CBH-Digital Sieve
Project Review and Closure Report

Industry Exchange

CBH Cooperative Bulk Handling Ltd

Ned Dorji, Curtin University

Supervisor: Sean Webb (CBH), Victor, Vinesha & Nick

11/11/2024 15/02/2025

Contents

DOCUMENT ACCEPTANCE

This <Project Title> Review and Closure Report is authorised for release once all signatures have been obtained.

PREPARED: ______________________________ **DATE:** ____________

(for acceptance) (<Intern Name>, <Engineering, etc.> Associate)

ACCEPTED: ______________________________ **DATE:** ____________

(for release) (<Industry Supervisor>)

1 Executive Summary

Screening is an important test to determine the quality of the grain. Using the mechanical method measures the weight of the screening and expressed as fraction of the sample weight. Higher screening indicates lower quality and vice-versa. However, the mechanical method is slow, cumbersome and does not consider the axial measurments of the kernel which could provide better representation of a sieve. An alternative method using the digital image processing of images of the samples caputred on Deimos machine is used to predict the screenings from the sample. The 2-d captured images were analysed to find third axis by considering shape and size of the each kernel using flakiness ratio(thickness-to-breath-ratio). The alternative method has been applied to both laboratory and harvest data and both showed strong correlation between actual weight and estimated weight ratio based on area fractions thus showing it perhaps the physical sieve could be replaced with digital sieve method.

2 Introduction

Cooperate Bulk Handling (CBH) uses several metrics to determine quality of the grain and one of the methods is the screenings test. Screenings test, in another word known as mechanical sieving, is one of the prominent techniques for grading aggregates in terms of particle size distribution[1]. In this test, a sample is placed on the physical sieve called the Agtator, and then shaken for specific number of rounds. For barley and wheat, it is shaken for 40 rounds and 20 rounds for oats. The grains, including dust or any other material, that pass-through sieve is the screenings and is expressed as percentage mass (ratio) of the weight of the grains passed relative to the weight of the sample. Higher screenings indicate lower quality of grade and consequently leads to downgrading of the grade for the grower - example malt barley goes to a feed grade.

Despite its massive significance, the conventional mechanical sieving methods have drawbacks. Due to its mechanical process, sieving is massively time consuming, and this puts massive pressure to samplers in sample sheds when there are huge deliveries by the growers to CBH during harvest seasons. It measures the mass of particles that pass-through sieve against total mass of the sample and does not provide axial dimensions of grains which would be a better determinant of whether kernel passes through or not. Additionally, it also depends on the shape of the particles, sample size, sample placement on the device, density, and method of shaking. An alternative method is required, a method that is efficient while also providing accurate representation of the conventional and mechanical methods.

Digital image processing methodologies for axial dimensional analysis could provide valuable information in determining the screening of samples and generate accurate screening ratios. CBH has access to a device called "Deimos" that performs image recognition and segmentation. A sample of grain is passed through conveyor belt in machine and the system captures images of the grains. The machine then identifies any defects, foreign materials along with the meta-data. Among the metadata collected, axial measurements (major axis and minor axis) provide valuable information about the shape and size of the grains, which is crucial for understanding their behaviour during sieve analysis. By integrating a digital sieve analysis feature into the existing digital system -Deimos, the screenings assessment process can be streamlined, like how other receival standards have been digitized, making the process more time-efficient for growers and CBH operations.

2.1 Background

The current model for sieve analysis in the Deimos machine deploys a binary approach. For each kernel in a sample, the model compares both its major and minor axes against the sieve aperture dimensions (for example, 2.0mm by 12.7mm for barley). The prediction logic is straightforward:

1. If the both the axes are smaller than the sieve apertures, the models predict the kernel is going to pass through

2. 2. If either axis exceeds the sieve dimensions, the model predicts that kernel will not pass through the sieve.

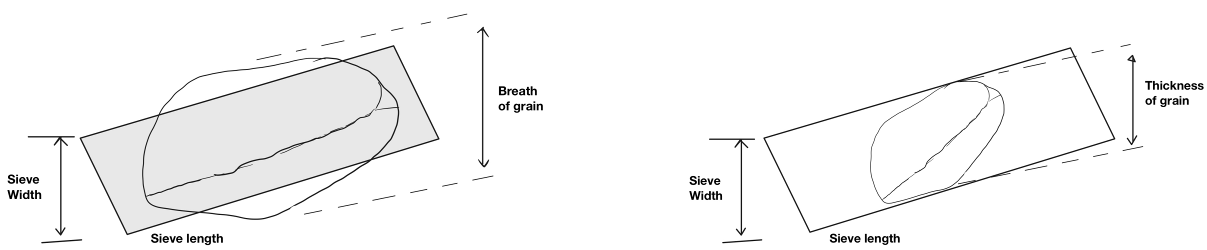
Through experimentation, we found that the major axis (longest dimension) of the grains is consistently shorter than the sieve aperture length, making this measurement irrelevant for predicting sieve passage. Thus, for the existing model, the only available measurement is the breadth of the kernel (minor axis). However, the breadth alone is not a sufficient determinant of whether the kernel will pass or not; the thickness (third axis) and the shape of the grain also play important roles.

The Deimos machines capture only two-dimensional projects of each kernel, but the grains are 3-dimensional, so thickness is currently not measured. One assumption during the capturing process is that the grains tend to lay flat due to vibration before capture. The orientation of the grain in the image renders unfeasible to capture the third dimension. The digital image processing (DIP) captures only the two-dimensional projections of the grains/particles and that the thickness of the grain is not measurable through this process [2]. This poses major problem to visual data captured on Deimos and the ability to the predict passage of grains through the sieve. Both breath and thickness of the grain is crucial in interpreting the results of screening because a grain can pass through the sieve as long as one of its dimensions is lesser than the sieve aperture.

