KOREAN JOURNAL OF PACKAGING SCIENCE & TECHNOLOGY
Vol. 29, No. 3 163~174 (2023)
https://doi.org/10.20909/kopast.2023.29.3.163

A Comparative Analysis of Artificial Neural Network (ANN) Architectures for Box Compression Strength Estimation

By Juan Gu,[1] Benjamin Frank,[2] and Euihark Lee[1]*

[1]School of Packaging, Michigan State University, East Lansing, MI 48824-1223, USA
[2]Packaging Corporation of America, Lake Forest, IL, 60045, USA

Abstract Though box compression strength (BCS) is commonly used as a performance criterion for shipping containers, estimating BCS remains a challenge. In this study, artificial neural networks (ANN) are implemented as a new tool, with a focus on building up ANN architectures for BCS estimation. An Artificial Neural Network (ANN) model can be constructed by adjusting four modeling factors: hidden neuron numbers, epochs, number of modeling cycles, and number of data points. The four factors interact with each other to influence model accuracy and can be optimized by minimizing model's Mean Squared Error (MSE). Using both data from the literature and "synthetic" data based on the McKee equation, we find that model estimation accuracy remains limited due to the uncertainty in both the input parameters and the ANN process itself. The population size to build an ANN model has been identified based on different data sets. This study provides a methodology guide for future research exploring the applicability of ANN to address problems and answer questions in the corrugated industry.

Key words: Corrugated box, Box Compression Strength (BCS), Artificial Neural Network (ANN)

Introduction

Research into the production and use of corrugated packaging has produced an enormous amount of literature exploring the structural dynamics of corrugated boxes in use and in failure[1)]. This history includes various studies[2-8)] modeling package failure based on fundamental mechanics of materials, including the best-known model still in active use, the McKee equation[9)]. While many factors can influence the strength of a package, these models primarily focus on package dimensions, the inherent material strength, and either board flexural stiffness or overall thickness of the box panels. Nonetheless, these models are limited in their predictive accuracy by the uncertainty in the measurements of package properties[10)]. The work by McKee and others was limited because it involved simplifying more general physical relationships. Fundamentally, these were linear regression analyses based on specific data sets, typically limited by processing constraints.

Improvements in processing speed and power have created opportunities for application of computer-based analysis methods even for applications as "simple" as the corrugated box. Finite element analysis (FEA), a powerful technique often used for simulation of engineering process, is finding a home in the corrugated industry. FEA models generate predictions by leveraging fundamental physical mechanics across different length scales, stitching together functional relationships to estimate the effect of changes in very basic material properties (e.g., paper elasticity) on the larger final system (e.g., box strength). The functional form is known; the propagation of parameters and their impacts produce a prediction of the result. Various studies have explored using an FEA approach to predict ECT[11-14)] or BCS[15-24)], allowing for detailed examination of the impact of moisture, perforations, holes, and openings, crushing, and more complex structures. The literature has grown so extensively that even review articles addressing the usefulness of FEA on broader topics have sections discussing corrugated paperboard packaging[25)]. Each of these studies requires detailed information on the material parameters to input into the models, typically producing reasonable agreement between the model and the limited number of physical samples evaluated. As such, they potentially contribute to our understanding of the impact of specific changes examined (i.e., hole size and placement.)[18)]. However, very few of these studies address or investigate how well their models work with boxes made of different, varied, or unknown materials. They also don't often discuss about how the varying physical and mechanical properties of paper or combined

*Corresponding Author: Euihark Lee
448 Wilson Road, East Lansing, MI 48824
Tel: +1-517-355-7613
E-mail: leeeuiha@msu.edu

board can affect the accuracy of their predictions. Typically, the input parameters required for an FEA are not properties regularly measured in the papermaking or box making process. Thus, existing (published) analyses cannot reasonably be used for a generalized assessment of a random box in the same way that we can use the McKee equation.

Artificial Neural Networks (ANNs) are a very different computing approach that can also be used to explore the underlying relationships in a set of data and generate predictive models. The concept of ANN was introduced in 1956[26] and the ANN approach has been utilized in many different applications over the past few decades[27]. ANNs have been employed in the marketing domain to enhance communication between companies and their customers[28]. In the areas of environment and water management, ANNs have been leveraged for data virtualization, real-time processing, and other related activities[29]. In the government sector, ANNs have been instrumental in supporting policy-making activities through impact prediction and resource allocation optimization, among other policy making activities[30]. Additionally, ANNs have found relevance in the renewable energy sector, specifically in wind speed prediction, load forecasting, and prediction activities[31]. ANNs even have begun to appear in assessing problems in packaging and supply chains[32,33].

ANNs have many advantages because they can strive to take whatever information we happen to know in terms of materials inputs and gather relationships to the outputs of interest. This inference process can take in a broader range of inputs, teasing out their connections (implicit or explicit) to "understand" their relationship to a given output. The goal of ANNs is to minimize the error of the predicted property. By mapping features in data, ANNs can substantially add to the power of exploratory data analysis[34]. Using ANNs can bring many benefits in scientific research[35], making more consistent decisions and shortening decision-making process[36]. Given the fundamentally non-linear relationship between fiber characteristics and the mechanical properties of paper, combined board, and boxes, this alternate approach is beginning to garner interest among researchers[37,38]. This potentially allows us to incorporate a large number of input parameters into a single prediction model, limited only by the size of our data set. ANN research to date has focused on specific areas or factors influencing box strength[39].

Given the limited existing research, it remains to be seen whether an ANN can estimate BCS any more accurately than our historical, closed-form approaches. A properly structured ANN might be able to identify additional parameters that contribute to BCS with similar level of impact as known existing factors (e.g., Edge crush test (ECT) value) and thus improve current models over the known levels of inherent variation in the input data. In order to leverage those opportunities, we need to clearly identify the size of the data set required. Compared with many ANN applications which automatically create the underlying data to build a model, collecting data point for BCS estimation model is comparatively expensive, necessitating a series of off-line tests. For ANN modeling of corrugated packaging, the challenges required to generate sufficient data sets may well be the limiting factor on the capability of the model.

This study investigates the data requirements for an ANN to estimate compressive strength and assesses the ANN approach's ability to overcome input variation limitations in the papermaking and box manufacturing process. To begin, we apply the ANN approach to the existing data from McKee's 1963 research. Although the McKee data set proves too small for a robust ANN study, its well-established nature enables us to define our ANN methodology. Moreover, it illustrates the process to readers who are familiar with box compression modeling but less acquainted with ANNs. Next, we employ the ANN approach to analyze a significantly larger "synthetic" data set, constructed using idealized data derived from the McKee equation. This dataset allows us to evaluate the potential accuracy of an ANN model when applied to an established large data set and physical relationship. Furthermore, we introduce variation to the input data of the idealized data set, enabling us to assess how this variation propagates through the ANN. This investigation addresses the fundamental question of data set size and evaluates whether the current data collection approaches in the corrugated industry are sufficiently advanced to support the application of ANN in assessing box performance. The conclusion has been appropriately presented at the end, encapsulating the main findings and providing a conclusive summary.

