A DEEP LEARNING MODEL FOR MICROPLASTICS DETECTION IN OPEN SEWER SYSTEMS

JOSEPH MICHAEL ODHIAMBO

A THESIS SUBMITTED TO THE INSTITUTE OF COMPUTING AND INFORMATICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN INFORMATION TECHNOLOGY OF THE TECHNICAL UNIVERSITY OF MOMBASA

2023

DECLARATION

This thesis is my original work and has not been presented for academic award in any other university.

Sign _____ Date _____

JOSEPH MICHAEL ODHIAMBO

MSIT/0012/2021

This thesis report has been submitted with our approval as University Supervisors.

Sign _____ Date _____

DR. MVURYA MGALA

Institute of Computing and Informatics

Technical University of Mombasa

Sign Date 03/06/2023

DR. ANTHONY LUVANDA

Institute of Computing and Informatics

Technical University of Mombasa

DEDICATION

I dedicate this work to my three daughters, and to my late wife, Vida. They are the inspiration for this work.

ACKNOWLEDGEMENT

I would like to thank Dr. Mgala Mvurya, Dr. Anthony Luvanda, Dr. Mwakondo, Dr. Kennedy Ondimu, and Dr. Kennedy Hadullo for their direct and indirect contributions to this work. I took a lot of inspiration from their lectures, especially from Dr. Luvanda, Dr. Hadullo, Dr. Mgala, and Dr. Mwakondo, who were the first objective critics of this topic and helped shape it up to this standard. I cannot forget Dr. Obadiah Musau who accompanied me during data collection exercise and provided very valuable insights into my research.

I would also like to thank Mr. Bernard Onyango, a classmate who encouraged me to pursue higher education. His faith in me has motivated me to take on this task. Onyi (that's his pet name), you're a true friend, and I thank you very much for seeing something good in me. I would also like to thank the other members of my master's class, especially Asnath Nyachiro. She has been by my side through thick and thin, even when we thought we couldn't make it. She encouraged me to keep on working hard and harder. I am grateful to you and others for your insightful comments, which have helped shape this thesis into what it is today.

In a special way, I want to appreciate the Technical University of Mombasa through Registrar PRI for grant that made my data collection and publication a smooth process. Last but not least, I want to thank the Almighty God for bringing me this far and for allowing me to complete this thesis and graduate with a masters of science degree in information technology. Thank you very much, Lord.

TABLE OF CONTENTS

DECLARATION ii
DEDICATION iii
ACKNOWLEDGEMENT iv
TABLE OF CONTENTS v
LIST OF TABLES ix
LIST OF FIGURES x
LIST OF ACRONYMS AND ABBREVIATION xi
DEFINITION OF KEY TERMS xiii
CHAPTER ONE 1
INTRODUCTION 1
1.1 Background of the Study 1
1.2 Statement of the Problem 4
1.3 Objectives of the Study 5
1.3.1 General Objective of the Study 5
1.3.2 Specific Objectives of the Study 6
1.4 Research Questions 6
1.5 Significance of the Study 6
1.6 Limitations of the Study 7
1.7 Scope of the Study 8
1.8 Organization of the Research 9
CHAPTER TWO 10
THE LITERATURE REVIEW 10
2.1 Microplastics 10
2.2 Microplastics in the Food Chain 12
2.3 Sources of Microplastics MPs 13
2.4 Identification of MPs Using Traditional Methods 14
2.5 Deep Learning 17
2.5.1 Convolutional Neural Network (CNN) 17
2.5.2 Recurrent Neural Network (RNN) 19
2.5.3 Deep Neural Network (DNN) 20
2.6 Related Work on Identification of MPs Using Machine Learning 20

