UM-SJTU JOINT INSTITUTE

ECE4500J 2024 Spring Term

Major Design Experience

Group 15's Design Review 3: Report

Replacing a cloud based computation tool on DrBoxOnline.com with faster running neural network

Instructed by

Dr. Jigang Wu

Sponsored by

Dr. Shane Johnson, Associate Professor of UM-SJTU JI

August 12, 2024

Name	**Email**
Yanzhuo Cao	yanzhuo@umich.edu
Fengyu Zhang	zhangfengyu@umich.edu
Keye Chen	keyechen@umich.edu
Shuo Deng	denso@umich.edu
Yukuan Zhu	yukuanz@umich.edu [1]

Abstract

This project tackles a critical issue in the packaging and delivery industry, focusing on structural behaviors of corrugated paper boxes. Although these boxes are widely used because of their lightweight, recyclable, and customizable properties, they are prone to buckling during storage and shipment, resulting in product damage, lost revenue, wasted resources, and customer complaints. Currently, the industry can refer to [2] the Dr. Box Calculator Pro [1], which employs Finite Element Analysis (FEA) to predict box performance under various conditions. While effective, FEA is time-consuming and demands significant computational resources. To address these limitations, we propose integrating Deep Neural Networks (DNN) into the Dr. Box Calculator Pro to replace part of the FEA calculation. DNN's ability to model complex non-linear relationships through multiple layers offers advantages such as automatic feature learning, scalability, and end-to-end learning, making them well-suited for big data applications. Our solution aims to provide instant predictions of buckling strength, reduce computational costs, and efficiently handle diverse box types and conditions. The expected outcome is enhanced prediction accuracy, faster analysis, and overall improved efficiency in packaging and logistics operations. The primary challenge is achieving accurate and efficient predictions of buckling strength, with specifications including high predictive accuracy ($\geq$ 90%), fast analysis rate ($\leq$ 5 minutes per simulation), the ability to handle various box configurations, and system reliability above 99.9% in practice. The project plan is methodically structured. In the initial two weeks, the project will commence with requirement analysis, followed by two weeks of data preparation and cleansing. The subsequent four weeks will focus on developing and initially training the DNN model. The final five weeks will involve optimizing the DNN model, integrating it with the Dr. Box Calculator Pro, conducting system testing, and validating it against traditional FEA methods. The integration of DNNs is anticipated to revolutionize the packaging industry by offering faster, more accurate, and scalable predictions of buckling strength. This advancement will reduce operational costs and enhance customer satisfaction by minimizing product damage during transit. The motivation behind this project is to apply modern machine learning techniques to solve long-standing industry problems, improving the efficiency and effectiveness of packaging and logistics operations.

[2] **SOLVE:** The website of Dr. Box should be provided. It is not the only solution.

Contents

1 Introduction

1.1 Background

Our project addresses critical challenges in the packaging and logistics industry, particularly concerning corrugated paper boxes. These boxes are extensively used due to their lightweight, recyclable, and customizable properties. However, they are prone to buckling during storage and shipment, leading to product damage, revenue loss, wasted resources, and customer complaints. One of [3] the current solution to this problem is the Dr. Box Calculator Pro, a website that provides users with an interface to predict the performance of corrugated boxes using Finite Element Analysis (FEA). As shown in figure 1 [4] FEA is a numerical method for predicting how a product reacts to real-world forces, vibrations, heat, and other physical effects by dividing a complex problem into smaller, simpler parts (finite elements) and solving them. Despite its effectiveness, FEA is time-consuming and demands extensive computational resources. Our proposed solution involves replacing FEA with Deep Neural Networks (DNN) in the Dr. Box Calculator Pro. DNNs are artificial neural networks with multiple layers between the input and output layers, allowing them to model complex non-linear relationships. DNNs offer several advantages, including automatic feature learning from raw data, scalability to handle large-scale data and complex models, and the capability for end-to-end learning, which simplifies the prediction pipeline. Our expected outcomes are improved accuracy, speed and lower inference cost. By incorporating DNNs into the Dr. Box Calculator Pro, we can achieve instant predictions of buckling strength, significantly reduce computational costs, and expedite analysis. Additionally, DNNs can handle diverse box types and conditions, enhancing the efficiency and effectiveness of packaging and logistics operations.

[3] **SOLVE:** It is not the only solution.

[4] **SOLVE:** In section 1.1, it will be nice to include one or two figures, showing the input and output of the problem. Already added as figure 1

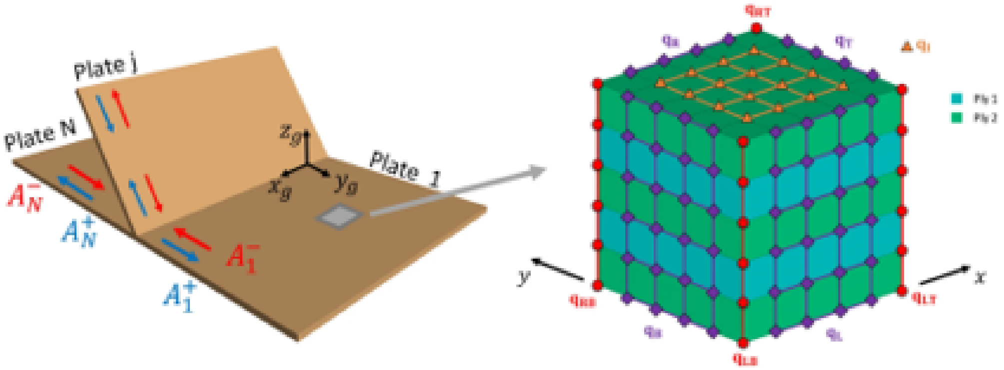


Figure 1: FEA for Corrugated Paper Box

1.2 Literature Review

1.2.1 Traditional Finite Element Analysis (FEA) Approaches

Finite Element Analysis (FEA) has been extensively used to predict the buckling strength of corrugated cardboard boxes, offering a robust and accurate method for such analyses. For instance, as shown in figure 2, Feher et al.[2] analyzed the compressive strength of corrugated paperboard packages with different cutout rates using FEA. Their study, which employed a homogenized linear elastic orthotropic material model with Hill plasticity and modeled only one-eighth of the box to simplify analysis, demonstrated high accuracy in predicting compression force from Box Compression Tests (BCT) despite a slight decrease in accuracy at higher cutout rates.

Traditional methods for analyzing the strength and failure of corrugated boxes include experiments, analytical models, and computational software. However, these methods require extensive experimentation and are often limited to specific box types and loading conditions. For example, analytical models like the McKee equation are limited to vertically loaded regular slotted containers (RSC).

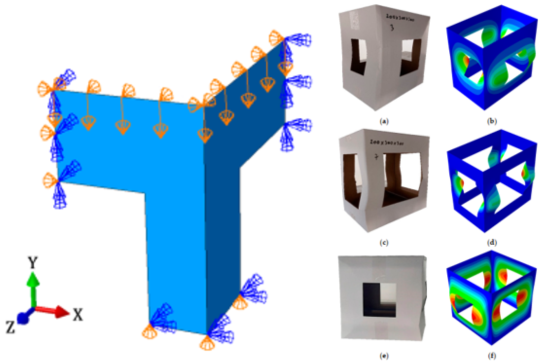


Figure 2: Numerical Analysis of FEA by Feher et al.[2]

1.2.2 Integration of Artificial Neural Networks (ANN) with FEA

A significant advancement in predictive modeling involves integrating Artificial Neural Networks (ANN) with FEA. This approach has been explored to improve prediction accuracy and computational efficiency, as evidenced by study of Hajializadeh [3] on predicting residual stresses in AISI 304 L parts created through the Direct Metal Deposition (DMD) process. Their innovative model integrated ANN with FEA, focusing on the thermal and mechanical response of materials. The ANN was trained using datasets generated from FE simulations, including spatial coordinates and temperature history, to predict residual stresses. As also shown in figure 3, this method not only enhanced prediction accuracy but also significantly reduced computational time.