1. A kernel that has breath and thickness both greater than 2.0mm for a sieve aperture of 2.00mm by 12.7mm will not pass through the sieve

2. A kernel that has breath greater than 2.0mm but thickness lesser than 2.0mm will pass through if there is any rotation

3. A kernel that has breath and thickness lesser than 2.0mm will pass through the sieve.



(a) (b)

Figure 1: fig 1. (a) A grain with breath greater than sieve width not passing through. (b) The same grain rotated with thickness lesser than sieve width passing through.

2.2 Requirements

There are two fundamental problems to replicating the conventional mechanical sieving with a digital sieve:

1. What determines whether a grain passes through the sieve? Is it the size, shape, or some other characteristic? Understanding the shape and size distribution of each kernel is crucial to addressing this.

2. Given that a grain is predicted to pass through, how do we compute the ratio of grains predicted to pass through against the total sample based on metrics other than weight? The ratio can then be compared to the actual screening ratio (weight of the screenings to the weight of the sample) to validate the alternative method.

Thus, the main goal for the project was to experiment and developed a visual Proof-of-Concept (PoC) of a digital sieve with an aim to replace the existing the mechanical one. To achieve this overarching goal, there were subsidiary goals:

1. Collect samples in the lab using the mechanical device and capture the sample on the Deimos machine for the visual data analysis.

2. Find the correlations between the screenings ratio of the physical samples (weight-to-weight ratio) and screening ratios of the visual samples (surface area to surface area) captured on Deimos.

3. Extrapolate the lab findings to the harvest dataset and analyse how two ratios correlate to each other.

2.2.1 Hardware

- **Agitator:** A mechanical sieving machine used to determine the screenings for a given sample.
- **Half Liter Chondrometer:** A device used for sampling.
- **Weighing Machine:** A precision scale used to measure the weight of samples and screenings.
- **Deimos Machine:** A device for capturing samples and segmenting each kernel to extract axial measurements (major and minor axes).

2.2.2 Software

- **VS Code:** The integrated development environment (IDE) of choice for the project.
- **Microsoft Excel:** Used for recording data collected from samples.
- **Jupyter Notebook:** An interactive environment used for its flexibility and versatility.
- **Python 3.12:** The programming language of choice.
- **NumPy:** A Python package for mathematical calculation.
- **Pandas:** A Python package for data manipulation and analysis.
- **Matplotlib:** A Python package for creating graphs and visualizations.
- **Seaborn:** A Python package for advanced graphical visualizations.
- **Plotly:** A Python package for interactive graphs and visualizations.

3 Methodology

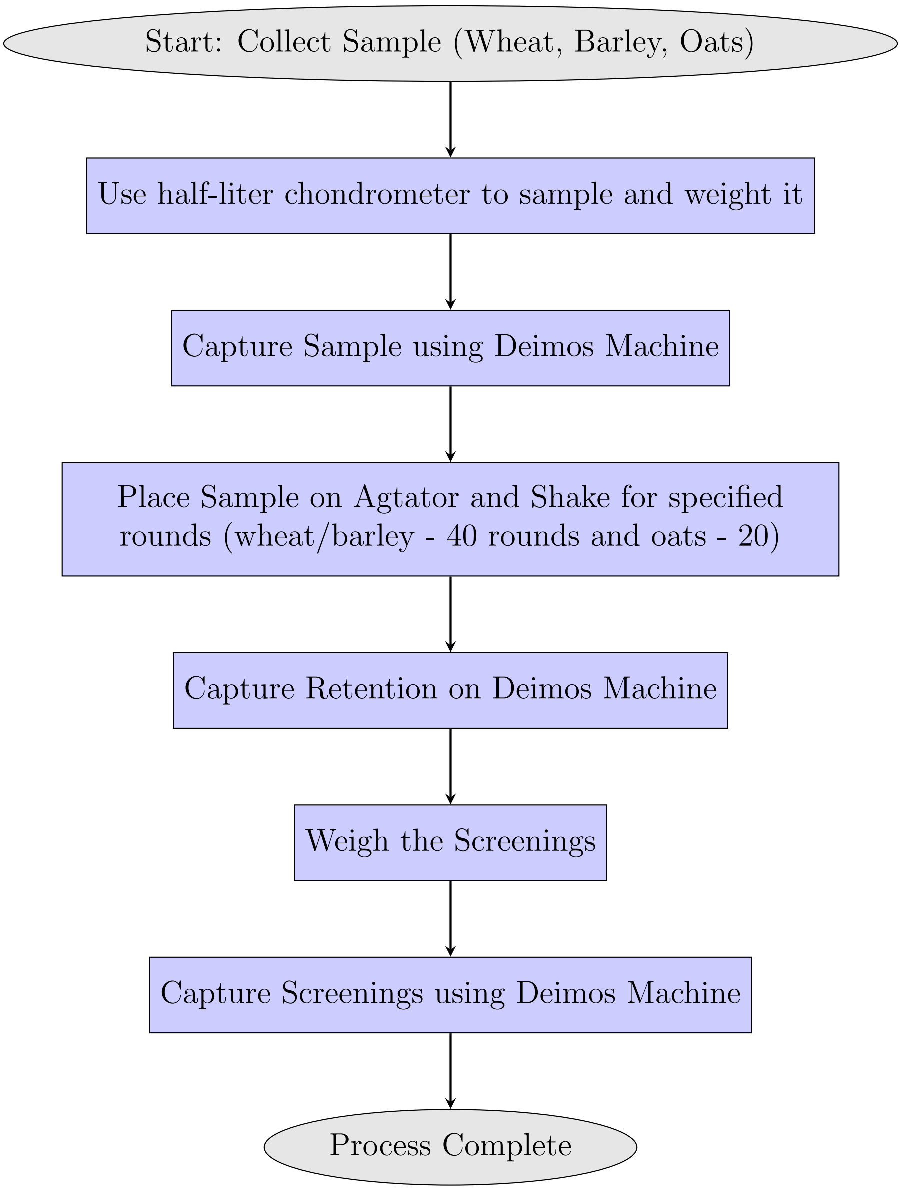
3.1 Data

3.1.1 Laboratory Data

A total of 71 cereal grain samples were collected in the laboratory using mechanical sieve (**agtator**), and images of the these samples were then captured on Deimos machines for the visual analysis.

