healthhub

Hyundai Motors

Subject 1 – Physical Properties Prediction of Gas Diffusion Layer (GDL) Material

List of Contents

Link to detailed document

healthhub

- Project Goal
- Introduction
- Literature Review
- Data Analysis
    - Phase 1: Shooting Conditions/Views
    - Phase 2: Company Analysis
    - Phase 3: Feature Extraction
    - Phase 4: Training Data
- Data Preprocessing
    - Alignment of Material
    - Removal of adhesive
    - Intensity Correction
    - Material segmentation
    - Material smoothing
    - Element segmentation and distribution curves
- Feature Engineering
- AI development
- App Development
- Improvement of AI
- Conclusion

Project Goal

The main goal of the project is to predict the values of Physical Properties (such as permeability and resistance for Gas Diffusion Layer (GDL) material from their CT scans with the assistance of AI.

- **Permeability** (mmH_2O):

  Defined as: Moisture permeability is the material's resistance to the water vapor diffusion through a unit of surface area.

- **Resistance** ($m\Omega cm^2$):

  This unit is of contact resistance.

  The term contact resistance refers to the contribution to the total resistance of a system which can be attributed to the contacting interfaces of electrical leads and connections as opposed to the intrinsic resistance.

Introduction

Introduction – Material

- Gas Diffusion layer (GDL) materials are used in PEM fuel cells as Anode and cathode GDLs.
- The GDL is typically wet-proofed and carbon based product comprising of:
  - Macroporous substrate (MPS) (Single layer GDL): Consists of horizontally and anisotropically stacked carbon fibres
  - Microporous layer (MPL): Comprised of carbon based powder

