

**Comparative Analysis of Supervised Learning Algorithms on
Image Classification of Saouropus Androgynus Domestic Food
Processing**

UNDERGRADUATE THESIS



**A. GREGORY QO'NITAH MICHELLE
1162001017**

**INFORMATICS STUDY PROGRAM
FACULTY OF ENGINEERING AND COMPUTER SCIENCE
BAKRIE UNIVERSITY
JAKARTA
2023**

Statement of Originality

The material in this Undergraduate Thesis is the result of my on work, and all source are quoted and cited properly.

Name : A Gregory Qo'nitah Michelle

NIM : 1162001017

Signature : 

Date : 15 February 2023

Statement of Approval

This research is prepared and submitted by :

Name : A. Gregory Qo'nitah Michelle
NIM : 1162001017
Departement : Informatics
Faculty : Engineering and Computer Science
Title : Comparative Analysis of Supervised Learning Algorithms on Image
Classification of Saouropus Androgynus Domestic Food Processing

has been approved by the Board of Examiners and accepted as partial fulfillment of the requirements to obtain a Bachelor degree in Informatics Departement, Faculty of Engineering and Computer Science, Bakrie University.

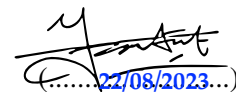
Jakarta, 15 February 2023

Approved By,

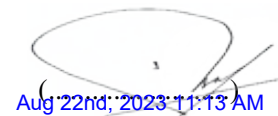
Supervisor : Irwan Prasetya Gunawan, S.T, M,Eng, Ph.D (.



Examiner 1 : Guson P. Kuntarto, S.T., M.Sc.


(.....22/08/2023...)

Examiner 2 : Berkah Iman Santoso, ST, MTI


Aug 22nd, 2023 11:13 AM

Acknowledgment

With gratitude and thanks to the Almighty Allah, the author extends their appreciation for His abundant grace and guidance, allowing the completion of the manuscript titled "Comparative Analysis of Supervised Learning Algorithms on Image Classification of Saouropus Androgynus Domestic Food Processing." This Final Project serves as a prerequisite for obtaining a Bachelor of Computer Science degree in the Computer Science Engineering Program at the Faculty of Engineering and Computer Science, Universitas Bakrie.

The author highly appreciates the support, guidance, assistance, and advice from all parties who have contributed to the research and compilation of this Final Project. Thus, with utmost humility, the author wishes to extend gratitude to:

1. Thank you to my beloved family, Mother, Father, and Sibling, who have always supported the author's journey throughout their academic pursuits. They have consistently provided encouragement and unwavering support.
2. I would like to extend my gratitude to my esteemed supervising advisor, Mr. Irwan Prasetya Gunawan, ST, MEng, PhD, who has significantly contributed to the development of this thesis. His expert guidance and mentorship have been instrumental from the initial stages to the culmination of this writing process.
3. Mr. Berkah Iman Santoso ST, MTI, and Mr. Guson Prasamuarso Kuntarto, S.T., M.Sc., as the examiners during the final project defense, who offered guidance and suggestions for improving the quality of this final project's research.
4. My support system Fiqih, who has been immensely helpful by providing advice, suggestions, and assistance in structuring and conducting this research from beginning to end,

and who has fought alongside in completing this final project.

5. My friend till Jannah Inshaa Allah, Elsismone, Syifa, Refa, Binya, Mifta, Helmi, and Lia, who have provided support in various aspects and consistently offered advice to ensure the completion of this thesis.
6. My partner Rizchy H and Khairunnisa, fellow warriors who provided motivation to complete this final project together.
7. My friends in informatics 2016, who have stood together until now.

May Allah SWT repay their kindness with even greater blessings. This Final Project is presented with full awareness of its imperfections; therefore, all criticism and suggestions will contribute to enhancing the quality of this final project. The hope is that this final project will be beneficial for various parties in the future.

Jakarta, 12 February 2023

Author



A. Gregory Qo' nitah Michelle

Declaration of Partial Copyright License

As student of Bakrie University, I here by:

Name : A Gregory Qo'nitah Michelle
NIM : 1162001017
Departement : Informatics
Fakulty : Engineering and Computer Science

The author, whose copyright is declared on the title page of this work, agree and grant Bakrie University a non exclusive royalty free right for educational and academic endeavour with this Undergraduate Thesis titled:

Comparative Analysis of Supervised Learning Algorithms on Image Classification of
Saouropus Androgynus Domestic Food Processing

With the granted permission to use this material, Bakrie University is allowed to keep or make digital copy, communicate, and publish this Undergraduate Thesis by providing full acknowledgement of the copyright and the source of the material. It is understood that copying and publicizing of this work for financial gain shall not be allowed without written permission. The original Partial Copyright License attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Bakrie University Archive.

Authorize in : Jakarta
Date : 15 February 2023



A Gregory Qo'nitah Michelle

Abstract

This research compared supervised machine learning algorithms that classify images. The focus is an image of food processing on katuk leaves in domestic food. Food processing is carried out food processing by boiling and steaming. This research began by collecting a dataset of images captured using a Fujifilm camera and an iPhone camera. After that, the image will be grouping based on the two categories of food processing. Each category will be divided into dataset modification. Before going to the classification stage, the author will preprocess the images that have been collected. In the classification process, the system will classify using supervised machine learning algorithms, namely CNN algorithms. This research applied CNN with few scenario that compared the number epochs, number of layer, and dataset modification. The result of comparing the scenario are the models with the number of epoch 100 have better accuracy than model with the number of epoch 50, the models with a higher number of layers tend to perform better than the ones with fewer layers and The models that use 90% and 80% of the data for training tend to perform slightly better than the ones that use 70%. The final outcome of the alternate scenario is that datasets subjected to cropping processing achieve better accuracy values compared to those that do not undergo the cropping process.

