**元 智 大 學**

資 訊 工 程 學 系

碩 士 論 文

**Going Deeper with Convolutional Neural Network for Stock Market Prediction**

研 究 生：郭祿丁

指導教授：歐昱言 博士

中華民國 107 年 7 月

**Going Deeper with Convolutional Neural Network for Stock Market Prediction**

研 究 生 ：郭祿丁 Student : Rosdyana Mangir Irawan Kusuma

指 導 教 授 ：歐 昱 言 博士 Advisor : Dr. Yu-Yen Ou

元 智 大 學

資 訊 工 程 學 系

碩 士 論 文

A Thesis

Submitted to Department of Computer Science and Engineering

Yuan Ze University

in Partial Fulfillment of the Requirements

for the Degree of Master of Science

in

Computer Science and Engineering

July 2018

Chungli, Taiwan, Republic of China.

中華民國 107 年 7 月

Going Deeper with Convolutional Neural Network for Stock Market Prediction

**Student: Rosdyana Mangir Irawan Kusuma Advisor: Dr. Yu-Yen Ou**

Submitted to Department of Computer Science and Engineering

College of Informatics

Yuan Ze University

# ABSTRACT

This thesis explores predictability in the stock market using Deep Convolutional Network and candlestick charts. The outcome is to design a decision support framework that can be used by traders to provide suggested indications of future stock price direction. Residual network method will be our machine learning algorithm to build deeper convolutional network. From stock market historical data, we converted to candlestick chart for the model analyzing the pattern. Using Taiwan and Indonesian stock market historical time series data we can achieve xx % for 5 days sliding windows period, xx % for 10 days sliding windows period and xx % for 20 days of binary classification accuracy.

This thesis also addresses problems specific to learning with stock market historical time series data, model performance due to sliding windows period days.

Keywords: *Stock Market Prediction, Neural Network, Residual Network*

# Acknowledgements

I would like to dedicate this thesis to Allah Subhanahu wa ta’ala, for your help through all the difficulties. And a special thanks to my family, my wife, my lab mates, all the Yuan Ze University teachers and friends that I have ever met during my study in Taiwan.

I would like to express my gratitude for the supervision of my advisor, Dr. Yu-Yen Ou who suggested an interesting research topic to me. And gave me clear direction to made this paper possible. You have been a tremendous teacher for me during my study especially for your support, guidance, corrections, encouragement and advices. I would also like to thank my thesis committee for the insightful comments and advice.

Finally, I would like to acknowledge the generous financial support of Yuan Ze University’s Department of Computer Science and Engineering. Which has illuminated my long-lasting dream of obtaining a Master Degree in Computer Science and Engineering.

# Table of Contents

[ABSTRACT iii](#_Toc509943964)

[Acknowledgements iv](#_Toc509943965)

[Table of Contents v](#_Toc509943966)

[Table of Figures vi](#_Toc509943967)

[Table of Tables vii](#_Toc509943968)

[Chapter 1 Introduction 1](#_Toc509943969)

[1.1. Background 1](#_Toc509943970)

[1.2. Related work 2](#_Toc509943971)

[Chapter 2 Data Collection 3](#_Toc509943972)

[2.1. Data Collection using Yahoo! Finance 3](#_Toc509943973)

[2.2. Feature Investigation 4](#_Toc509943974)

[2.2.1. Opening Price 4](#_Toc509943975)

[2.2.2. Closing Price 4](#_Toc509943976)

[2.2.3. Highest Price 5](#_Toc509943977)

[2.2.4. Lowest Price 5](#_Toc509943978)

[2.2.5. Volume 6](#_Toc509943979)

[2.3. Data Preprocessing 6](#_Toc509943980)

[Chapter 3 Methodology 8](#_Toc509943981)

[3.1. Chart Encoding 9](#_Toc509943982)

[3.2. Binary Classification 9](#_Toc509943983)

[3.3. Convolutional Neural Network 10](#_Toc509943984)

[3.3.1. CNN Architecture Overview 10](#_Toc509943985)

[3.3.2. Residual Network 11](#_Toc509943986)

[3.4. Performance Evaluation 13](#_Toc509943987)

[Chapter 4 Experimental Results 14](#_Toc509943988)

[4.1. Analysis Data 14](#_Toc509943989)

[4.2. Classification for Each Stock Market 15](#_Toc509943990)

[Chapter 5 Conclusion and Future Works 17](#_Toc509943991)

[References 18](#_Toc509943992)

# Table of Figures

**Figure 1**- Yahoo! Finance website, provide an API to download historical time series data 3

**Figure 2** - General way to visualizing the candlestick chart 7

**Figure 3** – Our methodology design 8

**Figure 4** - Our proposed candlestick chart have volume indicator, besides the open, high, low, close 9

**Figure 5** - Logic statement of our binary classification 10

**Figure 6** - A regular 3-layer Neural Network 11

**Figure 7** – ResNet Architecture shows many layers for different configuration 11

**Figure 8** - Residual learning: a building block 12

**Figure 9** – EWT from Taiwan stock market, candlestick chart for training and testing data 14

**Figure 10** – FTW from Taiwan stock market, candlestick chart for training and testing data 14

**Figure 11** – EIDO a stock market from Indonesia, candlestick chart for training and testing data 15

**Figure 12** – IDX a stock market from Indonesia, candlestick chart for training and testing data 15

# Table of Tables

**Table 1 -** The period time of our dataset, separated between the training and testing 4

**Table 2** – Number of dataset following their period of trading days 6

**Table 3** - Residual Network proved that this network have the most minimum error-rate 12

**Table 4** – Result from ResNet 50 for all stock market period days 16

**Table 5** – Result from ResNet 101 for all stock market period days 16

**Table 6** – Result from ResNet 152 for all stock market period days 16

# Chapter 1 Introduction

## 1.1. Background

Stock market is something that cannot be separated from modern human life. Investment in stock market is a natural thing done by people around the world. They set aside their income to try their luck by investing in stock market to generate more profit. Not a few who got benefit more not less also get big losses. This is because the movement of the daily stock market price is not easy to predict. This is also one of our reasons for doing research and try to make predictions into the field of stock market. Besides, research on stock market is still classified as a difficult thing to do considering the price changes in the stock market can happen suddenly. With the current technological advances, machine learning is a breakthrough in aspects of human life today. In this case we use machine learning to make predictions on the stock market. Convolutional neural network architecture will be our machine learning model for predicting in stock market.

In machine learning, we will find some kind of data input, such as text sequence, image, audio, video, from 2d to 3d. Taken as an example for the use of images as inputs from machine learning, not only as inputs to predict or classify an animal, items or the other thing, but also as an input to predict a condition, we take the example of Google Deep mind in their research results in alpha go, they are successfully get a lot of attention in the research field. By using the image as their input, where the image represents a Go game board, which later this image is used to predict the next step of the opponent in the Go game.