- 20 barley samples
- 24 oat samples
- 27 wheat samples

3.1.2 Data Collection



3.1.3 Data Scheme

- **Sample-ID:** A unique 8-digit reference ID for each sample. For screenings, -40B was added to the unique reference ID to denote that it is a screening and -40T for retention. For example, a reference ID **1181206283-40B** denotes visual screening data for sample with ID **1181206283**.
- **Grade:** Different grades of a particular commodity.
- **Half-liter-sample-weight:** Weight of the sample.
- **Retention:** Weight of the retained material.
- **Screenings:** Weight of the screenings.
- **Other Meta-data:** Fields to ensure that capturing processes were not missed.
  - **Deimos-captured:** Boolean
  - **Screenings-captured:** Boolean

3.1.4 Harvest Data

A total 1387 of harvest data from 2025 were retrieved from the CBH database. The data contained weight to weight ratio (actual screening ratio) and corresponding visual data was extracted from the cloud similar to the lab data.

- 474 barley samples
- 465 oats sample
- 448 wheat sample

3.2 Data Preprocessing

The Deimos machine employs computer vision and machine learning algorithms to segment and analyze individual kernels within each sample. This data is automatically stored in Microsoft Azure cloud storage. The visual dataset encompasses substantial metadata, including major and minor axes, which are important for sieve analysis. Each sample contains approximately 10,000 individual kernels, and for 30 samples, this amounts to roughly 300,000 kernels requiring analysis at the kernel level. A comprehensive data folder was provided, containing all these necessary information.

The data folder had the following hierarchy:

```
root_directory/
          sample_id (32-character UUID)/
              tray_id (32-character UUID)/
                  model (32-character UUID)/
                      instances.json
```

A Python script was developed to convert the raw meta-data to corresponding CSV file. And the pre-processing workflow involved:

1. **Download Dataset**
   - Obtain the complete dataset folder containing all configuration files and sample data.
2. **Sample ID Mapping**
   - Read the configuration file to get the physical sample ID.
   - Cross-reference this with VA data to find the corresponding **VA sample ID** (32-character UUID) universally unique identifier.
3. **Tray Processing**
   - Navigate to the sample ID folder (UUID format).
   - Identify and iterate through all tray folders within the sample ID directory.
4. **Model Folder and Reading Metadata**
   - Access the **model** folder within each tray directory.
   - Locate and read the `instances.json` file.
5. **Data Transformation**
   - Extract the following measurements/properties from `instances.json`:
     - **majorAxisMm** Major axis measurement in millimeters (length).
     - **minorAxisMm** Minor axis measurement in millimeters (breadth).
   - Export the data in **CSV format** for further analysis.

3.3 Data Analysis

3.3.1 Estimation of Third Axis - thickness

add some images
The third dimension of the kernel is an important determinant of digital screenings. To approximate the third dimension, it is essential to understand the morphology of each kernel. Grains have complex and irregular shapes, which makes it harder to estimate their size. However, two important propertieselongation and flakinessare key to determining shape and size. Elongation is the ratio of breadth to length, while flakiness is the ratio of thickness to breadth[2].

According to [2], particles originating from the same source are assumed to exhibit identical characteristics. This assumption has been applied to the present study, and based on this premise, the ratio linking the breadth and thickness of each kernel has been calculated. Thus, an average thickness can be approximated as follows:

$$\textbf{average thickness} = \lambda \times breath \tag{1}$$

Using equation 1, we can estimate volume of each grain

$$\begin{aligned} \textbf{Volume} &= \text{Average Thickness} \times \text{Projected Area} \\ &= \lambda \times \text{Breadth} \times \text{Projected Area} \end{aligned} \tag{2}$$

The projected area is the area of the flattened kernel captured by the Deimos machine, which can be assumed to be an **ellipse**. Thus, the area is calculated as:

$$A = \pi \times a \times b \tag{3}$$

where A represents the projected area, a is the semi-major axis, and b is the semi-minor axis.

