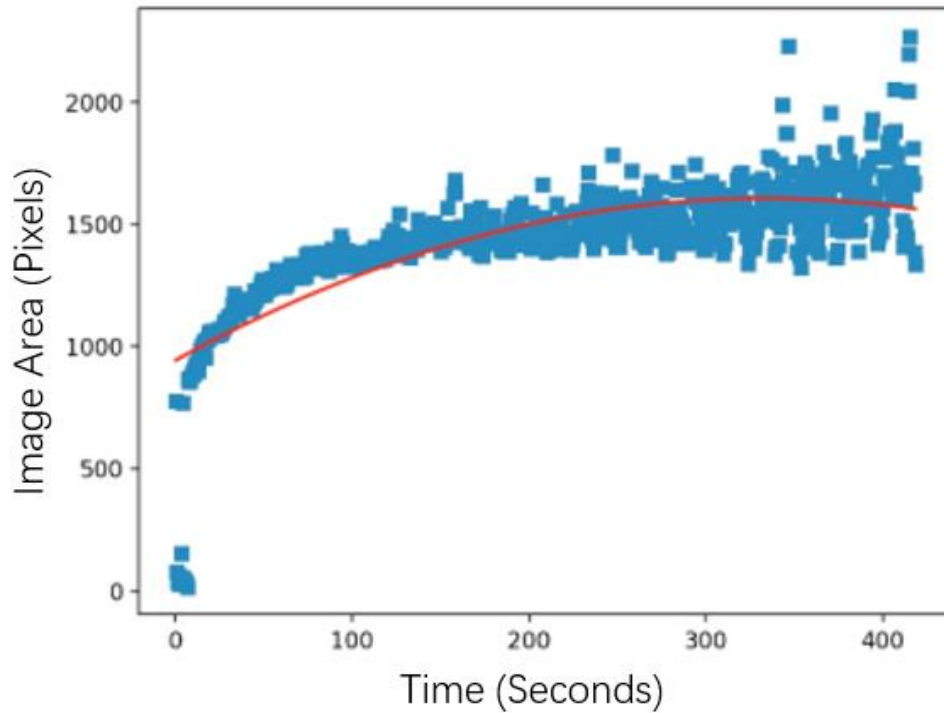


Figure 5.7 The experimental results of Large Size Disintegration Tablets (c)

As depicted in Figure 5.7, consistent with the trend line, it can be found that its volume becomes stable after 200 sec, indicating the disintegration phenomenon completed at the same time. The remaining identification images are all conical precipitates formed by water-insoluble inactive ingredients.

Summary

As depicted in Figure 5.5, 5.6, and 5.7, the liquid in the dissolution cup at the end of the disintegration is significantly turbid, i.e., almost indistinguishable by the naked eye.

However, the system can still effectively identify the aggregated objects, suggesting that it can be employed in practical applications. In brief, it is capable of replacing the experimenter to observe the dissolution behavior of large-size tablets.

5.3.3 Small Size Disintegration Tablets

Sample: Famotidine Tablets 20mg/Tablet (Guangdong Bidi Pharmaceutical Co., LTD.)

Appearance: White and round tablet

Prescription composition:

(a)Active ingredient: Famotidine

(b)Inactive ingredients: Lactose, low substituted hydroxypropyl cellulose, cassava starch, sodium carboxymethyl starch, magnesium stearate. Lactose and cassava starch are the fillers, hydroxypropyl cellulose and sodium carboxymethyl starch are using as disintegrator, and magnesium stearate is the glidants.

Figure 5.8, 5.9, and 5.10 present the experimental results. The respective group of figures comprises three figures. The first image is a scatter diagram of volume changes over time and a real-time trend line. The horizontal axis is in second, and the vertical axis is in pixels. A screenshot of the practical tablet is presented in the second figure, and a binarized image is illustrated in the third figure.

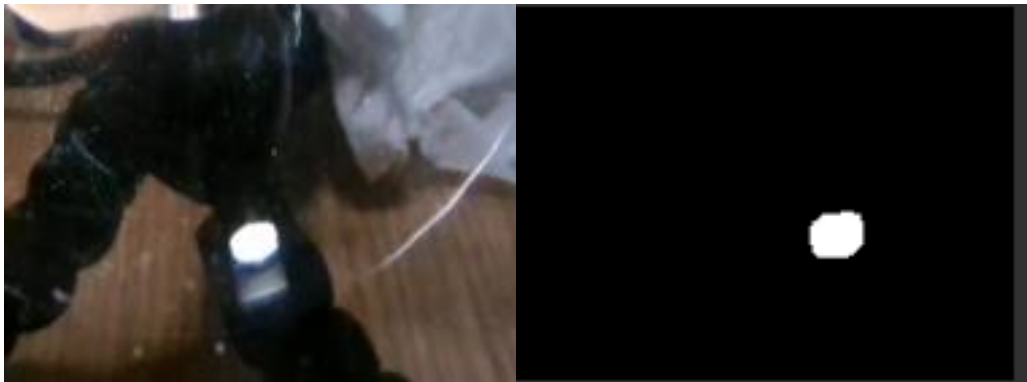
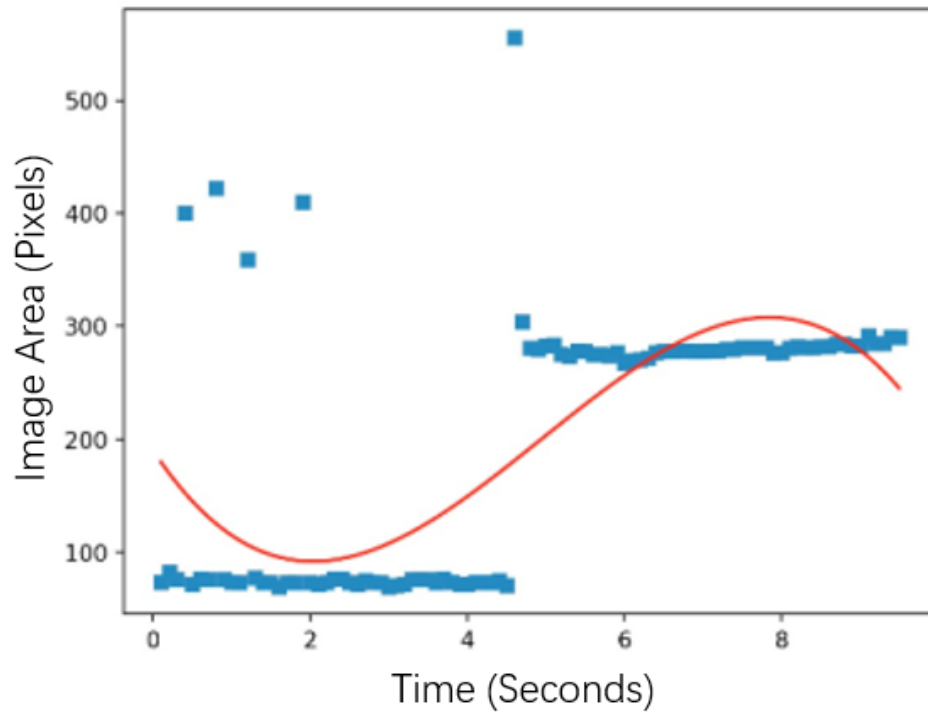


Figure 5.8 The experimental results of Small Size Disintegration Tablets(a)

As depicted in Figure 5.8, after the small tablet is placed into the dissolution cup, the image recognition program successfully recognizes them and starts to generate a scatter diagram of area changes.

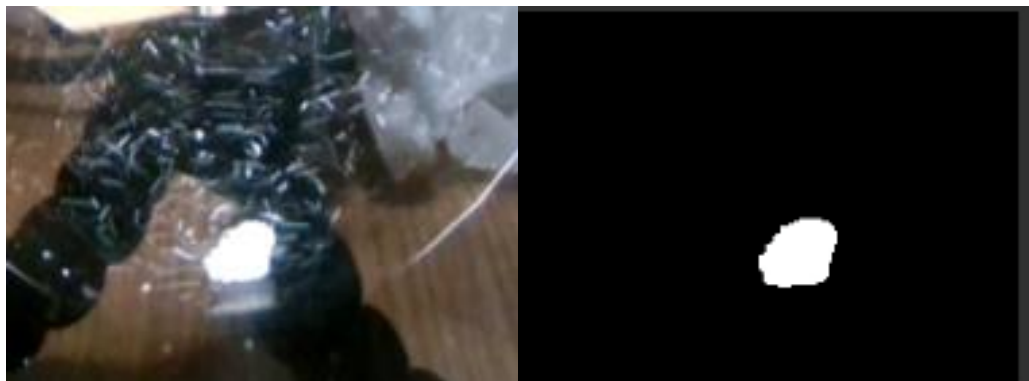
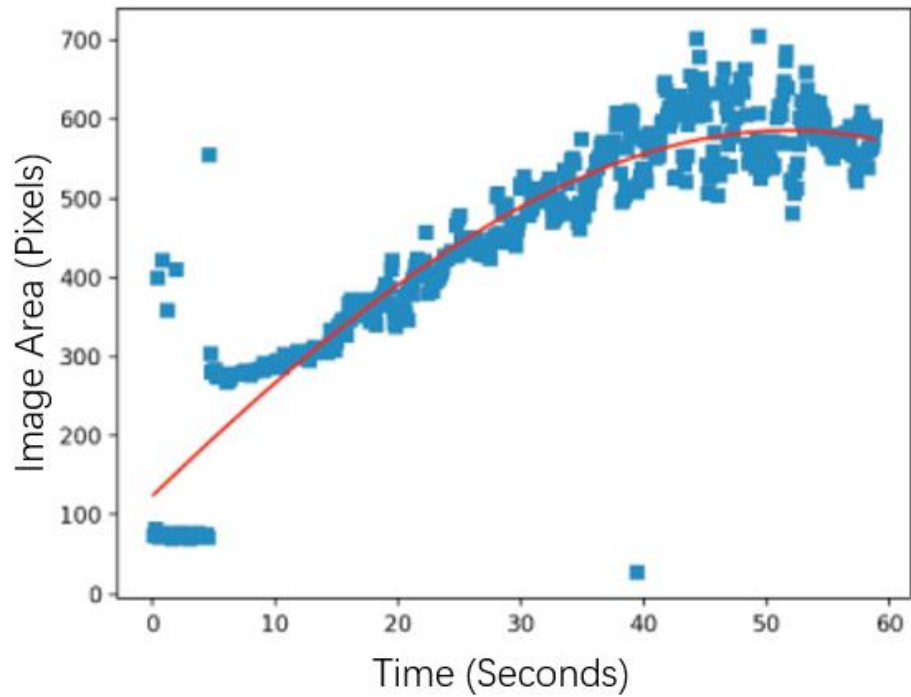


Figure 5.9 The experimental results of Small Size Disintegration Tablets(b)

As depicted in Figure 5.9, the small tablet starts to absorb water and get imbibition. Subsequently, particles spread in the respective part of the dissolution medium, whereas the image recognition program still clearly recognizes the tablet and paints an exemplary scatter diagram.