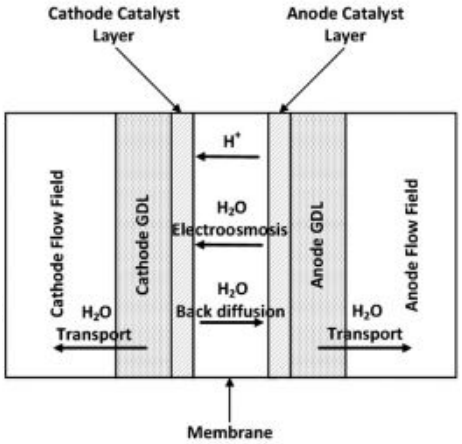


Figure 1: GDLs shown within Fuel cell

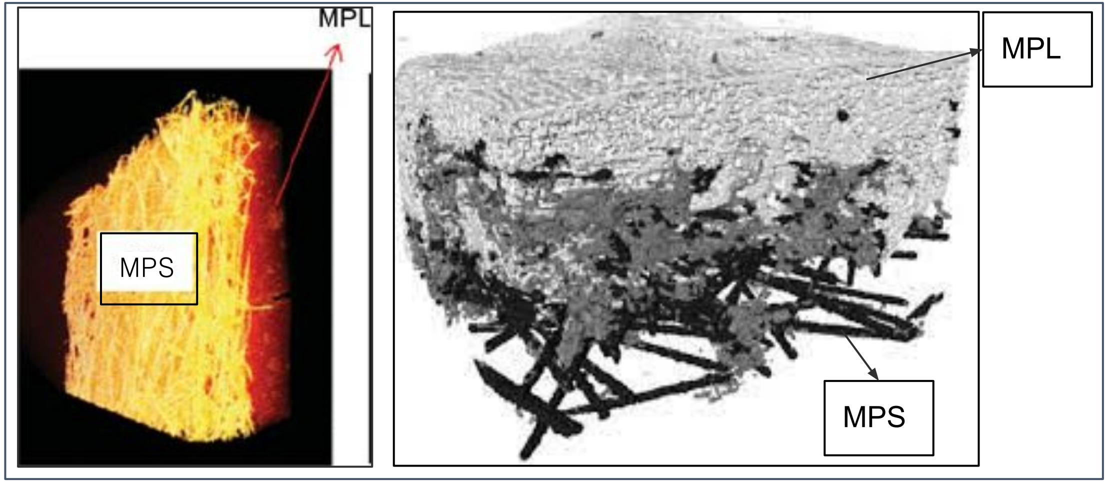


Figure 2: Both MPL and MPS layer shown in GDL

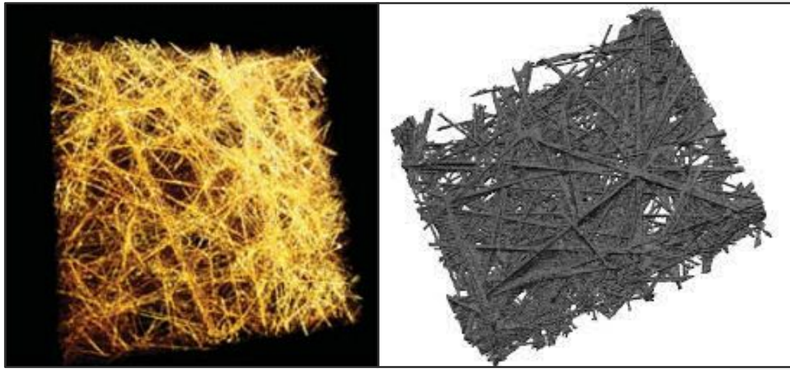


Figure 3: Carbon fibres within MPS layer

Introduction – Elements

- We identified the following elements:
  - Fiber
  - Air
  - Carbon

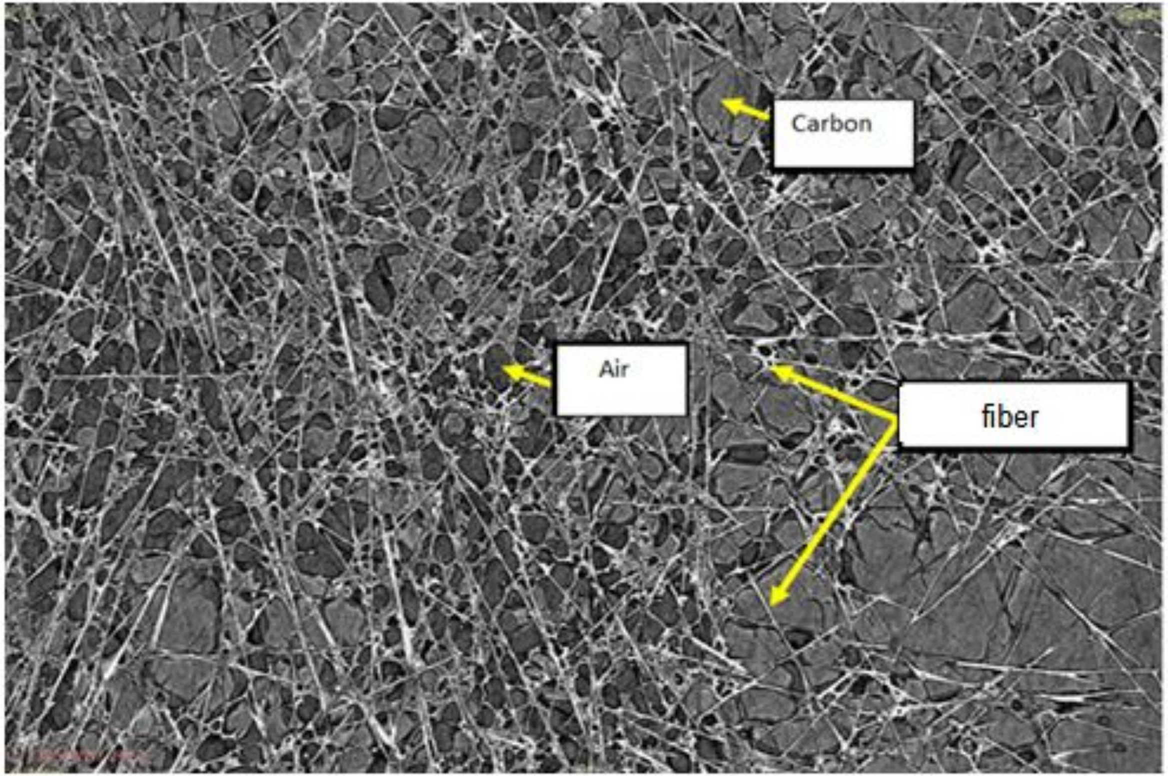


Figure 4: Elements of the materials

Literature Review

Literature Review

Insights:

A detailed literature review provided the following insights about the GDL materials:

- The **void region** controls the transport of reactants and by-product water while the remaining solid structure provides pathway for electron and heat transport.
- Thickness, morphology, uniformity, hydrophobic treatment, and pore characteristics are the critical parameters affecting GDL inherent characteristics, and hence internal resistance and permeability.
- The **carbon powder** particle size also influences the penetration behavior of the MPL. When it is significantly small, the MPL is more likely to penetrate deeply into the MPS. In some cases, intrusion of the carbon particle/hydrophobic agent agglomerates into the MPS can build up more direct and additional electron-carrying pathways between the carbon fibers, significantly affecting the resistance.The deposition of MPL onto the MPS causes a significant drop in the pore size range.
- Different Types of GDL fiber (Carbon Cloth– Woven fibres, Carbon Paper– straight stretched fibres and Carbon Paper– felt/spaghetti fibres) can affect the physical properties.

Literature Review

Possible features:

The insights from literature were used to list down possible features that can be used for the development of AI. The possible features are listed below as:

- Fiber, Air and carbon ratios in the material
- Distribution of elements in the material (Distribution curves)