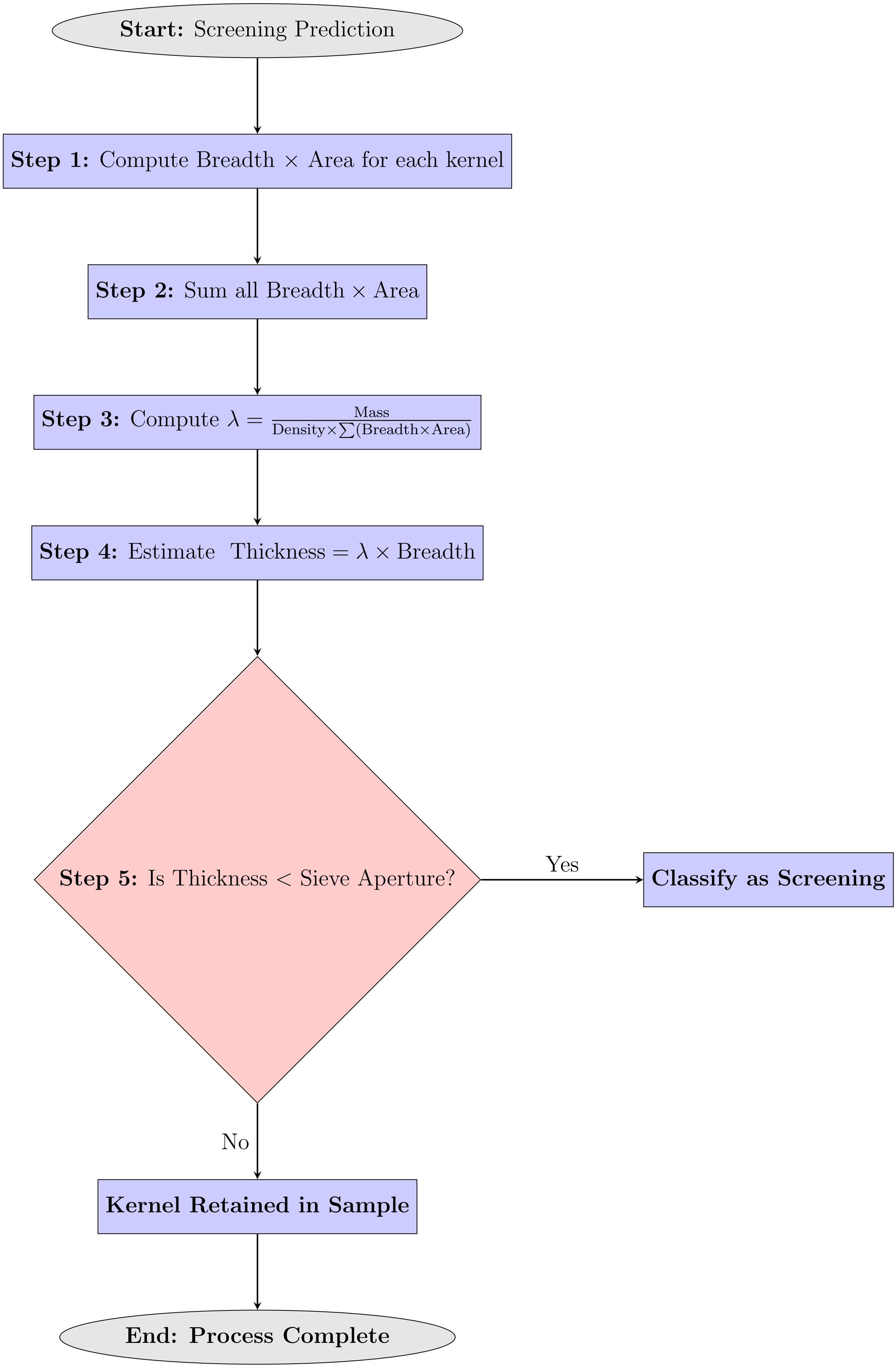
The calculation of the λ (lambda) coefficient is crucial for estimating the thickness of each kernel. This parameter is determined from the relationship between sample mass and density (ρ). The sample mass is measured directly during the sample collection process, while density is the average density extracted from harvest data.

$$\rho = \frac{\text{mass}}{\text{volume}} \tag{4}$$

$$\rho = \frac{\text{mass}}{\lambda \times \sum_{i=1}^{n}(\text{breadth}_i \times \text{area}_i)} \tag{5}$$

$$\lambda = \frac{\text{mass}}{\rho \times \sum_{i=1}^{n}(\text{breadth}_i \times \text{area}_i)} \tag{6}$$

Based on this calculation, a prediction can be made for any kernel whose thickness is less than the width of the sieve aperture. These kernels will pass through the sieve and are classified as **screening** as described in 3.3.1


Start: Screening Prediction
Step 1: Compute Breadth × Area for each kernel
Step 2: Sum all Breadth × Area
Step 3: Compute $\lambda = \frac{\text{Mass}}{\text{Density} \times \sum(\text{Breadth} \times \text{Area})}$
Step 4: Estimate Thickness = $\lambda \times$ Breadth
Step 5: Is Thickness < Sieve Aperture?
Yes
Classify as Screening
No
Kernel Retained in Sample
End: Process Complete

3.3.2 Estimate Weight Ratio Based on Area

Weight ratio fo reach sample can be estimated using the formula5. For a given sample, we now have estimated screenings from 3.3.1 of the analysis. Given that kernels in the sample have been classified as screenings, we need to calculate the ratio of screenings to the sample based on surface area. Using the relationship between mass, density, and volume, we can calculate the mass of the screenings and the mass of the entire sample.

The mass of the screenings as given in: 5

$$\text{mass of screening} = \rho \times \text{volume of the screenings} \tag{7}$$

Since volume can be expressed as:

$$\text{volume} = \text{thickness} \times \text{projected area} \tag{8}$$

and instead of using estimated thickness, we replace it with $\lambda \times$ breadth (as we have the actual measurement of breadth), thus the mass of the screenings becomes:

$$\text{mass of screenings} = \rho \times \lambda \times \sum_{i=1}^{n} (\text{Breadth}_i \times \text{Area}_i) \tag{9}$$

where the summation runs over all screening kernels in the sample.

Similarly, the **sample mass** can be calculated as:

$$\text{Mass of Sample} = \rho \times \lambda \times \sum_{i=1}^{N} (\text{Breadth}_i \times \text{Area}_i) \tag{10}$$

where the summation runs over all kernels in the sample.

