CORRUGATED BOX DAMAGE PREDICTION USING ARTIFICIAL NEURAL NETWORK IMAGE TRAINING

By

Sarah Holland

A THESIS

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Packaging – Master of Science

2023

ABSTRACT

This thesis proposes a novel packaging evaluation method using corrugated box images and an Artificial Neural Network (ANN). An Artificial Neural Network works in a way similar to that of neurons in a human brain: by making connections between a trained dataset and the new data provided after training. The ANN has been implemented in the industry in various ways but limited in the packaging evaluation. This paper is focused on the corrugated box damage prediction using ANN. By capturing the damaged corrugated box images with an Artificial Neural Network, damaged products can be identified allowing a decision to be made as to what type of package failure occurred. One of the benefits to using an Artificial Neural Network to evaluate corrugated box images is that it allows for the evaluation of package protection in a real distribution environment as compared to a controlled lab setting. In turn, this reduces the cost of testing, as the package failure will have been identified with the assistance of the Artificial Neural Network, rather than full retesting to identify where damage occurred. This process would also reduce costs associated with the usage of materials for testing, due to the lower number of test samples required.

This thesis is dedicated to the people who supported me through my education, and those who aren't around anymore to celebrate with me. To my family, my friends, my amazing cats, and my incredible professors: thank you for all of the support and encouragement along the way.

ACKNOWLEDGEMENTS

I would like to acknowledge and give my warmest thanks to my Major Professor Dr. Euihark Lee, who made this work possible. His guidance and advice carried me throughout all the stages of my project. I would also like to thank my committee members for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

I would also like to give special thanks to my partner Timothy Roback who knows my research almost as well as I do, and my family as a whole for their continuous support and understanding throughout this journey. I wouldn't have been able to complete this thesis without all of your support, and for that I am so grateful.

Finally, I would like to acknowledge my roommate and lifelong friend Cole Pauley, and my best friend Morgan Graham, who provided countless words of support during the difficult parts of completing this degree. To my friends in the industry, your passion is inspiring. For that, and your friendship, I am thankful for each of you.

iv

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ABBREVIATIONS	ix
CHAPTER 1: INTRODUCTION	
1.1 Objective	
CHAPTER 2: BACKGROUND	5
2.1 E-Commerce	
2.2 Package Evaluation: Lab Testing	
2.3 Package Evaluation: Field Testing	
2.4 Package Evaluation: Computer Simulation	
2.5 Artificial Neural Networks (ANN)	
2.6 Research Goal	
CHAPTER 3: METHODS	
3.1 Data Collection	
3.2 Data Preparation	
3.3 ANN Modeling	
3.4 Verification	
CHAPTER 4: CASE STUDY	
4.1 Data Collection	
4.2 Data Preparation	
4.3 ANN Modeling	
4.4 Results and Verification	53
CHAPTER 5: CONCLUSION	
BIBLIOGRAPHY	64

LIST OF TABLES

Table 1. ISTA 3A Drop Sequence (ISTA, 2018)
Table 2. Methodology Drop Orientations
Table 3. Image Modifications on Television Images from E-Commerce Platform
Table 4. Exhaustive Search Method Example for Neurons per Hidden Layer 38
Table 5. Exhaustive Search Method Example for Activation Function and Solver. 39
Table 6. Corrugated RSC Sample Dimensions
Table 7. Number of Images in Each Damage Category. 46
Table 8. Image Modifications on Corrugated RSC Images. 47
Table 9. Number of Images per Damage Category for Testing and Training Split. 50
Table 10. Neurons per Layer High Saturation Cropped. 51
Table 11. Neurons per Hidden Layer in Each Image Modification Category 52
Table 12. Activation Function and Solver for High Saturation Cropped. 53
Table 13. Prediction Table Output from ANN. 56

LIST OF FIGURES

Figure 1. Package Optimization in terms of cost
Figure 2. Retail e-commerce sales worldwide from 2014 to 2024 (in billion US dollars) (International Trade Administration, 2023)
Figure 3. Touchpoints of brick-and-mortar and e-commerce distribution (Skyline University College, 2016)
Figure 4. Drop testing machine from Lansmont (Lansmont, 2023)
Figure 5. Vibration test system from Lansmont (Lansmont, 2023) 10
Figure 6. Compression tester from Lansmont (Lansmont, 2023) 12
Figure 7. Shock and vibration data logger from Lansmont (Lansmont, 2023) 14
Figure 8. Environmental monitoring tool from Sensolus. (Sensolus, 2023) 15
Figure 9. Comparison of movement of unit load components under deflection for experimental (left) and FEM (right) for two layers and (a) two columns, (b) three columns, and (c) four columns of packages (Molina, 2021)
Figure 10. Conceptual diagram of an ANN. (UpGrad, 2022)
Figure 11. Conceptual diagram of calculations made within an ANN (Obuchowski, 2020)
Figure 12. Commonly utilized activation functions: Identity (a), ReLu (b), Tanh (c), and Logistic (d) (Šegota, 2020)
Figure 13. Overview of methodology
Figure 14. Conceptual diagram of the Web Scraping process
Figure 15. Sample labeling example
Figure 16. Box Compression Test Example (Rycobel, 2023)
Figure 17. ANN Hidden Layers

Figure 18. Conceptual Confusion Matrix (Draelos, 2019)
Figure 19. Prediction Workflow Overview
Figure 20. Sample labeling example
Figure 21. Examples of images from the "High Contrast" modification category
Figure 22. Contour Plot for High Contrast Cropped Predictive Accuracies 51
Figure 23. Predictive Accuracies of Corrugated RSC Images in each modification category
Figure 24. Images from the Highest Predicted Categories: "Black and White Original" (A), "High Saturation Original" (B), "Low Contrast Cropped" (C), and "High Contrast Cropped" (D)
Figure 25. Confusion Matrices for Black and White "Original" (A) and "Cropped" (B)
Figure 26. Incorrectly predicted edge images from the category "Black and White Cropped"
Figure 27. Confusion matrices for the highest predicted categories: Black and White Original (A), High Saturation Original (B), High Saturation Cropped (C), and Low Contrast Cropped (D)
Figure 28. Images that could fit in either "Edge" or "Corner" prediction categories. Image A: Corner Drop, Image B: Edge Drop

LIST OF ABBREVIATIONS

- ANN Artificial Neural Network
- RSC Regular Slotted Container
- FEM Finite Element Method
- ReLu Rectified Linear Unit
- SGD Stochastic Gradient Descent
- LB Lower Bound
- UB Upper Bound
- NN Neural Network
- B&W Black and White

CHAPTER 1: INTRODUCTION

Packaging is used in every industry to ensure protection, containment, convenience, and communication to consumers and manufacturing professionals. Protection and containment refer to the product itself by ensuring that the product will arrive at its destination intact and in a damage-free and manageable way. Convenience and communication refer to the use of each package. Consumers must be able to utilize packaging in a convenient way while also having all pertinent information provided to them. There are three layers of packaging commonly used. These include primary packaging, which is the packaging in direct contact with the product; secondary packaging, which is used outside of primary packaging to group products together; and tertiary packaging, which is used by wholesalers for shipping products to their destination while avoiding damage. These layers of packaging play a fundamental role through the supply chain. All products undergo distribution, which can involve rough shipping environments. Ensuring that a product's package will meet each of the four functions of packaging and will assist in facilitating smooth distribution. To ensure this, packaging evaluation is used before distribution.

Packaging evaluation is very important for packaging cost and optimization. When considering the type of packaging material to be used, there are many costs, including the material costs and manufacturing costs, to consider. For example, a

1

package that is created with a heavier-weight corrugate will cost more to produce than one with a lighter-weight corrugate. This also impacts the optimization of the package. Figure 1 below shows a chart of damage cost vs. package cost. If a package cost is low but the damage is high, the product is under packaged. On the opposite end of this, if the package cost is high but the damage cost is low, the product is over-packaged. Optimization occurs between these when the damage cost and package cost are equal, or close to equal. Ensuring that a package is within this optimal zone will result in a successfully protected product through the supply chain.



Package Optimization

Figure 1. Package Optimization in terms of cost.

Packaging evaluation is an important step in any product development process to ensure that a product will survive the various stressors of distribution. The types of distribution include brick-and-mortar and e-commerce. Brick-andmortar distribution refers to distribution to a physical store where customers browse and make purchasing decisions in person, while e-commerce distribution refers to the distribution of products that are purchased online and shipped directly to the consumer. Stressors in distribution can include shock and vibration, among others. Various physical tests, including vibration, shock, and compression testing are conducted on new package designs to ensure protection before distribution.