1. Artificial Neural Networks (ANN)

The ANN model space is structurally built in three types of layers: the input layer, the hidden layer and the output layer. Each layer is comprised of neurons connected to each of the inputs or neurons in the adjacent layers in the model, as shown in Figure 1.

The weights of the connections between neurons are the adjustable model parameters that govern how the model calculates the output from the given inputs. The ANN calculation process cycle includes a forward step computation of fitting input data into an ANN and a backward step computation of calculating errors and updating the weights in the model. A single iteration of this computational process is termed an epoch within an ANN.

At the beginning of the ANN process, all weights between nodes are randomly assigned. The squared difference between predicted BCS values from our training data and their actual BCS values is then calculated in equation (1), and the weights are adjusted via a backpropagation process.

Input	Output
Edge Crush Strength (lb/in) – ECT Flexural stiffness in the machine direction of the combined board (lb*in) – EI_x Flexural stiffness in the cross-machine direction of the combined board (lb*in) – EI_y Board Caliper (in) – BC Box Length (in) – BL Box Width (in) – BW Box Depth (in) –BD	Box Compression Strength (lb) – BCS

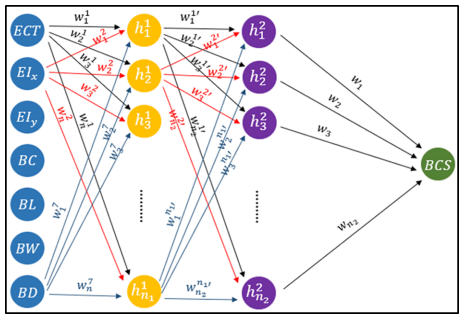

Fig. 1. A model of an Artificial Neural Network (ANN) structure for predicting box compression strength (BCS) using inputs provided by the McKee data set.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(BCS_{predicted} - BCS_{actual})^2 \quad (1)$$

where n is the number of samples in a trained dataset, and MSE represents mean squared error

An epoch signifies the process of feeding a dataset into the model and the model's weights are adjusted to reduce the overall MSE. This iterative repetition of the process, known as multiple epochs, continues until the MSE reduction rate falls below given criteria.

An ANN approach segments a given set of known data into two uneven groups, training data and testing data. The former is used to build and refine the model and the latter is used to evaluate the model accuracy. To assess our ANN, 67% of each data set were split into training data and the remaining 33% were split into testing data. Each node in the hidden layers could be defined based on a weighted sum of the parameters in the prior layer, as shown in equation (2).

$$h_j^i = f\sum_k (w_k^j \times x_k) \quad (2)$$

h_j^i is the value of j^{th} neuron in i^{th} hidden layer. $j = 1, 2, 3,\ldots, n_1$ when $i = 1$; $j = 1,2,3,\ldots, n_2$ when $i = 2$.

x_k is the value of k^{th} neuron in previous layer, $k = 1, 2, 3,\ldots, n$.

w_k^j is the weight from the k^{th} neuron in previous layer to j^{th} neuron.

f is the activation function.

To enhance the efficiency, an activation function was added in the hidden layers. The Rectified Linear Unit (ReLU) is the default activation function for hidden layers and perhaps the most common function used for hidden layers in Machine Learning studies[40]. The ReLU function returns the input itself when the input is positive and returns 0 when the input is 0 or less than 0. Since only a certain number of neurons are activated in this case, the ReLU function effectively mitigates the vanishing gradient problem. This problem occurs when the derivative exponentially diminishes across multiple layers, resulting in the gradient approaching zero and hindering weight updates during backpropagation. The ReLU function avoids the gradient vanishing problems and enables faster and more efficient learning processes[41]. The output layer typically uses a different activation function from the hidden layers. a sigmoid function were selected, which has a very convenient first derivative and is popular in neural network research[42]. The sigmoid function is an efficient way of producing an output p∈[0,1], which can be interpreted as a probability. Plots of the ReLU function and sigmoid function are shown in Figure 2.

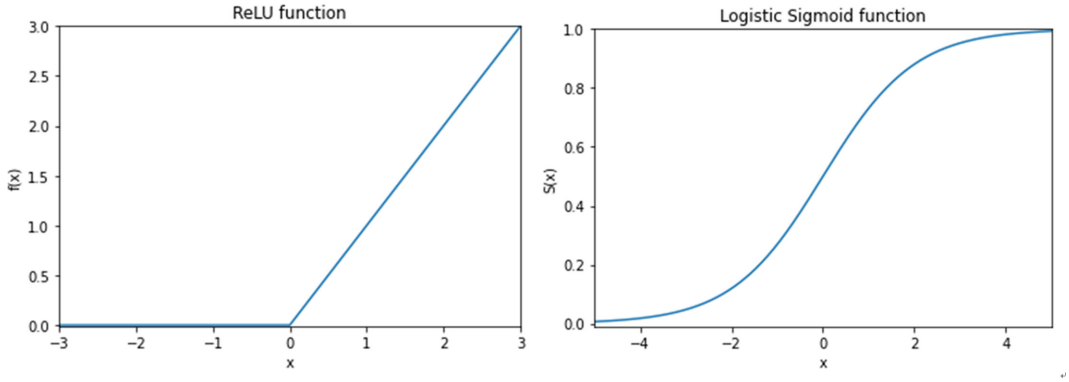

Fig. 2. ReLU function and Sigmoid function used.

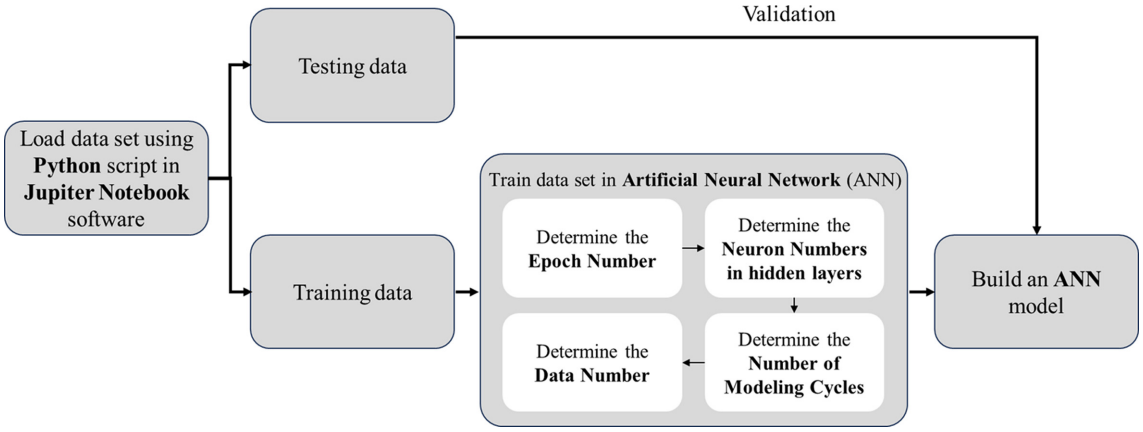

Fig. 3. Flow for building an ANN model for BCS estimation.