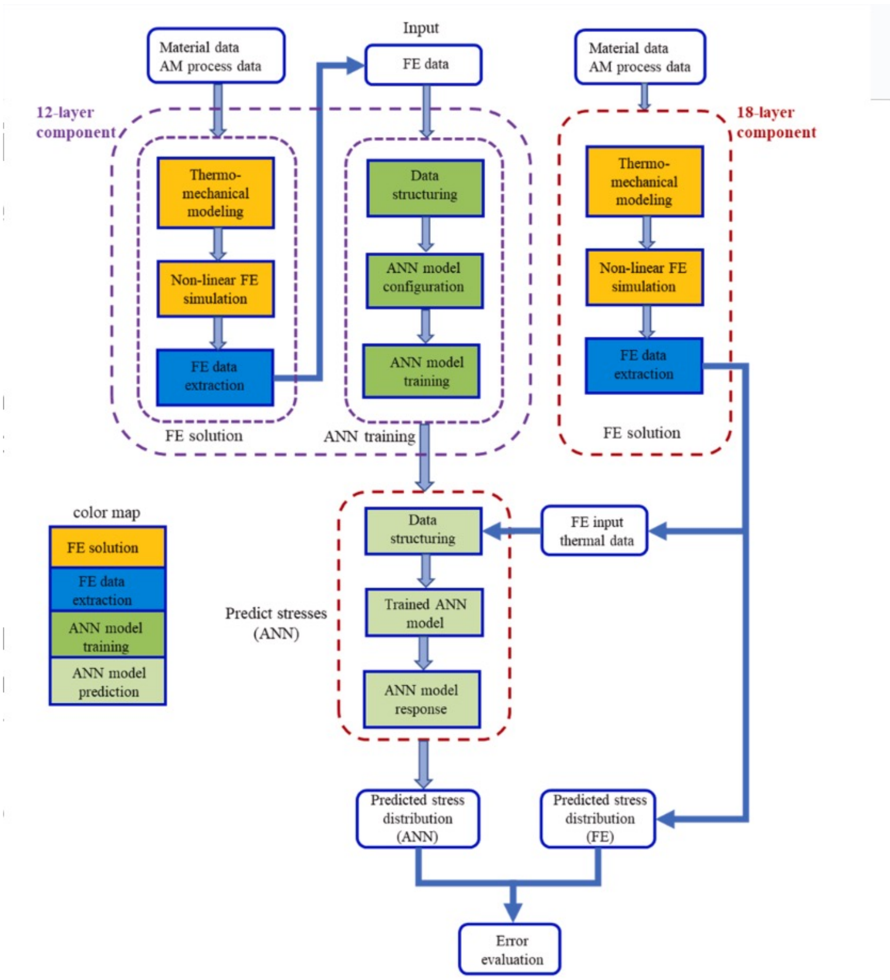


Figure 3: Training Process for DMD [3]

The novel aspect of integrating ANN with FEA lies in its ability to handle complex data relationships more effectively than traditional methods. This integration allows for quicker evaluations and can be adapted to different box types and materials, providing a flexible and scalable solution for packaging analysis.

1.2.3 Comparison and Novelty of the Proposed Design

In comparison to the aforementioned studies, our proposed design integrates the strengths of both traditional FEA and ANN to provide a comprehensive and efficient analysis tool for corrugated cardboard boxes. Unlike the study by Feher[2], which solely relied on FEA, and the work of Hajializadeh[3], which focused on metal materials, our design applies ANN-enhanced FEA specifically to corrugated cardboard boxes. This approach addresses the limitations of traditional methods by improving prediction accuracy and reducing computational time, making it suitable for a wider range of packaging solutions.

The novelty of our design lies in its ability to predict box performance based on paper properties for various box types and designs, including special features like hand holds and ventilation holes. Moreover, in contrast, our model is further designed for multi-task labeling, which includes nodal Deformation x, nodal Deformation y, nodal Deformation z, nodal Failure, displacement and force. Each dimension has more 128 outputs to generate. Like figure 4, it also shows a broader application in the physical field.

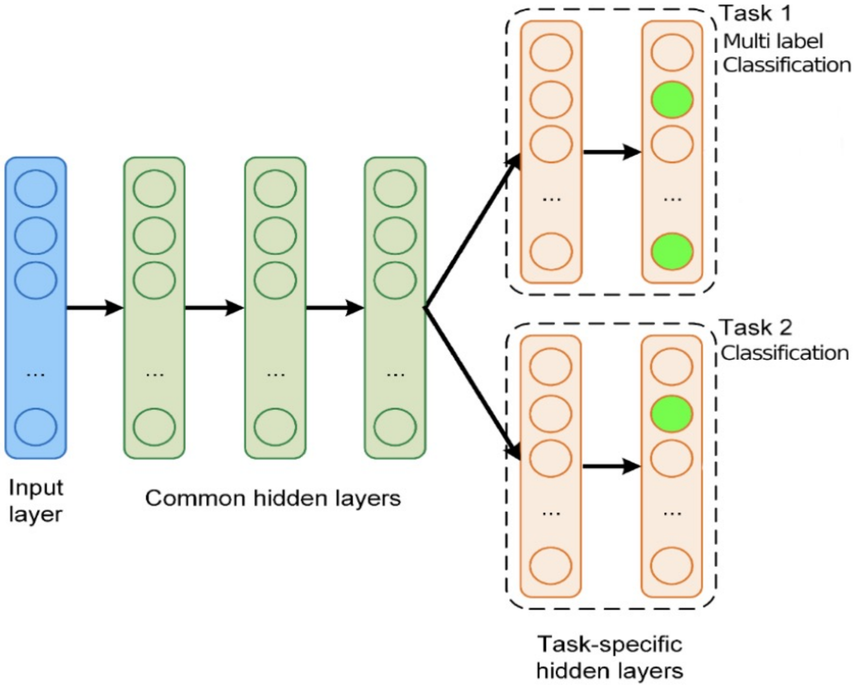


Figure 4: Multi-task Predictions [3]

This hybrid approach combines the accuracy of FEA with the efficiency of ANN, enabling non-specialized personnel to conduct robust and cost-effective packaging analyses.

2 Customer Requirements and Engineering Specifications

2.1 Customer Requirements

At the beginning, the customer requirements for us were only relatively high accuracy. After many offline discussions, we came up with the following detailed customer requirements. [5]

- **Good Simulation:** High-quality predictions that accurately reflect real-world outcomes.
- **Very Fast:** The system should deliver results rapidly to enable efficient decision-making.
- **Can Accept Various Inputs:** The system must be capable of handling diverse input variables to accommodate different user needs.
- **Very Reliable:** The system needs to operate with high reliability and minimal downtime.
- **User-Friendly :** The interface should be easy to use for all user levels, ensuring a low learning curve.
- **Inexpensive:** The solution should be cost-effective for the user.