In this thesis we propose methods using historical time series data from Taiwan stock market and also Indonesia stock market. Then, we convert the historical time series data into a sequence from the candlestick chart. For each candlestick chart will be represented the following trading days’ data with combination of open price, close price, high price, low price and also volume. Which will later become the input for our CNN architecture model, after more the result of predicting is predict the price will be going up or going down for the next day.

## 1.2. Related work

Basically, there are many researchers who already done doing the research on stock market predictions. E.Schöneburg conducted a study using data from a randomly selected German stock market, then using the back-propagation method for their machine learning architecture. State of the arts, in stock market data we know some kind of element like open price, close price, high price, low price and also volume. In addition to using historical time series data from the stock market, some researchers in this field of stock market predictions began to penetrate the method of sentiment analysis to predict and analyze movements in the stock market.

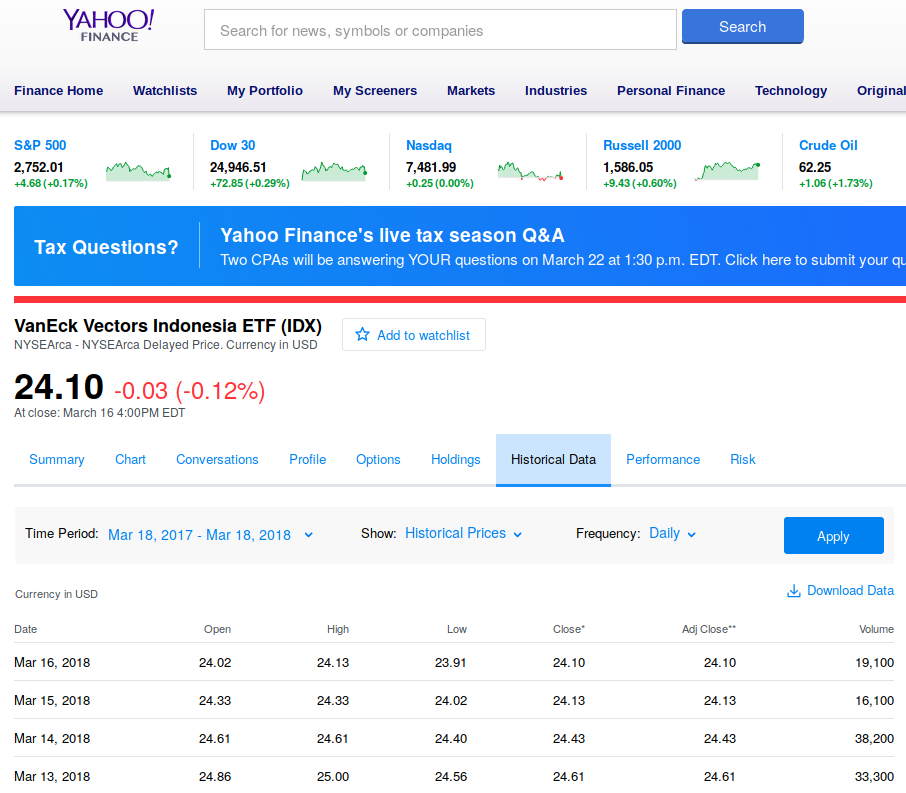
Johan Bollen used their sentiment analysis method by taking data from one of the famous microblogging site Twitter to predict the Dow Jones Industrial Average (DJIA) movements. There are more studies on stock market predictions, they use the input data not only by using elements of historical time series data, but also processing the data into other different forms. Anastasia Borovykh tried to use the deep convolutional wavenet architecture method to perform analysis and prediction using data from S & P500 and CBOE.

We also found some related work using candlestick charts in their research. Guosheng Hu et al in his research using candlestick chart to build a decision-making system in stock market investment. They use the convolutional encoder to learn the patterns contained in the candlestick chart. While Tsai using candlestick chart in his research to be combined with seven different wavelet-based texture to analyze candlestick chart. Next, Prado using candlestick chart to learn the pattern contained in Brazilian stock market by using sixteen candlestick pattern.

# Chapter 2 Data Collection

## 2.1. Data Collection using Yahoo! Finance

Data collection is the first step in this research. In the data collection we will use some stock markets such as FTW, EWT from Taiwan stock market, and EIDO and IDX from Indonesian stock market. We do not close the possibility to try global stock market like Amazon, Microsoft or Facebook. In this data collection we are using application program interface (API) service from Yahoo! Finance to get historical time series data for each stock market that we are using for the dataset (Figure 1). From the time period that we have been set in the following table 1 , we certainly get some trading days data, starting from Monday until Friday is the trading days period.



**Figure 1**- Yahoo! Finance website, provide an API to download historical time series data

**Table 1 -** The period time of our dataset, separated between the training and testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SYMBOL | TRAINING DATA | | TESTING DATA | |
| FTW | 2012/02/28 | 2016/12/30 | 2017/01/03 | 2018/03/16 |
| ETW | 2000/06/23 | 2016/12/31 | 2017/01/03 | 2018/03/16 |
| IDX | 2009/01/20 | 2016/12/30 | 2017/01/03 | 2018/03/16 |
| EIDO | 2010/05/07 | 2016/12/30 | 2017/01/03 | 2018/03/16 |

Segregation of data based on predetermined time for data training and data testing is important, where some studies make mistakes by scrambling data, this is certainly fatal because the data we use is time-series. In contrast to the iris dataset (Anderson, 1936; Fisher, 1936) where such data can be randomized.

## 2.2. Feature Investigation

The data we get from the data collection is in the form of time-series data. This time series of data contains several elements in daily stock market activity. Some of these elements are open price, close price, high price, low price and volume. The element can be said as feature set, where we will describe one by one feature set.

### 2.2.1. Opening Price

Opening price is the first price in daily activity of the stock market that is noted when the stock market is open in the specified time period. In this case for example Taiwan stock market and Indonesian stock market will open at 09:00.

The price of the first trade for any listed stock is its daily opening price. There are several day-trading strategies based on the opening of a market. Gap Fade and Fill. "Traders attempt to profit from the price correction that usually takes place subsequent to a sizable price gap at the opening. Another popular strategy is to fade a stock at the open that is showing strong pre-market indication contrary to the rest of the market, or to similar stocks in a common sector or index.

### 2.2.2. Closing Price

The closing price is the final price at which a security is traded on a given trading day. The closing price represent the most up-to-date valuation of a security until trading commences on the next trading day. Taiwan stock market close precisely at 3:00 p.m. while Indonesian stock market will be close precisely at 4:00 p.m.

Closing prices do not reflect corporate actions, which may skew returns significantly. For example, on June 9, 2014, Apple Inc. (NASDAQ: AAPL) issued a seven-for-one stock split. Therefore, Apple's shares outstanding was increased by a multiple of seven, while its closing share price was divided by seven. On June 6, 2014, prior to Apple's stock split, it had a closing price of $ 645.57 per share. After Apple's seven-for-one stock split, the stock had a closing price of $ 93.70 per share on June 9, 2014. Since the closing price does not include adjustments for corporate actions, the calculation of Apple's returns based on closing prices would have indicated a return of -85.49%, or ($ 93.70 - $ 645.56) / $ 645.57, in just one trading day.