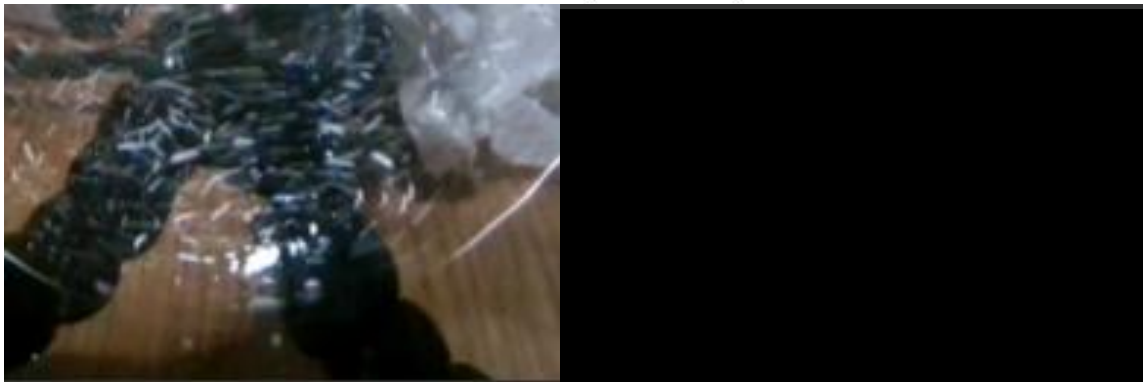
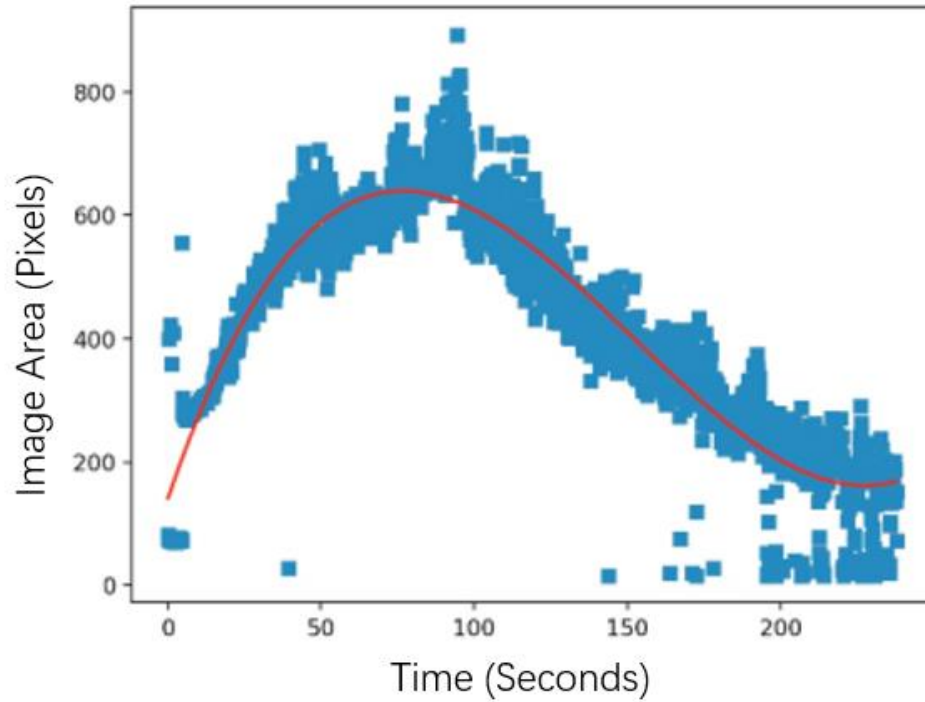


Figure 5.10 The experimental results of Small Size Disintegration Tablets(c)

Figure 5.10 presents the completion of small volume tablet disintegration, and most of the insoluble particles are suspended in the cup. Since the particles are smaller than the set limit value, they are no longer recognized. The area change of the whole dissolution process can be observed easily in accordance with the final curve.

This experiment examines the system's ability to recognize small-volume tablets, such that the needs of typical experiments can be met.

5.3.4 Erosion Tablets

Identification test of corrosion phenomena

The erosion phenomenon usually occurs in the dissolution experiments of standard formulations and sustained-release formulations with dissolution skeletons. Among them, two widely recognized soluble formulations are candy and lozenges. The sample use here is Lemon mints tablet with yellow color and double layer structure.

Sample: Lemon mints

Appearance: Yellow and white double layer oval tablet

Prescription composition: Sorbitol, DL-malic acid, magnesium stearate, sucralose, natural menthol, tartrazine, vitamin C, in which except magnesium stearate as a flow aid, all are flavor correction agent, dye or preservative.

Figure 5.11 presents the experimental results. The respective group of figures comprises three figures. A scatter diagram of volume changes over time and a real-time trend line are presented in the first figure. The horizontal axis is illustrated in the second figure, and the vertical axis is in pixels. A screenshot of the practical tablet is presented in the second figure, and the image after binarization is illustrated in the third figure.

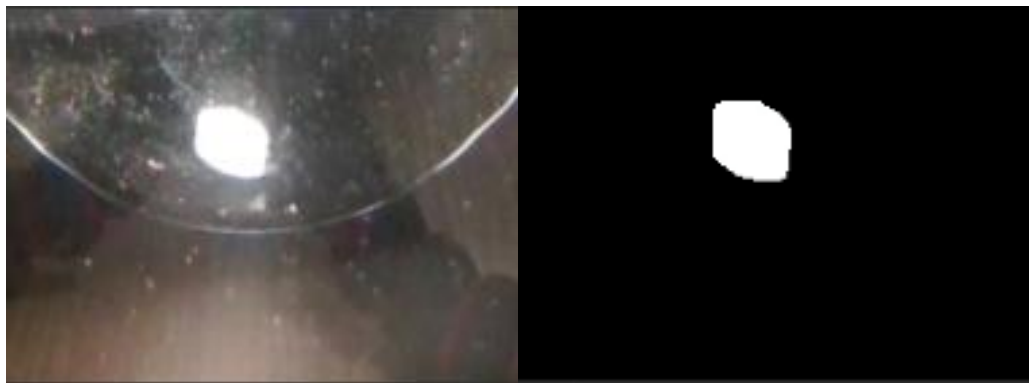
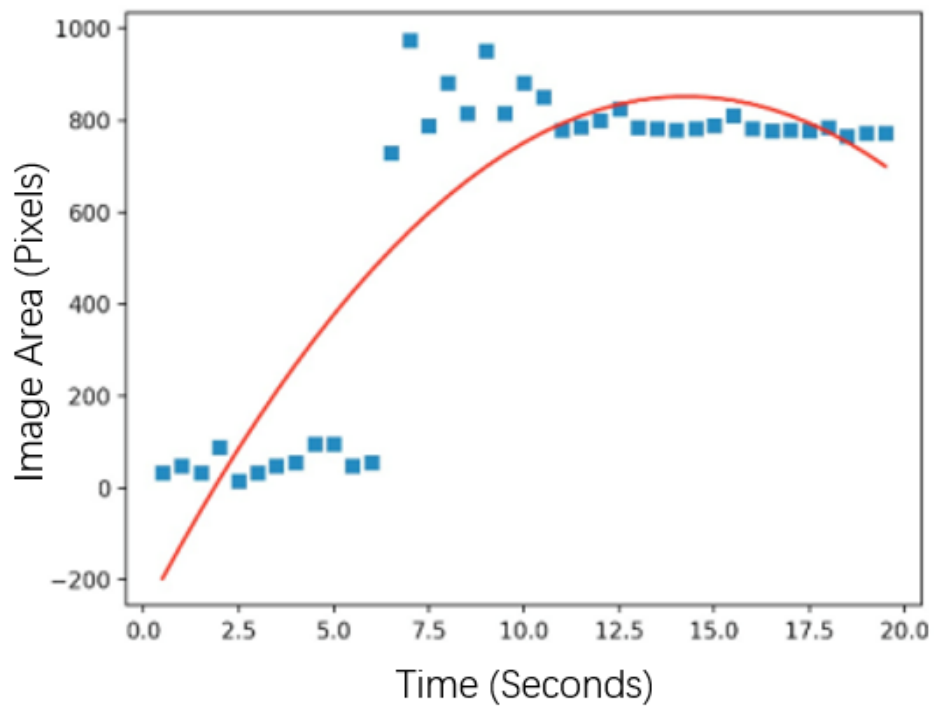


Figure 5.11 The experimental results of Erosion Tablets (a)

As depicted in Figure 5.11, the tablets are placed at the bottom of the dissolution cup, the image recognition program starts to identify them effectively, records the scatter diagram of volume changes while adding the real-time trend line.