2.2 Process to Determine Engineering Targets

After confirming the detailed customer requirements, we held several internal seminars to determine the engineering specifications and the granularity. [6]

- **Good Simulation (Accuracy, Feature Learning, End-to-End Learning)**
  - *Requirement Translation*: Convert the need for "high-quality predictions" into a specific accuracy metric.
  - *Specification*: Achieve at least 90% accuracy in simulation predictions.
  - *Metric*: Use confusion matrices, ROC curves, and precision-recall metrics from validation datasets to measure.

[5] **SOLVE:** n section 2.1, some description on how the CRs are obtained should be included.
[6] **SOLVE:** Similarly, in section 2.2, some descriptive sentences should be added before the list.

- *Process*: Involve feature learning capabilities in the DNN to automatically discern and utilize the most impactful features from input data, enhancing prediction accuracy through deep learning techniques.

- **Very Fast (Speed, System Optimization)**
  - *Requirement Translation*: Define what "deliver results rapidly" means in a measurable time frame.
  - *Specification*: Reduce simulation time to under 5 minutes per run.
  - *Metric*: Implement benchmarking tests to track and verify the response times across various operations within the system.
  - *Process*: Optimize the computational efficiency of the DNN model by refining algorithmic approaches and utilizing efficient computing resources.

- **Can Accept Various Inputs (Scalability)**
  - *Requirement Translation*: Specify the types and variety of inputs the system should handle.
  - *Specification*: Support at least 20 distinct types of input variables.
  - *Metric*: Test the system's ability to process and accurately analyze different data types and sources.
  - *Process*: Design the system architecture to be modular, allowing easy integration of new input variables without significant modifications.

- **Very Reliable (Server Reliability, System Optimization)**
  - *Requirement Translation*: Translate the need for high reliability into specific uptime and redundancy requirements.
  - *Specification*: Achieve 99.9% system uptime.
  - *Metric*: Use server monitoring tools to track uptime and automatically detect and resolve potential downtimes.
  - *Process*: Implement fault tolerance mechanisms such as redundancy and regular maintenance checks to ensure consistent performance.

- **User-Friendly (User Interface Design)**
  - *Requirement Translation*: Define what makes the interface "easy to use".

- *Specification*:Achieve a user satisfaction rating of at least 80% on usability.
- *Metric*: Conduct user testing sessions and gather feedback to assess the intuitiveness of the interface.
- *Process*: Apply principles of user-centered design during the development phase to ensure the interface is intuitive and accessible to all user levels.

- **Inexpensive (Cost)**
  - *Requirement Translation*: Quantify what ″cost-effective″ means in comparison to competing solutions.
  - *Specification*: Keep the cost per simulation below a competitive benchmark, say 20% less than similar services.
  - *Metric*: Regularly review cost structures and market prices to ensure competitiveness.
  - *Process*: Streamline operations and utilize cost-effective technologies and methodologies to reduce overheads and pass these savings on to users.

2.3 Quality Function Deployment (QFD) and Specific Engineering Requirements

2.3.1 Critical-to-Quality Engineering Specifications

[7]

In this section, we delve into the critical-to-quality engineering specifications derived from our QFD analysis. The QFD diagram, as depicted in table 1, illustrates the correlation between customer requirements and engineering specifications, which are pivotal in ensuring product quality and meeting customer expectations. [8]

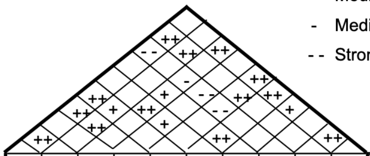
[7] **SOLVE:** In section 2.3.1, discussions on the CTQ ESs should be included.
[8] **SOLVE:** In section 2.3.1, ″... as below″ should be revised as ″as shown in Fig. 4″.

Table 1: QFD Diagram

Quality Function Development (QFD)

Correlation between eng. S

++ Strong Positive
+ Medium Positive
- Medium Negative
- - Strong Negative



Benchmarks

	Weight (1-10)	DNN Integration	Feature Learning	Scalability	End-to-End Learning	System Optimization	Server Reliability	User Interface Design	Cost	Accuracy	Speed	*competitor A*	*competitor B*	
Good simulation	10	10	8		5	10	7			10		9	8	
Very fast	8	8	7	7	7	10		7	7		10	5	4	
Accept various inputs	8	7	6	10	8	7		7		7		10	7	
Very reliable	7	5			4	8	10		7	5	6	9	7	
User friendly	5				3	6	6	10				5	4	
Inexpensive	7	7		5		10	8	8	10	6	7	5	4	
Measurement Unit		MSE	Score	GB	h	LOP	%	Score	¥	%	s			
Target Value							99	20		90	300			
Total		304	184	171	213	392	226	218	175	233	171			
Normalized		0.13	0.08	0.07	0.09	0.17	0.10	0.10	0.08	0.10	0.07			
competitor A														
competitor B														

Correlation between customer reqs and eng. Specs
9 = Strong Relationship
3 = Medium Relationship
1 = Small Relationship
(blank) = Not Related

The weights assigned to each engineering specification are indicative of their importance in the overall product development process. For instance, 'Good simulation' and 'Very fast' have been assigned high weights, reflecting their critical nature in satisfying customer needs.

For "Good simulation": This specification has a strong positive correlation with customer requirements, indicating that a high-quality simulation is essential for meeting customer expectations. The normalized weight of 0.13 underscores its significance. For "Very fast": Speed is another critical factor, with a normalized weight of 0.08, suggesting that the system's responsiveness is crucial for user satisfaction. For "Accept various inputs": This specification is also highly valued, with a normalized weight of 0.07, indicating the importance of flexibility in handling diverse input types. For "Very reliable": Reliability is a key concern, reflected in the normalized weight of 0.09, highlighting the need for a dependable system. For "User friendly": With a normalized weight of 0.10, user-friendliness is a significant factor in ensuring ease of use and overall customer satisfaction. For "Inexpensive": Cost-effectiveness, represented by a normalized weight of 0.07, is another important consideration, balancing quality with affordability.

2.3.2 Specific Engineering Requirements

Since our customers want to ensure all the requirements quantitatively and all the measuring method clearly,[9] after a detailed discussion with the customer, we made another table of specific engineering requirements as below as table 2.

Table 2: Table of Specific Engineering Requirements

Requirements	Specifications	Target	Measurement
Accuracy	Predictive accuracy within 90%	≥ 90% accuracy	Statistical analysis of prediction vs actual
Speed	Analysis time per simulation	≤ 5 minutes	Time tracking software
Customization	Support diverse conditions	Handle 20+ variables	System configuration checks
Reliability	System operational uptime	99.9% uptime	Server Monitoring tools

3 Concept Generation

3.1 Data Augmentation

From the perspective of customization, some sub-function includes stable prediction for boxes with different scales and the ability to capture environmental symmetry. We came up with several ideas to meet these requirements.

- Oversampling: Different sampling probability can be assigned to the data bins respectively based on their frequency to balance the distribution of training data.

[9] **SOLVE:** What is the relations between "Specific Engineering Specifications" and previous ESs? Why are the specific ESs described separately?