Currently, the packaging evaluation process is heavily reliant on mechanical testing in a controlled lab setting. Studies have been conducted to evaluate the importance of physical testing on packages in this manner (Nygårds M, 2019; Fadiji, 2016). Other methods of evaluation include field testing where a package will undergo the physical stressors of distribution, and computer simulation where a package is simulated and various properties are tested without physical testing. Lab testing will ensure that a product is protected under controlled conditions, but it does not account for the true stressors of distribution. Field testing, while conducted in the true stressors of distribution, cannot account for the many variables that are encountered in real-world environments. Computer simulations can decrease the cost of physical testing, but their results are not always reliable. Another method that has not yet been widely used for packaging evaluation is

Artificial Neural Networks (ANN). This tool can be used for many applications but is used in this thesis to predict the cause of corrugated Regular Slotted Container (RSC) damage from images.

1.1 Objective

Package evaluation using ANN has been implemented in the packaging industry, but not as widely for the visual analysis of a package. This thesis aims to develop an ANN model that can predict the cause of package damage from images of corrugated regular slotted containers (RSCs). To do this, various key objectives were identified as follows:

- 1. Develop damage prediction model by implementing machine learning tools.
- 2. Build ANN modeling process for corrugated RSC damage prediction.
- 3. Find optimal image modification for ANN modeling.

CHAPTER 2: BACKGROUND

Packaging has been utilized since ancient times to ensure that product quality is maintained in their route to the consumer. In the ancient era of packaging, reed baskets, wineskins, wooden boxes, pottery vases, and more natural material containers were utilized. The first set-up boxes were used in the 16th century (Twede, 2005). Since ancient times and the 16th century, countless advances have been made to improve the functions of packaging in every industry. Materials have been modified to better fit the needs of distribution, new technologies have been introduced, and testing standards have been created and accepted by the industry. These standards ensure that a package meets the needs of the product contained while withstanding various stressors of distribution. Innovation is constant, though, meaning that the widely accepted ways of developing packages must continue to be developed in order to fit the needs of each company and consumer.

2.1 E-Commerce

E-commerce is the term for the buying and selling of goods and services over the internet. This tool has recently become a heavily utilized resource in the supply chain for a variety of products. This style of distribution relates directly to the convenience function of packaging, as consumers can obtain a higher level of convenience by not having to go in person to search for the product they want. There are several advantages to utilizing e-commerce, including its global reach. This distribution style has been even further utilized since the start of the COVID- 19 global pandemic, as consumers faced closings of brick-and-mortar stores and a higher likelihood of avoiding public spaces (Jílková, 2021). The quarterly share of total U.S. e-commerce retail sales has grown from 9.8% in the second quarter of 2018 to almost 15% of total sales in the third quarter of 2022 (Coppola, 2022). Research by the International Trade Admission suggests that the global growth of e-commerce will continue, reaching a point of approximately \$7,000 billion in sales by 2024 (International Trade Administration, 2023). A chart for this growth is shown below in Figure 2.



Figure 2. Retail e-commerce sales worldwide from 2014 to 2024 (in billion US dollars) (International Trade Administration, 2023).

This method of order fulfillment endures greater stressors throughout the supply chain than brick-and-mortar because of the tougher environment and hazards that packages encounter. In brick-and-mortar distribution, there are a minimum of 4 handling points compared to a minimum of 11 handling points in ecommerce distribution (Skyline University College, 2016). Figure 3 below shows an example of the touchpoints for each of these distribution types.



Figure 3. Touchpoints of brick-and-mortar and e-commerce distribution (Skyline University College, 2016).

A greater number of touchpoints means there are more opportunities for damage to occur through e-commerce distribution. Product deformations as a result of shock through e-commerce were observed at a rate of 19.3% in a study conducted by Spruit et. al (Spruit, 2021). Additionally, product deformations as a result of impact, shock, and static load through e-commerce were observed at a rate of 10.5% (Spruit, 2021). These damages can result in a higher cost of replacing products, especially when considering that a new product will have to endure the same supply chain in its route to the consumer. Additionally, when damage is observed a package must undergo retesting to ensure that future damage is avoided. This

increases costs due to the number of samples required while testing, as well as the labor associated with these tests. As shown, it is very important that packaging is developed with protection through e-commerce distribution environments in mind. This can be achieved through package development and evaluation with a goal of ensuring the four functions of packaging.

2.2 Package Evaluation: Lab Testing

There are three main approaches to ensure the packaging functions: lab testing, field study, and computer simulation. The lab test conducted in a controlled laboratory environment follows standards that have been widely accepted in the packaging industry. These standards specify the exact methods for evaluating packages for various purposes. For example, ISTA 6-Amazon.com-SIOC is the standard test method for products that are meant to ship in their primary packaging through the Amazon e-commerce platform (ISTA, 2016). There are many standards for different distribution environments and each of these standards specifies the number of samples that should be used to ensure thorough results (ASTM International, 2016; ASTM International, 2022). These standards have been set up for all types of physical testing, including drop testing, vibration testing, and compression testing. A common standard for conducting drop testing is the ISTA 3A standard for Packaged-Products for Parcel Delivery System Shipment (ISTA, 2018). In this procedure, a series of drops and vibrations are conducted on a single

container using a drop tester (Figure 4) and vibration table. For this example, only the drop portions will be explained.



Figure 4. Drop testing machine from Lansmont (Lansmont, 2023).

Packages are loaded with the product including all primary packaging. Then, the package undergoes a series of drops from heights ranging from 18-36". The package is positioned for these drops so that it encounters impact on various faces, corners, and edges of the corrugated case to ensure that a thorough evaluation of damage can be performed.

A commonly followed standard for vibration testing is ASTM D4728, the Standard Test Method for Random Vibration Testing of Shipping Containers (ASTM International, 2022). In this procedure, packages undergo a series of random vibrations on a vibration (or shaker) tester to determine how they should perform through distribution. Figure 5 shows an example of what this vibration tester looks like.



Figure 5. Vibration test system from Lansmont (Lansmont, 2023).

Packages are loaded with the product and all associated primary packaging. The package is placed in either a horizontal or vertical orientation related to the direction of corrugate fluting, and supports are positioned to ensure that the package will not vibrate off of the table while still allowing space for movement. Then, random vibrations are applied for a predetermined amount of time which is specific to each individual or group performing the test. In this standard, the vibration rates and times can be adjusted to meet the needs of a specific distribution environment. For example, if a package must travel for 2 hours to meet its destination and the shipping route is known, the vibration test can be conducted for 2 hours under vibration rates similar to those on its route. Many additional types of vibration testing can be conducted. These include sine sweep testing, which involves subjecting a package to a vibration that gradually increases over time; fixed frequency testing, which involves subjecting a package to a constant level of vibration; and resonance search testing, which involves identifying the resonant frequency of a package and subjecting the package to vibrations around that frequency. Additionally, random vibration and sine sweep vibration testing can be combined for sine on random testing. In this method, a package is subjected to random vibrations that coincide with a single frequency of vibration. This test allows the professional to identify how a package will perform both at random vibrations and a set vibration over time.

Compression testing is commonly conducted following ASTM D642-20, the Standard Test Method for Determining Compressive Resistance of Shipping Containers, Components, and Unit Loads (ASTM International, 2020). In this testing method, packages are loaded with the product and any associated primary packaging and placed on a compression tester (Figure 6).



Figure 6. Compression tester from Lansmont (Lansmont, 2023).

The package is placed either horizontally or vertically relative to the alignment of corrugate fluting and undergoes one compression. The test stops in one of two ways: once the compression reaches a predetermined point, or once a point of failure has been achieved. Conducting this type of test can assist in determining the strength of a container when stacked for long periods of time.

Retesting that must occur if large amounts of damages are reported can become very expensive and time-consuming as a result of these standards. While the widely accepted standards for lab testing are thorough, they are not without drawbacks. A study by Frank et.al discusses the limitations of compression testing in-lab on corrugate boxes when compared to their real-world environments and found that lab testing alone is not enough to measure the true functionality of a package through the supply chain (Frank, 2010). It can be assumed that this holds true when comparing brick-and-mortar to e-commerce distribution. To account for this, many packaging professionals utilize field testing for further evaluation.