### 2.2.3. Highest Price

High price or today’s high is the highest price at which a stock traded during the course of the day. Today’s high is typically higher than the closing or opening price. More often than not this is higher than the closing price.

When you look at a stock quote, you can find today's high by looking at the second number listed next to "Range." One way that day traders and technical analysts use today's high, along with today's low, is to help them identify gaps or sudden jumps up or down in a stock's price with no trading in between those two prices. For example, if today's low is $25 and the previous day's high is $20, there is gap. The identification of a gap, along with other market signals such as changes in trading volume and overall bullish or bearish sentiment, helps market analysts generate buy and sell signals for particular stocks.

### 2.2.4. Lowest Price

Today’s low or low price is the lowest price at which a stock trades over the course of a trading day. Today’s low is typically lower than the opening or closing price.

When you look at a stock quote, you can find today's low by looking at the first number listed next to "Range." Today's low and today's high are important to day traders and technical analysts, who seek to earn profits from a security's short-term price movements and identify and track trends. One way that day traders use today's low along with today's high is to identify gaps, or sudden jumps up or down in a stock's price with no trading in between. Gaps are used in technical analysis to identify directional movement, average true range/price volatility, candlestick patterns and more. Traders then analyze these patterns to determine profitable entry and exit points.

### 2.2.5. Volume

Volume is the number of shares or contracts traded in a security or an entire market during a given period of time.

For every buyer, there is a seller, and each transaction contributes to the count of total volume. That is, when buyers and sellers agree to make a transaction at a certain price, it is considered one transaction. If only five transactions occur in a day, the volume for the day is five.

Volume is an important indicator in technical analysis as it is used to measure the relative worth of a market move. If the markets make a strong price movement, then the strength of that movement depends on the volume for that period. The higher the volume during the price move, the more significant the move.

## 2.3. Data Preprocessing

We are processing our time series data using library matplotlib in python programming to convert from the historical data that we have prepared into a candlestick chart. We divide the time period which used to create candlestick chart based on 5 days trading data, 10 days trading data and also 20 days trading data.

The amount of data that can be different because we will only make a candlestick chart that qualifies based on the time period that we have set in the following table 2.

**Table 2** – Number of dataset following their period of trading days

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 5 PERIOD | | 10 PERIOD | | 20 PERIOD | |
| SYMBOL | TRAINING | TESTING | TRAINING | TESTING | TRAINING | TESTING |
| FTW | 1215 | 298 | 1210 | 293 | 1200 | 283 |
| EWT | 4152 | 298 | 4147 | 293 | 4137 | 283 |
| IDX | 1998 | 298 | 1993 | 293 | 1983 | 283 |
| EIDO | 1671 | 298 | 1666 | 293 | 1656 | 283 |

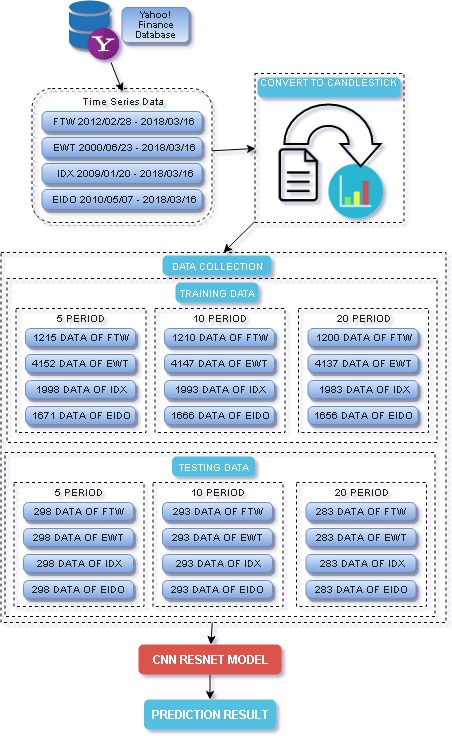
In our candlestick chart, we are not only uses historical time series data such as open price, close price, low price and high price like the most candlestick charts in general way (figure.2), but we also add new indicator in our candlestick chart, which is volume indicators to enrich our candlestick chart information.

In the following figures we illustrate the form of the candlestick chart we propose for this study. This we do because to add information in the candlestick chart. The more information contained in the data we prepare, the more likely it is for machine learning algorithms to learn the pattern hidden in the candlestick chart.



**Figure 2** - General way to visualizing the candlestick chart

# Chapter 3 Methodology

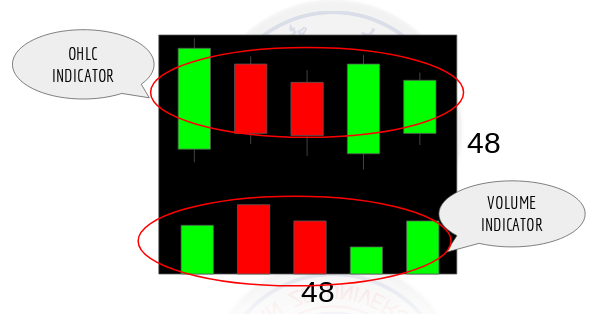


**Figure 3** – Our methodology design

Those figure below (Figure 3), shows about the design architecture of our proposed stock market prediction. The process is started with collecting data using Yahoo! Finance API and then using sliding window to generate the period data before using computer graphic technique to create candlestick chart images.

## 3.1. Chart Encoding

Our raw data for each stock market is consist of five channels from time series (lowest price, highest price, starting price, final price and volume) within the time frame according to the following table.2. We use computer graphics techniques with the help of a python library called matplotlib to convert this time series data into a candlestick image 48x48 dimension with RGB like in figure 4.



**Figure 4** - Our proposed candlestick chart have volume indicator, besides the open, high, low, close

In the following figure.1 we are using black color as our candlestick chart background, and then for each candlestick chart we are using green color for indicator if its closing price is higher than open price, and red color if its closing price is lower than opening price.

## 3.2. Binary Classification

Our goal here is to perform binary classification of stock market movements using candlestick chart. Where we will use the value of "1" as an indicator that there is an increase or closing price will rise in the future with t = n + 1, while "0" is the opposite, if the closing price will decrease in the future. The diagram for making this label is available in figure 5.



**Figure 5** - Logic statement of our binary classification

## 3.3. Convolutional Neural Network

Convolutional Neural Networks are very similar to ordinary Neural Networks from the previous chapter: they are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity.

The whole network still expresses a single differentiable score function: from the raw image pixels on one end to class scores at the other. And they still have a loss function (e.g. SVM/Softmax) on the last (fully-connected) layer and all the tips/tricks we developed for learning regular Neural Networks still apply.