- Customized Loss Function: We can assign larger weight to sparsely distributed data points when calculating the loss function. This can make minority data more important during back propagation.
- Data Augmentation: By generating new data with x-y spatial symmetry from the original ones, we allow the model to learn spatial symmetry conditions while relieving the imbalance by expanding the training data.

Finally, we proposed augmenting data with spatial symmetry, which would allow the model to learn symmetry conditions while relieving the imbalance by expanding the training dataset.

3.2 Activation Function

To achieve system reliability above 99.9%, our solution need to ensure stable converging performance during repeated model training sessions, acceptable variance of (MSE) loss across multiple validation sets, and maintaining high reliability when handling extremely large or small inputs. Since the only non-linearly is induced by the activation functions, we basically discuss the choices of common activation functions.

- Sigmoid
$$\sigma(x) = \frac{1}{1+e^{-x}}$$
- Hyperbolic Tangent
$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
- Relu Function
$$\mathrm{ReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$
- Leaky ReLU
$$\mathrm{LeakyReLU}(x) = \max(\alpha x, x)$$
- Gaussian Error Linear Unit
$$\mathrm{GELU}(x) = \frac{x}{2} \cdot (1 + \tanh(\sqrt{2/\pi} \cdot (x + 0.044715 \cdot x^3)))$$

After intense tests on the model, we founded that relu out-performs in regular regression model while sigmoid out-performing in probability models.

3.3 Dimension Reduction

As we deal with high-dimensional data outputs, such as 128-dimensional vectors representing various deformation parameters, dimension reduction becomes essential. By dimension reduction such as RCA and t-sne, we may be able to effectively reduce the dimensionality from 128 to a more manageable 1 to 20 dimensions. This not only aims to simplify the data, making it easier to analyze and interpret, but also try to preserve the global structure of the data. This is critical for understanding overarching patterns and trends in the data, which are fundamental for accurate modeling and prediction.

3.4 Optimizer Selection

The choice of an optimizer is integral to the training of neural networks, particularly in terms of convergence speed and the model's performance on new, unseen data. We have optimizer options of SGD, RMSprop as well as AdamW, with their adaptive learning rate and weight decay strategy, ensures that our model not only learns efficiently but also generalizes better to other data. This is especially important in a project where stability and reliability in predictions are paramount. By optimizing these parameters, we can have more stable and effective learning, which is critical in achieving high accuracy and performance in our neural network applications.

3.5 Overall Generated Concepts

As shown in table 3, it is our group's overall generated concepts.

4 Concept Selection Process

In the context of conceptual selection methodology, we will first introduce our concept and aim with respective to engineering specifications to qualitatively and quantitatively analyze our chosen subject and its advantages and disadvantages. This will enable our neural network[10] more accurate and robust, handling limited size inputs to output high dimension result.

[10]**SOLVE:**In this section, we will focus our selection process on different neural network concept chosen as response to comment in DR2 which suggested that we should pay more attention to choosing nn model during concept generation and select process.

4.1 Data Augmentation

We evaluated several strategies to enhance the training dataset for a neural network model predicting the structural behavior of corrugated paper boxes. The decision was made to select "Data Augmentation with x-y spatial symmetry." The following outlines the reasons for not selecting the other considered options: "Oversampling" and "Customized Loss Function."

- **Potential for Overfitting:** Replicating minority class samples excessively can lead the model to overfit these scenarios, which can degrade its performance on new, unseen data.
- **Skewed Data Distribution:** Oversampling does not introduce new information; it merely repeats what is already present, which might not be sufficient for learning complex, varied real-world scenarios.
- **Complexity in Implementation:** Designing and tuning a loss function that weights minority data more significantly is complex and can lead to an imbalance in learning priorities.
- **Balancing Act:** Improper weighting can cause the model to bias towards less frequent patterns at the expense of overlooking common, critical patterns.

As for the advantages of choosing data augmentation with x-y spatial symmetry, which is also shown in Concept Selection Matrix. This method introduces new, practical data variations, enhancing the model's ability to generalize from training to real-world applications. Additionally, augmenting data by simulating physical changes such as orientation relative to forces enriches the model's understanding of structural behaviors under varied conditions.

This approach allows the model to develop a nuanced understanding of box behaviors, leading to improved accuracy and performance on diverse datasets.

4.2 Activation Function

This section describes the rationale behind the selection of the activation function for a neural network model designed to predict the structural behaviors of corrugated paper boxes. The decision was made to employ "ReLU (Rectified Linear Unit)" enhanced with "Anomaly Detection and Exception Handling." Compared to more complex activation functions or models, the decision has following advantages:

- **Simplicity and Efficiency:** ReLU is preferred for its simplicity and efficiency in training deep neural networks. It accelerates the convergence of stochastic gradient descent compared to sigmoid or tanh functions due to its linear, non-saturating form.
- **Non-linearity:** While ReLU introduces non-linearity into the model, it maintains a range that allows for effective gradient propagation, crucial for learning complex patterns without the vanishing gradient problem.
- **Enhancing Robustness:** To safeguard the model against potential anomalies during training and inference, anomaly detection mechanisms will monitor activation outputs for unexpected behaviors, such as excessively high or zero values indicative of dying ReLU problems.
- **Exception Handling:** In cases where anomalies are detected, exception handling procedures are implemented to adjust computations or revert to safer states, ensuring model stability and reliability.

However, more complex activation functions like parametric ReLU or ELU (Exponential Linear Unit) could increase the model's capacity but also its susceptibility to overfitting and higher computational cost. For the project's scope, which requires balancing performance with computational efficiency, ReLU provides an optimal compromise without unnecessary complexity.

Thus, the choice of ReLU, enhanced with anomaly detection and exception handling, aligns with the project's goals to develop a robust, efficient, and reliable model. This approach not only simplifies the learning process but also ensures that the network remains stable under various operating conditions, crucial for the predictive analysis of box structures.

4.3 Diemnsion Reduction

Our model outputs features like x, y, z deformation, force, nodal failure, and displacement, each as 128-dimensional vectors containing deformation information of various nodes in a cardboard box. To manage and analyze this high-dimensional data, we need to reduce its dimension from 128 to between 1 and 20 dimensions. We have chosen Principal Component Analysis (PCA) over t-SNE for this purpose due to several reasons:

1. **Interpretability:** PCA simplifies the high-dimensional data by identifying the principal components that account for the most variance, making the outputs easier to understand.

2. **Linear Dimensionality Reduction:** PCA is suitable for our data assuming linear relationships among the deformation parameters. It effectively captures these linear or near-linear relationships.

3. **Preservation of Global Structure:** Unlike t-SNE, which focuses on local relationships, PCA maintains the global structure of the data, a critical aspect in understanding overall deformation patterns and trends.

4.4 Optimizer Selection

Moreover, in neural network training, selecting the right optimizer is crucial for model performance and convergence speed. AdamW is a popular optimizer that combines the benefits of Adam's approach with an improved weight decay strategy [4], enhancing generalization in some scenarios. Here we discuss the reasons for choosing AdamW over SGD or RMSprop from a mathematical perspective. The update rule for AdamW can be expressed as:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} - \eta \cdot \lambda \cdot \theta_t$$

where θ_t represents the parameters at step t, η is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected estimates of the first and second moments, ϵ is a small constant for numerical stability, and λ is the weight decay coefficient.