2.3 Package Evaluation: Field Testing

Field research on package design can show the truest functionality of a package as it moves through the supply chain. This method is commonly used after lab testing has been completed. In this evaluation method, products are packaged in the way that they are expected to enter the supply chain. They are then sent out with either a traditional shipping source, like USPS or UPS, or with company owned distribution tools. One benefit of conducting field testing with company owned tools is that the test can provide a view of how a package will perform in a more controlled environment than that of a typical distribution service. However, field testing is not always the closest view to actual distribution. When utilizing this method, all factors of distribution are set and controlled by the package producer. These factors are not always consistent with how distribution will actually occur, resulting in a higher chance for damage after validation testing. This is where it may be of benefit to use a traditional shipping source like USPS or UPS. By sending a package out into an uncontrolled distribution environment, packaging professionals can evaluate how a package will truly perform. Information can be

gathered from field testing in numerous ways, including vibration and shock records with data logging tools, and physical damage depending on the field test administered. An example of a shock and vibration data logger from Lansmont is shown in Figure 7 below.



Figure 7. Shock and vibration data logger from Lansmont (Lansmont, 2023). With this tool, shock and vibration data from a distribution route can be recorded and saved. This can be used in conjunction with lab testing to better simulate the environment that a package will undergo.

Certain products may have attributes that cannot exceed a set level of humidity or temperature. To ensure that a package is meeting the protection needs of a product like this, environmental monitoring tools can be utilized in conjunction with field testing. Figure 8 (below) shows an example of one of these tools that can be implemented in field studies.



Figure 8. Environmental monitoring tool from Sensolus. (Sensolus, 2023). A tool like the one shown in Figure 8 can store data about humidity, temperature, contact, and orientation. This information can be extracted with the specific route that was taken, allowing for evaluation of package needs for a product meant to follow that same route. Data collected from tools like those shown in Figures 7 and 8 can assist in optimizing packaging by ensuring that the cost of the package is viable when considering the type of product to be packaged. This is especially important when considering medical devices and electronics. The cost of each product is high, meaning the packaging must be of high enough quality to ensure

that losses are not observed in the form of product damage as a result of insufficient packaging.

Many researchers conducted their research through field study (Dunno, 2017; Jung, 2012) since studies like these are the closest possible test to a realworld application. A study by Zhong et.al focuses on corrugate boxes through shipping conditions in China and found that a majority of packages aren't placed in the correct position during shipping, resulting in a greater number of package drops (Zhong, 2016). The placement of a package through distribution cannot always be controlled, so adjusting packages to meet protection needs in uncertain environments can assist in ensuring package success. Conducting testing in the field quickly becomes costly and time consuming due to the number of resources required, especially when considering the cost of the product contained. Moreover, these studies cannot measure and evaluate the functionality and protection of the package due to the lack of consistency in shipping conditions. These environments can change as shipping routes change, as well as when the environment changes the conditions of a route naturally.

2.4 Package Evaluation: Computer Simulation

An alternative to lab and field testing is the use of computer models. Computer models have become a popular method for simulating and evaluating packaging test methods. There are many kinds of computer simulations, including computational fluid dynamics, which can be used to model the flow of air, liquid, or gas inside a package; moldflow analysis, which can be used to predict how a plastic package will be molded during manufacturing; drop and impact simulation, which can predict the damage on a package when dropped; and thermal analysis, which can be used to evaluate the temperature distribution inside a package through various environments. Another common form of computer simulation is the finite element method (FEM). Many studies have been conducted simulating various stages of the supply chain that a package must endure utilizing FEM, which is a computer modeling technique that can be used to simulate the stresses and strains placed on a package during distribution and storage. This tool has been used in previous research to simulate the behavior of multiple materials to evaluate their performance in various applications (Rowson, 2008; Mills, 2005; Hallbäck, 2014; Huang T. C., 2022). This technique can help to identify areas of a package that are most vulnerable to damage, allowing for a packaging professional to evaluate and adjust the package before production. A study by Molina et.al. details a frictiondriven FEM model that simulates the load bridging effect of unit loads stored in warehouse racks (Molina, 2021). Figure 9 shows a comparison of movement of unit load components under deflection for in-lab experiments (left) and with FEM (right). In this figure, FEM images are labeled based on deflection, with red meaning large displacement and blue meaning no movement. As shown with the figure, utilizing FEM to simulate the behaviors of packages in various environments can provide insightful data for how pallet loads or packages must be adapted to survive different conditions.



Figure 9. Comparison of movement of unit load components under deflection for experimental (left) and FEM (right) for two layers and (a) two columns, (b) three columns, and (c) four columns of packages (Molina, 2021).

A study by Biancolini et.al focuses on the buckling strength of corrugated boxes with FEM and found that the model provides negligible error when compared to a lab-tested value under the same conditions (Biancolini, 2003). Based on these results, adopting this approach would imply that the monetary costs associated with physical testing can be cut back. While the benefits of utilizing FEM have been observed, this process is not perfect. Materials have many properties that can be changed as their chemical composition is altered. Accounting for each of these properties involves a large variety of simulations to be performed. This can increase the computational cost, and in many cases, the results are not reliable due to the simulations being based on estimations of real-world conditions. Because of this, a new method is needed to evaluate packages that endure the stressors of e-commerce to ensure customer satisfaction.

2.5 Artificial Neural Networks (ANN)

Machine learning (ML), a subset of Artificial Intelligence, is a field of computer science that studies algorithms and techniques for automating solutions to complex problems that are difficult to program using conventional programming methods (Rebala, 2019). There are many types of machine learning algorithms, each with varying levels of supervision required by the user. The first of these is supervised learning, where the data has already been labeled or classified with the correct output. Unsupervised learning involves an algorithm that is trained on unlabeled data. Semi-supervised learning contains a combination of labeled and unlabeled data. Finally, reinforcement learning involves an algorithm that learns from positive and negative feedback for certain actions. Within ML exists Artificial Neural Networks (ANN). ANNs can be utilized to make assumptions and predictions about various datasets and work in a way similar to that of a human brain, by making connections between characteristics of data points and drawing conclusions. ANNs consist of an input layer which receives data, one or more hidden layers where neurons are contained and make connections within the data, and an output layer where a result is provided. A conceptual diagram of this is shown in Figure 10.



Figure 10. Conceptual diagram of an ANN. (UpGrad, 2022).

Each neuron in the hidden layer(s) applies mathematical functions to the data and passes the results of these functions to other neurons following the flow of the network between layers. These functions are a product of a neuron's value,

as determined by an activation function, multiplied by the weight of each neuron which is determined by the solver. All values are added to provide the output data. Figure 11 shows a conceptual diagram of the calculations made between neurons, as well as the equation used for providing an output.

Where: x: numerical value of the neuron from activation function w: weight of each connection from solver h: output value



$$h_{1}^{1} = f(w_{1}x_{1} + w_{2}x_{2} + w_{3}x_{3} + \dots + w_{n}x_{n} + b)$$
$$= f\left(\sum_{i=0}^{n} w_{i}x_{i} + b\right)$$

Figure 11. Conceptual diagram of calculations made within an ANN (Obuchowski, 2020).

Activation functions and solvers are contained between each neuron in the ANN. Activation functions introduce nonlinearity into the output of a neuron, allowing for the ANN to model complex nonlinear relationships between input and output data. Commonly used activation functions include identity, logistic (sigmoid), tanh (hyperbolic tangent), and rectified linear unit (ReLu). Graphs for each of these activation functions can be found below in Figure 12.



Figure 12. Commonly utilized activation functions: Identity (a), ReLu (b), Tanh (c), and Logistic (d) (Šegota, 2020).