This study involved running programming tasks on an HP Laptop 15t-dy100 featuring an Intel(R) Core(TM) i5-1035G1 CPU, operating at a processing speed of 1.00GHz. The coding process to train the ANN model was conducted using Jupiter Notebook software, an integrated development environment (IDE). Figure 3 illustrates the sequential steps in constructing an ANN model. The training duration varied depending on the dataset's size and characteristics, influenced by the combination of hardware and software. For instance, training a smaller data set of around 60 data points took approximately 3 minutes, whereas training a larger data set comprising approximately 3000 data points required about 30 minutes.

In this study, three datasets were trained to build an ANN model for BCS estimation. The three data sets include the McKee data set, an idealized data set and a data set with variation. The McKee data set is from literature presented by McKee in 1963[9], specifically compiled for BCS estimation. It consists of 63 data points derived from box compression testing. The idealized data set is a synthetic data set based on McKee equation[9]. This data set was generated by including the box dimensions, ECT values, and thicknesses of 3,009 boxes commonly used in commerce and substituting them into the McKee equation. The data set with variation was created by introducing random error to the parameters of the idealized data set's boxes. BCS values were then calculated using the McKee equation[9]. This process was carried out to achieve a variation of ±5.4% for BCS. It contains an equivalent number of data points as the idealized data set. Detailed descriptions of these three datasets are provided in the ANN training section, delineating the specifics for each dataset.

2. Artificial Neural Networks (ANN) and McKee Date Set

Like many of the modeling efforts in the industry, we begin our exploration of the applicability of ANNs on box compression estimation with the work of McKee et.al., Their model was built using 63 data points including A-, B-, and C-flute boxes. This data set captured information on ECT, flexural stiffness in the machine and cross-machine directions of the combined board (El_x and El_y), caliper of the corrugated board, and the length, width, and depth of the box. Those seven physical parameters serve as the input parameters for an ANN model with BCS as the output, as shown in Figure 1. Of note, these parameters are not independent - flexural stiffness depends in part on the caliper of the board. Including all the available parameters in the data allows the ANN to appropriately assess the relative importance of each parameter to BCS estimation.

To assess our ANN given the limited data presented by McKee et.al., the 63 data points were split into 42 training data points and 21 testing data points. Two hidden layers were implemented to generate the output value (BCS). We initially considered utilizing 200 epochs for conducting the calculations.

Model accuracy and consistency can be influenced by many modeling factors. The first task in developing an ANN model includes optimizing neuron number combinations in each of the hidden layers. We implemented an exhaustive search method[43] examining different neuron number combinations in the various layers, as shown in Figure 4: To assess the model accuracy, the neuron numbers for the first hidden layer were examined ranging from 80 to 184 (with an increment of 8). Similarly, for the second hidden layer, the neuron numbers were examined from 24 to 42 (with an increment of 3). The MSE was calculated for each combination to evaluate the model's accuracy. In each case a random selection of data points from the underlying data set was used, which has implications on the robustness of the minimum MSE. The minimum MSE occurred with 160 neurons in the first hidden layer and 36 in the second hidden layer. To confirm this result, the increments for both hidden layers were reduced. The increment of 8 in the first hidden layer decreased to 2, and the increment of 3 in the second hidden layer decreased to 1. Remarkably, even with these decreased increments, the min-

Mean Square Error		Neurons' number (2nd hidden layer)						
		24	27	30	33	36	39	42
Neurons' number (1st hidden layer)	80	1559	1837	1842	1533	1675	1681	1672
	88	1643	1742	1357	1459	1468	1555	1579
	96	1718	1379	1334	1477	1691	1343	1414
	104	2091	1886	1688	1601	1452	1556	1254
	112	1545	1846	1447	1530	1343	1512	1423
	120	1335	1925	1587	1236	1275	1463	1201
	128	1431	2123	1446	1515	1619	1447	1322
	136	1850	1712	1429	1311	1383	1180	1485
	144	1402	1322	1573	1427	1379	1301	1315
	152	1471	1213	1139	1236	1304	1481	1314
	160	1605	1597	1484	1376	1069	1363	1393
	168	1390	1512	1377	1326	1295	1457	1403
	176	1396	1267	1477	1253	1223	1424	1251
	184	1282	1473	1385	1318	1210	1479	1171

Mean Square Error		Neurons' number (2nd hidden layer)		
		35	36	37
Neurons' number (1st hidden layer)	152	1301	1304	1313
	154	1260	1154	1296
	156	1441	1236	1208
	158	1391	1327	1387
	159	1502	1452	1381
	160	1482	1069	1213
	161	1292	1237	1331
	162	1395	1318	1305
	164	1434	1395	1352
	166	1442	1285	1555
	168	1266	1295	1192

Fig. 4 Mean Squared Error (MSE) calculations for the model using McKee data with varying numbers of neurons in each of two hidden layers for the McKee data set. The error is minimized for 160 neurons in the first layer and 36 neurons in the second.

imum MSE still occurred with the same combination of neuron numbers. That structure was maintained for the analysis of the McKee data. Notably, this allows for substantially more degrees of freedom in the model than the available data in the McKee data set. Those degrees of freedom may account for the variation observed across different combinations in Figure 4.

To understand the computational load further, we explored how the number of epochs impacted model convergence. While again this calculation is not resource intensive for a small data set like the one provided by McKee et.al., it becomes critical to stop the process at convergence once the data set grows. Figure 5 illustrates that the MSE begins to converge in less than 50 epochs. As the number of epochs increases up to 200, the MSE reduction rate becomes increasingly slow, with minimal improvement. To strike a balance between computational time and accuracy, we select a stopping point when the MSE reduction rate falls below 3.0%. This threshold is typically reached at approximately 100 epochs, ensuring optimal computational efficiency without significantly compromising accuracy.