- **Adaptive Learning Rate:**
  - **SGD** (Stochastic Gradient Descent) uses the basic update rule $\theta_{t+1} = \theta_t - \eta \cdot g_t$, where g_t is the gradient. It lacks an adaptive learning rate mechanism, potentially leading to imbalanced training dynamics across different parameter dimensions.
  - **RMSprop** incorporates an exponential moving average of squared gradients to adjust learning rates per dimension, reducing oscillations during training. Its update rule is $\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t + \epsilon}} \cdot g_t$, where v_t is the moving average of squared gradients.

  AdamW combines advantages from both by using estimates of the first and second moments to adaptively adjust the learning rate for each parameter, promoting more stable and effective optimization, and offering superior convergence speeds and generalization compared to SGD and RMSprop.

4.5 Concept Selection Matrix

Based on our generated concepts, we present three different configurations for the neural network model designed to enhance the predictive analysis of structural behaviors in corrugated paper boxes. Each model incorporates a unique combination of data augmentation techniques, activation functions, dimension reduction methods, and optimizers to optimize performance for specific scenarios:

- **Model A**:
  - Data Augmentation with x-y spatial symmetry
  - Smoother Functions like ReLU and Anomaly Detection with Exception Handling
  - PCA Dimension Reduction Method
  - AdamW Optimizer
- **Model B**:
  - Oversampling
  - More Complex Model
  - t-SNE Dimension Reduction Method
  - SGD Optimizer
- **Model C**:
  - Customized Loss Function
  - Smoother Functions like ReLU
  - PCA Dimension Reduction Method
  - RMSprop Optimizer

After our simulation, we decided to use Model A (Data Augmentation with x-y spatial symmetry, Smoother Functions like ReLU and Anomaly Detection with Exception Handling, PCA Dimension Reduction Method, AdamW Optimizer) as our selected models as well as concepts since Model A obtained the highest score according to table 4.

Table 4: Concept Diagram Diagram

Design criterion	Weight Factor	Unit	Model A			Model B			Model C		
			Value	Score	Rating	Value	Score	Rating	Value	Score	Rating
high prediction accuracy	0.36	%	95	9	3.24	90	7	2.52	92	8	2.88
Feature Learning	0.07	Score	Good	8	0.56	Good	8	0.56	Good	7	0.49
Scalability	0.05	Score	Excellent	8	0.4	Good	6	0.3	Good	6	0.3
System Optimization	0.06	Score	Excellent	9	0.54	Good	6	0.36	Good	5	0.3
Server Reliability	0.12	%	99.9	9	1.08	99	9	1.08	90	6	0.72
User Interface Design	0.02	Score	Fair	6	0.12	Fair	6	0.12	Fair	6	0.12
Cost	0.12	¥	20	9	1.08	40	7	0.84	30	5	0.6
Speed	0.2	s	260	8	1.6	200	9	1.8	400	6	1.2
			8.62			7.58			6.61		

5 Selected Concept Description

According to our previous selection process, we decided to use a model with data augmentation with x-y spatial symmetry, smoother functions like ReLU and anomaly detection with exception handling, PCA dimension reduction method and AdamW optimizer. The detailed layout drawing is shown in figure 5.

5.1 Engineer Design Analysis

5.1.1 General Design Principles

Based on the chosen concepts, we designed models with various structure and test them intensely to tune the weights. During the process, we followed these principles and methods for engineer design:

- Structure: Our models' structure should be adapted to inputs and outputs. For example, we had an output dimension of either 128 representing the number of node or 10 representing the number of principal components when predicting nodal behaviors. Also, the structure should reflect real-world mechanisms.

- Complexity: For model complexity trade-off, adding complexity can better fit the relationship, while reducing complexity can prevent overfitting and boost inference speed. We should adjust the parameters according to the training loss and evaluation loss curves.

- Data: Cross validation is a common practice to evaluate the performance of a model. We randomly splited 20% of the data for cross-validation. Since the data is limited, we also performed data augmentation with spatial symmetry to improve the performance.

5.1.2 Model Design Development

We first tried models without PCA to explore how the models could capture the relationships between node outputs without dimensional reduction.

- Basic MLP model without PCA: We started with the most basic full connection structure, with the basic module of linear layer, relu function, batch normalization and dropout repeated for 3 times. The basic model only had a parameter number of about 270 thousand. It could be seen that the loss on the training set remained high during the training process, indicating that the parameters struggled to represent the relationship between the inputs and outputs. A more complex model was needed.

- Independent layer with self-attention: To capture independence while preventing the model complexity from intense increase, we use a single independent linear layer, followed by self-attention modules simulating the dependency between 128 nodes. The structure of the model accords with the essence of FEA: first assign the nodes and then calculate the interaction. We reduced the parameter num to around 800 thousand, while improving the accuracy to 840%. Nevertheless, we still noticed slight fluctuation at the end of the evaluation loss, which indicated that there might be overfitting.

- Independent layer with common MLP part: To further reduce the complexity, we reserve the independent layer part, while replacing the self-attention part with a common MLP module to capture the dependency. The parameter number was reduced to 430 thousand and the accuracy increased to 0.863

- MLP model with PCA: With PCA we successfully obtained a linear transformation to convert the 128 dimensions into 10 principal components while remaining a variance explanation ratio more than 99.8%. After experiment, we found that MLP model, along with batch normalization and dropout was capable enough for prediction. To reduce the significant scale difference between the value of each node, we normalize the matrix with column standardization before PCA, and use whiten algorithms to ensure balanced variance among the principal components.

5.2 Design Description

The chosen concept integrates a Deep Neural Network (DNN) within the Dr. Box Calculator Pro system, aimed at enhancing the prediction of buckling strength for corrugated paper boxes. This new approach replaces traditional Finite Element Analysis (FEA) with advanced neural network technologies that accelerate processing times and improve accuracy.

System Overview and Operation (Refer to Figure 5)

User Interaction Interface:

- **Input Data:** Users input box specifications, including dimensions, material type, and expected load conditions. This interface is designed for ease of use, ensuring that all necessary data is accurately captured for processing.
- **Result Display:** The system outputs predictions such as maximum load capacity and potential nodal failure points. These results are presented in an accessible format to aid users in decision-making processes.

Training Data:

The DNN is trained on a large dataset comprising historical data that details various box specifications and their performance outcomes. This dataset is vital for training the model to recognize complex relationships and patterns that influence box performance.

Neural Network Model:

- **Hidden Layers:** Specific hidden layers are designated for different prediction tasks—max force and nodal failure—allowing the model to handle diverse aspects of mechanical behavior in corrugated boxes.
- **Optimization:** The network utilizes optimization algorithms to adjust its parameters, aiming to minimize loss functions that directly impact prediction accuracy. Since we have to predict different outcomes with different features as well as different requirements, here we list the two most representative prediction models:
  - **Max Force:** Focuses on minimizing Mean Squared Error (MSE) Loss to refine force prediction.

- **Nodal Failure:** Targets a reduction in a combined MSE and cosine similarity loss, enhancing the accuracy of failure point predictions.

Implementation Details:

- **Data Processing:** Upon entry, user data is preprocessed to conform to the model's input requirements, ensuring optimal performance.
- **Prediction Processing:** The DNN processes the input data, applying learned patterns and behaviors to generate predictions which are then relayed back to the user interface.

Technical Specifications:

- **Back-end Frameworks:** The model operates on a Python-based architecture utilizing TensorFlow, which supports extensive model training and execution.
- **Front-end Development**: The user interface leverages ReactJS to provide a responsive and dynamic experience, facilitating straightforward user interactions.