Drawing on the state of research to date, Dubey et.al. posits the following explanations for various activation functions. Identity is a linear activation function where the output of the function is equal to the input. Logistic, or sigmoid, maps input to a value between 0 and 1 and is commonly used in probability models. Tanh is a hyperbolic function that maps the value of an input between -1 and 1. ReLu is a piecewise linear function that returns the value of an input if it is positive, and a 0 if it is not (Dubey, 2022). Solvers are used to optimize the weights between each neuron. Three commonly used solvers include stochastic gradient descent (SGD), L-BFGS-B, and Adam. SGD is an optimization algorithm that updates the weights and biases of the network based on the loss function with respect to each parameter (Bottou, 2012). L-BFGS-B is a quasi-Newton optimization algorithm that approximates the Hessian matrix of the loss function to update weights and biases within the network (Liu, 1989). Adam is a variant of SGD that adapts the learning rate of each weight based on the first and second moments of the gradient (Brownlee, 2017). Each activation function and solver provides its own benefits depending on the application. Benefits of ANN as a whole include optimization, predictive accuracy, and time savings. Applications of this tool can be found in various industries, including marketing (Bloom, 2005). ANNs can also be utilized for image-based evaluations. Kalyan et. al. details a method for diagnosing disease conditions with the use of ANN (Kaylan, 2014). While the functionality of ANN has been studied extensively, this application has not yet been widely brought to the packaging industry. There have been instances of this tool being utilized in terms of packaging analysis (Esfahanian, 2022; Xie, 2023; de Abajo, 2004), but little has been done in terms of visual analysis of a package with images. In this thesis, a method for utilizing the image analysis capabilities of ANN for corrugated RSC evaluation is introduced.

2.6 Research Goal

The goal of this research is to predict the cause of box damage using corrugated RSC images with an artificial neural network (ANN). It is important to know how package failure has occurred in an effort to improve protection throughout the supply chain. Knowing the cause of damage allows professionals to improve that area without conducting a full investigation, which involves total retesting of the package in a controlled lab environment. If the package professional utilizes field testing or simulations, largescale damage reports would also result in an additional round of these tests and simulations to be performed, increasing physical and computational costs further. By utilizing ANN for package evaluation, it will be possible for packages to be evaluated in a real-world context, as opposed to in-lab controlled testing. Additionally, the costs associated with testing will be reduced due to the smaller number of samples required for post-production evaluation. The ANN is able to predict what the cause of damage is for each corrugated RSC, meaning a full retesting process is not necessary when damage is reported. Industry professionals will be able to modify current packaging without identifying the issue manually.

CHAPTER 3: METHODS

This chapter introduces a novel methodology of corrugated box evaluation processing. Instead of using typical packaging test standard or numerical simulation, image training with ANN was implemented as a main tool for corrugated box evaluation. With this tool, the type of damage experienced by a corrugated RSC can be predicted, allowing for a more streamlined evaluation process. The packaging evaluation process using corrugated RSC images was composed of 4 main processes: (i) data collection, (ii) data preparation, (iii) ANN model development, and (iv) verification. Each process is important to ensure optimized predictive accuracies with the ANN. Figure 13 shows an overview of the methodology followed. The process begins with data collection, which involves gathering images from previous validation testing, web scraping, or creating new images in-lab. These images are then modified in various ways including cropping to remove the background and modifying various color properties to fit into 12 different categories. Data is split into 70% for training and 30% for testing before the ANN is modeled, which involves determining the number of hidden layers, the number of neurons in each hidden layer, and the activation function and solver for the model. Finally, predictions are verified using manual verification and confusion matrix evaluation to ensure that all predictions can be considered valid. The following sections explain the detailed process of each step.



Figure 13. Overview of methodology.

3.1 Data Collection

In the data collection portion of this method, there are three ways that images can be gathered. The first of these is by utilizing images previously taken in the validation stage of package development. Utilizing these images would cut down on the cost of new samples required when creating images, but they aren't always created in a way that highlights one form of damage.

When validation testing, industry standards like ISTA 3A allow for multiple tests to be performed on a single sample (ISTA, 2018). While this test is appropriate for validation of the package, it is not designed to highlight a single type of damage to the container. The second approach is by utilizing a web scraping tool that gathers information in various forms from a website. The forms of data that can be extracted in web scraping include product prices, consumer reviews, images, and more. Figure 14 (below) shows a diagram of the web scraping process.


Figure 14. Conceptual diagram of the Web Scraping process.

To follow the process of web scraping, unstructured data in the form of a link to a website is input to the web scraping tool. This tool must be programmed to extract data that fits the needs of the user. Structured data can be output in many forms, including images. In this method, images that are gathered must be labeled manually based on what it looks like the damage could be. Consumers are not always able to identify the specific kind of damage that their product has experienced and typically will not mention the damage type in reviews. Packaging professionals may be able to identify the possible cause of damage, but it may not always be the true cause. When utilizing web scraping, it is important to recognize that the legality and ethics of the tool have not yet been fully defined (Krotov, 2018). There is opportunity for this tool to extract data that can pose security risks to consumers. Additionally, web scraping can result in the gathering of data that is licensed to a corporation. If utilizing this method, care should be taken to ensure that all data gathered is done so in an ethical and legal way.

The third approach is creating new images via testing. Utilizing this method allows for specific kinds of damage to be tested and documented which ensures that the image label will be true to the kind of damage shown. For this method specifically, impact (drop) testing and compression testing are utilized. Vibration testing was not considered for this method due to the minimal visual damages observed through testing. To conduct impact testing, the ISTA 3A standard for packaged products for parcel delivery system shipment was modified and followed. In this standard, a series of drops and vibrations are conducted at various heights and vibration levels, focusing on different orientations of the package being tested (ISTA, 2018). For the purposes of this research, vibration portions of ISTA 3A were not followed. Table 1 below shows an overview of the drop sequence associated with the ISTA 3A test. Figure 15 below shows a sample labeling example followed in this testing. When following the drop sequence, the corrugated RSC is oriented so that the drop occurs on the edge or corner where the numbered faces meet. For example, if the drop orientation states "Edge 4-6," that would correspond to the edge between faces 4 and 6 on the RSC.

Drop #	Samples <70 lbs. (32 kg)	Orientation of Drop
1	18 in (460 mm)	Edge 3-4
2	18 in (460 mm)	Edge 3-6
3	18 in (460 mm)	Edge 4-6
4	18 in (460 mm)	Corner 3-4-6
5	18 in (460 mm)	Corner 2-3-5
6	18 in (460 mm)	Edge 2-3
7	18 in (460 mm)	Edge 1-2
8	36 in (910 mm)	Face 3
9	18 in (460 mm)	Face 3

Table 1. ISTA 3A Drop Sequence (ISTA, 2018).



Figure 15. Sample labeling example.

Face drops were not included due to the minimal visual damage observed through testing. Each drop orientation was conducted on a new sample from a height of 36" to ensure that the true extent of one type of damage was observed. These drops were also expanded to include each corner and edge orientation possible for the corrugated RSC. In total, 14 drop orientations were included in the drop testing process. Table 2 below shows the drop orientations for this methodology in detail.

Orientation # (new sample each time)	50 lbs. (22.7 kg) loaded sample drop height	Orientation of drop	
1	36 in (910 mm)	Corner 3-4-5	
2	36 in (910 mm)	Corner 1-2-6	
3	36 in (910 mm)	Corner 1-2-5	
4	36 in (910 mm)	Corner 1-4-5	
5	36 in (910 mm)	Corner 1-4-6	
6	36 in (910 mm)	Corner 2-3-5	
7	36 in (910 mm)	Corner 2-3-6	
8	36 in (910 mm)	Corner 3-4-6	
9	36 in (910 mm)	Edge 1-2	
10	36 in (910 mm)	Edge 4-5	
11	36 in (910 mm)	Edge 4-6	
12	36 in (910 mm)	Edge 3-6	
13	36 in (910 mm)	Edge 1-5	
14	36 in (910 mm)	Edge 3-4	

Table 2. Methodology Drop Orientations.

Prior to each drop, a load of 50 lbs. (22.7 kg) was inserted into the case. While this load may be greater than the expected weight of some products, it will ensure that damage is shown for each drop administered. Images were taken from multiple angles after the drop to capture the full extent of this damage.

To conduct compression testing, ASTM D642-20 was modified and followed (ASTM International, 2020). No weight was added to the case for this test. Samples were loaded into a compression tester in either a horizontal or vertical orientation in relation to the direction of the corrugate fluting with a preload of 50 lbs. and compression was applied to the case at a rate of 0.5 in/min until a yield of 50% from the maximum point of compressive load was achieved. This level of yield may exceed the expected compression strength of many packages in use, but it will ensure that the fullest extent of possible damage is shown. Photos were again taken after each test from multiple angles to capture the full extent of damage. Figure 16 shows an example of what this testing looks like in the lab with a vertical case orientation. From here, collected data was prepared for modeling the ANN.



Figure 16. Box Compression Test Example (Rycobel, 2023).