- Thickness of material
- Density of each element in MPL and MPS layer separately e.g. Carbon in MPS layer and porosity (air/void region) in MPL and MPS
- Maximum Intensity projections (MIPs) of fiber, air and carbon to provide information of structure of material (structure of fiber etc)

Data Analysis

Data Analysis – Phase 1

Phase 1: Shooting Conditions/Views

Hyundai Motors initially gave us 4 samples:

- Full view
- Half View
- Quarter view
- Half Quarter View

Sample	Initial Mode	Slice in Axial	Pixel Spacing (µm)	Volume of material Covered $(mm)^3$
Full View	Axial		16.269	510.71879468
Half View	Axial		6.257	106.68166581
Quarter view	Axial		3.754	46.973589445
Half-Quarter View	Axial		1.877	8.0839564927

Figure 5: Comparison of data samples

Based on our initial analysis, we **concluded** that even though there is better resolution in Half Quarter View, the volume covered by this view is very less, thus the distribution patterns are different from other 3 views. To decide on a view from the rest of the samples and to check if it will be more important to have more volume covered or to have a better resolution**we asked for more samples**.

Data Analysis – Phase 1

Phase 1: Shooting Conditions/Views

For further analysis, Hyundai Motors provided us with 4 new samples.

- JNTG 1X1 (zoomed in)
- JNTG 3x4 (zoomed out)
- SGL 1x1 (zoomed in)
- SGL 4x4 (zoomed out)

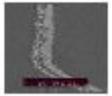
Sample	Initial Mode	Slice in Axial	Pixel Spacing (μm)	Volume of material Covered (mm)^3
JNTG 1x1	Sagittal		5.246	24.943547501
JNTG 3x4	Sagittal		22.527	352.85976077
SGL 1x1	Sagittal		5.006	20.972894075
SGL 4x4	Sagittal		22.527	501.85141312

Figure 6: Comparison of data samples

The resolution of new samples was poor as compared to previously provided samples. From our analysis, we **recommended** to use quarter view with the resolution and initial mode of old samples for further work.