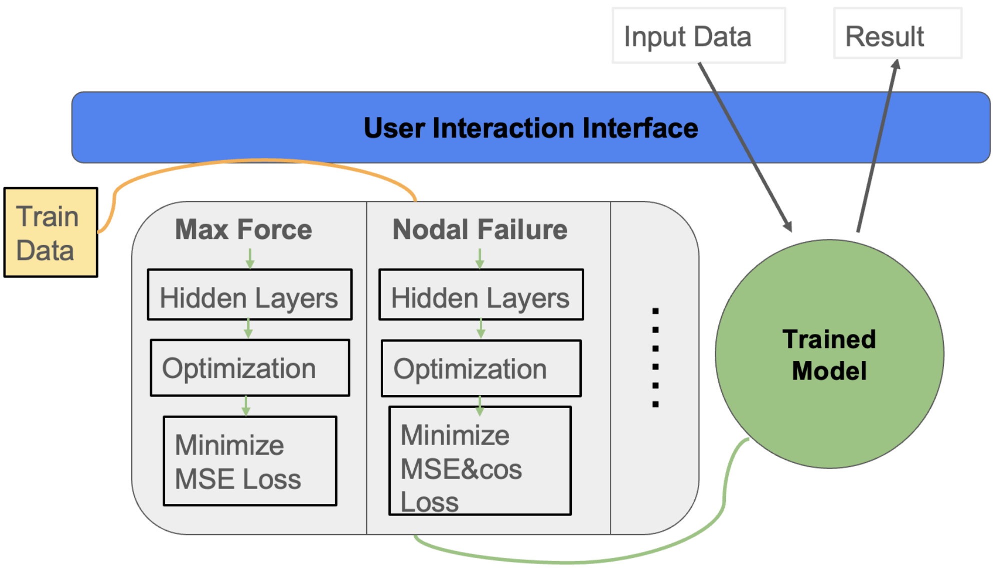


Figure 5: Concept Diagram Diagram

5.3 Manufacturing Plan

5.3.1 Resource Procurement

Hardware Requirements

- **High-Performance Servers:** Acquire high-performance servers with GPU capabilities to handle the computational demands of training and running the DNN models.
- **Storage Solutions:** Ensure ample storage capacity for large datasets and model checkpoints.
- **Networking Equipment:** High-speed networking equipment to facilitate data transfer and communication between servers.

Software Requirements

- **Development Tools:** Set up development environments with necessary software such as Python, TensorFlow, PyTorch, and other relevant machine learning libraries.
- **Database Management Systems:** Implement robust database management systems for storing and managing training data, model parameters, and simulation results.
- **Version Control:** Use version control systems like Git for managing code and collaborative development.

5.3.2 System Integration

DNN Model Integration

- **Model Training:** Utilize the procured hardware to train the DNN models using prepared and cleaned data. Ensure that the training process includes extensive testing to achieve the desired accuracy and performance.
- **API Development:** Develop APIs to allow the DNN models to interface seamlessly with the Dr. Box Calculator Pro. These APIs will handle input data, run predictions, and return results to the users.

User Interface (UI) Development

- **Front-End Design:** Create a user-friendly interface that allows users to input data and view prediction results. Ensure the interface is intuitive and accessible to users with varying levels of technical expertise.
- **Back-End Integration:** Integrate the back-end systems with the front-end interface. Ensure that the data flow between the UI and the DNN models is smooth and efficient.

5.3.3 Testing and Validation

Unit Testing

- Test individual components of the system, including the DNN models, APIs, and UI elements, to ensure they function correctly in isolation.

System Testing

- Conduct comprehensive testing of the integrated system. Simulate various scenarios to ensure the system performs well under different conditions and inputs.
- Perform stress testing to assess the system's ability to handle high loads and maintain performance.

User Acceptance Testing (UAT)

- Involve end-users in testing the system to gather feedback on its functionality, usability, and performance. Make necessary adjustments based on user feedback to enhance the overall user experience.

5.3.4 Deployment

Staging Environment

- Deploy the system in a staging environment that mirrors the production environment. Conduct final testing and validation in this environment to ensure everything works as expected.

Production Deployment

- Gradually roll out the system to the production environment. Monitor the deployment closely to identify and resolve any issues that may arise.
- Ensure that there is minimal disruption to users during the transition from the old FEA-based system to the new DNN-based system.

Documentation and Training

- Provide comprehensive documentation for system administrators, developers, and end-users. This documentation should cover installation, configuration, usage, and troubleshooting.
- Conduct training sessions for end-users to familiarize them with the new system and its features.

5.3.5 Maintenance and Support

Ongoing Monitoring

- Implement monitoring tools to continuously track the system's performance, reliability, and usage. Set up alerts to quickly address any issues that may arise.

Regular Updates

- Schedule regular updates to improve the system's functionality, security, and performance. This includes updating the DNN models with new data and enhancements.

User Support

- Establish a support system to assist users with any questions or issues they encounter. Provide multiple channels for support, including email, chat, and phone.

5.4 Validation Plan

5.4.1 Precision of our Results

As for the real number output ″max force″, we just decide to measure it using precision defined by the following equation on the validation data:

$$Precision = \frac{1}{n} \times \sum_{i=1}^{n} \frac{PredictedMaxForce_i}{ActualMaxForce_i} \tag{1}$$

As for the vector outputs which have 128 dimensions showing the detailed analysis of sample points on the box, just calculate the differences of the predicted value and original value is meaningless. Therefore, we decide to measure it using both cosine similarity (proved by Lahitani et al. [5]) and length accuracy, which are defined as the following equations:

$$\text{Cosine Similarity} = \frac{1}{n} \times \sum_{i=1}^{n} \frac{x_{i1}y_{i1} + x_{i2}y_{i2} + \cdots + x_{in}y_{in}}{\sqrt{x_{i1}^2 + x_{i2}^2 + \cdots + x_{in}^2} \cdot \sqrt{y_{i1}^2 + y_{i2}^2 + \cdots + y_{in}^2}} \tag{2}$$

$$\text{Length Accuracy} = \frac{1}{n} \times \sum_{i=1}^{n} \frac{\sqrt{x_{i1}^2 + x_{i2}^2 + \cdots + x_{in}^2}}{\sqrt{y_{i1}^2 + y_{i2}^2 + \cdots + y_{in}^2}} \tag{3}$$

where the ith predicted result is $x = x_{i1}, x_{i2}, \ldots, x_{in}$ and the ith actual result is $y = y_{i1}, y_{i2}, \ldots, y_{in}$.

Here is our current results, and to make the table more concise, we use ″MF″ to replace ″Max Force″, ″Nx″ to replace ″Nodal Deformation x″, ″Ny″ to replace ″Nodal Deformation y″, ″Nz″ to replace ″Nodal Deformation z″, ″NF″ to replace ″Nodal Failure″:

Table 5: Caption

	MF	Nx	Ny	Nz	NF	Displacement	Force
Precision	93%	92%	95%	82%	39%	70%	85%
Cosine Similarity	NA	99%	99%	99%	90%	99%	99%

As shown in the table5, the cosine similarity is very high for different output parameters, indicating that the directions of our predictions are very close to the actual data. However, the precision of Nodal Failure and Displacement is still not satisfactory. Therefore, we plan to focus on improving our neural networks for these two output parameters. If time permits, we will also aim to enhance the precision of Nodal Deformation z and Force to 90%.