3.2 Data Preparation

In the data preparation process, images were sorted into categories respective of their damage and modified. This sorting step is important in order to provide the best chance at an accurate damage prediction by the ANN. Modifications are important to ensure that the ANN can capture all features associated with the damage observed. First, images were sorted into a category that corresponds to the type of damage experienced. All drop images were split into "Edge" or "Corner," while all compression images were placed in a category labeled "Compression." Labeling and sorting these images in this manner allows for simple data validation as well as organization before image modification. When modifying, copies of each sorted damage type were made and placed into another

category that specifies the type of image modification done. All modifications were done using the computer's included image modification program. The first modification included was to crop each image manually. This was done to determine how the background of each image impacts prediction results. By conducting this process manually, it can be ensured that the minimum amount of background is shown in each image. Once cropping was completed, five additional categories of image modifications were performed. These modifications included black and white, high contrast, low contrast, low exposure, and high saturation. When utilizing the image modification platform included on the computer, modification sliders for these modifications were moved to the most extreme version of each category. All categories of image modifications are important to include because it is not known how the ANN sees and evaluates images. The ANN draws information from each pixel associated with the image, making it difficult to determine which modification would be best for the prediction. Including and testing each modification ensures that at least one modification category will provide an accurate prediction. The modification category that predicts most accurately will vary between sample properties, meaning this process will need to be followed for each application. The following Table 3 shows an example of each modification performed on images using a television image gathered from an ecommerce platform. From here, the ANN can be modeled for use.

	Original	Cropped
Original		
Black and White		
High Contrast		2- Co
Low Contrast	3	2-3
High Saturation		
Low Exposure		

Table 3. Image Modifications on Television Images from E-Commerce Platform.

3.3 ANN Modeling

Modeling the ANN involves splitting data, determining the neuron structure, and determining the activation functions and solvers. Data were split into 70% for training and 30% for testing. Furthermore, each category of damage was split with the same 70/30 setup to ensure that examples of each category are included in both the testing and training portions. This was done manually to ensure that each image modification category contained the same data split, providing a true view of how the ANN predicts damage. The next step of ANN modeling was determining neuron structure, which includes the number of hidden layers within the network and the number of neurons in each hidden layer. Figure 17 below shows an example of how hidden layers are set up within the ANN.



Figure 17. ANN Hidden Layers.

Following Figure 17, images are input to the first neuron on the left. These images are embedded, and numerical values are extracted for each pixel associated with the image that are pushed to the first hidden layer. Within this first layer, activation functions are used to determine if the neuron should be activated for evaluation or not. Then, a solver determines the weight associated with each neuron. Values from the activation function and solver are multiplied to provide the value for the next hidden layer connection point. This step repeats in the second hidden layer, until all values are summed, and a prediction is output in the final neuron on the right. This prediction is provided in the form of a damage category.

The number of hidden layers that should be used has been researched extensively and it has been found that two layers are sufficient for many applications of an ANN since the potential number of neurons in each layer can be large (Hecht-Nielsen, 1987; Kůrková, 1992; Huang G. B., 2003). Limiting the number of hidden layers to two will allow for the prediction to be made with high accuracy while limiting the computational cost of determining neuron numbers per layer. Determining the number of neurons in each hidden layer was done with an exhaustive search method which tests the predictive accuracy of the ANN at each increment of neurons per hidden layer. An example of this process is shown below in Table 4.

NN in 1 st /2 nd Layer	LB_1	LB1+A	$LB_1+2\Delta$			$LB_1+n\Delta$	UB_1
LB ₂	0.881	0.795	0.863	0.885	0.795	0.843	0.874
$LB_2+\Delta$	0.787	0.642	0.899	0.867	0.885	0.778	0.850
$LB_2+2\Delta$	0.849	0.733	0.881	0.831	0.776	0.687	0.841
•	0.871	0.873	0.798	0.698	0.851	0.881	0.735
•	0.777	0.647	0.805	0.805	0.795	0.770	0.815
$LB_2 + n\Delta$	0.825	0.842	0.756	0.823	0.801	0.795	0.787
UB ₂	0.632	0.881	0.793	0.776	0.884	0.856	0.739

Table 4. Exhaustive Search Method Example for Neurons per Hidden Layer.

KEY:

 Δ : Increment NN: Neuron Number

LB: Lower Bound UB: Upper Bound

When utilizing this method, the upper and lower bounds of possible neuron numbers must first be determined. The minimum and maximum bounds were determined using the rule of thumb method which states that the number of neurons per layer should be within the range of the number of datapoints available (Karsoliya, 2012). Increments for the exhaustive search were decided according to the minimum/maximum range. The combination to be utilized was then selected. This method is being used to ensure that the neuron combination does not fall into the local minimum. While this method can have high computational costs, it will ensure that maximum predictive accuracy is obtained.

The final step of ANN modeling was to determine the activation function and solver to be used. The available activation functions were identity, logistic, tanh, and ReLu. Activation functions decide whether or not a neuron should be activated for prediction while applying a mathematical function to the data. The available solvers include L-BFGS-B, SGD, and Adam. Solvers are used to optimize the parameters used in predictions by applying weights to each data value. The combination to be used was determined using the same exhaustive search method. Combinations of each activation function and solver were tested on the ANN and the predictive accuracy was recorded. Then, the combination with the highest predictive accuracy was selected. The combination selected for this method is tanh, Adam. An example of this exhaustive search method is shown below in Table 5.

	Activation Functions								
L		Identity	Logistic	Tanh	ReLu				
Solve	L-BFGS-B	0.774	0.742	0.865	0.752				
	SGD	0.612	0.467	0.805	0.854				
	Adam	0.826	0.889	0.899	0.827				

Table 5. Exhaustive Search Method Example for Activation Function and Solver.

3.4 Verification

The final step in this methodology was to verify that the results obtained are accurate to the true labels of each image. This was done utilizing two methods: manual verification and confusion matrix evaluation. When verifying manually, image prediction labels were compared to the true label of their respective image. By comparing each photo manually, the user can identify which image is not being predicted correctly and identify any possible cause of the incorrect prediction. Additionally, the ANN's confidence rate in each prediction can be calculated as a result. A strong confidence rate for an incorrect prediction could mean that there is an issue with the data, allowing the user to identify this. The second method of verification is utilizing confusion matrices. These show the breakdown of predicted categories vs. the actual categories of data, which can also assist in determining which images are not being predicted correctly. An example of a confusion matrix is shown below in Figure 18.

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 18. Conceptual Confusion Matrix (Draelos, 2019).

If a predicted positive datapoint is actually negative, the confusion matrix would show this datapoint as a "false positive." Evaluating these matrices can also assist in determining which image modification category will work best for the damage the user is looking to predict. For example, if one image modification category is predicting at a higher rate than others for the category "corner," it may be best to focus on that image modification for future corner damage predictions.

Figure 19 below shows an overview of the prediction workflow using the Orange Data Mining software with Image Analytics add on (Orange Data Mining Platform, 2023). As a reminder, Training data consists of 70% of the data while testing data consists of the remaining 30%. Following this figure, training images are input to the widget labeled "Training Images." These images are embedded to extract numerical data which is pushed into the ANN widget to train the model. The trained ANN model is connected to the prediction widget. Then, testing images are uploaded to the "Testing Images" widget. These images are also embedded, and numerical data for each is pushed to the prediction widget as data. Predictions are made with the ANN's trained dataset on the testing images, providing a table of prediction results with the confidence rates, predicted damage category, and error rates. A confusion matrix widget is attached to the prediction widget which can show the overview of how images were predicted, both correct and incorrect data. The image viewer widget connected to the confusion matrix widget can show selected categories from the confusion matrix, allowing for simple verification of the images predicted.



Figure 19. Prediction Workflow Overview.

The testing dataset, training dataset, and ANN widgets are also connected to a "Test and Score" widget. This test and score widget provides the predictive accuracy without a breakdown of each image, as well as the option to split data randomly to evaluate how the ANN would perform without manual splitting of data. This widget is helpful when determining how data should be split to provide the most accurate predictions for each case, as well as when determining the neuron structure as the breakdown of each predicted image is not necessary during that step. The process of this ANN can be summarized as consisting of four steps: data collection, data preparation, ANN modeling, and result verification.