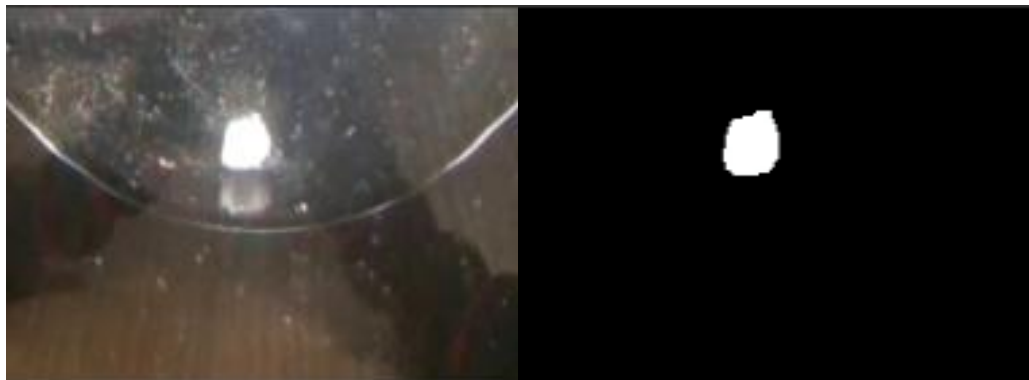
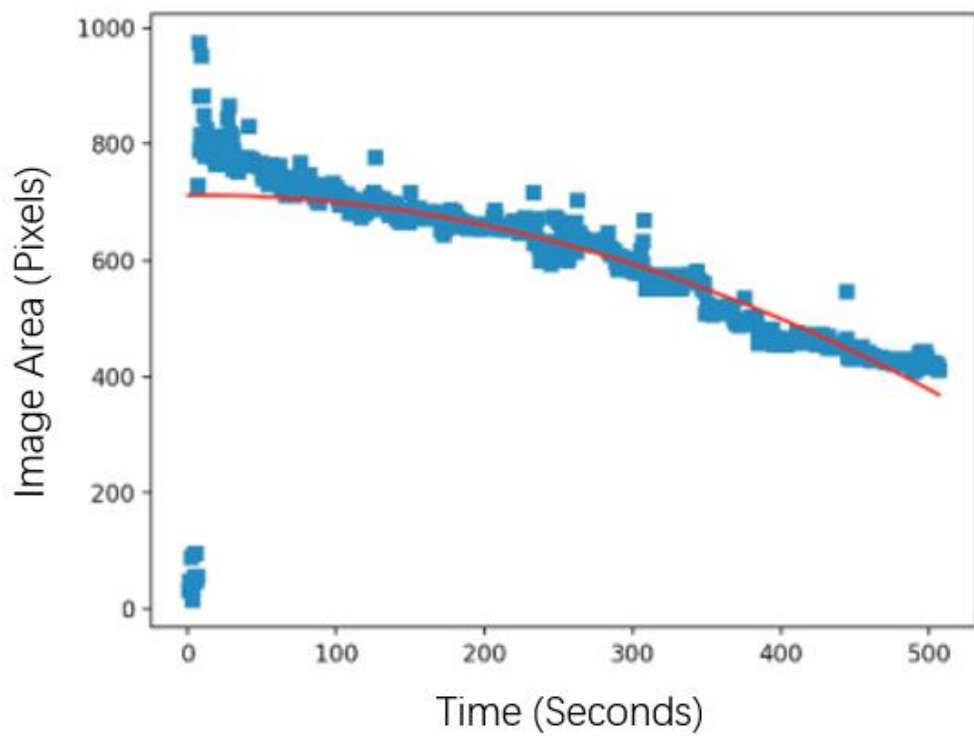


Figure 5.12 The experimental results of Erosion Tablets (b)

As depicted in Figure 5.12, the tablet gradually dissolved and became a smaller size. The image recognition program is also operating normally, and the scatter diagram is plotted well, and the real-time trend line goes well.

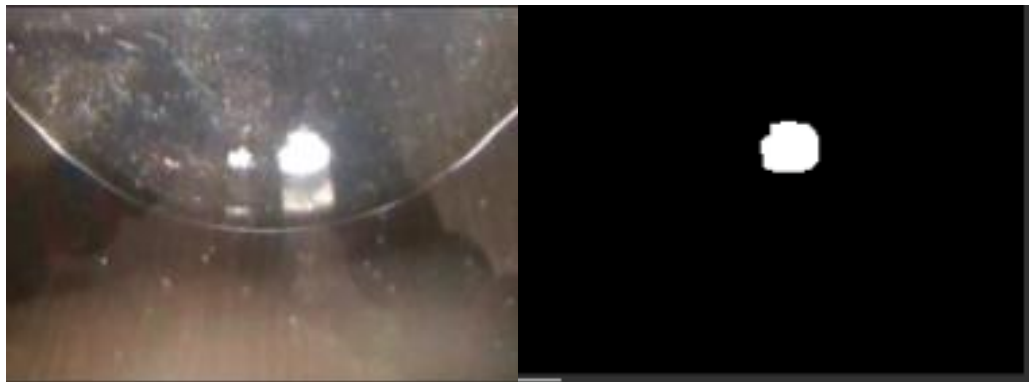
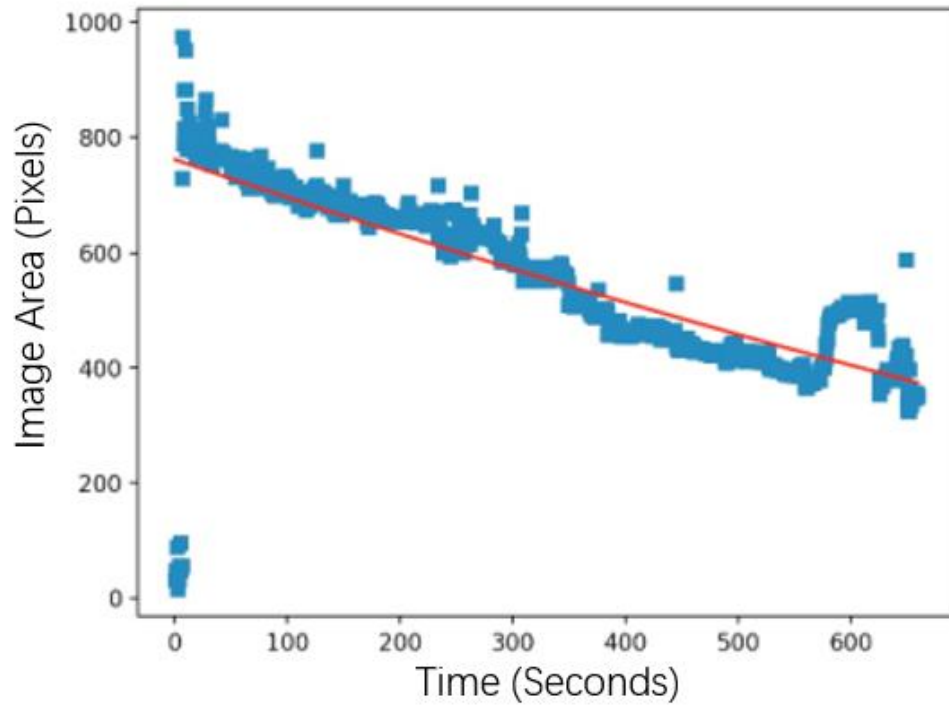


Figure 5.13 The experimental results of Erosion Tablets (c)

As depicted in Figure 5.13, due to the fragmentation phenomenon of the tablet during the dissolution process, resulting from the data and the scatter diagram anomaly, but as the fragments continue to become smaller, the data returns to the trustworthy interval.

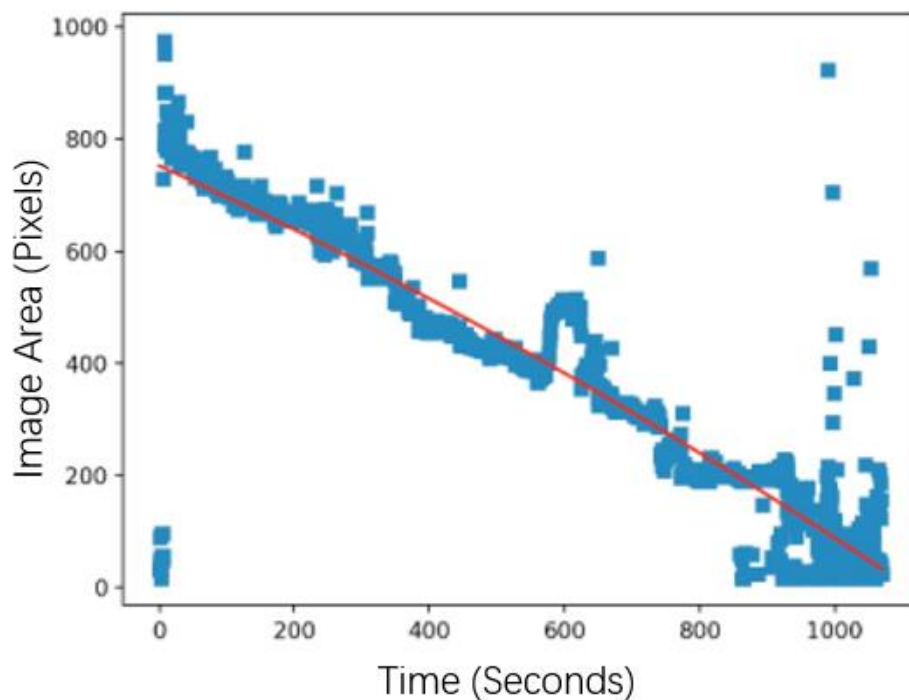


Figure 5.14 The experimental results of Erosion Tablets (d)

As depicted in Figure 5.14, the tablet is completely dissolved. The volume (recognition area) changes during the eroding period can be accurately analyzed in accordance with the scatter diagram.

Summary

In this dissolution experiment, unexpected situations (e.g., tablet debris) may cause short-term analysis errors. However, with the gradual decrease of the large pieces of debris, the system returns to normal, accurately depicts the scattered points of the identified area

image while reasonably adding real-time trend lines.

5.3.5 Different Color Tablets

The coating of tablets has a variety of colors. Figure 5.15 presents the coating color card provided by Shanghai Colorcon Coating Technology Co., LTD. There are 200 colors from dark to light, thus posing a significant challenge to the sensitivity of image recognition in the dissolution experiment.