2.7 Related Work on Identification of MPs Using Deep Learning 22
2.8 Lessons Learned from Literature Reviewed 25
2.9 Identified Gap from the Literature Reviewed 26
2.10 Chapter Summary 27
CHAPTER THREE 29
METHODOLOGY 29
3.1 Methodology Used 29
3.2 Study Approach 31
3.3 Research Design 33
3.4 Research Framework 35
3.4.1 Microplastics Physical Feature Determination 35
3.4.2 Data Selection 37
3.4.3 Data Collection 38
3.4.4 Data Preprocessing 39
3.4.5 Algorithm Selection on TensorFlow 43
3.5 Details of the Considered Algorithms 44
3.5.1 The Single-Shot Detector (SSD) 44
3.5.2 Convolutional Neural Networks (CNN) 45
3.6 Model Building 49
3.6.1 Sing-Short Detector (SSD) Model 52
3.6.2 Convolutional Neural Network (CNN) Model 55
3.7 Models Training and Testing 58
3.7.1 Training Sing-Short Detector (SSD) 59
3.7.2 CNN Model Training 60
3.8 Models Evaluation 62
3.9 Chapter Summary 65
CHAPTER FOUR 67
RESULTS AND DISCUSSION 67
4.1 Identifying Microplastics Features 67
4.1.1 Physical Characteristics of Microplastics Features 68
4.1.2 Feature Extraction from Images 69
4.2 A Deep Learning Model for Identifying Microplastics 73
4.2.1 Convolutional Neural Network (CNN) Model 74
4.2.2 Single Shot Detector (SSD) Model 78
4.3 Test The Model to Establish its Applicability 82

4.3.1 Testing Results for CNN Model 82
4.3.2 Testing Results for SSD Model 83
4.4 Comparison of the CNN & SSD Models 84
4.5 Discussion of the Results 85
4.6 Chapter Summary 87
CHAPTER FIVE 88
CONCLUSION AND RECOMMENDATIONS 88
5.1 Synthesis of Deep Learning process findings 88
5.1.1 Characteristics Indicating Presence of Microplastics 89
5.1.2 Model for Detecting Microplastics in Photos 90
5.1.3 Model Testing to Determine its Applicability 91
5.2 Summary of the Conclusion 91
5.2.1 Framework Used 92
5.2.2 The Process 93
5.2.3 Contribution to the Knowledge 94
5.3 Limitations of the Study 94
5.4 Recommendation for Further Research 95
REFERENCES 97
APPENDICES 101
Appendix 1 101
Code snippet for SSD Model 101
Appendix 2 109
Publication Certificate 109
Appendix 3 110
School of Graduate Studies (SGS) Certifications 110
Appendix 4 111
Ethical Review Committee (ERC-TUM) Certificate 111
Appendix 5 112
NACOSTI Certificate 112

LIST OF TABLES

Table 3.1 Attributes of Object Detection Algorithms 48

Table 4.1: Extracted Features in CSV Format 72

Table 4.2: CNN Accuracy, Loss & Time Results 83

Table 4.3: SSD Accuracy, Loss & Time Results 84

Table 4.4: mAP Comparison of the CNN & SSD Models 84

LIST OF FIGURES

Figure 3.1: SEMMA Process 30

Figure 3.2: A Deep Learning Stepwise Flow 34

Figure 3.3: Research Framework 35

Figure 3.4: Sample of Primary Data 40

Figure 3.5: Code for Image Renaming 41

Figure 3.6: Code for Feature Extraction 42

LIST OF ACRONYMS AND ABBREVIATION

AI	Artificial Intelligence
ANN	Artificial Neural Networks
API	Application Programming Interface
ATR	Attenuated Total Reflectance
CNN	Convolutional Neural Networks
CONV	Convolutional layers
ConvNet	Convolutional Neural Network
CRISP-DM	Cross Industry Standard Process in Data Mining
CSO	Combined Sewer Overflow
CSV	Comma Separated Value
DNN	Deep Neural Networks
Dr.	Doctor
DSC	Dice Similarity Coefficient
FTIR	Fourier Transform Infrared
GRU	Gated Recurrent Units
IoU	Intersection over Union
KDD	Knowledge Discovery in Databases
k-NN	k Nearest Neighbor
LSTM	Long Short Term Memory
ML	Machine Learning
MPs	Microplastics
NOAA	National Oceanic and Atmospheric Administration
OCCAM	Open Center for the Characterization of Advanced Materials
PP	Polypropylene
QA/QC	Quality Assurance and Quality Control
RDF	Random Decision Forest

ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SEMMA	Sample, Explore, Modify, Model and Assess
SIFT	Scale Invariant Feature Transform Algorithm
SVM	Support Vector Machine
TUM	Technical University of Mombasa
USR	Untreated Surface Runoff
UV	Ultra Violet
WWF	Waste Water Effluent
WWTPs	Waste Water Treatment Plants