### 3.3.1. CNN Architecture Overview

In figure 6, Neural Networks receive an input (a single vector), and transform it through a series of hidden layers. Each hidden layer is made up of a set of neurons, where each neuron is fully connected to all neurons in the previous layer, and where neurons in a single layer function completely independently and do not share any connections. The last fully-connected layer is called the “output layer” and in classification settings it represents the class scores.



**Figure 6** - A regular 3-layer Neural Network

### 3.3.2. Residual Network

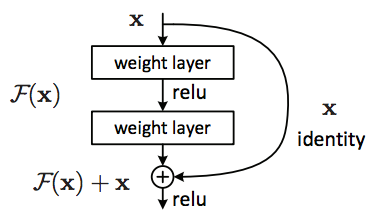
Developed by Kaiming He et al. was the winner of ILSVRC 2015. It features special skip connections and a heavy use of batch normalization. The architecture is also missing fully connected layers at the end of the network.



**Figure 7** – ResNet Architecture shows many layers for different configuration

As per what we have seen so far, increasing the depth should increase the accuracy of the network, as long as overfitting is taken care of. But the problem with increased depth is that the signal required to change the weights, which arises from the end of the network by comparing ground-truth and prediction becomes very small at the earlier layers, because of increased depth. It essentially means that earlier layers are almost negligible learned. This is called vanishing gradient.

The second problem with training the deeper networks is, performing the optimization on huge parameter space and therefore naively adding the layers leading to higher training error. Residual networks allow training of such deep networks by constructing the network through modules called residual models as shown in the figure 8. This is called degradation problem.



**Figure 8** - Residual learning: a building block

By comparing with other CNNs architecture, residual network have been prove with the most minimum error rate according to table 3.

**Table 3** - Residual Network proved that this network have the most minimum error-rate

|  |  |  |  |
| --- | --- | --- | --- |
| Year | CNN | Developed by | Top-5-error-rate |
| 2012 | AlexNet | Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever | 15.3 % |
| 2013 | ZFNet | Matthew Zeller, Rob Fergus | 14.8 % |
| 2014 | GoogLeNet | Google | 6.67 % |
| 2014 | VGG Net | Simonyan, Zisserman | 7.3 % |
| 2015 | ResNet | Kaiming He | 3.57 % |

## 3.4. Performance Evaluation

There are some statistical measure of the performance evaluation to evaluate the result of all the classifiers (binary classification test) by measuring the sensitivity (true positive rate or recall), specificity (true negative rate), accuracy and Matthew's correlation coefficient (MCC) for each class of efflux protein families. In general, TP is true positive or correctly identified, FP is false positive or incorrectly identified, TN is true negative or correctly rejected and FN is false negative or incorrectly rejected. Formulated as follows:

Sensitivity is called true positive rate or recall measures the performance of positives data are correctly identified.

Otherwise, to measure the proposition of negative rate the specificity formula is used during the prediction result and performance all classifiers.

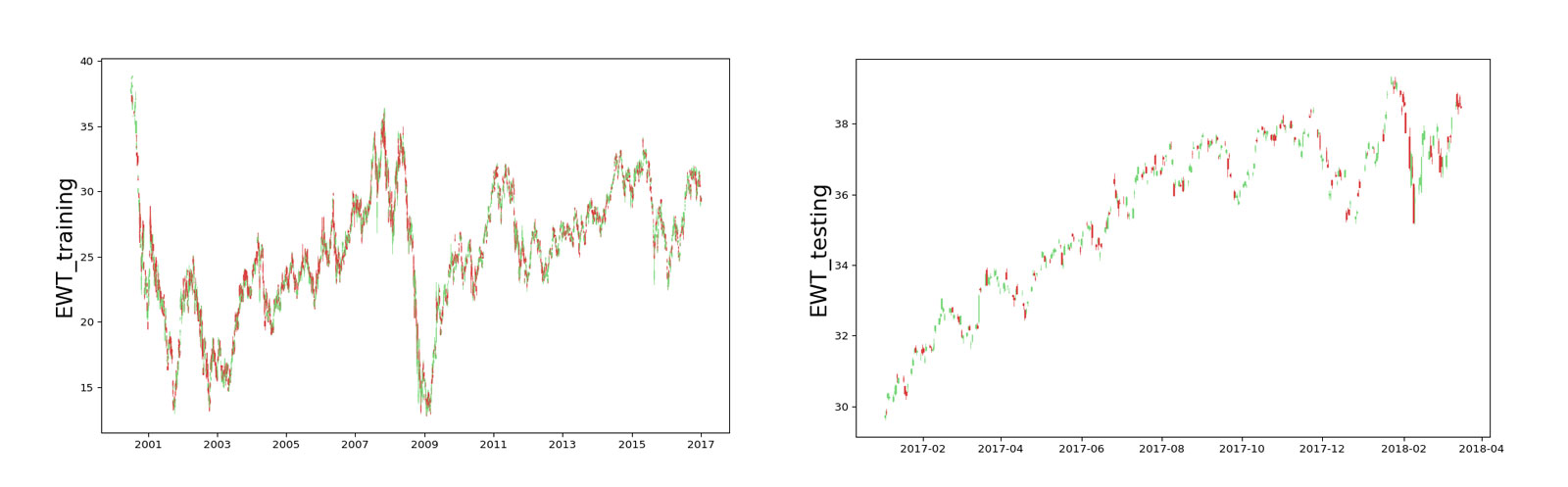
The accuracy formula measures the quality all classifiers with based on the true value or maximum predicted values compared with measurement results.

Then, Matthews’s correlation coefficient or MCC is used to predict binary (two class) classifications and focus on the quality of predicted binary. During the prediction results MCC returns a value between -1 and +1. If the correlation value closer to +1 indicates perfect prediction, and otherwise if the correlation value closer to -1 indicates total disagreement between prediction and observation.