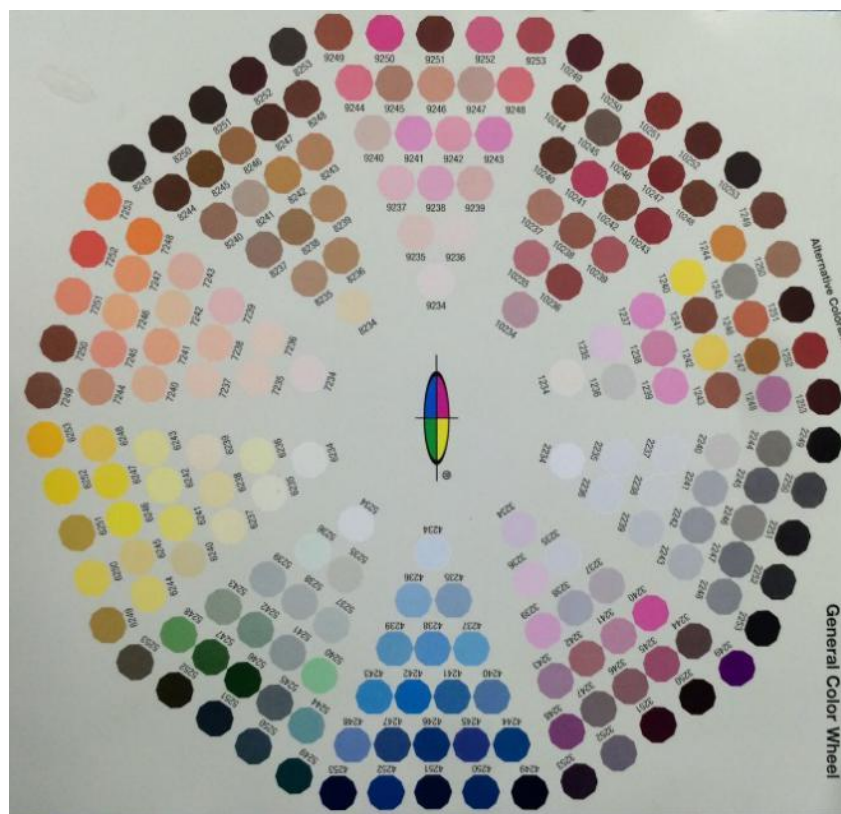


Figure 5.15 General Color Wheel from Shanghai Colorcon Coating Technology Co.LTD

White, yellow, bright red, bright blue, deep blue, dark red, green and purple small round plastic plates are selected to test the recognition ability of the system to increase and test the adaptability of the system. Eight groups of figures are presented in Figure 5.16(a-h).

The first figure presents a practical photo of the respective set of plastic sheets. Moreover, second figure represents the photo of the plastic plates at the bottom of the dissolution cup. Image after binarization is presented in the third figure. The fourth figure illustrates the photo after noise cancellation.



Figure 5.16a White small round plastic plate



Figure 5.16b Yellow small round plastic plate



Figure 5.16c Bright red small round plastic plate

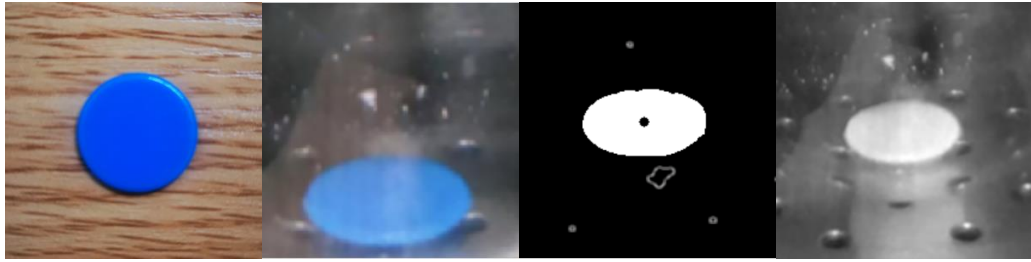


Figure 5.16d Bright blue small round plastic plate



Figure 5.16e Deep blue small round plastic plate

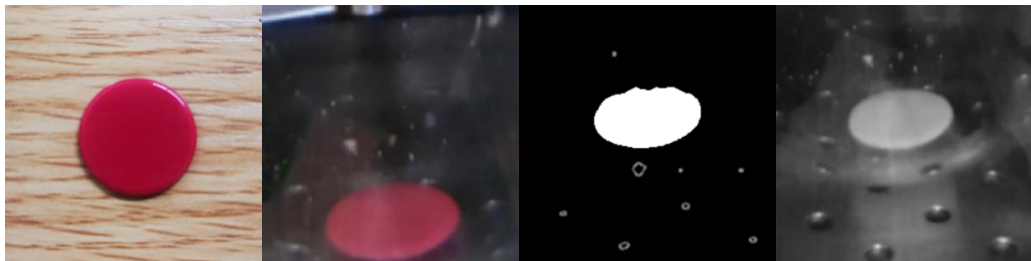


Figure 5.16f Dark red small round plastic plate



Figure 5.16g Green small round plastic plate



Figure 5.16h Purple small round plastic plate

Summary

In this test, eight colors cover the spectrum from long optical to short optical wavelength while receiving a good result that the image recognition program can successfully recognize the above-mentioned color plates. As revealed by the above result, the program exhibits high suitability in tablets with standard colors.

5.3.6 Infrared Condition

For some photosensitive drugs, shading treatment should be performed for the dissolution experiment. Since the water bath completely blocks light, visible light cameras cannot be employed, resulting in the necessity of adopting the infrared camera to overcome the deficiency. Moreover, the image recognition of the infrared camera is optimized. The advantage of infrared cameras is not just their applicability to the situation without visible light. In general, the images are black and white, such that the color of the tablet is ignored in this experiment. The IR-CUT camera module is adopted using this system, in which the main structure serves as an LDR to control the switch between infrared camera and visible camera. Six infrared light sources automatically turn on when the camera is switched into infrared camera mode, infrared mode, and visible light mode. Figure 5.17(a-c) present the system practicability test of the infrared camera under infrared light source. The sample applied is consistent with those employed in the small volume tablet experiment.

The first image depicts a scatter diagram of volume changes over time and a real-time trend line. The unit of the horizontal axis is the second, and the vertical axis is in pixels. A

screenshot of the practical tablet is presented in the second figure, while the image after binarization is presented in the third figure.

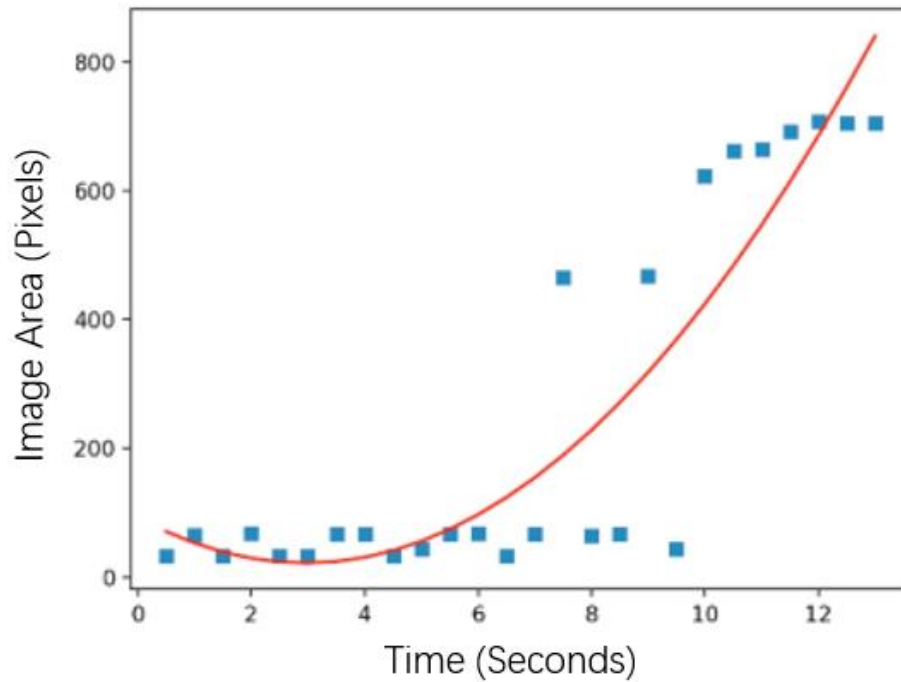


Figure 5.17a System practicability test(a)

As depicted in Figure 5.17a, when the tablet is located at the bottom of the dissolution cup, the infrared camera successfully captures the tablet, and the image recognition program successfully recognizes it.

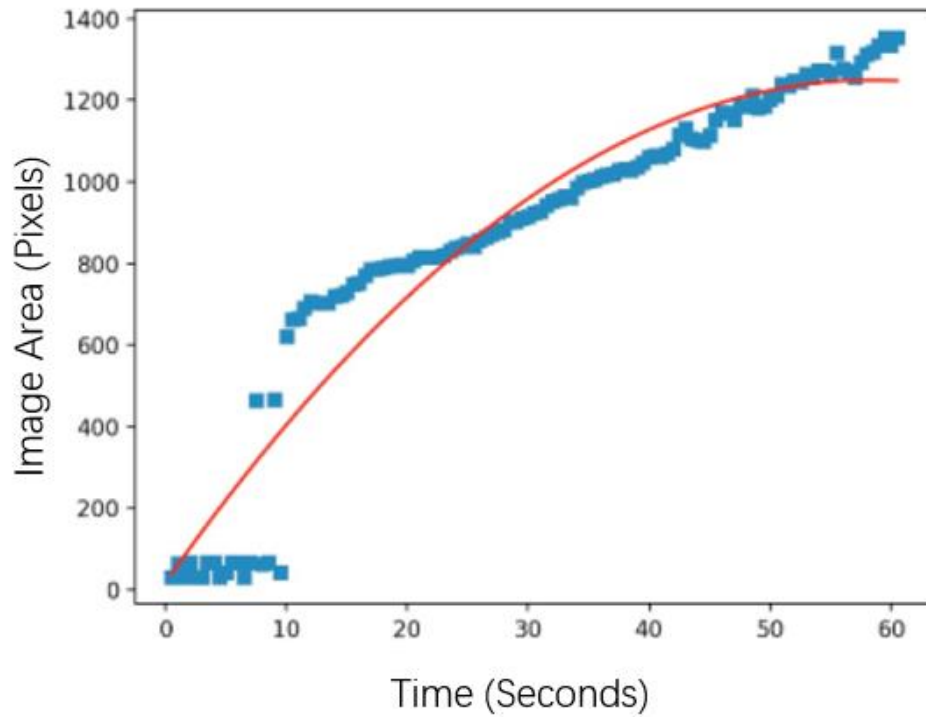


Figure 5.17b System practicability test(b)

As depicted in Figure 5.17b, the tablet begins to absorb water and expand. Moreover, the reflection of the stirring paddle in the second figure does not interfere with the image recognition presented in the third figure. The program runs well and plots the ideal scatter diagram, and real-time trend lines are introduced.