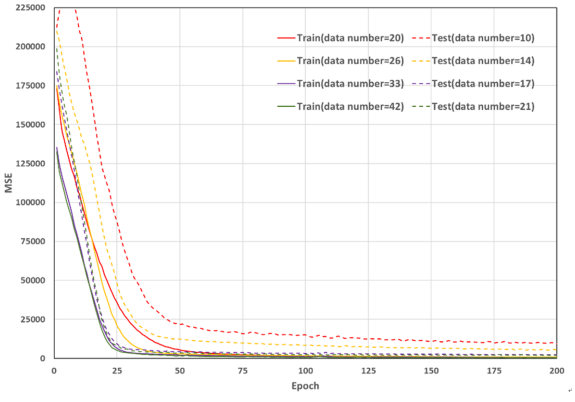

Fig. 5. MSE versus epoch plot of different data numbers (McKee data set).

The number of data points plays a very important role in the ANN accuracy. Figure 6 displays how the number of data points influences the ANN accuracy. Exploring different total population sizes from 30 to the full data set of 63 points, the chosen population was randomly divided into training data (2/3 of the points) and testing data (1/3 of the points). In the modeling process, partitions of underlying data vary from modeling cycle to cycle. This can have a significant impact on the model accuracy due to some special data points may fall into training data in a given cycle and fall into testing data in the subsequent cycle. To assess the impact of variation in the input data on model results and predictions, the process of partitioning a data set into training and testing data was regularly repeated. Multiple modeling cycles were performed. For the McKee data set, a sufficient number of "modeling cycles" of 60 were performed in each population size from 30 data points

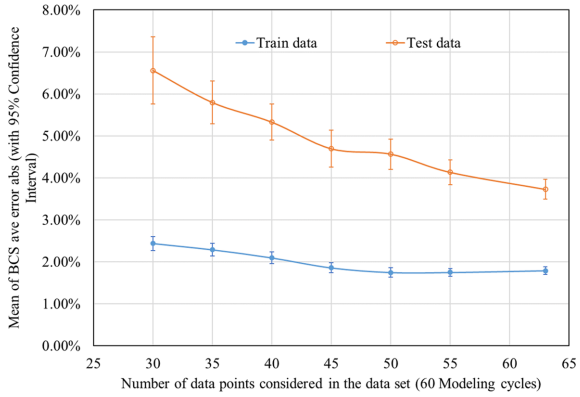

Fig. 6. Average error in estimated box compression strength given different subsets of the data (McKee data set), each run through 60 modeling cycles. Note error bars indicate 95% confidence intervals on the mean values.

to the full data set. Overall, the calculation process reflects a confidence interval on the model output related to the breadth of variation in potential input data sets. The confidence intervals around the mean error reflected the ANN training accuracy, as shown in Figure 6. As expected, ANN accuracy increases with population size. For this small data set like the McKee data set, ANN accuracy of the whole data set is notably higher than that of the partial data set, which concludes that the whole data set is needed for the McKee data set to minimize the error.

Since ANN randomly splits the data into training and testing data in the modeling process, each modeling cycle can have different underlying data partitions. As a result, each modeling cycle can generate a unique model that optimally fits the training data provided but can produce very different values for the error when assessing the testing data. Therefore, it is important to understand how many modeling cycles are required to have results converge to a "typical" reliability. To investigate the impact of different underlying data partitions on the accuracy of the ANN, various numbers of modeling cycles were examined. Figure 7 shows that when we partitioned the full McKee data set (all 63 points), the training data accuracy remained relatively consistent as the number of evaluation cycles increased. The average testing data accuracy converged after roughly 60 rounds of testing, very similar to the total number of data points in the database.

We have explored four modeling factors common in the ANN process using the data presented by McKee: the combination of neuron numbers in hidden layers, the number of epochs, the number of modeling cycles, and the number of data points in a data set. An optimal combination of neuron numbers in hidden layers can minimize the MSE and increase the ANN accuracy for BCS estimation. As the epoch number increases, the MSE reduction rate becomes increasingly slow. To strike a balance between computational time and accuracy, a stopping point when the MSE reduction rate falls below 3.0% was selected to ensure optimal computational efficiency without significantly compromising accuracy. Consistency for the ANN model is realized when the number of modeling cycles reaches a critical number for a given population size, and a minimum number of data points can be identified at which the MSE is minimized, and the ANN is most robust. We carry these observations forward into our analysis of larger, more generalizable data sets.

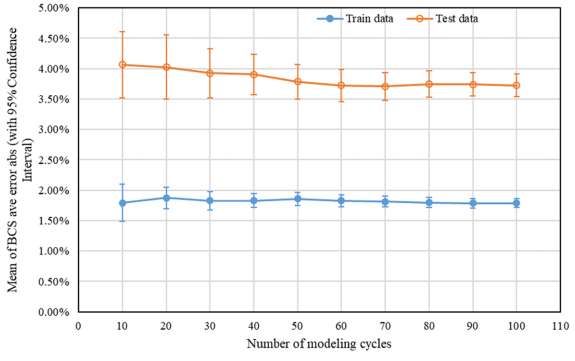

Fig. 7. Mean of average error in estimated box compression strength given different numbers of modeling cycles. Note that the testing data values converge after 60 modeling cycles (McKee data set). Error bars indicate 95% confidence intervals on the mean values.

3. Artificial Neural Networks (ANN) and An Idealized Data Set

McKee et.al.'s simplified model for box compression strength can be used to explore the applicability of ANN to compression strength estimation. However, the limited size of their data set constrains the ANN approach as noted above. Therefore, a larger data set is desirable. Using the simplified McKee equation[9] as shown in equation (3), a synthetic data set could be generated.

$$BCS = 5.87 \times ECT \times \sqrt{Caliper \times Perimeter} \quad (3)$$

In this way, the "idealized" data set was created with 3,009 data points. These data points represent boxes with ranges in length, width, aspect ratio, ECT, caliper and flute types (B- & C-flute) commonly used in North America (Table 1). Note that each "data point" discussed in this section is a specific set of information defining the physical properties of the box (lengths, width, caliper, and ECT) and the associated BCS calculated using equation (3).

Given the "perfect" nature of the fabricated data set, it is obvious that a simple least-squares fit of equation (3) to the input parameters reproduces the BCS values with 100% accuracy and 0% error. With a data set this large, one might also hope to overcome the ANN challenges experienced in fitting the much more limited data from McKee, and potentially reproduce the expected values in a test data subset perfectly, with close to no variation from the actual values.