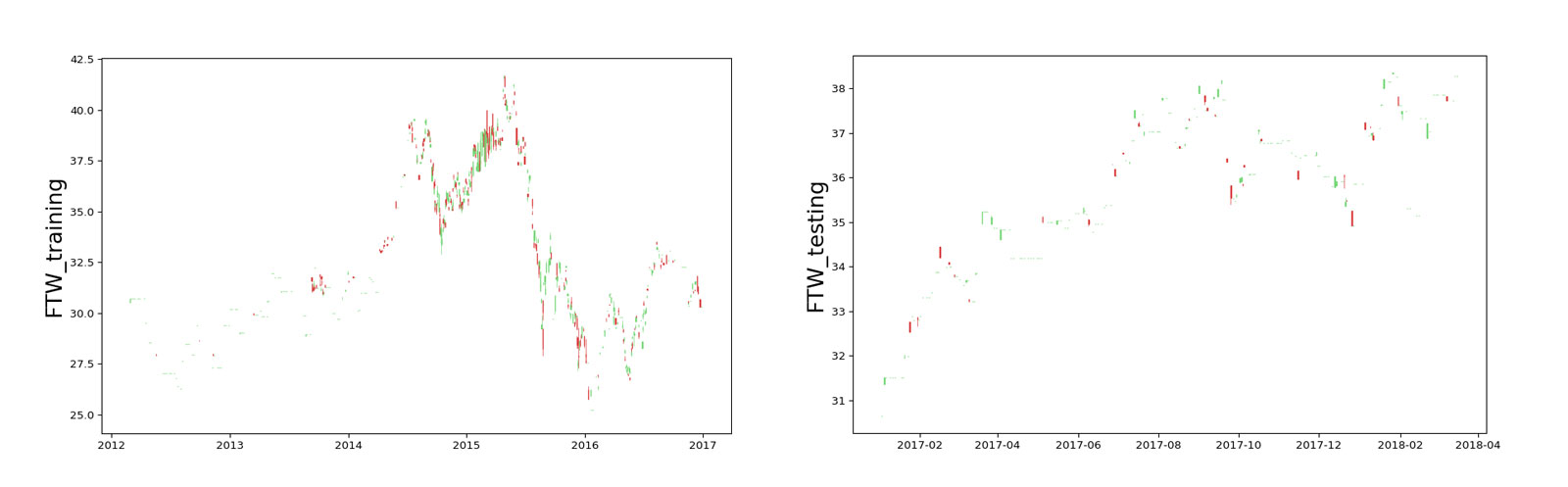
# Chapter 4 Experimental Results

## 4.1. Analysis Data

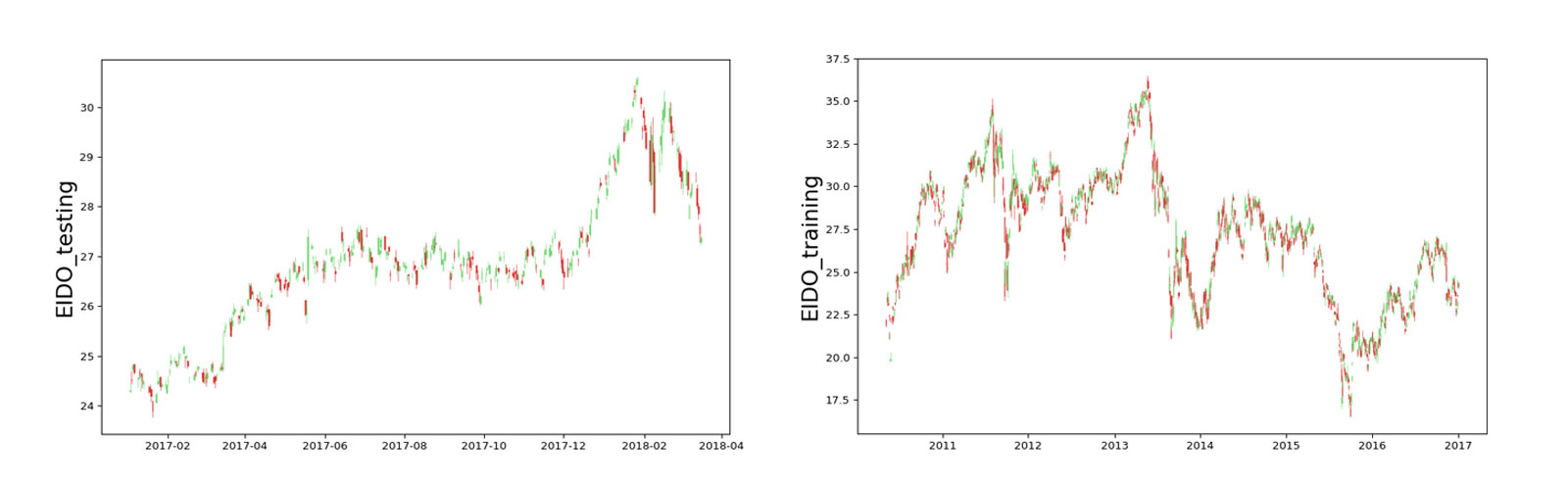
Since we have a different stock market time series data sources, we would like to visualizing the fully candlestick chart with all of the time span duration. This a whole candlestick chart should help us to understand our time series data behavior.



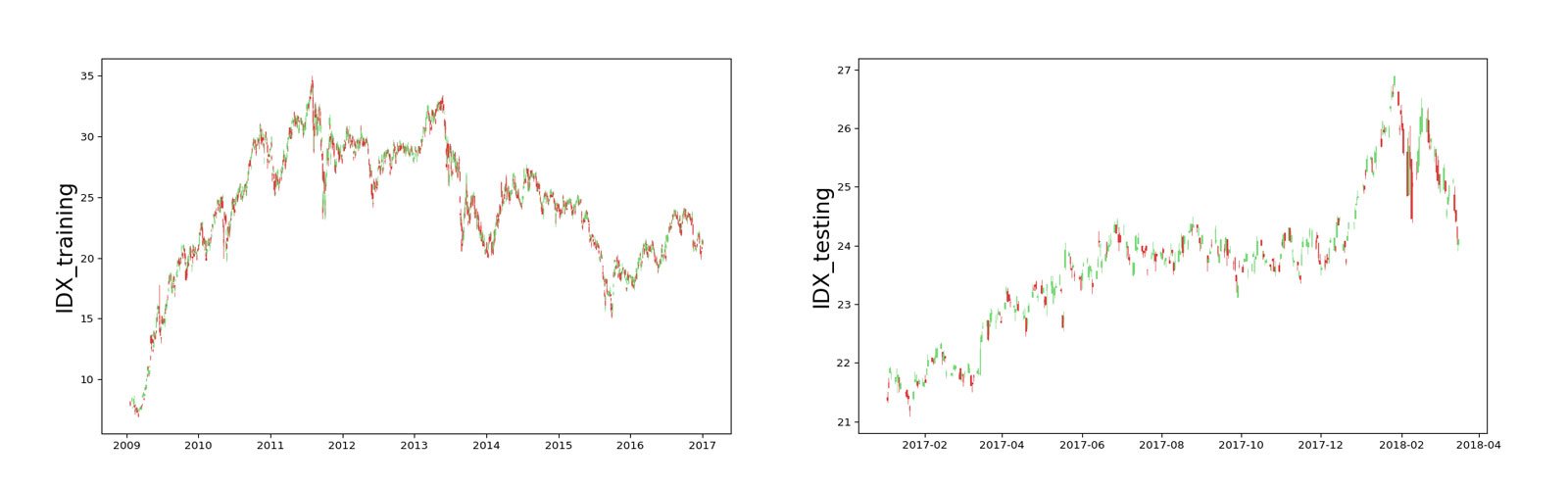
**Figure 9** – EWT from Taiwan stock market, candlestick chart for training and testing data



**Figure 10** – FTW from Taiwan stock market, candlestick chart for training and testing data



**Figure 11** – EIDO a stock market from Indonesia, candlestick chart for training and testing data



**Figure 12** – IDX a stock market from Indonesia, candlestick chart for training and testing data

In the data which we visualized, it can be seen that there are some parts that may affect the quality of the prediction. Recalling that a successful prediction with reliable results is one of the major factors of the dataset.