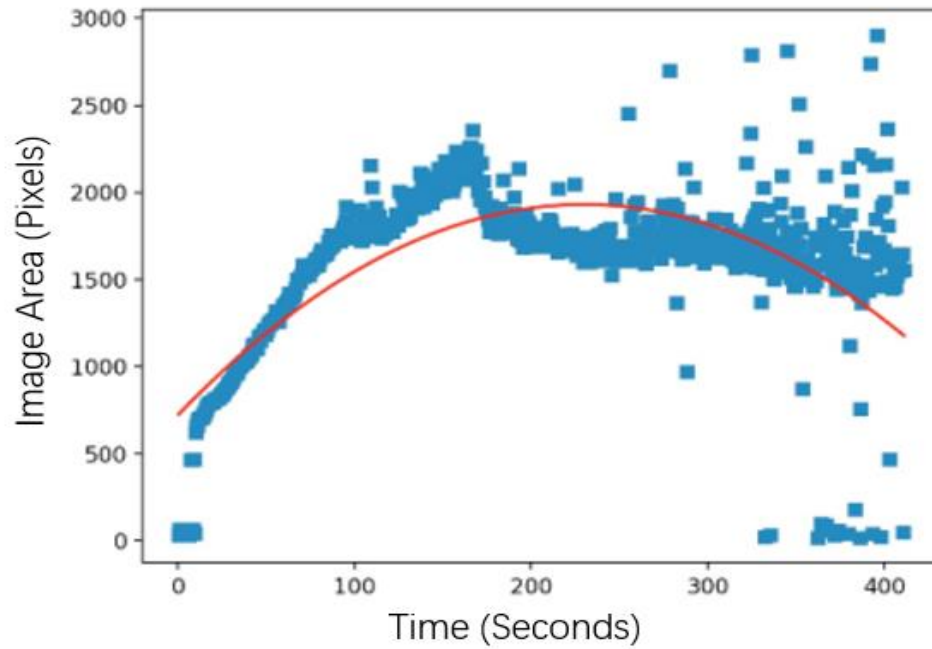


Figure 5.17c System practicability test(c)

As depicted in Figure 5.17c, the disintegration of the tablet is completed, and the water-insoluble inactive ingredients precipitate and accumulate. The scatter diagram in the first figure indicates that the area tends to be constant after nearly 300 sec. As a result, the total disintegration time can be judged at nearly 300 sec based on the scatter diagram and the trend line.

Summary

The light source is closer to idealization since there is no external light interference, such that infrared imaging has less environmental noise.

5.4 Chapter Summary

In result, a summarize table was shown in Table 5.1 below.

Table 5.1 A summarize table of experimental result and discussion

Experiment	Results	Comment
Large-size Disintegration	Tablets According to the change of measured area, the dissolution behavior was recorded by the scatter diagram clearly.	N/A
Small Volume Disintegration	Tablets The program successfully recognizes the small volume tablets from the beginning till the end.	N/A
Erosion	The output of scatter diagram gives a trend of tablet volume reduce.	A short noise result from a small tablet debris can influence the scatter diagram plotting, but it has no effect on the trend line.
Different Color	White, yellow, bright red, bright blue, deep blue, dark red, green and purple small round plastic were used to test the program suitability and received a good result.	N/A
Infrared Condition	Because of the stable light environment, the infrared condition experiment showed a good result.	N/A

This project shows that the real-time dissolution phenomenon monitoring (RTDPM) system provides solution to capture the moving tablets and provide a camera system to reduce the risk of light unstable tablets. This has been proven in the tests where it can satisfy the disintegration phenomenon of both small and regular size tablets. Besides, the volume

change curve of the corrosion plate can also be recorded. For some tablets that are less stable in visible light, IR method can give a perfect solution (all substances above absolute zero emit infrared light. The infrared light source of the IR camera can slightly enhance the illumination intensity but not luminous energy). To sum up, this system can solve most dissolution test image recognition problems.

However, the color of the tablets and the light jam of the environment in this system still affect the test result, which should be noted in future research.

Chapter 6 Conclusions and Future Works

6.1 Concluding Remarks

In Chapter 3, an ANN prediction model based on BPNN and RBFN is adopted to predict the dynamic solubility of Indometacin powder in water. In accordance with the database of five experiments, the model completes the training and provides a prediction result under different conditions with high accuracy. The F2 similarity between the experiment result and the prediction result range from 83.29 to 93.5 for the BPNN model, as well as between 67.23 and 91.46 for the RBFN model. Besides, a data optimization module is also presented in this chapter based on the relative standard deviation of experiment data groups. Moreover, the theory trend of the dynamic solubility curve is determined using a core equation derived from Hixson-Crowell cube-root and Nernst-Brunner expression. The application of this method limits the frequency of dynamic solubility experiment as 5. Compared with DDD Plus, the ANN model required less input elements and provided an easier operating mode.

A model is proposed based on ANN and non-linear regressions in Chapter 4 to predict the dissolution result of different prescriptions tablets. The research on Tinidazole tablet is adopted to test the prediction model, while the result of dissolution test is predicted using ANN, and two novel non-linear regression methods (i.e., EDRM and RLRM) are adopted to calculate the final value of prediction output. As indicated in the test results, the prediction results over 50 of EDRM and RLRM meet the guidance of F2 similarity. Compared with Arthur Manda's work[32], EDRM and RLRM can reduce the requirement of input data from 29 to 6 with over 50 F2 value. In contrast, EDRM provides more stable prediction accuracy, whereas the RLRM method can offer a high prediction accuracy under the premise of highly accurate reference data. Besides, an input data screening function is

achieved by the decision tree in accordance with the value of F2 similarity, such that the frequency of tablet experiment can be limited in a designed number. As a result, the time is saved to perform at least half the experiments.

Chapter 5 offers an image recognition system consisting of an IR camera module and image recognition program to monitor the phenomenon of tablet dissolution. In this system, an IR camera with infrared light source was adopted to record the image of tablets in visible light and dark conditions. Based on the region growth and HSV, the system can recognize and calculate the pixel areas of different colors tablets. The camera can also automatically capture the moving tablet image using the image segmentation method, thereby saving at least 8-72 h of working time for a single experiment.

6.2 Future Works

The proposed model can accurately predict the dynamic solubility of the API and dissolution of tablets, but its function still needs to be optimized. First, although the demand for input data is significantly reduced in the model, the required training data should still be obtained through experiments. The optimal method to tackle down the above-mentioned problem is to build a large database of formulation dissolution in cooperation with pharmaceutical companies. On that basis, relevant data will serve as the inputs to train the model, such that the experimental results can be predicted without experimental data. This method exhibits high prediction accuracy, whereas the data leakage should be strictly avoided in the establishment of the database since most of the experimental data are confidential information originating from a company. Moreover, the properties of all excipients can be digitally modeled to develop a database. When a novel formulation process is substituted into the model, predictions are generated based on the dynamic solubility of the drug substance, the physicochemical properties of the excipients, as well as the process parameters. The advantage of the above-described model lies in that it does

not require a large dissolution database. Given the absence of practical experimental data, the predicted results may not be always accurate, whereas they still take on guiding significance to a certain extent.

Nonetheless, the image recognition function of the model still has a lot of room for improvement. First, image recognition utilizing neural network techniques (e.g., convolutional neural networks) may work more efficiently under the sufficient database available. Second, the tablet volume in this model is presented by a 2D pixel image, thus revealing the tablet volume changes, whereas the calculating accuracy cannot be ensured. In future research, multiple cameras can be used to monitor the pills and draw the 3D images to accurately calculate the volume changes. Third, with the development of mobile technologies like the 5G technology, the internet of everything (IoE) will also be realised among dissolution equipment. The images and analysis results should be shared on mobile devices like mobile phones through the IoE to achieve real-time experimental monitoring.

List of Research Outputs

Journal Article Publications

- (1) **Haoyu Wang**, Chiew Foong Kwong, Qianyu Liu, Zhixin Liu, Zhiyuan Chen, "*A Novel Artificial Intelligence System in Formulation Dissolution Prediction*", Computational Intelligence and Neuroscience, vol. 2022, Article ID 8640115, 11 pages, 2022. <https://doi.org/10.1155/2022/8640115> (SCI, JCR Q1)
- (2) **Haoyu Wang**, Chiew Foong Kwong, Qianyu Liu, Junyao Liu, Zhixin Liu, Boon Giin Lee, Liang Huang, "*Monitoring and Analysis Solid Formulation Dissolution Phenomenon with Image Recognition Technologies*", Computational Intelligence and Neuroscience, vol. 2022, Article ID 3997870, 16 pages, 2022. <https://doi.org/10.1155/2022/3997870> (SCI, JCR Q1)

Invention Patents

- (1) **Haoyu Wang**, David Siu-Yeung Cho, Zhongyan Wang, Ningyuan Qin, "*Methods, equipments and systems for predicting dynamic solubility of active pharmaceutical ingredients by artificial intelligence*", China, Invention Patent, Patent Number 201910991714.6.
- (2) **Haoyu Wang**, David Siu-Yeung Cho, Boon Giin Lee, Zhiyuan Chen, "*A system, method and equipment for monitoring and analyzing dissolution behavior using artificial intelligence image recognition technology*", China, Invention Patent, Patent Number 202110842211.X.
- (3) **Haoyu Wang**, Chiew-Foong Kwong, Zhiyuan Chen, David Siu-Yeung Cho, Boon Giin Lee, Liang Huang, "*A system and method for artificial intelligence to predict the dissolution curve of formulation*", China, Invention Patent, Patent Number 202111663521.1.
- (4) **Haoyu Wang**, Chiew-Foong Kwong, Zhiyuan Chen, David Siu-Yeung Cho, Boon Giin Lee, Liang Huang, "*A system and method for artificial intelligence to screen abnormal data*", China, Invention Patent, Patent Number 20211657874.0