To start this process, 67% of the data set (2,016 randomly selected samples) were used for the ANN training process and the rest were used for evaluation of the resulting model. With 200 epochs, the optimal neuron number combination in the

Table 1. Minimum and maximum values of the data incorporated in the idealized data set

Property	Min	Max
Length (cm)	19.05	99.38
Width (cm)	12.70	76.96
L/W (aspect ratio)	1.00	4.00
Perimeter (cm)	71.12	346.71
Caliper (cm)	0.26	0.44
ECT (lbs/inch)	64.77	228.35

SUM MSE of train data and test data		Neurons' number in second hidden layer								
		24	27	30	33	36	39	42	45	48
Neurons' number in first hidden layer	128	17	18	16	15	16	16	16	16	14
	136	19	14	16	16	17	15	15	15	16
	144	16	14	17	17	16	14	13	12	15
	152	18	18	18	17	14	15	13	14	13
	160	16	14	16	14	17	16	13	14	15

SUM MSE of train data and test data		Neurons' number in second hidden layer						
		36	39	42	44	45	46	48
Neurons' number in first hidden layer	136	17	15	15	16	15	15	16
	138	14	13	15	14	17	15	18
	140	14	13	16	15	13	16	16
	141	14	15	15	17	16	14	15
	142	14	15	14	13	12	17	16
	143	14	14	14	17	14	15	16
	144	16	14	13	14	12	15	15

Fig. 8. Mean Squared Error (MSE) calculations for the model using idealized data with varying numbers of neurons in each of two hidden layers for the idealized data set. The error is minimized for 142 neurons in the first layer and 45 neurons in the second.

hidden layers was again explored using an exhaustive search method. Figure 8 displays the examination of neuron numbers in the first hidden layer, ranging from 128 to 160 (with an increment of 8), as well as the examination of neuron numbers in the second hidden layer, ranging from 24 to 48 (with an increment of 3). The MSE was calculated for each combination to assess the model's accuracy. The MSE was minimized with 144 neurons in the first hidden layer and 45 neurons in the second hidden layer. To validate this result, the increments for both hidden layers were reduced to 1. Notably, the minimum MSE was observed with 142 neurons in the first hidden layer and 45 neurons in the second hidden layer. As with our McKee analysis above, this number of neurons provides more degrees of freedom in our modeling space than data sets in our model population. Note that the MSE is much lower than for the McKee data set because the data is perfect. However, the values are not zero indicating some residual uncertainty in the estimation of BCS even for this idealized data.

In exploring the idealized dataset, a similar approach to the McKee dataset analysis in ANN was followed. The impact of the number of epochs on the model MSE was investigated, as depicted in Figure 9-a). As the number of epochs increased up to 200, a decreasing trend in MSE with significant fluctuations was observed. However, the rate of MSE reduction started slowing down. To provide a clearer view without the disturbance of MSE fluctuations, the Moving Average technique[44] was applied to create a smoothed graph, as illustrated in Figure 9-b). This revealed that the MSE experienced a rapid decrease before reaching 50 epochs. From 50 to 200 epochs, the MSE decreased steadily, and the large fluctuations disappeared after 140 epochs. Similar to the study of the McKee data set, to strike a balance between computational time and accuracy, a stopping point was selected when the MSE reduction rate falls below 3.0% after applying the Moving Average technique. This threshold is typically reached at approximately 140 epochs, ensuring optimal computational efficiency without significantly compromising accuracy.

When examining the full data set of 3,009 data points, the ANN accuracy remained relatively consistent while the confidence interval around the mean error decreased as the number of modeling cycles increased (Figure 10). To better understand why the error in the model was not zero as might be expected for a model fitting "perfect" data, the specific results from each cycle were examined. It was observed that the BCS errors of four data points in particular always showed higher estimation error than other data points. Those four data points are at the limits of the data set (or boundary data points). Figure 11 shows the frequency of the actual BCS values. As is typical for data at the end points of a distribution, these four points have excessive leverage in the modeling. Their impact on model accuracy in test data depends on what adjacent points happen to be in the training data. When the boundary data points are randomly selected to be part of the testing data and thus do not appear in the training data, the result tends to show higher BCS average error. The average error across multiple cycles is impacted by this contribution.

To see the influence of population size on the ANN accuracy for the perfect model (similar to Figure 6 above), we examined populations from 600 to 3,009 data points using 10 modeling cycles. The results show that the mean of BCS average errors fluctuates notably when we consider a limited number of data subsets (Figure 12). Even for this larger population, the influence of limiting population size can have a meaningful impact if we don't iterate the process sufficiently. For 50 cycles, the mean of BCS average error decreased steadily as the data included increased, and the BCS average error levels out at ~1,500 data points. This suggests that the influence of modeling cycles and population size need to be considered together.

The combination of neuron numbers, the number of epochs, the number of modeling cycles, and the number of data points impacts the accuracy of the ANN prediction. Even when using the full data population (>3000 data points) and many modeling cycles on a perfect data set generated by a closed form equation, the average relative error of the BCS prediction is not zero. From Figure 12, in conjunction with Figure 10, this analysis identifies the error contribution of the ANN approach