CHAPTER 4: CASE STUDY

This case study was conducted utilizing corrugated regular slotted containers (RSC). Images were gathered by conducting new testing in-lab with a focus on one type of damage for every sample, ensuring that each image label was accurate to the type of damage experienced. From there, images were cropped and various color properties were modified, resulting in 12 image categories for predictions. Results were obtained, and verification was conducted with two methods: manual verification and confusion matrix evaluation to ensure that all predictions were valid to consider. The following sections explain this process in detail.

4.1 Data Collection

In this data collection process, images were created from in-lab testing on corrugated RSCs, which were selected given that they are one of the most common shipping containers. Creating data in a controlled environment allows for accurate labels of the damage that occurred, as well as a true example of each category of damage. During this process, a total of 69 corrugated RSCs underwent compression and impact testing. During impact testing, 43 samples were used in various sizes shown in Table 6 below. These samples were labeled in uniform fashion before testing following the example in Figure 20 to ensure consistency across samples. Corrugated RSCs were loaded with 50 lbs. of weight and dropped from a height of 36" one time per sample, following a modified ISTA 3A test detailed in the

methodology section. The drops were structured to ensure that one sample would show damage from one component of the case: corner or edge. Additionally, each of the 12 edges and 8 corners of a case were represented in the data. Images were taken after the drop from multiple angles to ensure that all effects of damage are shown. During the compression testing phase, 26 samples were used in various sizes shown in Table 6.



Figure 20. Sample labeling example.

Corrugated RSC Dimensions	Impact Testing Samples	Number of Images	Corrugated RSC Dimensions	Compression Testing Samples	Number of Images
12"x12"x12"	5	17	12"x8"x8"	1	3
14"x10"x8"	1	3	12"x10"x8"	1	3
14"x10"x12"	9	25	14"x6"x8"	1	2
14"x12"x12"	5	15	14"x8"x8" 2		6
14"x14"x8"	1	3	14"x10"x8"	1	3
14"x14"12"	2	6	14"x14"x8" 6		18
16"x8"x8"	1	3	16"x8"x8" 1		3
16"x12"x10"	1	3	16"x14"x8"	8	24
16"x14"x8"	1	3	16"x16"x8"	5	15
16"x14"x12"	6	18	-	-	-
16"x16"x12"	11	34	-	-	-

Table 6. Corrugated RSC Sample Dimensions.

The preload associated with compression testing was 50 lbs. Compression stopped once the cases reached a yield of 50% from the maximum point of compression. This stopping point was set in the machinery but can also be determined by evaluating the graph provided during testing. After the compression load was applied, images were taken from multiple angles to ensure that all effects of compression were recorded. From here, the collected data was prepared before the ANN was modeled.

4.2 Data Preparation

In the data preparation process, images were first labeled to represent the type of damage experienced. All drop images were labeled either "Edge" or "Corner", while all compression images were labeled into "Compression." Table 7 below shows a distribution breakdown of how many images were sorted into each category.

Damage Category	# of Images
Edge	70
Corner	62
Compression	77
Total:	209

Table 7. Number of Images in Each Damage Category.

Labeling images in this manner allows for simple data verification as well as organization before image modification. When modifying, copies of each label were created and placed into another category that specifies the type of image modification done. All modifications were performed using the computer's included image modification program. Images were first manually cropped to determine how the background will impact prediction results. Once this was completed, five additional categories of modifications were performed on both the cropped and original versions of images. These include black and white, high contrast, low contrast, low exposure, and high saturation. Including every modification shown will ensure that at least one of the workflows will capture details needed for the ANN to make a prediction. Table 8 below shows an example of each modification performed on the lab created images.

	Original	Cropped
Original		
Black and White		
High Contrast		
Low Contrast		
High Saturation		
Low Exposure		

Table 8. Image Modifications on Corrugated RSC Images.

Images were modified to the most extreme version of each category. In some cases of image modification, it appeared as though all features that could assist in predicting the cause of damage were muted. Figure 21 (below) shows an example of three images from the "High Contrast" category of image modification. In this modification, portions of the corrugated RSC can appear to be "blacked out." This can be beneficial for the ANN evaluation as the impacted areas become highlighted. Image "A" of this figure shows one corner of the corrugated RSC highlighted at the location of damage from the drop. Similarly in image "B," a majority of the frontmost face of the RSC is muted with the exception of the damaged corner. In image "C," the muted portions of the RSC bring more attention to the bright face of the box as well as the split on the corner. These examples show that while an image modification may not seem like the best representation, there are still useful details that can assist in predicting the damage type.



Figure 21. Examples of images from the "High Contrast" modification category.

4.3 ANN Modeling

To begin modeling the ANN for use, data was first split into training and testing portions. In total, 167 training images and 42 testing images were used. When splitting the data, each drop and compression category was split into 70% for training and 30% for testing to ensure that each damage category had images in both 'training' and 'testing' portions. These data splits were performed manually to guarantee that each remained the same across each image modification category. This ensured that each accuracy was a consistent view of how the ANN made predictions on the images provided. Table 9 below shows a breakdown of how each damage category was split between training and testing.

Domogo Cotogomy	# of Images			
Damage Category	Training	Testing		
Edge	56	14		
Corner	50	12		
Compression	61	16		
Total:	167	42		

Table 9. Number of Images per Damage Category for Testing and Training Split.

The next step of this modeling was determining neuron structure, including the number of hidden layers and number of neurons per hidden layer. This was done using an exhaustive search method with a minimum bound of 30 and a maximum bound of 100. The increment for testing combinations was chosen to be 10 and all combinations were tested and evaluated for the local maximums. Table 10 below shows an example of this exhaustive search for the category "High Saturation Cropped." As shown in this table, numerous occurrences of the local maximum were observed, which indicate the highest predictive accuracies of the model. For this research, the combination (40,30) was used as it was the first instance of this maximum in the "High Saturation Cropped" category. Figure 22 shows a contour plot of the exhaustive search for this category. The same exhaustive search method was performed across all image modification categories, providing a different number of neurons in each hidden layer for every category. A summary of the results of these exhaustive searches can be found in Table 11 below.

	30	40	50	60	70	80	90	100
30	0.863	0.844	0.863	0.863	0.883	0.905	0.863	0.863
40	0.905	0.863	0.905	0.863	0.863	0.883	0.883	0.905
50	0.816	0.863	0.883	0.883	0.844	0.881	0.883	0.905
60	0.844	0.863	0.883	0.883	0.883	0.835	0.883	0.884
70	0.883	0.826	0.905	0.853	0.863	0.879	0.844	0.863
80	0.863	0.883	0.863	0.863	0.826	0.883	0.863	0.905
90	0.883	0.816	0.883	0.863	0.883	0.863	0.863	0.844
100	0.905	0.905	0.835	0.905	0.905	0.883	0.863	0.883

Table 10. Neurons per Layer High Saturation Cropped.



Figure 22. Contour Plot for High Contrast Cropped Predictive Accuracies.

		Original	Black & White	High Saturation	High Contrast	Low Exposure	Low Contrast
ginal	1 st Layer	90	40	70	60	70	70
Orig	2 nd Layer	30	40	60	40	70	40
pped	1 st Layer	100	50	40	30	40	50
Crop	2 nd Layer	70	30	30	30	30	30

Table 11. Neurons per Hidden Layer in Each Image Modification Category.

Activation functions and solvers were determined using the same exhaustive search method. All combinations of the available activation functions and solvers were tested, and the predictive accuracies recorded for each modification category. Table 12 shows the results of this search method for the category "High Saturation Cropped." As shown, the activation function 'tanh' provided high predictive accuracy in conjunction with two solvers, L-BFGS-B and Adam. The solver 'Adam" was selected as this combination provided high predictive accuracy throughout multiple categories of modification. This activation function and solver combination was used across all image modification categories. Once the ANN was modeled, two steps were taken to verify that the results obtained were accurate.

	Activation								
Solver		Identity	Logistic	Tanh	ReLu				
	L-BFGS-B	0.835	0.785	0.905	0.742				
	SGD	0.835	0.338	0.801	0.878				
	Adam	0.863	0.844	0.905	0.834				

Table 12. Activation Function and Solver for High Saturation Cropped.

4.4 Results and Verification

The ANN prediction accuracy results from this case study can be found in Figure 23 below. The highest predictive accuracies were found in the categories "Black and White Original," "High Saturation Original," "Low Contrast Cropped," and "High Saturation Cropped" at a rate of 91%. These accuracy rates show that cropping the images does not always ensure higher predictive accuracy. In fact, the two "High Saturation" categories predicted at the same rate. Throughout the whole data, there were four instances where the cropped images predicted at an equal or higher rate than their original counterparts.