References

- [1] S. Rennane, L. Baker, and A. Mulcahy, "Estimating the Cost of Industry Investment in Drug Research and Development: A Review of Methods and Results," *Inquiry-J Health Car*, vol. 58, Nov 11 2021, doi: Artn 00469580211059731, 10.1177/00469580211059731.
- [2] M. Guarnieri, J. Kedda, and B. Tyler, "Buprenorphine implants: a model for expedited development and approval of new drugs," *Curr Med Res Opin*, vol. 37, no. 1, pp. 83-88, Jan 2 2021, doi: 10.1080/03007995.2020.1840971.
- [3] R. G. Frank, T. G. Mcguire, and I. Nason, "The Evolution of Supply and Demand in Markets for Generic Drugs," *Milbank Q*, vol. 99, no. 3, pp. 828-852, Sep 2021, doi: 10.1111/1468-0009.12517.
- [4] CFDA (2010). Consistency Evaluation of Generic Drug Guidelines. CFDA. China, Chinese Pharmacopoeia. [Online] Available: <https://www.cde.org.cn/>
- [5] P. Attuel *et al.*, "Prevention of atrial fibrillation by DDD plus atrial overdrive pacing: final results of a randomized crossover study," *Eur Heart J*, vol. 23, pp. 139-139, Aug-Sep 2002. [Online]. Available: <Go to ISI>://WOS:000179753300523.
- [6] S. Saeed, T. Mehmood, and M. Irfan, "Statistical optimization of cultural parameters for the optimized production of alginic acid using apple (*Malus domestica*) peels through solid-state fermentation," *Biomass Convers Bior*, Nov 10 2020, doi: 10.1007/s13399-020-01143-9.
- [7] K. Wang *et al.*, "In Vitro Predictive Dissolution Test Should Be Developed and Recommended as a Bioequivalence Standard for the Immediate-Release Solid Oral Dosage Forms of the Highly Variable Mycophenolate Mofetil," *Mol Pharmaceut*, May 23 2022, doi: 10.1021/acs.molpharmaceut.1c00792.
- [8] T. Felicijan, A. Krese, A. Mrhar, and M. Bogataj, "Applicability of Bottom-View Cameras for Evaluation of Tablet Performance During Dissolution Testing," *Dissolut Technol*, vol. 23, no. 4, pp. 24-32, Nov 2016, doi: 10.14227/Dt230416p24.
- [9] C. Sticherling, M. Niehaus, and M. Zabel, "Comparison of single-lead ICD system capable of atrial sensing (A plus ICD) and a DDD-ICD system in patients without antibradycardia pacing indications," *Eur Heart J*, vol. 30, pp. 278-279, Sep 2009. [Online]. Available: <Go to ISI>://WOS:000208702602514.
- [10] Almukainzi M, Okumu A, Wei H, Lobenberg R. "Simulation of in vitro dissolution behavior using DDDPlus™," *AAPS PharmSciTech*, vol. 16,1, Nov 2015, doi:10.1208/s12249-014-0241-5.
- [11] S. Kalepu and V. Nekkanti, "Insoluble drug delivery strategies: review of recent advances and business prospects," *Acta Pharm Sin B*, vol. 5, no. 5, pp. 442-453, Sep 2015, doi: 10.1016/j.apsb.2015.07.003.

- [12] P. Zhang, J. Forsgren, and M. Stromme, "Stabilisation of amorphous ibuprofen in Upsalite, a mesoporous magnesium carbonate, as an approach to increasing the aqueous solubility of poorly soluble drugs," *Int J Pharmaceut*, vol. 472, no. 1-2, pp. 185-191, Sep 10 2014, doi: 10.1016/j.ijpharm.2014.06.025.
- [13] H. D. Williams *et al.*, "Strategies to Address Low Drug Solubility in Discovery and Development," *Pharmacol Rev*, vol. 65, no. 1, pp. 315-499, Jan 2013, doi: 10.1124/pr.112.005660.
- [14] S. A. Damiati, L. G. Martini, N. W. Smith, M. J. Lawrence, and D. J. Barlow, "Application of machine learning in prediction of hydrotrope-enhanced solubilisation of indomethacin," *Int J Pharmaceut*, vol. 530, no. 1-2, pp. 99-106, Sep 15 2017, doi: 10.1016/j.ijpharm.2017.07.048.
- [15] S. J. Burns *et al.*, "Development and Validation of an in-Vitro Dissolution Method for a Floating Dosage Form with Biphasic Release Characteristics," *Int J Pharmaceut*, vol. 121, no. 1, pp. 37-44, Jul 6 1995, doi: Doi 10.1016/0378-5173(94)00431-4.
- [16] C. N. Magnan, A. Randall, and P. Baldi, "SOLpro: accurate sequence-based prediction of protein solubility," *Bioinformatics*, vol. 25, no. 17, pp. 2200-2207, Sep 1 2009, doi: 10.1093/bioinformatics/btp386.
- [17] P. Smialowski, G. Doose, P. Torkler, S. Kaufmann, and D. Frishman, "PROSO II - a new method for protein solubility prediction," *Febs J*, vol. 279, no. 12, pp. 2192-2200, Jun 2012, doi: 10.1111/j.1742-4658.2012.08603.x.
- [18] H. W. Z.J.Cao, and J.Y.Zhu, "Application of deep learning in gravitational wave data processing," *Journal of Henan Normal University (Natural Science)*, vol. 46, pp. 26-39, 2018.
- [19] M. F. Simoes *et al.*, "Artificial neural networks applied to quality-by-design: From formulation development to clinical outcome," *Eur J Pharm Biopharm*, vol. 152, pp. 282-295, Jul 2020, doi: 10.1016/j.ejpb.2020.05.012.
- [20] J. Janke, M. Castelli, and A. Popovic, "Analysis of the proficiency of fully connected neural networks in the process of classifying digital images. Benchmark of different classification algorithms on high-level image features from convolutional layers," *Expert Syst Appl*, vol. 135, pp. 12-38, Nov 30 2019, doi: 10.1016/j.eswa.2019.05.058.
- [21] W. Q. Wu, Y. J. Yin, X. G. Wang, and D. Xu, "Face Detection With Different Scales Based on Faster R-CNN," *Ieee T Cybernetics*, vol. 49, no. 11, pp. 4017-4028, Nov 2019, doi: 10.1109/Tcyb.2018.2859482.
- [22] T. Tuncer, S. Dogan, and F. Ertam, "Automatic voice based disease detection method using one dimensional local binary pattern feature extraction network," *Appl Acoust*, vol. 155, pp. 500-506, Dec 1 2019, doi: 10.1016/j.apacoust.2019.05.023.
- [23] S. Khurana, R. Rawi, K. Kunji, G. Y. Chuang, H. Bensmail, and R. Mall, "DeepSol: a deep learning framework for sequence-based protein solubility prediction,"

Bioinformatics, vol. 34, no. 15, pp. 2605-2613, Aug 1 2018, doi: 10.1093/bioinformatics/bty166.

- [24] M. Garcia Arenal, "The lead tablets from the Sacromonte (I) - Introduction," *Al-Qantara*, vol. 23, no. 2, pp. 342-345, 2002. [Online]. Available: <Go to ISI>://WOS:000180512000005.
- [25] Y. X. Chen, T. W. McCall, A. R. Baichwal, and M. C. Meyer, "The application of an artificial neural network and pharmacokinetic simulations in the design of controlled-release dosage forms," *Journal of Controlled Release*, vol. 59, no. 1, pp. 33-41, May 1 1999, doi: Doi 10.1016/S0168-3659(98)00171-0.
- [26] J. G. K. a. G. E. Peck, "Pharmaceutical granulation and tablet formulation using neural networks," *Pharmaceutical Development and Technology*, vol. 1, pp. 391-404, 1996.
- [27] U. Mandal *et al.*, "Optimization of metformin HCl 500 mg sustained release matrix tablets using artificial neural network (ANN) based on multilayer perceptrons (MLP) model," *Chem Pharm Bull*, vol. 56, no. 2, pp. 150-155, Feb 2008, doi: DOI 10.1248/cpb.56.150.
- [28] A. Thakur, A. P. Mishra, B. Panda, D. C. S. Rodriguez, I. Gaurav, and B. Majhi, "Application of Artificial Intelligence in Pharmaceutical and Biomedical Studies," *Curr Pharm Design*, vol. 26, no. 29, pp. 3569-3578, 2020, doi: 10.2174/1381612826666200515131245.
- [29] M. Bulbule and D. Gangrade, "Simultaneous Estimation of Norfloxacin and Tinidazole in Combined Tablet Dosage Form by Using Rp-Hplc Method," *Int J Pharm Sci Res*, vol. 8, no. 1, pp. 236-243, Jan 2017, doi: 10.13040/Ijpsr.0975-8232.8(1).236-43.
- [30] S. Khurana *et al.*, "AS03-adjuvanted H5N1 vaccine promotes antibody diversity and affinity maturation, NAI titers, cross-clade H5N1 neutralization, but not H1N1 cross-subtype neutralization," *Npj Vaccines*, vol. 3, Oct 1 2018, doi: ARTN 40, 10.1038/s41541-018-0076-2.
- [31] S. Khurana, S. Johnson, A. Karimaghloo, and M. H. Lee, "Effect of Sintering Process with Co₃O₄ on the Performance of LSCF-Based Cathodes for Solid Oxide Fuel Cells," *Int J Pr Eng Man-Gt*, vol. 5, no. 5, pp. 637-642, Oct 2018, doi: 10.1007/s40684-018-0066-x.
- [32] A. Manda, R. B. Walker, and S. M. M. Khamanga, "An Artificial Neural Network Approach to Predict the Effects of Formulation and Process Variables on Prednisone Release from a Multipartite System," *Pharmaceutics*, vol. 11, no. 3, Mar 2019, doi: ARTN 109, 10.3390/pharmaceutics11030109.
- [33] Y. Chen, T. W. McCall, A. R. Baichwal, and M. C. Meyer, "The application of an artificial neural network and pharmacokinetic simulations in the design of controlled-release dosage forms," *J Control Release*, vol. 59, no. 1, pp. 33-41, May 1 1999, doi: 10.1016/s0168-3659(98)00171-0.
- [34] J. G. Kesavan and G. E. Peck, "Pharmaceutical granulation and tablet formulation