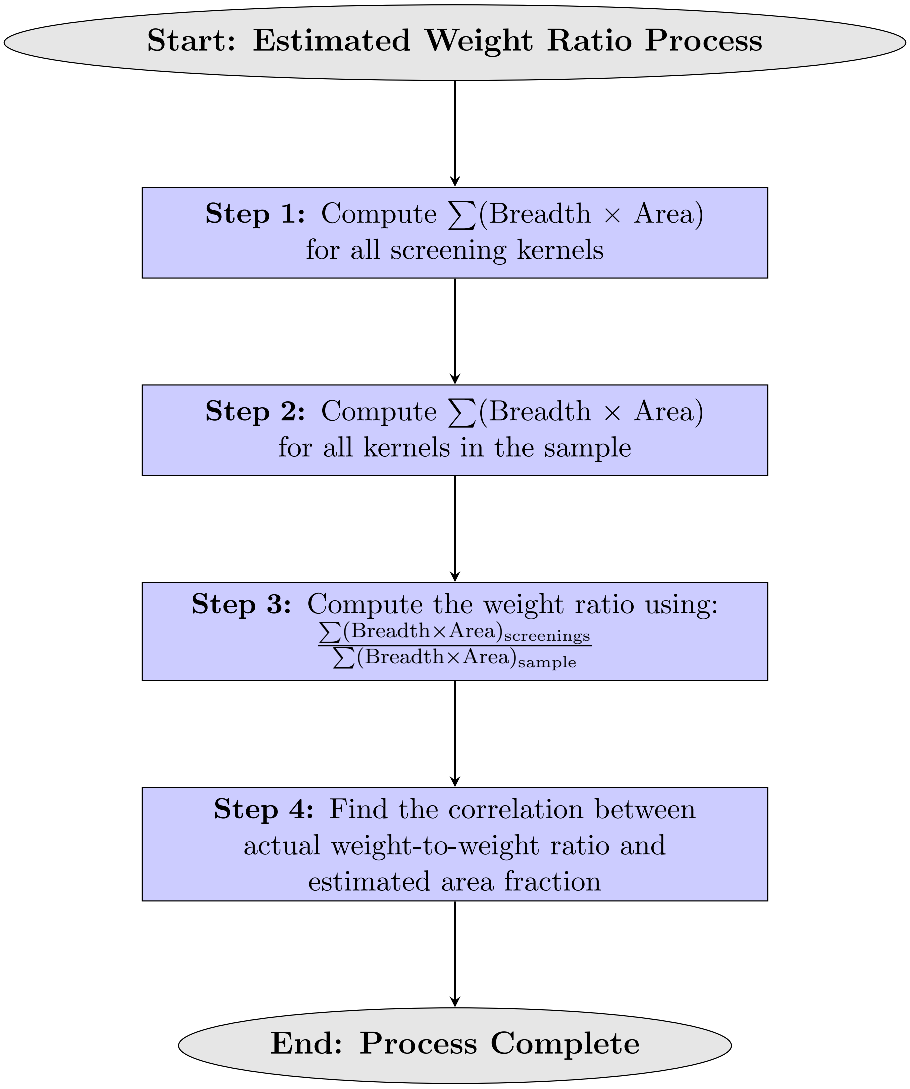
Finally, the estimated weight ratio based on area (i.e., the ratio of the screenings area to sample area) is calculated as:

$$\begin{align} \text{estimated weight ratio} &= \frac{\text{estimated screening mass}}{\text{estimated sample mass}} \tag{11} \\ &= \frac{\rho \times \lambda \times \sum_{i=1}^{p} (\text{Breadth}_i \times \text{Area}_i)}{\rho \times \lambda \times \sum_{i=1}^{N} (\text{Breadth}_i \times \text{Area}_i)} \tag{12} \end{align}$$

Since both the density ρ and λ cancel out from the equation, the mass fraction simplifies to:

$$\textbf{estimated weight ratio} = \frac{\sum_{i=1}^{p} (\text{Breadth}_i \times \text{Area}_i)}{\sum_{i=1}^{N} (\text{Breadth}_i \times \text{Area}_i)} \tag{13}$$

Thus, **density and λ have no effect on the ratio.**


Start: Estimated Weight Ratio Process
Step 1: Compute $\sum(\text{Breadth} \times \text{Area})$ for all screening kernels
Step 2: Compute $\sum(\text{Breadth} \times \text{Area})$ for all kernels in the sample
Step 3: Compute the weight ratio using: $\frac{\sum(\text{Breadth}\times\text{Area})_{\text{screenings}}}{\sum(\text{Breadth}\times\text{Area})_{\text{sample}}}$
Step 4: Find the correlation between actual weight-to-weight ratio and estimated area fraction
End: Process Complete

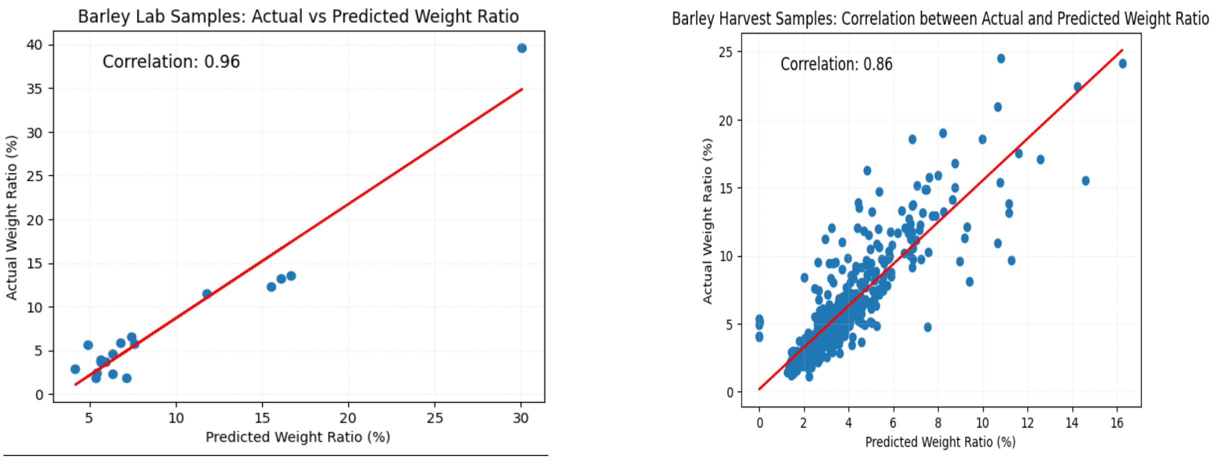
4 Results

4.1 Correlation

The analysis from this study demonstrated a strong correlation between screenings from both physical samples and digital (visual) data. The physical screenings ratio was calculated as the fraction of screening weight to total sample weight, while the estimated digital weight ratio was determined as the fraction of screening area to total sample area. For the laboratory data, the screening area was based on the axial measurements of the screenings from Deimos, as screenings were captured separately. No prediction was required. However, for the harvest data, the screenings were predicted using the method discussed in section 3.3.1.
Despite the differences in the analysis of laboratory and harvest data, the correlation remained consistent across both datasets, validating the strong relationship between the two as shown the correlational table 1 and correlation graphs for barley3, oats4, and wheat3.