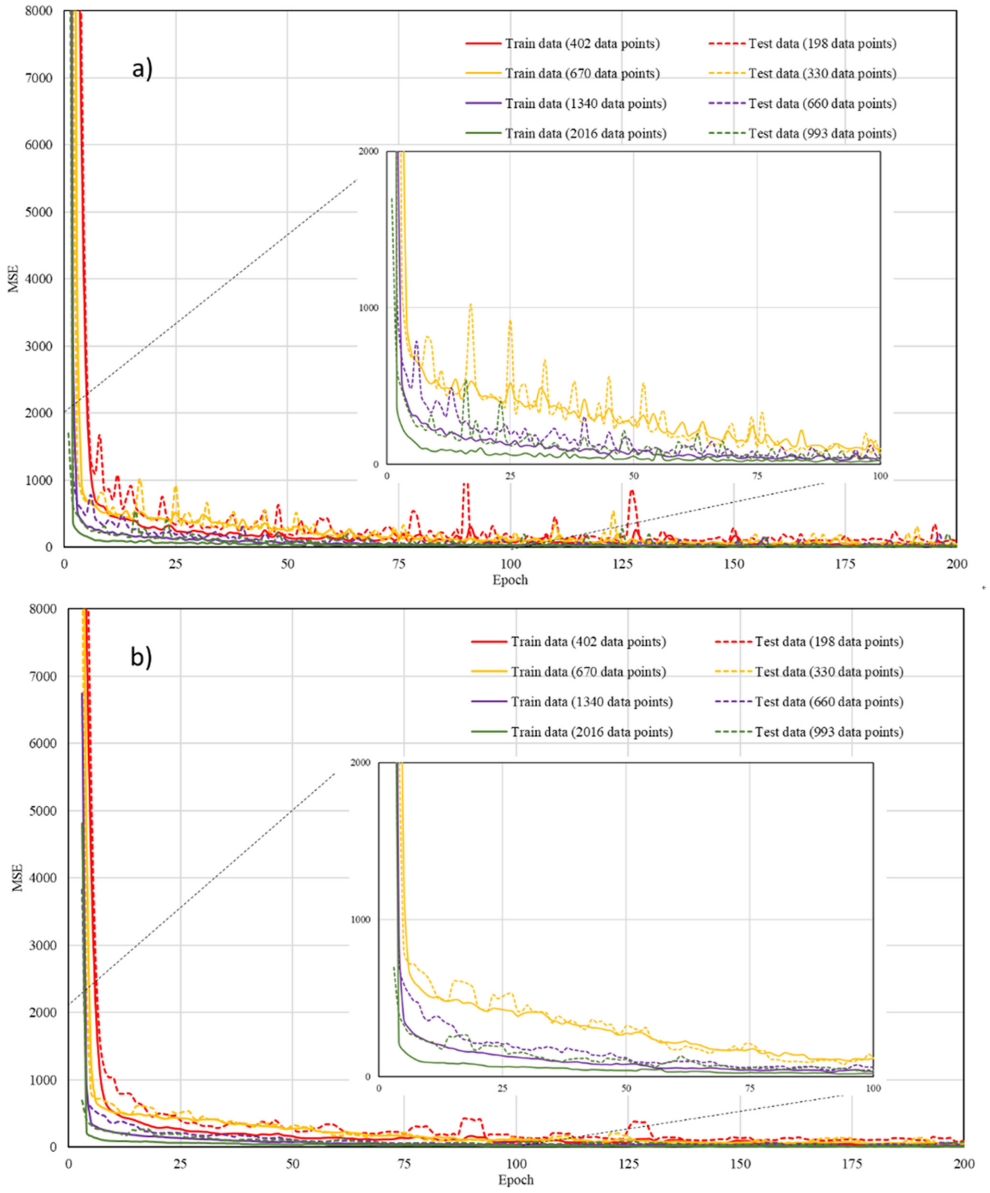

Fig. 9. Mean Squared Error (MSE) of the fits as a function of epochs for different sized data subsets from the idealized data set. 6a displays the raw MSE calculated for each epoch, while 6b presents smoothed data, more clearly displaying the asymptotic nature of the functional relationships.

itself at around 0.4% when estimating BCS from this type of data and data sets of this size. This residual error is independent of any physical properties; rather, it arises from the ANN process itself. As such, we would expect it to be additive[10] to any other errors that may arise in using a model for prediction, including measurement errors of the input parameters to the model as well as fundamental errors in the model functional form.

4. Artificial Neural Networks (ANN) and A Data Set with Variation

Variation occurs naturally in all processes. Typical variation in measurement of inputs associated with the performance of a corrugated box are on the order of 4-5% for measured quantities like ECT and BCS. To further study if and how the ANN model works while handling a data set incorporating variation, we modified the ideal set to represent boxes that might appear in commerce. We added fluctuations to each input value, using randomized, normally distributed values on the order of the variation observed in the different test methods. As with the idealized data set, a "data point" represents a specific set of information defining the physical properties of the box (lengths, width, caliper, and ECT) and the associated BCS calculated by equation (3). The average absolute difference

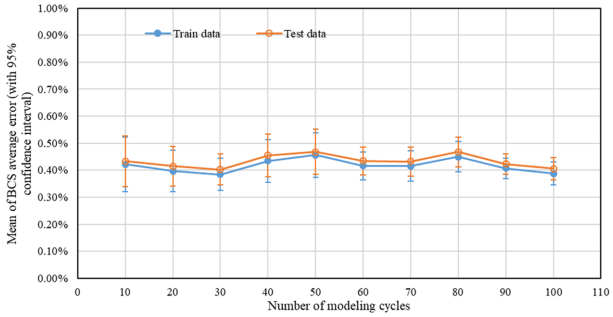

Fig. 10. Mean of average error in estimated box compression strength given different numbers of modeling cycle (Idealized data set). Note error bars indicate 95% confidence intervals on the mean values.

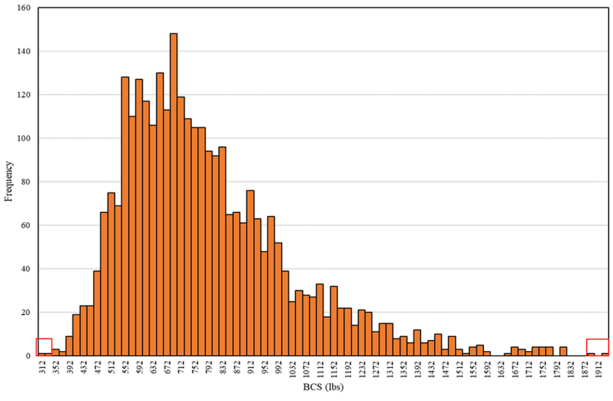

Fig. 11. BCS distribution of 3,009 data points (Idealized data set).

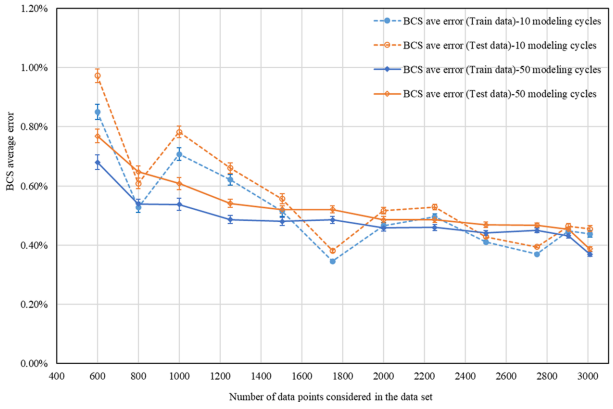

Fig. 12. Average error in estimated box compression strength given different numbers with same modeling cycles of 10 and 50 (Idealized data set). Note error bars indicate 95% confidence intervals on the mean values.

between the new predicted BCS values for the 3,009 data points and the "actual" BCS of the idealized model was obtained by adding variations to the input parameters and calculated by equation (3). This process was carried out to achieve a variation of ±5.4% for BCS. We then followed the same process as for the idealized data set: 67% of the data set (2,016 randomly selected samples) were used for the ANN training process and the remaining were used for evaluation of the model. We used the same number of epochs and neuron numbers in hidden layers as in the idealized modeling.

To explore the impact of the epoch number on the ANN model convergence, the model was run on different subsets of data to 250 epochs. Again, the MSE decreased rapidly in the first several epochs for all data subsets (Figure 13) and the largest data set had the lowest MSE for any epoch. As with the earlier modeling, Figure 13-a) displays the raw MSE calculated for each epoch and Figure 13-b) presents smoothed data to highlight the asymptotic behavior. As expected, since we added variation to the input data, the MSE values are all much larger than the idealized data set of Figure 9.