- using neural networks," *Pharm Dev Technol*, vol. 1, no. 4, pp. 391-404, Dec 1996, doi: 10.3109/10837459609031434.
- [35] S. Horkovics-Kovats, D. L. Galata, P. Zlatos, B. Nagy, L. A. Meszaros, and Z. K. Nagy, "Raman-based real-time dissolution prediction using a deterministic permeation model," *Int J Pharmaceut*, vol. 617, Apr 5 2022, doi: ARTN 121624, 10.1016/j.ijpharm.2022.121624.
- [36] E. Benko *et al.*, "Predicting Drug Release Rate of Implantable Matrices and Better Understanding of the Underlying Mechanisms through Experimental Design and Artificial Neural Network-Based Modelling," *Pharmaceutics*, vol. 14, no. 2, Feb 2022, doi: ARTN 228, 10.3390/pharmaceutics14020228.
- [37] S. Obeid, M. Madzarevic, M. Krkobabic, and S. Ibric, "Predicting drug release from diazepam FDM printed tablets using deep learning approach: Influence of process parameters and tablet surface/volume ratio," *Int J Pharmaceut*, vol. 601, May 15 2021, doi: ARTN 120507, 10.1016/j.ijpharm.2021.120507.
- [38] M. A. Mrad, K. Csorba, D. L. Galata, Z. K. Nagy, and B. Nagy, "Spectroscopy-Based Prediction of In Vitro Dissolution Profile Using Artificial Neural Networks," *Lect Notes Artif Int*, vol. 12854, pp. 145-155, 2021, doi: 10.1007/978-3-030-87986-0_13.
- [39] D. L. Galata *et al.*, "Real-time release testing of dissolution based on surrogate models developed by machine learning algorithms using NIR spectra, compression force and particle size distribution as input data," *Int J Pharmaceut*, vol. 597, Mar 15 2021, doi: ARTN 120338, 10.1016/j.ijpharm.2021.120338.
- [40] O. K. Saracoglu, M. O. Uludag, E. D. Ozdemir, and I. T. Degim, "Development of controlled release dexketoprofen tablets and prediction of drug release using Artificial Neural Network (ANN) modelling," *Braz J Pharm Sci*, vol. 56, 2020, doi: ARTN e18540, 10.1590/s2175-97902019000418540.
- [41] A. M. Khan *et al.*, "Artificial Neural Network (ANN) Approach to Predict an Optimized pH-Dependent Mesalamine Matrix Tablet," *Drug Des Dev Ther*, vol. 14, pp. 2435-2448, 2020, doi: 10.2147/Dddt.S244016.
- [42] M. Pishnamazi, H. Y. Ismail, S. Shirazian, J. Iqbal, G. M. Walker, and M. N. Collins, "Application of lignin in controlled release: development of predictive model based on artificial neural network for API release," *Cellulose*, vol. 26, no. 10, pp. 6165-6178, Jul 2019, doi: 10.1007/s10570-019-02522-w.
- [43] B. Nagy *et al.*, "Application of artificial neural networks for Process Analytical Technology-based dissolution testing," *Int J Pharmaceut*, vol. 567, Aug 15 2019, doi: ARTN 11846410.1016/j.ijpharm.2019.118464.
- [44] M. Madzarevic *et al.*, "Optimization and Prediction of Ibuprofen Release from 3D DLP Printlets Using Artificial Neural Networks," *Pharmaceutics*, vol. 11, no. 10, Oct 2019, doi: ARTN 544, 10.3390/pharmaceutics11100544.
- [45] S. Lefnaoui, S. Rebouh, M. Bouhedda, and M. M. Yahoum, "ANN Optimization Using Ant Colony Algorithm for Predicting the Valsartan Sustained Release from

Polyelectrolyte Complexes Matrix Tablets," *2019 3rd International Conference on Applied Automation and Industrial Diagnostics (Icaaid 2019)*, 2019. [Online]. Available: <Go to ISI>://WOS:000527444900042.

- [46] D. L. Galata *et al.*, "Fast, Spectroscopy-Based Prediction of In Vitro Dissolution Profile of Extended Release Tablets Using Artificial Neural Networks," *Pharmaceutics*, vol. 11, no. 8, Aug 2019, doi: ARTN 400, 10.3390/pharmaceutics11080400.
- [47] S. Lefnaoui, S. Rebouh, M. Bouhedda, M. M. Yahoum, and S. Hanini, "Artificial Neural Network Modeling of Sustained Antihypertensive Drug Delivery using Polyelectrolyte Complex based on Carboxymethyl-kappa-carrageenan and Chitosan as Prospective Carriers," *Proceedings of the 2018 International Conference on Applied Smart Systems (Icass)*, 2018. [Online]. Available: <Go to ISI>://WOS:000468404600041.
- [48] T. Y. Kolisnyk, O. A. Ruban, N. Y. Fil, and S. A. Kutsenko, "Application of an Artificial Neural Network for Design of Sustained-Release Matrix Tablets Containing Vaccinium Myrtillus Leaf Powder Extract," *Asian J Pharm*, vol. 12, no. 2, pp. 137-145, Apr-Jun 2018. [Online]. Available: <Go to ISI>://WOS:000440212000009.
- [49] W. Li, C. H. Shi, Y. L. Sheng, P. Cui, Y. Q. Zhao, and X. R. Zhang, "In vitro dissolution and in vivo gamma scintigraphic evaluation of press-coated salbutamol sulfate tablets," *Acta Pharmaceut*, vol. 63, no. 4, pp. 545-551, Dec 2013, doi: 10.2478/acph-2013-0035.
- [50] Y. T. Yutaka Morita, Masanobu Yasui, Ryoji Termoz, Junko Ajioka, Kozo Takayama, "Evaluation of the disintegration time of rapidly disintegrating tablets via a novel method utilizing a CCD camera," *Chem Pharm Bull*, vol. 50, 9, 2002, doi: 10.1248/cpd.50.1181.
- [51] Y. S. Wenbin Yang, Kai Liu, Qingwei Jiang, "Dissolution tester monitored control system," *China Patent Appl.* 201721009959.7, 2017.
- [52] S. G. Kazarian and J. van der Weerd, "Simultaneous FTIR spectroscopic imaging and visible photography to monitor tablet dissolution and drug release," *Pharm Res*, vol. 25, no. 4, pp. 853-860, Apr 2008. [Online]. Available: <Go to ISI>://WOS:000254748900013.
- [53] Y. Morita, Y. Tsushima, M. Yasui, R. Termoz, J. Ajioka, and K. Takayama, "Evaluation of the disintegration time of rapidly disintegrating tablets via a novel method utilizing a CCD camera," *Chem Pharm Bull*, vol. 50, no. 9, pp. 1181-1186, Sep 2002. [Online]. Available: <Go to ISI>://WOS:000177740100007.
- [54] F. Tres *et al.*, "Real time Raman imaging to understand dissolution performance of amorphous solid dispersions," *Journal of Controlled Release*, vol. 188, pp. 53-60, Aug 28 2014. [Online]. Available: <Go to ISI>://WOS:000340699300006.
- [55] Q. R. Cao, H. G. Choi, D. C. Kim, and B. J. Lee, "Release behavior and photo-image of nifedipine tablet coated with high viscosity grade

- hydroxypropylmethylcellulose: effect of coating conditions," *Int J Pharmaceut*, vol. 274, no. 1-2, pp. 107-117, Apr 15 2004. [Online]. Available: <Go to ISI>://WOS:000220837400009.
- [56] K. Tieu, A. Salt, J. Wirges, and G. N. Grove, "Regulatory Considerations for the Classification of Video Monitoring in Dissolution Testing," *Aaps Pharmscitech*, vol. 15, no. 6, pp. 1611-1618, Dec 2014. [Online]. Available: <Go to ISI>://WOS:000346787700027.
- [57] P. Gao, J. W. Skoug, P. R. Nixon, T. R. Ju, N. L. Stemm, and K. C. Sung, "Swelling of hydroxypropyl methylcellulose matrix tablets. 2. Mechanistic study of the influence of formulation variables on matrix performance and drug release," *J Pharm Sci*, vol. 85, no. 7, pp. 732-40, Jul 1996, doi: 10.1021/js9504595.
- [58] P. Colombo, R. Bettini, P. Santi, A. DeAscentiis, and N. A. Peppas, "Analysis of the swelling and release mechanisms from drug delivery systems with emphasis on drug solubility and water transport," *Journal of Controlled Release*, vol. 39, no. 2-3, pp. 231-237, May 1996. [Online]. Available: <Go to ISI>://WOS:A1996UT43200013.
- [59] K. Tahara, K. Yamamoto, and T. Nishihata, "Application of model-independent and model analysis for the investigation of effect of drug solubility on its release rate from hydroxypropyl methylcellulose sustained release tablets," *Int J Pharmaceut*, vol. 133, no. 1-2, pp. 17-27, May 14 1996. [Online]. Available: <Go to ISI>://WOS:A1996UP35500003.
- [60] I. Caraballo, "Factors affecting drug release from hydroxypropyl methylcellulose matrix systems in the light of classical and percolation theories," *Expert Opin Drug Del*, vol. 7, no. 11, pp. 1291-1301, Nov 2010. [Online]. Available: <Go to ISI>://WOS:000283491700006.
- [61] R. Bettini, P. L. Catellani, P. Santi, G. Massimo, N. A. Peppas, and P. Colombo, "Translocation of drug particles in HPMC matrix gel layer: effect of drug solubility and influence on release rate," *Journal of Controlled Release*, vol. 70, no. 3, pp. 383-391, Feb 23 2001. [Online]. Available: <Go to ISI>://WOS:000167379700012.
- [62] L. B. Yang and R. Fassihi, "Examination of drug solubility, polymer types, hydrodynamics and loading dose on drug release behavior from a triple-layer asymmetric configuration delivery system," *Int J Pharmaceut*, vol. 155, no. 2, pp. 219-229, Sep 26 1997. [Online]. Available: <Go to ISI>://WOS:A1997XX71100007.
- [63] G. Falzon, D. W. Lamb, and D. Schneider, "The Dynamic Aerial Survey Algorithm Architecture and Its Potential Use in Airborne Fertilizer Applications," *Ieee J-Stars*, vol. 5, no. 6, pp. 1772-1779, Dec 2012, doi: 10.1109/Jstars.2011.2179020.
- [64] C. Leave, "Neural Networks - Algorithms, Applications and Programming Techniques - Freeman,Ja, Skapura,Dm," *J Oper Res Soc*, vol. 43, no. 11, pp. 1106-1106, Nov 1992, doi: Doi 10.2307/2584109.
- [65] W. S. Mcculloch and W. Pitts, "A Logical Calculus of the Ideas Immanent in

Nervous Activity (Reprinted from Bulletin of Mathematical Biophysics, Vol 5, Pg 115-133, 1943)," *B Math Biol*, vol. 52, no. 1-2, pp. 99-115, 1990, doi: Doi 10.1016/S0092-8240(05)80006-0.