## 4.2. Classification for Each Stock Market

In this study we try to make stock market predictions by using binary classification. Where the value 1 on the label means there is a price increase on the next day, while the value 0 is the reverse of it.

We try to apply this binary classification to each stock market data we have prepared. There are four sources of stock market data we use. We also divide the retrieval period based on the duration of the five days, ten days and also twenty days of trading days to create a sequence of sliding windows that will be converted to candlestick chart. Here is a table of results from binary classification predictions that we do with CNN architecture more precisely residual network.

**Table 4** – Result from ResNet 50 for all stock market period days

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | 5 PERIOD | | | | 10 PERIOD | | | | 20 PERIOD | | | |
|  | EWT | FTW | EIDO | | IDX | EWT | FTW | | EIDO | IDX | EWT | | FTW | EIDO | IDX |
| Acc | 92.90% | 84.80% | 78.80% | | 82.00% | 90.80% | 87.40% | | 81.90% | 84.00% | 82.20% | | 86.20% | 81.90% | 81.50% |
| Spec | 90.60% | 78.30% | 73.50% | | 77.70% | 86.50% | 88.50% | | 77.80% | 80.20% | 80.50% | | 86.80% | 81.90% | 78.60% |
| Sens | 93.90% | 88.70% | 81.60% | | 84.10% | 93.40% | 86.50% | | 85.00% | 86.60% | 83.40% | | 85.70% | 81.90% | 84.30% |
| MCC | 83.50% | 67.40% | 54.20% | | 60.50% | 80.30% | 74.40% | | 63.00% | 66.90% | 63.60% | | 72.50% | 63.70% | 63.10% |

**Table 5** – Result from ResNet 101 for all stock market period days

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | 5 PERIOD | | | | 10 PERIOD | | | | 20 PERIOD | | | |
|  | EWT | FTW | EIDO | | IDX | EWT | FTW | | EIDO | IDX | EWT | | FTW | EIDO | IDX |
| Acc | 92.90% | 84.80% | 78.80% | | 82.00% | 90.80% | 87.40% | | 81.90% | 84.00% | 82.20% | | 86.20% | 81.90% | 81.50% |
| Spec | 90.60% | 78.30% | 73.50% | | 77.70% | 86.50% | 88.50% | | 77.80% | 80.20% | 80.50% | | 86.80% | 81.90% | 78.60% |
| Sens | 93.90% | 88.70% | 81.60% | | 84.10% | 93.40% | 86.50% | | 85.00% | 86.60% | 83.40% | | 85.70% | 81.90% | 84.30% |
| MCC | 83.50% | 67.40% | 54.20% | | 60.50% | 80.30% | 74.40% | | 63.00% | 66.90% | 63.60% | | 72.50% | 63.70% | 63.10% |

**Table 6** – Result from ResNet 152 for all stock market period days

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | 5 PERIOD | | | | 10 PERIOD | | | | 20 PERIOD | | | |
|  | EWT | FTW | EIDO | | IDX | EWT | FTW | | EIDO | IDX | EWT | | FTW | EIDO | IDX |
| Acc | 92.90% | 84.80% | 78.80% | | 82.00% | 90.80% | 87.40% | | 81.90% | 84.00% | 82.20% | | 86.20% | 81.90% | 81.50% |
| Spec | 90.60% | 78.30% | 73.50% | | 77.70% | 86.50% | 88.50% | | 77.80% | 80.20% | 80.50% | | 86.80% | 81.90% | 78.60% |
| Sens | 93.90% | 88.70% | 81.60% | | 84.10% | 93.40% | 86.50% | | 85.00% | 86.60% | 83.40% | | 85.70% | 81.90% | 84.30% |
| MCC | 83.50% | 67.40% | 54.20% | | 60.50% | 80.30% | 74.40% | | 63.00% | 66.90% | 63.60% | | 72.50% | 63.70% | 63.10% |

# Chapter 5 Conclusion and Future Works

After doing some experiments to do the stock market prediction, among other experiments it is using several different data periods. The data period is five days trading data, ten days trading data, and also twenty days trading data. In addition we also use some kind of residual network that is distinguished by the number of layers in our architecture. Among other is the residual network with the number 50 layers, one 101 layers, and also 152 layers. With some experiments, the conclusion with the best result is X. It shows that experiment with X can produce X.

Given that the results may not be satisfactory enough, for the future we want to try to improve those results and also add features to this research. Using candlestick chart image on input data is new, the possibility of doing more research on image processing handling will further improve prediction results. To make it easier for readers other than the source code available in github, the possibility of building a web server would be great.

# References

1. Enger, D.E., C.F. Ross, and B.D. Bailey, *Concepts in Biology, 11th edition.* New York: McGraw-Hill, 2005: p. 66.

2. Campbell, N.A., Reece, J. B., Simon, E. J., *Essential Biology 3th edition.* San Francisco: Pearson Benjamin Cummings, 2007: p. 60.

3. Lodish H, B.A., Zipursky SL, et al., *Molecular Cell Biology. 4th edition.* New York: W. H. Freeman, 2000.

4. Santoni, V., M. Molloy, and T. Rabilloud, *Membrane proteins and proteomics: un amour impossible?* Electrophoresis, 2000. **21**(6): p. 1054-1070.

5. SIPOS, L. and G. HEIJNE, *Predicting the topology of eukaryotic membrane proteins.* European Journal of Biochemistry, 1993. **213**(3): p. 1333-1340.

6. Hayat, M. and A. Khan, *Prediction of membrane protein types by using dipeptide and pseudo amino acid composition-based composite features.* Communications, IET, 2012. **6**(18): p. 3257-3264.

7. Bose, S.K., et al. *Presenting a novel neural network architecture for membrane protein prediction*. in *Intelligent Engineering Systems, 2006. INES'06. Proceedings. International Conference on*. 2006. IEEE.

8. Ou, Y.y., S.a. Chen, and M.M. Gromiha, *Prediction of membrane spanning segments and topology in β‐barrel membrane proteins at better accuracy.* Journal of computational chemistry, 2010. **31**(1): p. 217-223.

9. Chou, K.-C. and D.W. Elrod, *Protein subcellular location prediction.* Protein engineering, 1999. **12**(2): p. 107-118.

10. Emanuelsson, O., et al., *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.* Journal of molecular biology, 2000. **300**(4): p. 1005-1016.

11. Park, K.-J. and M. Kanehisa, *Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs.* Bioinformatics, 2003. **19**(13): p. 1656-1663.