5.4.2 Speed and Cost of our Results

Compared to the original FEA method, which requires 600 cores of computational resources and an average of 30 seconds to generate one result, our method only needs a few milliseconds once the neural network is trained. Additionally, the cost per computation can be reduced from 3 CNY to nearly 0 CNY. After the model is trained, the cost of computing one result becomes negligible. Therefore, we believe there is no need to improve the cost-efficiency of our method further.

5.4.3 Satisfaction of the Customers

Since this model is intended for use by companies (such as courier service companies), ensuring their satisfaction with our product is crucial. Therefore, if time permits after training our neural network, we plan to invite companies that have used the original FEA-based website to try our new model-based website. We will then seek their feedback to gauge their satisfaction with our new website.

6 Project Timeline (refer to figure 6) and Plan

[11]

- In the first two weeks, we initiated the project and analyzing requirements, which we have already completed.
- Moving into weeks three, we concentrated on preparing the necessary data and conducting data cleansing. We cleaned and pre-processed this data to ensure it is of high quality, ready for use in our model development.
- From week four, we started developing the initial Deep Neural Network model. After the initial construction of the modeling framework, we trained this model using the data we have prepared. We also conducted preliminary tests to evaluate the model's performance and accuracy. Based on these tests, we tried to adjust model parameters to optimize performance.
- From week eight till now, we met with sponsor, discussed about current progress and adjusted certain techniques to improve our accuracy and

[11] **SOLVE:** In section 3, it will be better to include a figure such as a Gantt Chart.

simplify our model. As current-stage result, our prediction accuracy has reached to satisfactory outcomes. However, we also observe our current model fail to capture and predict the features of nodal failure, which is one of the most direction for us to resolve before the final presentation.

- At the same time, we try to integrate our DNN model into the DrBoxOnline.com platform. We will establish a complete web front end for our final expo, and perform some system integration tests to ensure compatibility between the model and the platform.
- Finally, in week twelve, we will summarize our project outcomes and showcase our progress and achievements to stakeholders on presentation.

7 Analysis of Potential Problems

Currently, our model struggles to effectively capture and predict the dimensions (x, y, z) and the failure possibility of 128 test nodes, known as *nodal failure*. The challenges primarily stem from the model's limited capability to handle the complexity and variability of the data associated with these nodes.

To address the shortcomings identified in our current approach, we propose the following strategic modifications and enhancements to our predictive model:

- **Modifying Prediction Targets:** We plan to refine our prediction targets by transforming how we approach the problem of predicting nodal failure. Instead of predicting unbounded numerical values, we will shift to a probability prediction model that outputs values between 0 and 1, indicating the likelihood of failure at each node.
- **Data Transformation:** We will implement transformations on the original dataset to better suit our new predictive approach. This includes normalizing the data or applying other statistical techniques to stabilize the variance and scale of the input features.
- **Model Retraining:** With the new prediction targets and transformed data, we will retrain our model. This retraining will focus on optimizing the neural network to better understand and predict the transformed data, thereby increasing the accuracy of predicting nodal failure probabilities.

The planned adjustments and improvements are expected to enhance our model's ability to accurately predict nodal failures, contributing significantly to the overall reliability and effectiveness of our predictive analytics. These changes are pivotal for advancing our understanding of the structural dynamics and integrity of the systems we analyze.

8 Conclusions

This project aims to address the critical issue of buckling in corrugated paper boxes within the packaging and logistics industry by integrating Deep Neural Networks (DNN) into the Dr. Box Calculator Pro, replacing the traditional Finite Element Analysis (FEA) method. The proposed solution leverages the advanced capabilities of DNNs to model complex non-linear relationships, offering significant improvements in prediction accuracy, computational speed, and cost-effectiveness. Through our comprehensive project plan, we have structured the development process into distinct phases: requirement analysis, data preparation and cleansing, initial DNN model development and training, optimization of the DNN model, system integration, and final validation. Each phase is meticulously designed to ensure the successful implementation and integration of the DNN model into the existing system, aiming for high predictive accuracy, rapid analysis, and system reliability. The anticipated outcome of this project is a revolutionary tool that will provide instant predictions of buckling strength, reduce computational costs, and handle diverse box types and conditions with high reliability and user-friendliness. This advancement is expected to enhance the efficiency and effectiveness of packaging and logistics operations, ultimately leading to reduced operational costs and improved customer satisfaction by minimizing product damage during transit. In conclusion, by applying modern machine learning techniques to solve long-standing industry problems, our project not only aims to improve the performance and reliability of the Dr. Box Calculator Pro but also to set a new standard in the packaging industry for predictive analysis and operational efficiency.

References

[1] DR Box Online, "J. shane and k. liping," 2024, accessed: 2024-07-23. [Online]. Available: https://drboxonline.com/

[2] L. Fehér, D. Mrówczyński, R. Pidl, and P. Böröcz, "Compressive

strength of corrugated paperboard packages with low and high cutout rates: Numerical modelling and experimental validation," *Materials*, vol. 16, no. 6, p. 2360, 2023, accessed: 2023-03-15. [Online]. Available: https://doi.org/10.3390/ma16062360

[3] F. Hajializadeh and A. Ince, "Integration of artificial neural network with finite element analysis for residual stress prediction of direct metal deposition process," *Materials Today Communications*, vol. 27, p. 102197, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352492821001896

[4] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017. [Online]. Available: https://arxiv.org/abs/1711.05101

[5] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," in *2016 4th International Conference on Cyber and IT Service Management*, 2016, pp. 1–6, accessed: 2023-03-15. [Online]. Available: https://doi.org/10.1109/CITSM.2016.7577578
[12][13]

[12] **SOLVE:** References should be numbered as [1], [2], ..., so as to be consistent with the format requirements of final report.

[13] **SOLVE:** There are only two references. At least, the link to the website Dr. Box should be included. And it will be better to have more if possible.

9 Biographical Sketch

9.1 Yanzhuo Cao

9.1.1 Future Plans

Study Plan: After graduation from UM-SJTU Joint Institute, I will need to go back to the School of Information, University of Michigan to complete the Global Degree Program. If Everything Goes Well, I will gain a master degree of Big Data Analytic in May, 2025.

Work Plan: After graduation from University of Michigan, I decided to go back to China and find an internet company to work. My job preferences will be in algorithm and data fields rather than development fields since I am really poor at the basic data structure hahaha.

Life Plan: I have a lot of hobbies, and in near future, I would like to read as many politic books as possible since I really think that very important political events will happen in 5 years. Therefore, I will need to get prepared. Besides, I have decided to lose weight – I am 5 kilograms heavier than freshman year /cry.

9.1.2 Self Picture

Figure 7: Yanzhuo Cao

9.2 Fengyu Zhang

9.2.1 Future Plans:

- I am Fengyu Zhang, currently pursuing a Master's degree in Information Science at the University of Michigan. Concurrently, I am completing my senior year in Electrical and Computer Engineering at the Joint Institute (JI), where my academic pursuits are deeply intertwined with my passion for machine learning, including areas such as traditional supervised learning, deep learning, and natural language processing. My aim is to utilize my knowledge to significantly contribute to my team and achieve a successful outcome in our capstone design project.