Table 1: Correlation between Lab and Harvest Data for Different Commodities

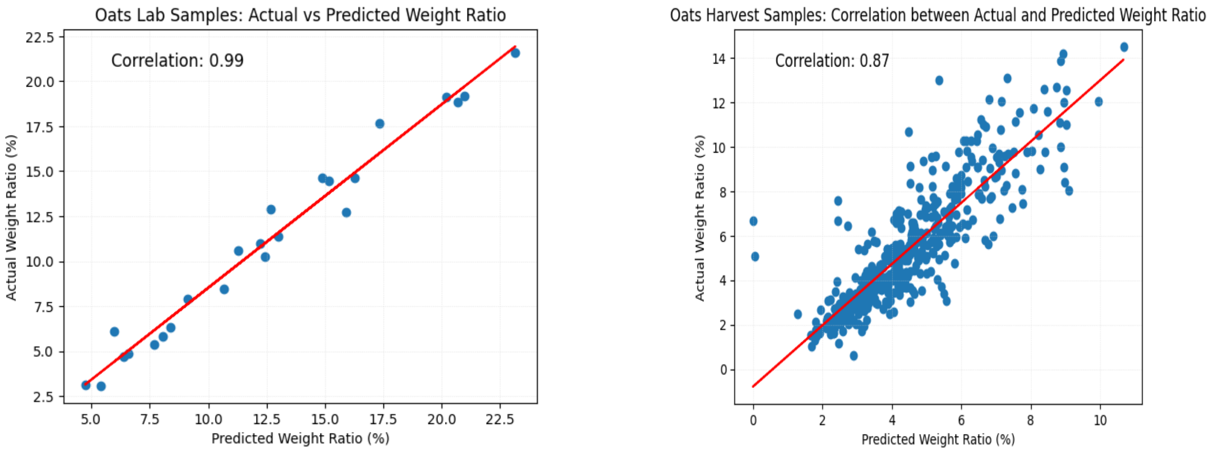
Commodity	**Lab Correlation**	**Harvest Correlation**
Barley	0.96	0.86
Oats	0.99	0.87
Wheat	0.90	0.85



(a) lab correlation

(b) harvest correlation

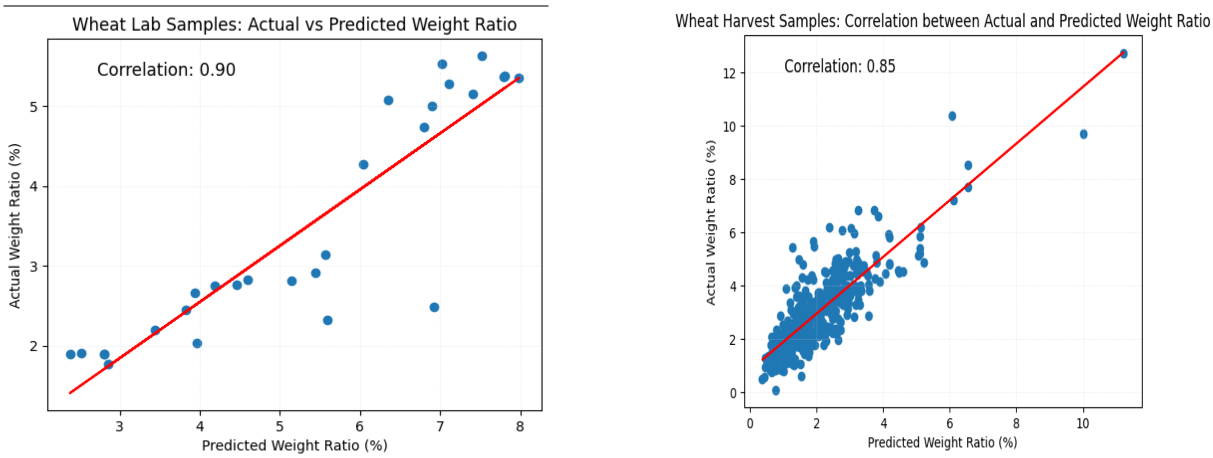
Figure 2: Correlation coefficients between laboratory and harvest data for Barley