12. Hsiao, H.C., et al., *Predicting Subcellular Locations of Eukaryotic Proteins Using Bayesian and/k/-Nearest Neighbor Classifiers.* J. Inf. Sci. Eng., 2008. **24**(5): p. 1361-1375.

13. Cai, Y.-D., et al., *Application of SVM to predict membrane protein types.* Journal of Theoretical Biology, 2004. **226**(4): p. 373-376.

14. Jia, P., et al., *Prediction of membrane protein types in a hybrid space.* Journal of proteome research, 2008. **7**(3): p. 1131-1137.

15. Pierleoni, A., P.L. Martelli, and R. Casadio, *MemLoci: predicting subcellular localization of membrane proteins in eukaryotes.* Bioinformatics, 2011. **27**(9): p. 1224-1230.

16. Li, H., X. Dai, and X. Zhao, *A nearest neighbor approach for automated transporter prediction and categorization from protein sequences.* Bioinformatics, 2008. **24**(9): p. 1129-1136.

17. Ou, Y.-Y. and S.-A. Chen, *Using Efficient RBF Networks to Classify Transport Proteins Based on PSSM Profiles and Biochemical Properties*, in *Bio-Inspired Systems: Computational and Ambient Intelligence*. 2009, Springer. p. 869-876.

18. Chen, S.-A., et al., *Prediction of transporter targets using efficient RBF networks with PSSM profiles and biochemical properties.* Bioinformatics, 2011. **27**(15): p. 2062-2067.

19. Jain, S., et al., *TpPred: A Tool for Hierarchical Prediction of Transport Proteins Using Cluster of Neural Networks and Sequence Derived Features.* International Journal for Computational Biology (IJCB), 2014. **1**(1): p. 28-36.

20. Aplop, F. and G. Butler. *On predicting transport proteins and their substrates for the reconstruction of metabolic networks*. in *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2015 IEEE Conference on*. 2015. IEEE.

21. Mishra, N.K., J. Chang, and P.X. Zhao, *Prediction of Membrane Transport Proteins and Their Substrate Specificities Using Primary Sequence Information.* 2014.

22. Liou, Y.-F., et al., *SCMMTP: identifying and characterizing membrane transport proteins using propensity scores of dipeptides.* BMC genomics, 2015. **16**(Suppl 12): p. S6.

23. Li, L., et al., *Prediction the substrate specificities of membrane transport proteins based on support vector machine and hybrid features.* 2015.

24. Webber, M. and L. Piddock, *The importance of efflux pumps in bacterial antibiotic resistance.* Journal of Antimicrobial Chemotherapy, 2003. **51**(1): p. 9-11.

25. Levy, S.B., *Active efflux mechanisms for antimicrobial resistance.* Antimicrobial Agents and Chemotherapy, 1992. **36**(4): p. 695.

26. Saier, M.H., C.V. Tran, and R.D. Barabote, *TCDB: the Transporter Classification Database for membrane transport protein analyses and information.* Nucleic acids research, 2006. **34**(suppl 1): p. D181-D186.

27. Poole, K., *Efflux pumps as antimicrobial resistance mechanisms.* Annals of medicine, 2007. **39**(3): p. 162-176.

28. Johnson, J.M. and G.M. Church, *Alignment and structure prediction of divergent protein families: periplasmic and outer membrane proteins of bacterial efflux pumps.* Journal of molecular biology, 1999. **287**(3): p. 695-715.

29. van Amsterdam, K., A. Bart, and A. van der Ende, *A Helicobacter pylori TolC efflux pump confers resistance to metronidazole.* Antimicrobial agents and chemotherapy, 2005. **49**(4): p. 1477-1482.

30. Boundless, *ATP: Adenosine Triphosphate.* Boundless Biology. Boundless, 08 Jan. 2016. Retrieved 26 Jan. 2016 from <https://www.boundless.com/biology/textbooks/boundless-biology-textbook/metabolism-6/atp-adenosine-triphosphate-71/atp-adenosine-triphosphate-349-12938/>.

31. DuPont, F., *Salt-induced changes in ion transport: regulation of primary pumps and secondary transporters*, in *Transport and receptor proteins of plant membranes*. 1992, Springer. p. 91-100.

32. Mahamoud, A., et al., *Antibiotic efflux pumps in Gram-negative bacteria: the inhibitor response strategy.* Journal of Antimicrobial Chemotherapy, 2007. **59**(6): p. 1223-1229.

33. Moreira, M.A.S., E.C.d. Souza, and C.A.d. Moraes, *Multidrug efflux systems in Gram-negative bacteria.* Brazilian Journal of Microbiology, 2004. **35**(1-2): p. 19-28.

34. Martinez, J.L., et al., *Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems.* FEMS microbiology reviews, 2009. **33**(2): p. 430-449.

35. Langton, K.P., P.J. Henderson, and R.B. Herbert, *Antibiotic resistance: multidrug efflux proteins, a common transport mechanism?* Natural product reports, 2005. **22**(4): p. 439-451.

36. Kourtesi, C., et al., *Suppl 1: Microbial Efflux Systems and Inhibitors: Approaches to Drug Discovery and the Challenge of Clinical Implementation.* The open microbiology journal, 2013. **7**: p. 34.

37. Pao, S.S., I.T. Paulsen, and M.H. Saier, *Major facilitator superfamily.* Microbiology and Molecular Biology Reviews, 1998. **62**(1): p. 1-34.

38. Paulsen, I.T., et al., *The SMR family: a novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs.* Molecular microbiology, 1996. **19**(6): p. 1167-1175.

39. Kuroda, T. and T. Tsuchiya, *Multidrug efflux transporters in the MATE family.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2009. **1794**(5): p. 763-768.

40. Schmitt, L. and R. Tampé, *Structure and mechanism of ABC transporters.* Current opinion in structural biology, 2002. **12**(6): p. 754-760.

41. Moussatova, A., et al., *ATP-binding cassette transporters in Escherichia coli.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2008. **1778**(9): p. 1757-1771.

42. Pedersen, J.M., *ATP-Binding-Cassette Transporters in Biliary Efflux and Drug-Induced Liver Injury.* 2013.

43. Blair, J.M. and L.J. Piddock, *Structure, function and inhibition of RND efflux pumps in Gram-negative bacteria: an update.* Current opinion in microbiology, 2009. **12**(5): p. 512-519.

44. Nikaido, H. and Y. Takatsuka, *Mechanisms of RND multidrug efflux pumps.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2009. **1794**(5): p. 769-781.

45. Borges-Walmsley, M., K. McKeegan, and A. Walmsley, *Structure and function of efflux pumps that confer resistance to drugs.* Biochem. J, 2003. **376**: p. 313-338.

46. Fluman, N., et al., *Dissection of mechanistic principles of a secondary multidrug efflux protein.* Molecular cell, 2012. **47**(5): p. 777-787.