We modeled different population sizes as above to again identify the influence of population size on ANN accuracy (Figure 14). The accuracy of both the training data and testing data remained relatively consistent as the number of modeling cycles increased. While the accuracy of the training data was in line with expectations from the variation built into the data set (~5.4%), the influence of limiting population size can have a meaningful impact if we don't iterate the process sufficiently. Notably, the ANN approach was not working with any more information than the closed form equation itself, and so the prediction accuracy did not improve upon what we would

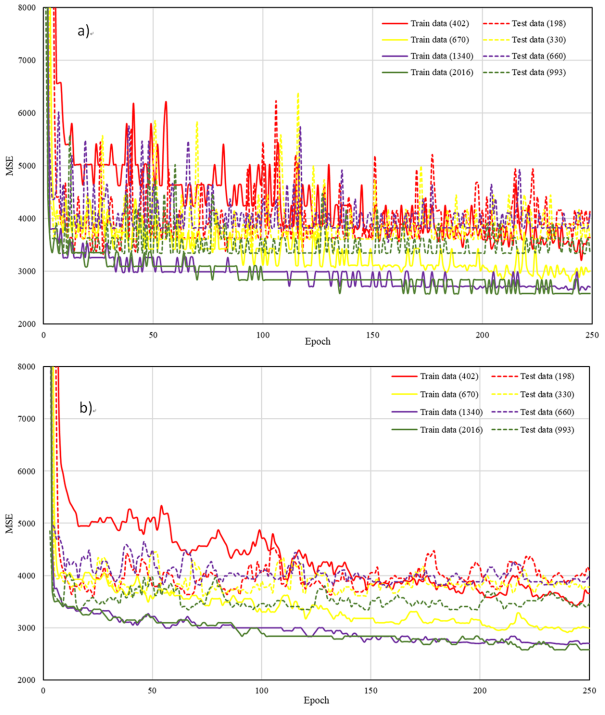

Fig. 13. Mean Squared Error (MSE) of the fits as a function of epochs for different sized data subsets from the variation data set. 10a displays the raw MSE calculated for each epoch, while 10b presents smoothed data, more clearly displaying the asymptotic nature of the functional relationships.

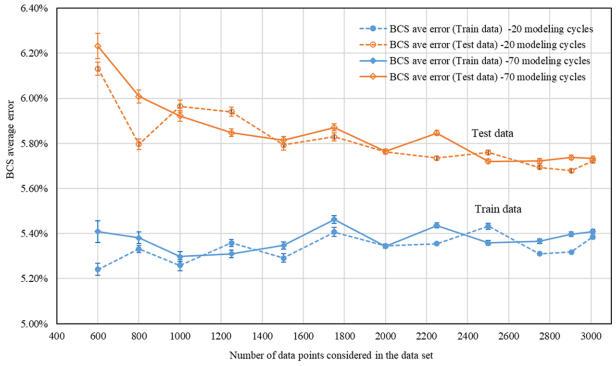

Fig. 14. Average error in estimated box compression strength given different numbers with same modeling cycles of 10 and 50 (Data set with Variation). Note error bars indicate 95% confidence intervals on the mean values.

get from the closed form equation. Test data accuracy didn't begin to converge until around 1500 data points when we used 20 modeling cycles, yet accuracy occurred slightly sooner (~1250 data points) when we used 70 modeling cycles. The BCS average error levels out at 2,500 data points, nearly the entire data set, at a value combining the inherent uncertainty in the input data and the uncertainty of the ANN process itself, identified above. This is notably larger than for the idealized data, because of the influence of variation in the input parameters. As with the idealized data above, the influence of modeling cycles and population size need to be considered together. The minimum data population size to get a robust result is also larger for the variation data set.

Conclusions

In this study we explored BCS estimation using Artificial Neural Networks across input data sets that included both actual data from the literature and data based on literature models. Partitioning each data set into test and training subsets and running multiple modeling cycles on different partitions provides an analysis of average model estimation accuracy that can be expected when the resulting models encounter new data. An ANN model with high accuracy and consistency can be built by adjusting four modeling factors: the combination of neuron numbers in hidden layers, the number of epochs, the number of modeling cycles, and the size of the data set. All four interact to influence model accuracy and can be optimized by minimizing model MSE. The combination of neuron numbers in the two hidden layers was determined as 160 and 36 for the McKee data set, and 142 and 45 for the idealized data set. Employing the same stopping criterion, where the MSE reduction rate is required to be below 3.0%, the epoch numbers were established as 100 for the McKee data set and 140 for the idealized data set. To ensure a robust result with high consistency in the ANN, it was found that 60 modeling cycles are needed for the McKee data set, 50 modeling cycles are required for the idealized data set, and 70 modeling cycles are necessary for the data set with variation. The data size needed to get a robust result varies based on the input data variations and can be identified by minimizing average BCS error: For the McKee data set, 63 data points are not enough for an ANN to predict the BCS reasonably. The other two data sets (idealized data set and data set with variation) need at least 1000 data points to get a robust result for ANN prediction. The data size needed is significant and data collection can be expensive considering the physical testing required. Our ANN models had more degrees of freedom than the number of underlying data sets, which might lead us to expect that we could perfectly fit the underlying data and achieve BCS estimations very close to "measured" values. Instead, we found that model estimation accuracy remains limited by the uncertainty or error in the input parameters combined with uncertainty from the ANN process itself. The variation of input parameters had a positive correlation with an ANN process (high variation increases the training error and vice versa). By identifying the challenges of small data sets and the interrelationship between modeling parameters and the estimation error in the data space, this study provides a methodological guide for future research exploring the applicability of ANN approaches to address problems and answer questions in the corrugated industry.

Author Contributions: Conceptualization, E.L. and J.G.; Methodology, J.G. and E.L.; Software, J.G.; Formal Analysis, J.G., E.L. and B.F.; Investigation, J.G.; Resources, B.F.; Data Curation, J.G. and B.F.; Validation, E.L. and B.F.; Visualization, J.G., E.L. and B.F.; Writing—original draft, J.G., E.L. and B.F.; Writing—review and editing, J.G., E.L. and B.F.; Supervision, E.L. and B.F. All authors have read and agreed to the published version of the manuscript.

References

1. Frank, B., Corrugated Box Compression—A Literature Survey. Packaging Technology and Science 2014. 27(2): 105.
2. Kellicutt, K.Q., & Landt, E. F, Basic design data for the use of fiberboard in shipping containers. *Fibre Containers*, 1951. 36(12): 62-80.
3. Batelka, J.J. and C.N. Smith, Package compression model. Project 3746, final report to the Containerboard and Kraft Paper Group of the American Forest and Paper Association. 1993.
4. Nordstrand, T., Analysis and testing of corrugated board panels into the post-buckling regime. *Composite Structures*, 2004. 63(2): 189-199.
5. Zhou, Y.H., B.Q. Zhong, and R.H. Guo. The multiple linear regression model on compression strength of corrugated boxes. in *Applied Mechanics and Materials*. 2012. Trans Tech Publ.