- In addition to my academic endeavors, I am an intern at a leading technology and e-commerce company. This role provides me with a profound insight into the practical applications of AI in the industry, nurturing my aspiration to one day excel as an algorithm engineer where I can actively engage in innovative AI solutions.
- Outside the realm of technology and academics, I have a profound love for music and am dedicated to mastering the guitar. I believe that music not only enhances my creative expression but also provides a balance to my rigorous academic and professional schedule. Furthermore, I am committed to bodybuilding, which I find essential for maintaining both physical and mental wellness.

9.2.2 Self Picture

Figure 8: Fengyu Zhang

9.3 Keye Chen

9.3.1 Future Plans

Study Plan: After graduation from UM-SJTU Joint Institute in this August, I will need to go back to the School of Information, University of Michigan to complete my master's degree in Big Data Analytics. If Everything Goes Well, I will gain a master's degree in May, 2025.

Work Plan: After graduation from University of Michigan, I decided to go back to China and find a job relevant to Information Technology. Since it would be a really tough experience working in a routine of 996, I will consider pursuing another degree or switching to a more relaxing job after several years.

Life Plan: I am devoted to traveling around the world with my cameras, and drones. I love the photograph and natural scenery! Additionally, I keep a hobby in the gym and I believe that adequate body training can make me always in a positive mode.

9.3.2 Self Picture

Figure 9: Keye Chen

9.4 Shuo Deng

9.4.1 Future Plans

Study Plan: After graduation from UM-SJTU Joint Institute in this August, I will need to go back to the School of Information, University of Michigan to complete my master's degree in Big Data Analytics. If Everything Goes Well, I will gain a master's degree in May, 2025.

Work Plan: After graduation from University of Michigan, I may go back to China and find a job relevant to AI agent and LLM. I'd like to spend several years to participate in the information technology to see how far AI can go.

Life Plan: I am interested in digital painting and my dream is to express the world with my own interpretation. I want to practice more while traveling around during my leisure time.

9.4.2 Self Picture

Figure 10: Shuo Deng

9.5 Yukuan Zhu

9.5.1 Future Plans

Study Plan: After graduation from UM-SJTU Joint Institute in this August, I will need to go back to the School of Information, University of Michigan to complete my master's degree in Big Data Analytics. If Everything Goes Well, I will gain a master's degree in May, 2025.

Work Plan: After graduating from the University of Michigan, I plan to return to China and work for a technology company, maybe working in NLP. If I will be able to earn enough before 40, I plan to become a high school or junior high school teacher after that.

Life Plan: My biggest hobby is traveling. I like to go to scenic places with my friends and then record the beauty with my camera. This is also the reason why I want to become a teacher after 40. I want to travel around the world during summer and winter vacation time.

9.5.2 Self Picture

Figure 11: Yukuan Zhu

Table 3: Overall Generated Concepts

Function	Options	Advantages	Disadvantages
Data Augmentation	• Oversampling • Customized Loss Function • Data Augmentation with x-y spatial symmetry	• Balances training data distribution • Increases importance of minority data • Enhances model's understanding of spatial symmetry	• May introduce bias • Could overfit on minority samples • May not generalize well beyond trained symmetries
Activation Function	• More complex models • Smoother functions like ReLU • Anomaly detection with exception handling	• Prevents extreme values • Stable performance • Warns users of anomalies	• May limit model capacity • Simple functions may not capture complex patterns • Additional computational overhead
Dimension Reduction	• PCA • t-SNE	• Simplifies high-dimensional data • Preserves global data structure	• t-SNE may not scale well with larger datasets • PCA assumes linearity
Optimizer Selection	• SGD • RMSprop • AdamW	• Adapts learning rate • Efficient learning and generalization	• SGD may converge slowly • RMSprop might not perform well on non-stationary problems • AdamW could lead to overfitting in some cases

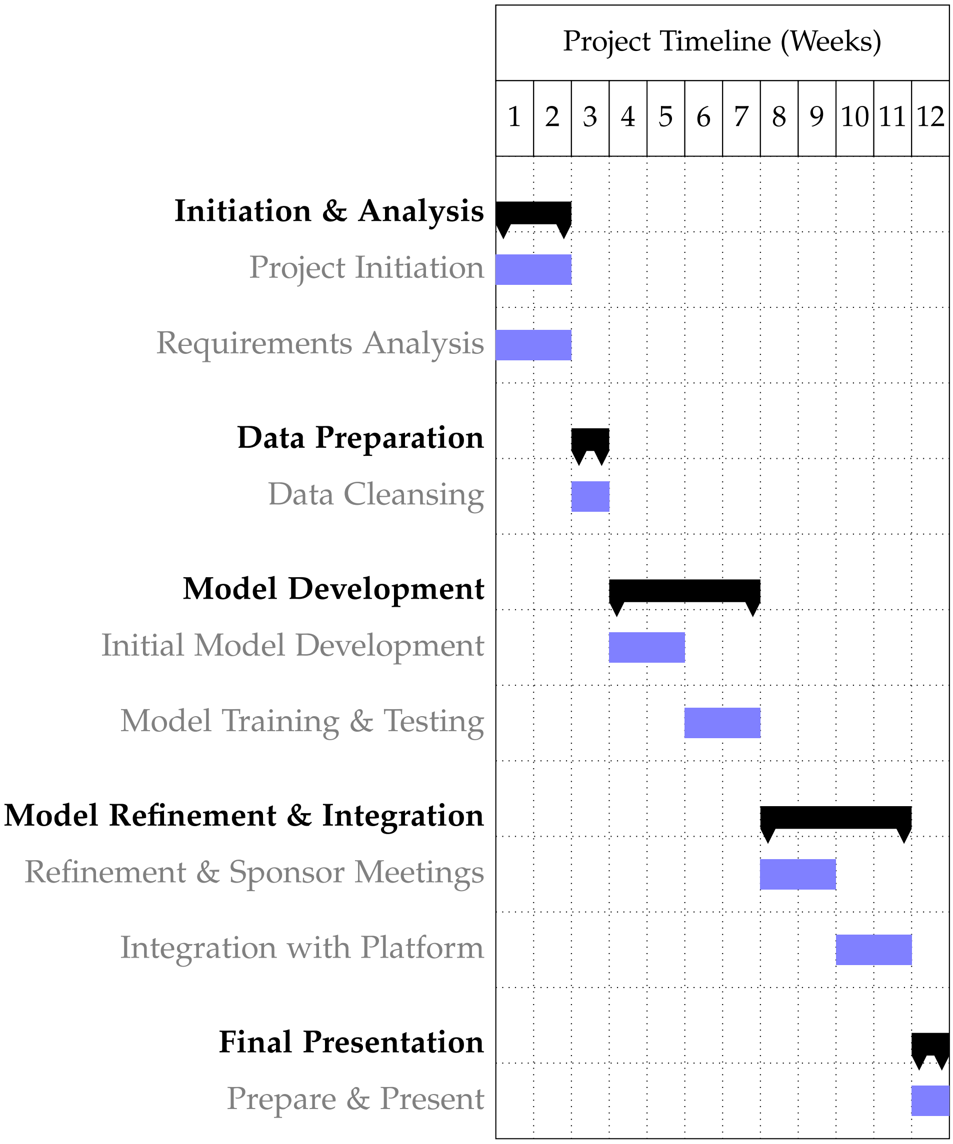


Figure 6: Project Gantt Chart