

**PUMP AND DUMP DETECTION ON CRYPTO  
CURRENCIES USING COMPUTER VISION**

Jayasundara Abeykoon Waruna Priyankara Wickramasingha

199371P

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2021

**PUMP AND DUMP DETECTION ON CRYPTO  
CURRENCIES USING COMPUTER VISION**

Jayasundara Abeykoon Waruna Priyankara Wickramasingha

199371P

Dissertation submitted in partial fulfilment of the requirements for the  
Degree Master of Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2021

## Declaration

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).



2021/05/29

J.A.W.P Wickramasingha

Date

The above candidate has carried out research for the Master's dissertation under my supervision.



2021/05/29

.....  
Dr. Uthayasanker Thayasivam

.....  
Date

## **Abstract**

Inspired by the immense success shown by artificial neural networks in computer vision on images classification, we propose a novel framework to detect one of the rife fraudulent financial manipulations in crypto currency trading world known as pump and dump. The representation of crypto currency financial charts was re-imagined ameliorating the classification by taking advantage of some of the very recent advancements of time series to spatial encoding techniques of Gramian Angular Field (GAF), Markov Transition Field (MTF) and Recurrence plots (RP) that are capable of spatially encoding the temporal financial time series data in the form of images. Encoded images were then used to train several convolutional neural network architectures which have been able to achieve a very high precision, recall and F1 values close to 99% over the unseen data for the above classification task. This is one of the first of such researches in pump and dump detection in crypto currencies using computer vision. This approach has the potential to be extended in detecting predefined shapes of time series charts.

## **Key words**

*cryptocurrency, pump and dump, imbalanced time series classification, spatial encoding temporal data, Gramian angular field, Markov transition field, recurrence plot, cnn, machine learning, market surveillance, class imbalance problem, synthetic minority class oversampling technique*

## **Acknowledgement**

My utmost gratitude and appreciation goes to Mrs. Kaushalya Kularatnam - Head of Quantitative Surveillance & Technology London Stock Exchange in making this project a success by providing the initial idea of this research and facilitating all the business, technical and required cloud resources for this research from the very beginning.

I would like to give my utmost gratitude to my supervisor Dr. Uthayasanker Thayasivam - Senior Lecturer of department of Computer Science and Engineering of University of Moratuwa, giving the outstanding consultation and supervision throughout the project.

I am really grateful to Dr. Rasika Withanawasam - Senior Software Architect at Millennium Surveillance systems, sharing all the expertise and knowledge in financial surveillance world and providing the golden opportunity to conduct this research with the help of London Stock Exchange's (LSE) market supervision team.

I would like to extend my gratitude to all the wonderful people from London Stock Exchange Quantitative Surveillance team and LSEG Technology who supported verifying the classifications by these models and assisted me expand the knowledge and exposure to capital market and market surveillance domains.

Last but not least, I would like to extend my gratitude to my beloved wife Ranmali and beloved baby Kiara who were behind me all the time giving me all the strength, courage and dedication for the time required for this research.

# TABLE OF CONTENTS

Declaration	i
Abstract	ii
Key words	ii
Acknowledgement	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vi
LIST OF TABLES	viii
LIST OF ABBREVIATIONS	ix
LIST OF APPENDICES	x
1. INTRODUCTION	2
1.1 Problem Definition	2
1.2 What are Crypto Currencies?	5
1.3 Motivation	6
1.4 Pump and Dump alerting logic	8
2 LITERATURE REVIEW	11
2.1 NLP techniques to classify financial frauds	11
2.2 Converting time series classification into image classification domain	13
2.3 LSTM techniques for time series classification	15
2.4 Clustering and automatic thresholding techniques	15
2.5 Threshold-based approaches	16
3 METHODOLOGY	18
3.1 Downloading and encoding data	18
3.2 Binance python API	20
3.3 Data cleaning and enriching new features	21
3.4 Noise in the downloaded data	25
3.5 Labelling the data	25
3.6 Temporal to spatial encoding	33
3.6.1 Gramian Angular Field (GAF)	33
3.6.2 Markov Transition Field (MTF)	35
3.6.3 Recurrence Plots (RP)	36
3.7 Evaluating labelling logic	38
3.8 Generating a synthetic data set using SMOTE	40

3.8.1	General SMOTE	41
3.8.2	Borderline SMOTE	42
3.8.3	Adaptive Synthetic (ADASYN) SMOTE	43
4	Experiments	48
4.1	CNN hyper parameter tuning	49
4.2	CNN Model training	51
4.2.1	Price and Volume encoded with GADF CNN3	52
4.2.2	Price and Volume encoded with GADF CNN5	53
4.2.3	Price and Volume encoded with GASF CNN5	54
4.2.4	Price and Volume encoded with MTF CNN3	55
4.2.5	Price and Volume encoded with RP CNN3	56
4.3	Benchmark Experiments	56
5	RESULTS	60
6	DISCUSSION	63
6.1	Contribution	64
7	CONCLUSIONS	66
7.1	Future work	66
	References	67
	Appendices	69