6. Urbanik, T.J. and B. Frank, Box compression analysis of world-wide data spanning 46 years. *Wood and fiber science*, 2006: 399-416.
7. Maltenfort, G., Revision of top to bottom compression equations for doublewall corrugated. *Paperboard Packaging* 48 (11), 1963.
8. Coffin, D.W., Some observations towards improved predictive models for box compression strength. *TAPPI J*, 2015. 14: 537-545.
9. McKee, R., J. Gander, and J. Wachuta, Compression strength formula for corrugated boxes. *Paperboard packaging*, 1963. 48(8): 149-159.
10. Frank, B. and K. Kruger, Assessing variation in package modeling. *TAPPI JOURNAL*, 2021. 20(4): 231-238.
11. Park, J., et al., Finite element-based simulation for edgewise compression behavior of corrugated paperboard for packaging of agricultural products. *Applied Sciences*, 2020. 10(19): 6716.
12. Park, J., S. Chang, and H.M. Jung, Numerical prediction of equivalent mechanical properties of corrugated paperboard by 3D finite element analysis. *Applied Sciences*, 2020. 10(22): 7973.
13. Park, J.-M., T.-Y. Park, and H.-M. Jung, Prediction of Deflection Due to Multistage Loading of a Corrugated Package. *Applied Sciences*, 2023. 13(7): 4236.
14. Molina, E. and L. Horvath, Development of a Gaussian Process Model as a Surrogate to Study Load Bridging Performance in Racked Pallets. *Applied Sciences*, 2021. 11(24): 11865.
15. Haj-Ali, R., et al., Refined nonlinear finite element models for corrugated fiberboards. *Composite Structures*, 2009. 87(4): 321-333.
16. Hua, G., et al., Experimental and numerical analysis of the edge effect for corrugated and honeycomb fiberboard. *Strength of Materials*, 2017. 49(1): 188-197.
17. Garbowski, T., T. Gajewski, and J.K. Grabski, Estimation of the compressive strength of corrugated cardboard boxes with various perforations. *Energies*, 2021. 14(4): 1095.
18. Jiménez-Caballero, M., et al. Design of different types of corrugated board packages using finite element tools. in *SIMULIA customer conference*. 2009.
19. Garbowski, T., T. Gajewski, and J.K. Grabski, Estimation of the compressive strength of corrugated cardboard boxes with various openings. *Energies*, 2020. 14(1): 155.
20. Han, J. and J.M. Park, Finite element analysis of vent/hand hole designs for corrugated fibreboard boxes. *Packaging Technology and Science*: An International Journal, 2007. 20(1): 39-47.
21. Biancolini, M. and C. Brutti, Numerical and experimental investigation of the strength of corrugated board packages. *Packaging Technology and Science*: An International Journal, 2003. 16(2): 47-60.
22. Garbowski, T., et al., Crushing of single-walled corrugated board during converting: Experimental and numerical study. *Energies*, 2021. 14(11): 3203.
23. Marin, G., et al., Experimental and finite element simulated box compression tests on paperboard packages at different moisture levels. *Packaging Technology and Science*, 2021. 34(4): 229-243.
24. Kobayashi, T., Numerical Simulation for Compressive Strength of Corrugated Fiberboard Box. *JAPAN TAPPI JOURNAL*, 2019. 73(8): 793-800.
25. Fadiji, T., et al., The efficacy of finite element analysis (FEA) as a design tool for food packaging: A review. *Biosystems Engineering*, 2018. 174: 20-40.
26. Kleene, S.C., Representation of events in nerve nets and finite automata. *Automata studies*, 1956. 34: 3-41.
27. Abiodun, O.I., et al., State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 2018. 4(11): e00938.
28. Mitić, V. Benefits of artificial intelligence and machine learning in marketing. in *Sinteza 2019-International scientific conference on information technology and data related research. 2019*. Singidunum University.
29. Sun, A.Y. and B.R. Scanlon, How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environmental Research Letters*, 2019. 14(7): 073001.
30. Pi, Y., Machine learning in governments: Benefits, challenges and future directions. *JeDEM-eJournal of eDemocracy and Open Government*, 2021. 13(1): 203-219.
31. Kalogirou, S.A., Artificial neural networks in renewable energy systems applications: a review. *Renewable and sustainable energy reviews*, 2001. 5(4): 373-401.
32. Aylak, B.L., et al., Application of machine learning methods for pallet loading problem. *Applied Sciences*, 2021. 11(18): 8304.
33. Garbowski T, Knitter-Piątkowska A, Grabski JK. Estimation of the Edge Crush Resistance of Corrugated Board Using Artificial Intelligence. *Materials,* 2023, 16, 1631.
34. Lisboa, P.J., A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural networks*, 2002. 15(1): 11-39.
35. Smith, J., Advances in neural networks and potential for their application to steel metallurgy. *Materials Science and Technology*, 2020. 36(17): 1805-1819.
36. Sheikhtaheri, A., F. Sadoughi, and Z. Hashemi Dehaghi, Developing and using expert systems and neural networks in medicine: a review on benefits and challenges. *Journal of medical systems*, 2014. 38: 1-6.
37. Adamopoulos, S., et al., Predicting the properties of corrugated base papers using multiple linear regression and artificial neural networks. *Drewno: prace naukowe, doniesienia, komunikaty*, 2016. 59.
38. Malasri, S., P. Rayapati, and D. Kondeti, Predicting Corrugated Box Compression Strength Using an Artificial Neural Network. *International Journal*, 2016. 4(1): 169-176.
39. Archaviboonyobul, T., et al., An analysis of the influence of hand hole and ventilation hole design on compressive strength of corrugated fiberboard boxes by an artificial neural network model. *Packaging Technology and Science*, 2020. 33(4-5):

171-181.
40. Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 2016: MIT press.
41. Roodschild, M., J. Gotay Sardiñas, and A. Will, A new approach for the vanishing gradient problem on sigmoid activation. Progress in *Artificial Intelligence*, 2020. 9(4): 351-360.
42. Kubat, M. and M. Kubat, Artificial neural networks. An Introduction to Machine Learning, 2021: 117-143.
43. Hang, H.M. and Y.M. Chou, Chapter 5 - Motion Estimation for Image Sequence Compression. This work was supported in part by the NSC grant 83-0408-E009012, in *Handbook of Visual Communications*, H.-M. Hang and J.W. Woods, Editors. 1995, Academic Press: San Diego. 147-188. ISBN: 9780123230508.
44. Hyndman, Rob J. Moving Averages. *International Encyclopedia of Statistical Science*. 2nd ed.; Editors: Lovric, Miodrag. Publisher: Springer. (2011); 866-869.