Figure 23. Predictive Accuracies of Corrugated RSC Images in each modification category.

Figure 24 below shows images from the four highest predicted categories

for comparison of cropped images vs. original.



Figure 24. Images from the Highest Predicted Categories: "Black and White Original" (A), "High Saturation Original" (B), "Low Contrast Cropped" (C), and "High Contrast Cropped" (D).

Evaluating Figure 24 provides some possible explanations for why cropped images do not always predict at a higher rate than original images. When cropping images, it is common that the resolution decreases. This can make it more difficult for the ANN to identify details and distinguish between similar attributes in an image. Additionally, removing a majority of the background of each image can cause loss of context. This is important for distinguishing between the area of interest in an image and the background. In this case, the smaller background area could cause the ANN to not know that the corrugated RSC in question is the majority of the image, resulting in a prediction based on portions of the RSC instead of the RSC in its entirety. While this wasn't the case in the two cropped categories that predicted at a high rate, it could explain the lower predictive accuracies in other cropped categories. This further proves that it is important to include all image modification categories when modeling the ANN since it is not completely certain how the ANN views and analyzes images. To dive deeper into these results, two verification methods were utilized.

The first method of verification was manual. In this method, a table from the data mining software's prediction, shown below in Table 13, was utilized.

	Confide						
Image #	Compression	Corner	Edge	Prediction	Error	Actual Category	Image Name
1	0.00	1.00	0.00	Corner	0.001	Corner	Sample 42.2
2	0.00	1.00	0.00	Corner	0.002	Corner	Sample 44.3
3	0.00	0.76	0.23	Corner	0.238	Corner	Sample 25.3
4	0.04	0.00	0.96	Edge	0.998	Corner	Sample 22.1
5	0.00	1.00	0.00	Corner	0.001	Corner	Sample 50.2
6	0.00	0.99	0.00	Corner	0.005	Corner	Sample 35.2
7	0.01	0.01	0.98	Edge	0.990	Corner	Sample 48.1
	•					•	
	•		•	•	•	•	
42	0.07	0.93	0.01	Corner	0.073	Corner	Sample 2.3

Table 13. Prediction Table Output from ANN.

Table 13 provided the confidence level in the prediction made, shown in the confidence rates columns where three values ranging from 0-1 are shown. The first confidence level corresponds to the category "Compression," the second to "Corner," and the third to "Edge." For example, for row 4 of this Table, the ANN had a confidence level of 0.04 that the image was compression damage, 0.00 that it was corner damage, and 0.96 that it was edge damage. As shown from the "Prediction" column compared to the "Actual Category" column, this prediction was incorrect. This is also shown by the "Error" column which contains an error

rate of 0.998 for this specific prediction. Analyzing confidence levels for each prediction can show if the prediction is reliable to consider. If the ANN is not very confident in a prediction, correct or incorrect, the image can be checked to ensure that all modifications are accurate as well as that the sorting is correct. If the confidence level is high, the prediction can be considered valid. Performing manual verification is useful to ensure that each image was modified and sorted correctly before confusion matrix verification.

The second verification method used involved evaluating confusion matrices. These were used to see how the ANN performed in a broader sense. To evaluate the impact of the image background on prediction results, confusion matrices for the categories "Black and White" original (A) and cropped (B) are shown below in Figure 25. In both categories, all compression images were predicted correctly. For the "Black and White Original" category (A), two corner images were predicted to be edge damage, while two edge images were predicted to be corner damage. In the "Black and White Cropped" category (B), one corner image and one edge image were predicted to be compression damage, and 5 edge images were predicted to be corner damage. In these cases, it is obvious that the cropped category performed worse than the original category. A portion of the incorrectly predicted images from the "Black and White Cropped" category can be seen in Figure 26, below.



Figure 25. Confusion Matrices for Black and White "Original" (A) and "Cropped" (B).



Figure 26. Incorrectly predicted edge images from the category "Black and White Cropped".

The images shown in Figure 26 are edge damage images predicted to be corner damage. Image A has been predicted incorrectly across multiple modification categories. In this case, it can be assumed that the background is not what causes the ANN to predict incorrectly, as it was predicted to be corner damage consistently. For image B, it could be assumed that the ANN chose the category "corner" because the entire edge impacted is not visible in the image. This raises questions about image C, though. The entire edge is visible in the image, but it was still predicted to be corner damage. This could be due to the largest amount of visible damage being shown in the corner closest to the camera. The black-and-white image modification does not highlight the damage shown on the edge very strongly in this case. As shown in the confusion matrices for the "Black and White" category, it is not always the case that cropped images are predicted at a higher accuracy rate than original photos of the same modification.

Figure 27 shows confusion matrices from the highest predictive accuracy categories Black and White Original (A), High Saturation Original (B), High Saturation Cropped (C), and Low Contrast Cropped (D). All of these categories were predicted at a rate of 91% accuracy.



Figure 27. Confusion matrices for the highest predicted categories: Black and White Original (A), High Saturation Original (B), High Saturation Cropped (C), and Low Contrast Cropped (D).

Three of these categories, Black and White Original (A), High Saturation Original (B), and High Saturation Cropped (C) predicted all compression images correctly. In the category Low Contrast Cropped (D), all but one compression image was predicted correctly, with one being predicted to be in the category 'corner.' This implies that compression images are consistently predicted correctly, regardless of the image modification performed. Most of the differences between each matrix lie within the corner and edge image predictions. In the categories Black and White Original (A) and High Saturation Cropped (C), two corner images were predicted to be 'edge,' while two edge images were predicted to be 'corner.' For the category High Saturation Original (B), one corner image was predicted to be 'edge,' one edge image was predicted to be 'corner,' and two edge images were predicted to be 'compression.' In Low Contrast Cropped (D), one edge image was predicted to be corner damage, and two corner images were predicted to be edge damage. Drop testing a corrugated RSC can provide damage that extends beyond the point of impact, resulting in a higher chance that the damage observed could fit into multiple categories. Figure 28 below shows an example of one corner drop image and one edge drop image from the "Original" category with damage that could result in a prediction for either category. Image A in this figure is a corner drop image and image B is an edge drop image. Both of these examples show damage that extends beyond the point of impact. In image A, the left edge associated with the corner drop shows signs of impact. In image B, the corner associated with damage is closest to the camera, showing the corner damage more than the edge damage. Both of these images were part of the training group, but they could have been predicted as either "Edge" or "Corner" damage if included in the testing group. Additionally, there are many more options for RSC positioning in impact testing compared to compression testing. In total, there were 14 orientations utilized in drop testing compared to the 2 orientations for compression testing. Most of the damage observed in compression testing is in the form of a line at the buckling point of the RSC that extends throughout the case. This kind of damage is more consistent than impact testing, which could explain the difference in predictions between the three categories of damage.



Figure 28. Images that could fit in either "Edge" or "Corner" prediction categories. Image A: Corner Drop, Image B: Edge Drop.