Keywords: Food Procesing, Image Classification, Image processing, CNN Algorithm , Confusion Matrix.

Contents

Statement of Originality	i
Statement of Approval	ii
Acknowledgment	iii
Declaration of Partial Copyright License	v
Abstract	vi
Contents	vii
List of Figures	ix
List of Tables	x
List Of Abbreviations	xi
1 Introduction	1
1.1 Background	1
1.2 Problem Statement	2
1.3 Purpose and Benefit	2
1.3.1 Purpose of Research	2
1.3.2 Benefit of Research	3
1.4 Scope Of Research	3
1.5 Outlines Of Proposal	3
1.6 Summary	4
2 Literature Review	5
2.1 Image Processing	5
2.1.1 RGB Color Image	6
2.2 Machine Learning	7
2.3 Deep Learning	8
2.4 Image Classification	8
2.4.1 CNN Algorithm Implementation	9
2.5 Performance Measure	13
2.6 Summary	14
3 Research Methodology	15

3.1	Research Phase	15
3.1.1	Literature Study	15
3.1.2	Research Problem	16
3.1.3	Collecting Dataset	16
3.1.4	Conducting Research	18
3.1.5	Results and Analysis	21
3.1.6	Report	22
3.2	Research Framework	22
3.3	Research Tools	23
3.4	Summary	24
4	Result and analysis	25
4.1	Simulation A	25
4.1.1	Number of Epoch	28
4.1.2	Number of Layer	32
4.1.3	Datasets split simulation	36
4.2	Simulation B	40
4.2.1	Performance Comparison	43
4.3	Summary	45
5	Conclusion	46
5.1	Conclusions	46
5.2	Future Work	46
	Bibliography	48

List of Figures

2.1	Feature, technique and analysis procedure of image processing [19]	6
2.2	Image Classification Process [19]	9
2.3	Convolution layer [3]	10
2.4	Pooling layer [3]	10
2.5	Fully Connected layer [3]	11
3.1	Research Phase	15
3.2	Katuk leaf datasets	17
3.3	Modify color profile with auto White Balanced	18
3.4	Example of auto White Balanced setting	19
3.5	CNN Architecture 1 Layer	19
3.6	CNN Architecture 2 Layer	20
3.7	CNN Architecture 3 Layer	20
3.8	CNN Architecture 4 Layer	21
3.9	Research Framework	23
4.1	Comparison of model accuracy with different numbers of epochs simulation	28
4.2	Comparison of model precision with different numbers of epochs simulation	29
4.3	Comparison of model recall with different numbers of epochs simulation	30
4.4	Comparison of model f-1 score with different numbers of epochs simulation	31
4.5	Comparison of model accuracy with different numbers of layer simulation	33
4.6	Comparison of model precision with different numbers of layer simulation	33
4.7	Comparison of model recall with different numbers of layer simulation	34
4.8	Comparison of model f-1 score with different numbers of layer simulation	35
4.9	Comparison of model accuracy with dataset split simulation	36
4.10	Comparison of model precision with dataset split simulation	37
4.11	Comparison of model recall with dataset split simulation	38
4.12	Comparison of model f-1 score with dataset split simulation	39
4.13	crop and non-crop datasets	40
4.14	Plot Accuracy, Validation Accuracy, Loss, and Validation Loss of the compare models	42
4.15	Precision comparison	43
4.16	Recall comparison	43
4.17	F-1 Score comparison	43
4.18	Confusion matrix comparison	44

List of Tables

2.1	<i>Previous research conducted CNN Algorithm research</i>	12
2.2	<i>Confussion Matrix</i>	13
3.1	<i>Grouping of katuk leaf images based on the method</i>	17
4.1	Performance of classification accuracy based on difference in the number of epochs.	28
4.2	Performance of classification precision based on difference in the number of epochs.	29
4.3	Performance of classification recall based on difference in the number of epochs.	30
4.4	Performance of classification f-1 Score based on difference in the number of epochs.	31
4.5	Performance of classification accuracy based on difference in the number of layer.	32
4.6	Performance of classification precision based on difference in the number of layer.	33
4.7	Performance of classification recall based on difference in the number of layer.	34
4.8	Performance of classification f-1 Score based on difference in the number of layer.	34
4.9	Performance of classification accuracy based on difference in the dataset split.	36
4.10	Performance of classification precision based on difference in the dataset split.	37
4.11	Performance of classification recall based on difference in the dataset split.	38
4.12	Performance of classification f-1 Score based on difference in the dataset split.	39
4.13	Performance of classification accuracy based on dataset preprocessing	42

List Of Abbreviations

ANN	: Artificial Neural Network
CNN	: Convolutional Neural Network
DT	: Decision Tree
GLCM	: Gray-Level Co-Occurrence Matrix
HSI	: Hue Saturation Intensity
KNN	: K-Nearest Neighbor
LTSM	: Long Short-Term Memory Networks
LR	: Logistic Regression
MLP	: Multi-layer Perceptron Classifier
RF	: Random Forest
RGB	: Red Green Blue
SOM	: Self Organizing Maps
SVM	: Support Vector Machine