## LIST OF FIGURES

Figure 1.1 Schematic abstraction of the three phases of a pump-and-dump operation	3
Figure 1.2 Converting continuous time series to OHLC candlestick	4
Figure 1.3 Price chart of a crypto currency against time	7
Figure 2.1 Use of obfuscation to make it hard for bots to parse	12
Figure 3.1 Sample chart to be encoded in different techniques	19
Figure 3.2 Price 10% pumping during 10 minutes example	22
Figure 3.3 Price 15% pumping during 10 minutes example	22
Figure 3.4 Price 20% pumping during 10 minutes example	23
Figure 3.5 Turnover 75% significance during 5 minutes	23
Figure 3.6 Left window representing an overall price increase	24
Figure 3.7 Window to the right represents overall price decrease	24
Figure 3.8 Noise in HOTBTC full view Price and Turnover	25
Figure 3.9 Basic stages of a pump and dump	26
Figure 3.10 Real pump and dump instance of POLY	27
Figure 3.11 Initial labelling logic	27
Figure 3.12 Directory hierarchy within s3 bucket	29
Figure 3.13 PND Price variation of APPBTC 2018-07-03_233400	29
Figure 3.14 Director hierarchy within s3 bucket	30
Figure 3.15 Volume variation during PND of APPBTC 2018-07-03_233400	31
Figure 3.16 Turnover variation during PND of APPBTC 2018-07-03_233400	31
Figure 3.17 Price variation during non PND of APPCBTC 2018-05-04_231600	32
Figure 3.18 Volume variation during non PND of APPCBTC 2018-05-04_231600	32
Figure 3.19 Turnover variation during non PND APPCBTC 2018-05-04_231600	33
Figure 3.20 Normalizing time series data	33
Figure 3.21 GAF angular cosine representation	34
Figure 3.22 Gramian angular field matrix	34
Figure 3.23 Gramian Angular Field encoding summary	35
Figure 3.24 Markov transition field	35
Figure 3.25 Markov Transition Field encoding summary	36
Figure 3.26 Recurrence matrix equation	37
Figure 3.27 Heaviside function	37
Figure 3.28 Sample Recurrence plots; totally random noise (left); random walk (middle); periodic composition of sine and cosine (right)	37
Figure 3.29 Thresholded (left) and an unthresholded (right) recurrence plots, generated from a same time series	38
Figure 3.30 Calculating expected agreement	39
Figure 3.31 Calculating kappa value	39
Figure 3.32 Class distribution before applying SMOTE	40
Figure 3.33 Class distribution before SMOTE after PCA	41
Figure 3.34 Class distribution after SMOTE	42
Figure 3.35 Class distribution visualized with PCA after general SMOTE	42



Figure 3.36 Class distribution visualized with PCA after borderline SMOTE	43
Figure 3.37 Class distribution visualized with PCA after ADASYN SMOTE	43
Figure 3.38 Applying SMOTEd instances to target data points	44
Figure 3.39 Before and after applying SMOTE cases over the span	45
Figure 3.40 Class distribution for a coin after applying synthetic samples 3hrs apart	46
Figure 3.41 Balanced class distribution set for model training	46
Figure 4.1 Variables of the experiment	48
Figure 4.2 Summary of experiment lifecycle	49
Figure 4.3 Hyper parameter tuning example	50
Figure 4.4 CNN3 architecture	51
Figure 4.5 CNN5 architecture	52
Figure 4.6 GADF CNN3 averaged confusion matrix	52
Figure 4.7 GADF CNN3 model fitting graph with epochs	53
Figure 4.8 GADF CNN5 averaged confusion matrix	54
Figure 4.9 GADF CNN5 model fitting graph with epochs	54
Figure 4.10 GASF CNN5 model fitting graph	55
Figure 4.11 MTF CNN3 model fitting graphs	55
Figure 4.12 RP CNN3 model fitting graphs	56
Figure 4.13 Rolling window average of close price	57
Figure 4.14 Price anomaly detection	58
Figure 4.15 Volume anomaly detection	58
Figure 6.1 GADF CNN5 predictions confusion matrix	63

## **LIST OF TABLES**

Table 1.1 Top 10 coins by percentage of market cap	5
Table 1.2 Common approached in PND detection	6
Table 1.3 Example data point of OHLCV data stream	7
Table 3.1 Same time window when encoded in different spatial techniques	19
Table 3.2 Kappa values for the labels	39
Table 4.1 10-fold cross validation results of GADF CNN3 model	52
Table 4.2 10-fold cross validation results of GADF CNN5 model	53
Table 4.3 10-fold cross validation results of GASF CNN5 model	54
Table 4.4 10-fold cross validation results of MTF CNN3 model	55
Table 4.5 10-fold cross validation results of RP CNN3 model	56
Table 5.1 CNN models 10-fold cross validation results summary	60
Table 5.2 Model evaluation	60

## **LIST OF ABBREVIATIONS**

<b>Abbreviation</b>	<b>Definition</b>
CNN	Convolutional Neural Network
GAF	Gramian Angular Field
GASF	Gramian Angular Summation Field
GADF	Gramian Angular Difference Field
MTF	Markov Transition Field
RP	Recurrence Plot
PND	Pump and Dump
OHLCV	Open High Low Close Volume
MTS	Multivariate Time Series Sensory
SMOTE	Synthetic Minority Oversampling Technique

## **LIST OF APPENDICES**

Appendix A Downloading data amidst of disconnections	69
Appendix B GADF CNN 3	70
Appendix C GADF CNN 5	71
Appendix D Spatial to temporal encoding	72
Appendix E ICNBTC expert comments and label	73
Appendix F MODBTC expert comments and label	74
Appendix G QKCBTC expert comments and label	75
Appendix H SCBTC expert comment and label	76