47. Ou, Y.Y., et al., *Identification of efflux proteins using efficient radial basis function networks with position‐specific scoring matrices and biochemical properties.* Proteins: Structure, Function, and Bioinformatics, 2013. **81**(9): p. 1634-1643.

48. Gromiha, M.M., et al., *Mutational studies to understand the structure–function relationship in multidrug efflux transporters: Applications for distinguishing mutants with high specificity.* International journal of biological macromolecules, 2015. **75**: p. 218-224.

49. Reddy, V.S., et al., *The major facilitator superfamily (MFS) revisited.* FEBS Journal, 2012. **279**(11): p. 2022-2035.

50. Madej, M.G., et al., *Evolutionary mix-and-match with MFS transporters.* Proceedings of the National Academy of Sciences, 2013. **110**(15): p. 5870-5874.

51. Law, C.J., P.C. Maloney, and D.-N. Wang, *Ins and outs of major facilitator superfamily antiporters.* Annual review of microbiology, 2008. **62**: p. 289.

52. Marger, M.D. and M.H. Saier, *A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport.* Trends in biochemical sciences, 1993. **18**(1): p. 13-20.

53. Saier Jr, M.H., et al., *The major facilitator superfamily.* J Mol Microbiol Biotechnol, 1999. **1**(2): p. 257-279.

54. Lomovskaya, O. and K. Lewis, *Emr, an Escherichia coli locus for multidrug resistance.* Proceedings of the National Academy of Sciences, 1992. **89**(19): p. 8938-8942.

55. Lee, E.H. and W.M. Shafer, *The farAB‐encoded efflux pump mediates resistance of gonococci to long‐chained antibacterial fatty acids.* Molecular microbiology, 1999. **33**(4): p. 839-845.

56. Colmer, J.A., J.A. Fralick, and A.N. Hamood, *Isolation and characterization of a putative multidrug resistance pump from Vibrio cholerae.* Molecular microbiology, 1998. **27**(1): p. 63-72.

57. Yan, N., *Structural advances for the major facilitator superfamily (MFS) transporters.* Trends in biochemical sciences, 2013. **38**(3): p. 151-159.

58. Abramson, J., et al., *Structure and mechanism of the lactose permease of Escherichia coli.* Science, 2003. **301**(5633): p. 610-615.

59. Huang, Y., et al., *Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli.* Science, 2003. **301**(5633): p. 616-620.

60. Yin, Y., et al., *Structure of the multidrug transporter EmrD from Escherichia coli.* Science, 2006. **312**(5774): p. 741-744.

61. Dang, S., et al., *Structure of a fucose transporter in an outward-open conformation.* Nature, 2010. **467**(7316): p. 734-738.

62. Newstead, S., et al., *Crystal structure of a prokaryotic homologue of the mammalian oligopeptide–proton symporters, PepT1 and PepT2.* The EMBO journal, 2011. **30**(2): p. 417-426.

63. Solcan, N., et al., *Alternating access mechanism in the POT family of oligopeptide transporters.* The EMBO journal, 2012. **31**(16): p. 3411-3421.

64. Sun, L., et al., *Crystal structure of a bacterial homologue of glucose transporters GLUT1-4.* Nature, 2012. **490**(7420): p. 361-366.

65. Kumar, S., et al., *Bacterial antimicrobial efflux pumps of the MFS and MATE transporter families: A review.* Recent Research Developments in Antimicrobial Agents & Chemotherapy, 2013: p. 1-21.

66. Bay, D.C., K.L. Rommens, and R.J. Turner, *Small multidrug resistance proteins: a multidrug transporter family that continues to grow.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2008. **1778**(9): p. 1814-1838.

67. Kolbusz, M.A., et al., *Orientation of small multidrug resistance transporter subunits in the membrane: correlation with the positive-inside rule.* Journal of molecular biology, 2010. **402**(1): p. 127-138.

68. Omote, H., et al., *The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations.* Trends in Pharmacological Sciences, 2006. **27**(11): p. 587-593.

69. Alvarez-Ortega, C., J. Olivares, and J.L. Martínez, *RND multidrug efflux pumps: what are they good for?* Frontiers in microbiology, 2013. **4**.

70. Fernando, D.M. and A. Kumar, *Resistance-nodulation-division multidrug efflux pumps in gram-negative bacteria: role in virulence.* Antibiotics, 2013. **2**(1): p. 163-181.

71. Li, X.-Z., K. Poole, and H. Nikaido, *Contributions of MexAB-OprM and an EmrE homolog to intrinsic resistance of Pseudomonas aeruginosa to aminoglycosides and dyes.* Antimicrobial agents and chemotherapy, 2003. **47**(1): p. 27-33.

72. Davidson, A.L. and J. Chen, *ATP-binding cassette transporters in bacteria.* Annual review of biochemistry, 2004. **73**(1): p. 241-268.

73. Kumar, A. and H.P. Schweizer, *Bacterial resistance to antibiotics: active efflux and reduced uptake.* Advanced drug delivery reviews, 2005. **57**(10): p. 1486-1513.

74. Davidson, A.L. and P.C. Maloney, *ABC transporters: how small machines do a big job.* Trends in microbiology, 2007. **15**(10): p. 448-455.

75. Golin, J., et al., *Studies with novel Pdr5p substrates demonstrate a strong size dependence for xenobiotic efflux.* Journal of Biological Chemistry, 2003. **278**(8): p. 5963-5969.

76. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic acids research, 1997. **25**(17): p. 3389-3402.

77. 富井健太郎 and 金久實, *AAindex: A Database of Amino Acid Indices and Mutation Matrices.* Genome Informatics, 1995. **6**: p. 142-143.

78. Tang, J., S. Alelyani, and H. Liu, *Feature selection for classification: A review.* Data Classification: Algorithms and Applications, 2014: p. 37.

79. Ou, Y., *QuickRBF: a package for efficient radial basis function networks.* Software available at <https://www.csie.ntu.edu.tw/~yien/quickrbf/>, 2005.

80. Chang, C.-C. and C.-J. Lin, *LIBSVM: a library for support vector machines.* ACM Transactions on Intelligent Systems and Technology (TIST), 2011. **2**(3): p. 27.

81. Holmes, G., A. Donkin, and I.H. Witten. *Weka: A machine learning workbench*. in *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*. 1994. IEEE.

82. Du, K.-L. and M. Swamy, *Radial basis function networks*, in *Neural Networks and Statistical Learning*. 2014, Springer. p. 299-335.

83. Ou, Y.-Y., Y.-J. Oyang, and C.-Y. Chen. *A novel radial basis function network classifier with centers set by hierarchical clustering*. in *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*. 2005. IEEE.

84. BLAST, G., *PSI-BLAST: a new generation of protein database search programs Altschul.* Stephen F: p. 3389-3402.

85. Consortium, U., *The universal protein resource (UniProt) in 2010.* Nucleic acids research, 2010. **38**(suppl 1): p. D142-D148.