- [66] L. Lin and J. S. Su, "Anomaly detection method for sensor network data streams based on sliding window sampling and optimized clustering," *Safety Sci*, vol. 118, pp. 70-75, Oct 2019, doi: 10.1016/j.ssci.2019.04.047.
- [67] A. Machireddy and S. S. Garani, "Guessing the Code: Learning Encoding Mappings Using the Back Propagation Algorithm," *Ieee Ijcn*, 2019. [Online]. Available: <Go to ISI>://WOS:000530893804007.
- [68] D. X. Gong, C. N. Wei, L. Wang, L. C. Feng, and L. D. Wang, "Adaptive Methods for Center Choosing of Radial Basis Function Interpolation: A Review," *Lect Notes Comput Sc*, vol. 6377, pp. 573-+, 2010. [Online]. Available: <Go to ISI>://WOS:000289612900073.
- [69] F. Montini *et al.*, "HSV encephalitis associated with off-label rituximab treatment of multiple sclerosis," *Neurol Sci*, vol. 43, no. 3, pp. 2095-2097, Mar 2022, doi: 10.1007/s10072-021-05803-0.
- [70] Y. Lu, X. D. Wang, H. D. Gu, and M. Gao, "Morphological transformation assisted switchable deep eutectic solvents combined with HPLC-DAD for the detection of six UV-filters in surface and bathing waters," *Microchem J*, vol. 169, Oct 2021, doi: ARTN 106626, 10.1016/j.microc.2021.106626.
- [71] C. Gendre *et al.*, "Comprehensive study of dynamic curing effect on tablet coating structure," *Eur J Pharm Biopharm*, vol. 81, no. 3, pp. 657-665, Aug 2012, doi: 10.1016/j.ejpb.2012.04.006.
- [72] Food and H. H. S. Drug Administration, "International Conference on Harmonisation; guidance on Q8(R1) Pharmaceutical Development; addition of annex; availability. Notice," *Fed Regist*, vol. 74, no. 109, pp. 27325-6, Jun 9 2009. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/19588572>.
- [73] F. F. Xie, S. Ji, and Z. N. Cheng, "In vitro dissolution similarity factor (f2) and in vivo bioequivalence criteria, how and when do they match? Using a BCS class II drug as a simulation example," *Eur J Pharm Sci*, vol. 66, pp. 163-172, Jan 23 2015, doi: 10.1016/j.ejps.2014.10.002.
- [74] R. E. Stevens, V. Gray, A. Dorantes, L. Gold, and L. Pham, "Scientific and Regulatory Standards for Assessing Product Performance Using the Similarity Factor, f2," *Aaps J*, vol. 17, no. 2, pp. 301-306, Mar 2015, doi: 10.1208/s12248-015-9723-y.
- [75] R. S. Gabbay, R. S. Kenett, R. Scaffaro, and A. Rubinstein, "Synchronizing the release rates of salicylate and indomethacin from degradable chitosan hydrogel and its optimization by definitive screening design," *Eur J Pharm Sci*, vol. 125, pp. 102-109, Dec 1 2018, doi: 10.1016/j.ejps.2018.09.020.
- [76] R. D. Oparin, A. Idrissi, M. V. Fedorov, and M. G. Kiselev, "Dynamic and Static Characteristics of Drug Dissolution in Supercritical CO2 by Infrared Spectroscopy:

- Measurements of Acetaminophen Solubility in a Wide Range of State Parameters," *J Chem Eng Data*, vol. 59, no. 11, pp. 3517-3523, Nov 2014, doi: 10.1021/je500456s.
- [77] FDA. (2010). *FDA Dissolution Guidance*. USA FDA: FDA. [Online] Available: <https://www.fda.gov/drugs>
- [78] T. Matsuo *et al.*, "Effects of yogurt as a deglutition aid on disintegration and dissolution of oral tablets," *J Texture Stud*, Feb 19 2022, doi: 10.1111/jtxs.12665.
- [79] A. Berardi, L. Bisharat, J. Quodbach, S. A. Rahim, D. R. Perinelli, and M. Cespi, "Advancing the understanding of the tablet disintegration phenomenon - An update on recent studies," *Int J Pharmaceut*, vol. 598, Apr 1 2021, doi: ARTN 120390, 10.1016/j.ijpharm.2021.120390.
- [80] Z. Golshani, S. M. A. Hosseini, M. Shahidizandi, and M. J. Bahrami, "Increase corrosion resistance of mild steel in sulfuric acid and hydrochloric acid solutions by metoclopramide tablet," *Mater Corros*, vol. 70, no. 10, pp. 1862-1871, Oct 2019, doi: 10.1002/maco.201910896.
- [81] M. R. Siddiqui, Z. A. AlOthman, and N. Rahman, "Analytical techniques in pharmaceutical analysis: A review," *Arab J Chem*, vol. 10, pp. S1409-S1421, Feb 2017, doi: 10.1016/j.arabjc.2013.04.016.
- [82] L. Soovali, E. I. Room, A. Kutt, I. Kaljurand, and I. Leito, "Uncertainty sources in UV-Vis spectrophotometric measurement," *Accredit Qual Assur*, vol. 11, no. 5, pp. 246-255, Jun 2006, doi: 10.1007/s00769-006-0124-x.
- [83] V. A. Gray, "Celebration of Twenty Years from Dissolution Tester Equipment Manufacturers and Other Providers of Dissolution Related Services," *Dissolut Technol*, vol. 21, no. 3, pp. 29-36, Aug 2014, doi: Doi 10.14227/Dt210314p29.
- [84] Haoyu Wang, David Siu-Yeung Cho, Zhongyan Wang, and N. Qin, "Methods, Equipments and Systems for Predicting Dynamic Solubility of Active Pharmaceutical Ingredients by Artificial Intelligence," *China Patent Appl*. 201910991714.6, 2019.
- [85] E. Eross, "The efficacy of gliacin, a specialized boswellia serrata extract, on indomethacin responsive headache syndromes," *Cephalalgia*, vol. 31, no. 1, pp. 47-47, Jul 2011. [Online]. Available: <Go to ISI>://WOS:000572463400107.
- [86] Japanese FDA. (2010). Japanese Orange Book. Japanese FDA. Japanese Orange Book Website, Japanese Orange Book. [Online] Available: <http://www.jp-orangebook.gr.jp/>
- [87] A. A. W. Noyes, W.R., "The rate of solution of solid substances in their own solutions," *JJournal of the American Chemical Society*, pp. 930-934, 1897.
- [88] A. W. Hixson and J. H. Crowell, "Dependence of reaction velocity upon surface and agitation II - Experimental procedure in study of surface," *Ind Eng Chem*, vol. 23, pp. 1002-1009, 1931, doi: DOI 10.1021/ie50261a009.

- [89] J. Z. Wang and D. R. Flanagan, "General solution for diffusion controlled dissolution of spherical particles. 1. Theory," *J Pharm Sci*, vol. 88, no. 7, pp. 731-738, Jul 1999, doi: DOI 10.1021/js980236p.
- [90] D. Brooke, "Dissolution Profile of Log-Normal Powders - Exact Expression," *J Pharm Sci*, vol. 62, no. 5, pp. 795-798, 1973, doi: DOI 10.1002/jps.2600620519.
- [91] Haoyu Wang, Chiew Foong Kwong, Zhiyuan Chen, David Siu-Yeung Cho, Boon Giin Lee, and L. Huang, "A system and method for artificial intelligence to predict the dissolution curve of formulation," China Patent Appl. 202111663521.1, 2021.
- [92] Haoyu Wang, Chiew-Foong Kwong, Zhiyuan Chen, David Siu-Yeung Cho, Boon Giin Lee, and L. Huang, "A system and method for artificial intelligence to screen abnormal data, Chinese Patent," China Patent Appl. 20211657874.0, 2021.
- [93] H. Wang, C. F. Kwong, Q. Liu, Z. Liu, and Z. Chen, "A Novel Artificial Intelligence System in Formulation Dissolution Prediction," *Comput Intell Neurosci*, vol. 2022, p. 8640115, 2022, doi: 10.1155/2022/8640115.
- [94] Q. Q. Qin, H. C. Qin, K. Li, R. L. Tan, X. Y. Liu, and L. C. Li, "The adsorption characteristics and degradation mechanism of tinidazole on an anatase TiO₂ surface: a DFT study," *Rsc Adv*, vol. 10, no. 4, pp. 2104-2112, Jan 12 2020, doi: 10.1039/c9ra06665a.
- [95] Japanese FDA. (2010). Japanese Pharmacopoeia. Japanese FDA. Japanese Pharmacopoeia Website, Japanese Pharmacopoeia. [Online] Available: <http://www.fpmaj-saihyoka.com/>
- [96] EMA (2010). European Pharmacopoeia. EMA. EU EMA, EMA. [Online] Available: <https://store.edqm.eu/index.html>
- [97] FDA, B. (2013). British Pharmacopoeia. B. FDA. UK. [Online] Available: <https://www.pharmacopoeia.com/>
- [98] P. J. S. Raymond C Rowe, Marian E Quinn, *Handbook of Pharmaceutical Excipients*. Royal Pharmaceutical Society of Great Britain, 2009, p. 917.
- [99] S. Sarraf, E. Marzbanrad, and H. Mobedi, "Mathematical Modeling for Predicting Betamethasone Profile and Burst Release From In Situ Forming Systems Based On PLGA Application of Artificial Neural Networks," *Can Con El Comp En*, 2014. [Online]. Available: <Go to ISI>://WOS:000361017900110.
- [100] Haoyu Wang, David Siu-Yeung Cho, Boon Giin Lee, and Z. Chen, "A system, method and equipment for monitoring and analyzing dissolution behavior using artificial intelligence image recognition technology," China Patent Appl. 202110842211.X, 2021.
- [101] H. Y. Wang *et al.*, "Monitoring and Analysis Solid Formulation Dissolution Phenomenon with Image Recognition Technologies," *Computational Intelligence and Neuroscience*, vol. 2022, Sep 14 2022. [Online]. Available: <Go to ISI>://WOS:000864450000011.

- [102] T. Katayama *et al.*, "In Vivo Drug Dissolution in Human Oral Cavity from Orally Disintegrating Tablet and Comparability with in Vitro Testing," *Chem Pharm Bull*, vol. 66, no. 10, pp. 999-1005, Oct 2018, doi: DOI 10.1248/cpb.c18-00492.