(a) lab correlation

(b) harvest correlation

Figure 3: Correlation coefficients between laboratory and harvest data for Oats



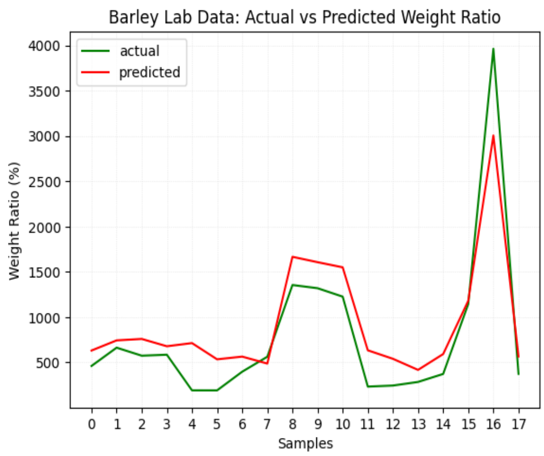
(a) lab correlation

(b) harvest correlation

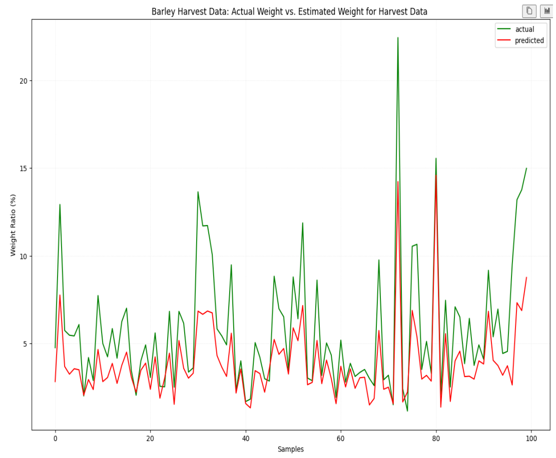
Figure 4: Correlation coefficients between laboratory and harvest data for Wheat

4.2 Weight Ratio Prediction

The equivalent weight ratio was calculated for both laboratory and harvest data based on the method described in Section 3.3.2. A comparison between the actual weight ratio and the predicted weight ratio is presented for barley5, oats6, and wheat7. The results indicate a strong similarity between actual and estimated values across different commodities, despite some observed variations.

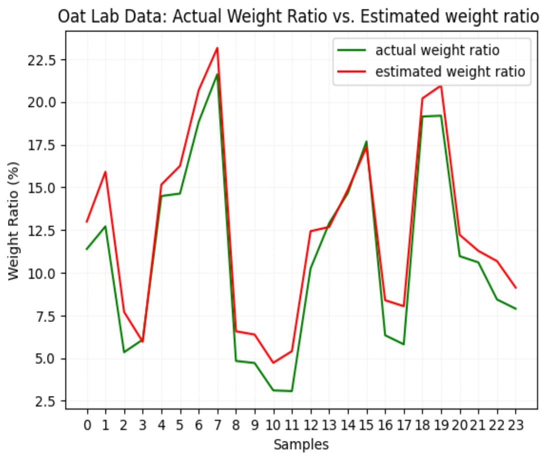


(a) lab prediction (20 samples)

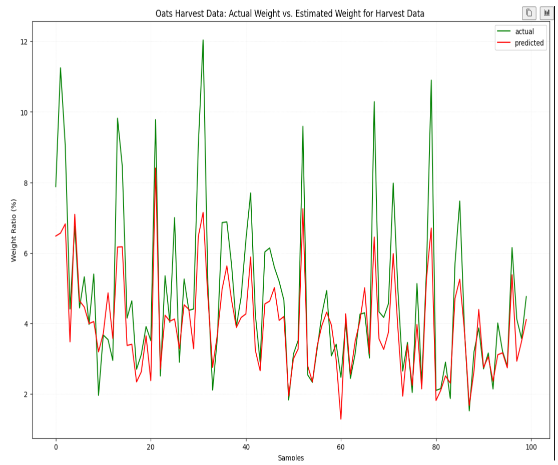


(b) harvest prediction (100 samples)

Figure 5: Actual weight vs. predicted weight ratio for Barley

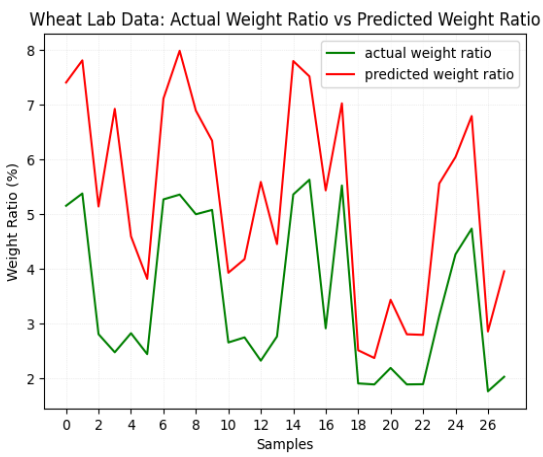


(a) lab prediction (24 samples)

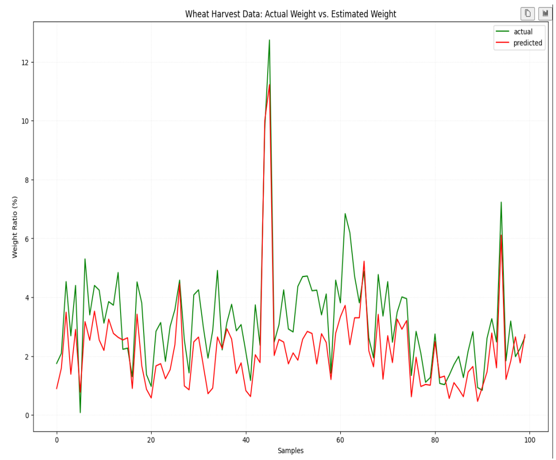


(b) harvest prediction (100 samples)

Figure 6: Actual weight vs. predicted weight ratio for Oats



(a) lab prediction (24 samples)



(b) harvest prediction (100 samples)

Figure 7: Actual weight vs. predicted weight ratio for Wheat

As shown in Table 2, the laboratory predictions tend to be overestimated although not by much, whereas the harvest predictions tend to be underestimated for all commodities.

Table 2: Comparison of Laboratory and Harvest Predictions for Different Commodities

Commodity	Lab Prediction	Harvest Prediction
Barley	Over	Under
Oats	Over	Under
Wheat	Over	Under

4.3 Other findings

Density and flakiness ratio (λ) are two critical measurements for estimating the third axis of a kernel and predicting screenings for a given sample. These parameters play a significant role in determining kernel passage through the sieve. The average density and λ values for different commodities are summarized in Table 3.