DEFINITION OF KEY TERMS

Artefact	Artifact is a piece of creativity showing human workmanship. In this study artifact means a model.
Artificial Intelligence	The creation of computer systems with capabilities of activities that often require human intelligence, such as visual perception, is what is referred to as artificial intelligence, or simply AI.
Artificial Neural Networks	An artificial neural network simulates the network of neurons or nodes that comprise the human brain, allowing the computer to learn and make decisions in the same manner humans make decisions. It is a method of programming ordinary computers to behave like interconnected brain cells.
Blue Economy	The use and preservation of the marine environment are considered to be part of the "blue economy" in terms of economics. It is typically used to represent a sustainable development approach to coastal resources in the context of international development.
Convolution Neural Networks	Convolutional Neural Network is an algorithm of Deep Learning method that receives input images, assigns relevance in terms of learnable weights and biases to distinct parts of the image, and may extract differences from those features (ConvNet or CNN).
Deep Learning	A subtype of artificial intelligence known as deep learning, comprises networks that can learn unsupervised from unstructured data or random data. Deep Learning is often referred to as Deep Neural Networking or Deep Learning.
Deep Neural Networks	Deep Neural Networks (DNN) consist of neurons, synapses, weights, biases, and functions as components just like neural network of the human brain. Between the input and output levels, these construct several layers, resulting in an artificial neural

network (ANN).

Ecosystem	An ecosystem means an environment within which a complex coexistence of a community of interacting organisms and their relationships in unit space.
Indian Ocean	The Indian Ocean boarders Africa on the eastern side, Australia to the western side and Asia on the southern side. It is among the largest of the five oceans in the world, with a total of 70,560,000 km^2 (19.8%) of water coverage on the surface of the Earth.
Machine Learning	An area of artificial intelligence is called machine learning (ML). Making machinery capable of mimicking intelligent human behavior is what this technique is all about. Machines with ML capabilities can carry out complex jobs almost as well as people can (expert.ai).
Macroplastics	Macroplastics are large particles of plastics ranging in different sizes, basically, more than 5 mm, which can be seen and picked easily. Macroplastics are in the marine environment and on land.
Microplastics	Microplastics (MPs) are small and tiny particles of big plastics not more than 5 mm in length. The breaking apart of the Macroplastics is attributable to natural causes like wind, sun, and other causes. Most of the products we use at home like facial scrubs, toothpastes, etc also contain tiny pieces of Microplastics, known as microbeads.
Open Sewer	The main sources of sewage waste water include the kitchen, dishwashers, washers, bathrooms, and toilets. Drains, pipes, and manholes are all part of the sewerage system, which transfers waste water to a sewage treatment facility.

Recurrent Neural Networks	Recurrent Neural Network (RNN) is a type of Artificial Neural Network and a Deep Learning technique. This approach is frequently used in object detection and classification on images. Recurrent neural networks recognize the sequential properties of data to forecast the next likely scenario.
TensorFlow	TensorFlow, an open-source platform for Machine Learning, accepts inputs as a complex structure of data called Tensor. It uses data flow plots to construct models, enabling designers to create large-scale neural networks with many layers.

ABSTRACT

Microplastics (MPs) are small plastic particles that pose a threat to aquatic organisms and human health. Detecting MPs in bodies of water is critical for controlling their flow and limiting their negative effects. This study proposes a Deep Learning algorithm for detecting MPs in photos taken from open sewer systems that flow into the ocean. The research adopted the Sample, Explore, Modify, Model, and Assess (SEMMA) framework, a comprehensive data mining process. A dataset comprising 1000 photos was constructed from locations in Kilifi, Mombasa, and Kwale counties in Kenya, by employing the Scale-Invariant Feature Transform (SIFT) algorithm for feature extraction. The researchers compared the performance of two object detection models: Sing-Shot Detector (SSD), and Convolutional Neural Network (CNN), and discovered that SSD performed best with a mean average precision (mAP) score of 100%, while CNN performed worst with 96.5%. A model for detecting MPs in photos taken from open sewer systems that flow into the Indian Ocean along Kenya's coast was developed using the best-performing SSD model. The model can be used to detect MPs in other open sewer systems, assisting in the implementation of effective management and control measures. Future research could look into creating a mobile app that captures images and provides information about MPs in open sewer systems.