CHAPTER 5: CONCLUSION

This thesis proposes a method that uses Artificial Neural Networks (ANN) to evaluate corrugated box damage using images. The process begins with collecting data from previous validation testing images, e-commerce platforms, or new images created in-lab using a web scraping tool. The data is then prepared for ANN modeling by sorting each image into a damage category and modifying it to fit into 12 image modification categories. The ANN is modeled through data splitting, determining the neuron structure, and selecting an appropriate activation function and solver. Data is verified through manual verification and confusion matrix evaluation to ensure that all predictions are valid for consideration. The results show that modifying images in various ways provides high predictive accuracy in multiple categories of modifications. Adopting this approach offers several benefits, including reduced expenses for retesting packages in case of damage following the initial validation phase, exceptional predictive accuracy, and streamlined processes due to the significant reduction in time required to assess damaged packages. This method is novel to the packaging industry and can be expanded upon. Future research ideas include expanding from lab-created images to review images from an e-commerce platform and predicting damage from a combination of validation test images and e-commerce review images. Additionally, this methodology could be expanded to account for the numerous packaging types used in the industry.
BIBLIOGRAPHY

- Šegota, S. B. (2020). Frigate Speed Estimation Using CODLAG Propulsion System Parameters and Multilayer Perceptron. *NAŠE MORE: znanstveni časopis za more i pomorstvo*, 117-125.
- ASTM International. (2016, Dec 27). Standard Test Methods for Vibration Testing of Shipping Containers. Retrieved from ASTM Compass: https://compass.astm.org/document/?contentCode=ASTM%7CD0999-08R15%7Cen-US

ASTM International. (2020, Oct 22). Standard Test Method for Determining Compressive Resistance of Shipping Containers, Components, and Unit Loads. Retrieved from ASTM Compass: https://compass.astm.org/document/?contentCode=ASTM%7CD0642-20%7Cen-US

- ASTM International. (2022, Feb 18). *Standard Practice for Performance Testing* of Shipping Containers and Systems. Retrieved from ASTM Compass: https://compass.astm.org/document/?contentCode=ASTM%7CD4169-22%7Cen-US
- ASTM International. (2022, May 13). Standard Test Method for Random Vibration Testing of Shipping Containers. Retrieved from ASTM Compass: https://compass.astm.org/document/?contentCode=ASTM%7CD4728-17R22%7Cen-US
- Biancolini, M. E. (2003). Numerical and experimental investigation of the strength of corrugated board packages. *Packaging Technology and Science*, 821-832.
- Bloom, J. Z. (2005). Market Segmentation: A neural network application. *Annals* of tourism research, 93-111.
- Bottou, L. B. (2012). The Tradeoffs of Large Scale Learning. *Optimization for Machine Learning*, pp. 351-368.
- Brownlee, J. (2017, July 3). *Gentile Introduction to the Adam Optimization Algorithm for Deep Learning*. Retrieved from Machine Learning Mastery:

https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/

- Coppola, D. (2022). *E-commerce as share of total U.S. retail sales from 1st quarter 2010 to 3rd quarter 2022*. Retrieved from Statistica: https://www-statista-com.proxy1.cl.msu.edu/statistics/187439/share-of-e-commerce-sales-in-total-us-retail-sales-in-2010/
- de Abajo, N. D. (2004). ANN quality diagnostic models for packaing manufacturing: an industrial data mining case study. *Tenth ACM SIGKDD international conference on Knowledge Discovery and data mining*, (pp. 799-804).
- Draelos, R. (2019, Feb 17). *Measuring Performance: The Confusion Matrix*. Retrieved from Glass Box Medicine: https://glassboxmedicine.com/2019/02/17/measuring-performance-theconfusion-matrix/
- Dubey, S. S. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 92-108.
- Dunno, K. (2017). Effects of transportation hazards on high barrier flexible packaging films. *Journal of Applied Packaging Research*.
- Esfahanian, S. &. (2022). A novel packaging evaluation method using seniment analysis of customer reviews. *Packaging Technology and Science*, 903-911.
- Fadiji, T. C. (2016). Susceptibility to impact damage of apples inside ventilated corrugated paperboard packages: Effects of package design. *Postharvest Biology and Technology*, 286-296.
- Frank, B. G. (2010). Compression testing to simulate real-world stresses. *Packaging Technology and Science*, 275-282.
- Hallbäck, N. K. (2014). Finite element analysis of hot melt adhesive joints in carton board. *Packaging Technology and Science*, 701-712.
- Hecht-Nielsen, R. (1987). Kolmogorov's mapping neural network existence theorem. *International conference on Neural Networks* (pp. 11-14). New York, NY: IEEE Press.

- Huang, G. B. (2003). Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE transactions on neural networks*, 274-281.
- Huang, T. C. (2022). Investigations of structure strength and ventilation performance for agriproduct corrugated cartons under long-term transportation trip. *Packaging Technology and Science*, 821-832.
- International Trade Administration. (2023). *Ecommerce sales & size forecast*. Retrieved from International Trade Administration: Trade.gov
- ISTA. (2016). *Ships in Own Container (SIOC) for Amazon.com Distribution System Shipment*. Retrieved from International Safe Transit Association: https://ista.org/docs/6AmazoncomSIOCOverview.pdf
- ISTA. (2018). Packaged-Products for Parcel Delivery System Shipment 70 kg (150 lb) or less. Retrieved from International Safe Transit Association: https://ista.org/docs/3Aoverview.pdf
- Jílková, P. &. (2021). Digital consumer behaviour and ecommerce trends during the COVID-19 crisis. *International Advances in Economic Research*, 83-85.
- Jung, H. M. (2012). Effects of vibration fatigue on compression strength of corrugated fiberboard containers for packaging of fruits during transport. *Journal of Biosystems Engineering*.
- Kůrková, V. (1992). Kolmogorov's theorem and multilater neural networks. *Neural networks*, 501-506.
- Karsoliya, S. (2012). Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. . *International Journal of Engineering Trends and Technology*, 714-717.
- Kaylan, K. J. (2014). Artificial neural network application in the diagnosis of disease conditions with liver ultrasound images. *Advances in bioinformatics*.
- Krotov, V. &. (2018). Legality and ethics of web scraping. Twenty-fourth Americas Conference on Information Systems. New Orleans: Emergent Research Forum.

- Lansmont. (2023). *Model 1800 Vertical vibration test system*. Retrieved from Lansmont: https://www.lansmont.com/products/vibration/vertical/lansmont-standard-1800
- Lansmont. (2023). *PDT 80 Precision Drop Tester*. Retrieved from Lansmont: https://www.lansmont.com/products/drop/lansmont-pdt-80
- Lansmont. (2023). *Saver(TM) AM-Asset Monitor Shock and vibration data logger*. Retrieved from Lansmont: https://www.lansmont.com/products/data_loggers/saver_am
- Lansmont. (2023). *SqueezerPro Compression tester*. Retrieved from Lansmont: https://www.lansmont.com/product/compression-testers/squeezerpro
- Liu, D. C. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 503-528.
- Mills, N. J.-M. (2005). Finite element analysis (FEA) applied to polyethylene foam cushions in package drop tests. *Packaging Technology and Science: An International Journal*, 29-38.
- Molina, E. H. (2021). Development of a friction-driven finite element model to simulate the load bridging effect of unit loads sotred in warehouse racks. *Applied Sciences*.
- Nygårds M, S. S. (2019). Simulation and experimental verification of a drop test and compression test of a gable top package. *Package Technology and Science*, 32(7):325-333.
- Obuchowski, A. (2020, April 16). Understanding neural networks 2: The math of neural networks in 3 equations. Retrieved from Becoming Human: Artificial Intelligence Machine: https://becominghuman.ai/understanding-neural-networks-2-the-math-of-neural-networks-in-3-equations-6085fd3f09df
- Orange Data Mining Platform. (2023). *Orange*. Retrieved from Orange Data Mining: https://orangedatamining.com/

- Rebala, G. R. (2019). *Machine learning definition and basics. An introduction to machine learning*. Retrieved from Springer: https://www.springer.com/journal/10994
- Rowson, J. Y. (2008). Modelling capping of 28 mm beverage closures using finite element analysis. *Packaging Technology and Science, An International Journal*, 287-296.
- Rycobel. (2023). *Box compression tester* (*BCT*) *Compression strength tester*. Retrieved from Rycobel: https://www.rycobel.com/products/boxcompression-tester
- Sensolus. (2023). Use cases: Condition Monitoring. Retrieved from Sensolus: https://www.sensolus.com/use-cases/transport-condition-monitoring/
- Skyline University College. (2016). *Packaging for a new era of e-commerce*. Retrieved from Skuline University College: https://www.skylineuniversity.ac.ae/pdf/ecommerce/Bemis-eBookeCommerce.pdf
- Spruit, D. A. (2021). First market study in e-commerce food packaging: Resources, performance, and trends. *Food Packaging and Shelf Life*.
- Twede, D. &. (2005). *Cartons, Crates, and Corrugated Board*. Seoul, South Korea: DESIGN HOUSE Incorporated.
- UpGrad. (2022, Sep 22). *Neural Network: Architecture, Components & Top Algorithms*. Retrieved from UpGrad: https://www.upgrad.com/blog/neural-network-architecture-componentsalgorithms/
- Xie, Y. C. (2023). Methodology selecting and packaing materials combined multi-sensory experience and fuzzy three-stage network DEA model. *Packaging Technology and Science*, 125-134.
- Zhong, C. L. (2016). Measurement and analysis of shocks on small packages in the express shipping environment of China. *Packaging Technology and Science*, 437-449.