Table 3: Average Density and Flakiness Ratio (λ) for Different Commodities

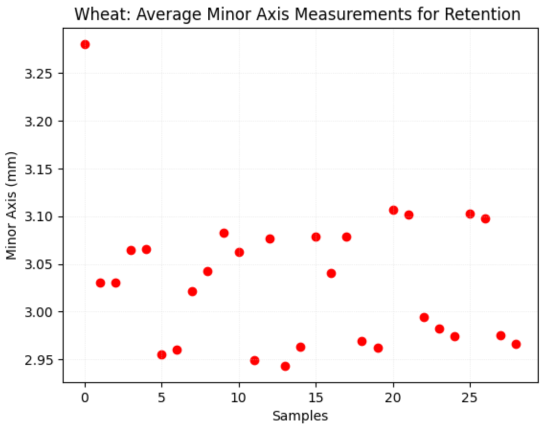
Commodity	Density (kg/hl)	Density (g/mm^3)	Average harvest λ
Barley	67.20	0.0000672	0.7412
Oats	54.74	0.0000547	0.8479
Wheat	80.07	0.0000800	0.8531

An exploratory analysis of the axial dimensions indicates that there is no significant difference in major axis measurements between screenings and retentions as shown in Figure 8. This supports the earlier statement that the major axis, or the longest dimension of the kernel, has little to no impact on determining kernel passage through the sieve. However, as shown in Figure 9, there is a noticeable difference in the minor axis (breadth) measurements between screenings and retentions across all three commodities.

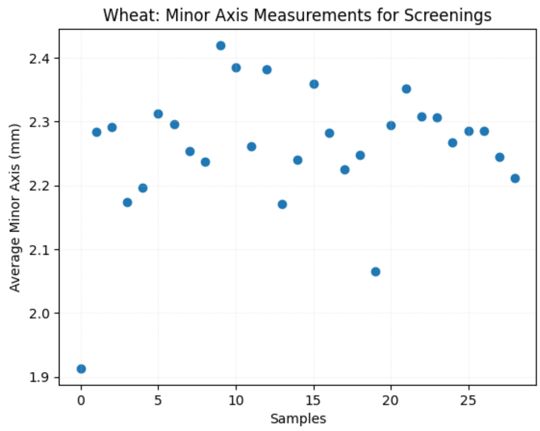
(a) retention major axis

(b) screening major axis

Figure 8: Average major axis measurements for Wheat



(a) retention minor axis



(b) screening minor axis

Figure 9: Average minor axis measurements for Wheat

5 Discussion

5.1 Conclusion

Although the conventional sieving method is crucial for assessing the quality of samples, it has become outdated and has several drawbacks. The mechanical process is slow, which can delay deliveries on-site, and it also does not account for the axial dimension, which could provide a more accurate representation of sieving. This correlational study introduces an alternative method based on area fractions using metadata (axial dimensions) measured by Deimos, which showed a strong correlation between the physical screening ratio and the estimated digital screening ratio. The results also indicated that the minor axis (breadth) and the flakiness ratio λ play significant roles in determining kernel passage through the sieve, while the major axis (length) has little to no impact. The prediction model based on axial measurements demonstrated reliable predictions across all commodities.

A comparison between the laboratory data and harvest data revealed differences in prediction trends. The method tends to overestimate the screening ratio for laboratory data, while it underestimates the screening ratio for harvest data across all commodities, as shown in Figures 5, 6, and 7. The underestimation for harvest data can be explained by the fact that the digital sieving method used in the study only considers kernels, not dust or other foreign materials, which can also be classified as screenings and affect the weight ratio. The overestimation is likely due to issues with the image capturing process, as several machines encountered problems during laboratory image capture. Through a physical examination, it was found that kernels were not sufficiently dispersed during image capture, and the axial measurements were also likely erroneous.

The study served as a preliminary study with an aim to replacing conventional and mechanical sieving method with a digital sieve. Through comprehensive analysis of kernels, screenings were predicted from a sample without using a mechanical sieve. The predictions were based on axial dimensions, density and flakiness ratio λ which helps estimate the third axis and thus predict screenings. The findings from the study showed strong correlation between the actual weight ratio and predicted weight ratio across all commodities. While there are some discrepancies between laboratory and harvest data, the model still provides reliable predicted values across all the commodities.

5.2 Future Work

Although the study demonstrated strong correlations between the actual weight ratio and the predicted weight ratio, several considerations should be addressed.

1. **Consideration of dust and other foreign materials in the screenings.** The screenings contained not only kernels with smaller axial dimensions but also dust and other foreign materials, which contributed to the weight and influenced the weight ratio.

2. **Enhancing classification with machine learning and computer vision.** More robust models based on machine learning and computer vision could improve the classification of screenings and retentions more accurately, rather than relying solely on the third axis.

3. **Variability in kernel shape properties.** Each kernel differs in shape and structure. Instead of assuming that grains from the same source exhibit similar characteristics and, consequently, similar axial dimensions, it is important to account for individual variability.

4. **Limitations of using average density and flakiness ratio.** The study used average density and the average flakiness ratio λ, but these values are not consistent across all kernels. Relying on mean values may not provide an accurate representation and could introduce errors in calculations. A more precise method is needed to incorporate kernel-specific variations.

References

[1] C. Mora, A. Kwan, and H. Chan, "Particle size distribution analysis of coarse aggregate using digital image processing," *Cement and Concrete Research*, vol. 28, no. 6, pp. 921–932, Jun. 1998. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S000888469800043X

[2] A. Kwan, C. Mora, and H. Chan, "Particle shape analysis of coarse aggregate using digital image processing," *Cement and Concrete Research*, vol. 29, no. 9, pp. 1403–1410, Sep. 1999. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0008884699001052