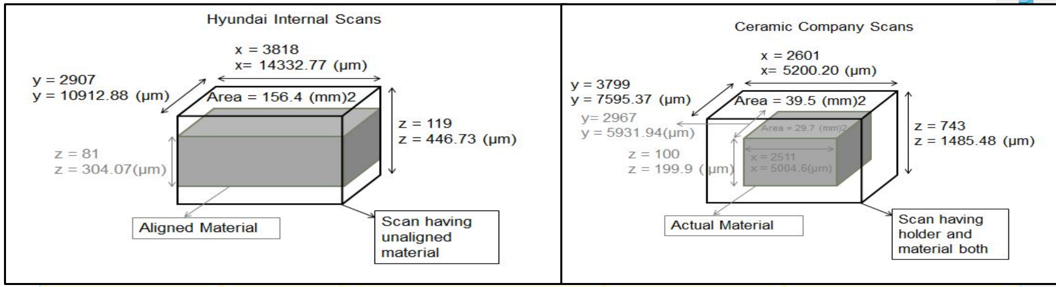
Data Analysis – Phase 2

Phase 2: Company Analysis

In second phase , Hyundai Motors sent us two scans:

- Hyundai's internal scans in dicom format (Quarter view)
- Ceramic Company scans in vgl format

We **concluded** that, even though the resolution is better in Ceramic Company scans, there is a chance that the material is being compressed by the holder and because we cannot fully exploit vgl format scans, it's **better to use Hyundai's Internal scans** since features can be extracted from both the samples and Hyundai's internal scans are in DCM format which is easy to use for our analysis.



Sample	Initial Mode	Slice in Axial	Pixel Spacing (μm)	Volume of material Covered $(mm)^3$
Hyundai's Internal	Axial		3.754	46.973589445
Ceramic Company	Axial		1.999307	5.9344286861

Figure 7: Comparison of data samples

Data Analysis – Phase 3

Phase 3: Feature Extraction

We were given 1 sample to analyse and check if the scanning conditions are appropriate for future work on the project. We compared the new scan with that of ceramic company.

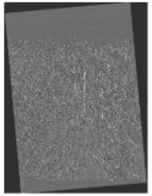
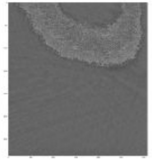
Sample	Initial Mode	Slice in Axial	Pixel Spacing (µm)	Volume of material Covered $(mm)^3$	File Format	Status of Holder
Ceramic Company	Axial		1.9993	5.934428686	VGL	Present
New Sample	Sagittal		4.501	35.94754439	DICOM	Absent

Figure 8: Comparison of samples

From Comparing these new samples with the scans provided by ceramic company earlier, the new scan appeared to be a different material. We were, however, able to extract out all the elements in new scan as well.

We **recommended** providing ceramic company scans in dicom format instead of vgl format and axial view as compared to sagittal as ceramic company scans have better resolution and contrast. We also recommend providing the scan of the same material as before so that better analysis can be performed.

Phase 4: Training Data

In the last phase, We were provided scans in sagittal view with a pixel spacing of 0.001995mm.

Following are the number of provided scans along with the company (manufacturer) wise distribution:

Company	No. of Materials	Samples per material	Total No. of scans
JNTG Ca	8	3	24
JNTG An	8	3	24
FRB	8	3	24
SGL Ca	8	3	24
SGL An	8	3	24
Total No. of scans provided			120

A sample sagittal view slice from the data can be seen in fig.:

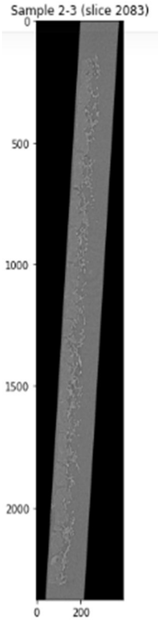


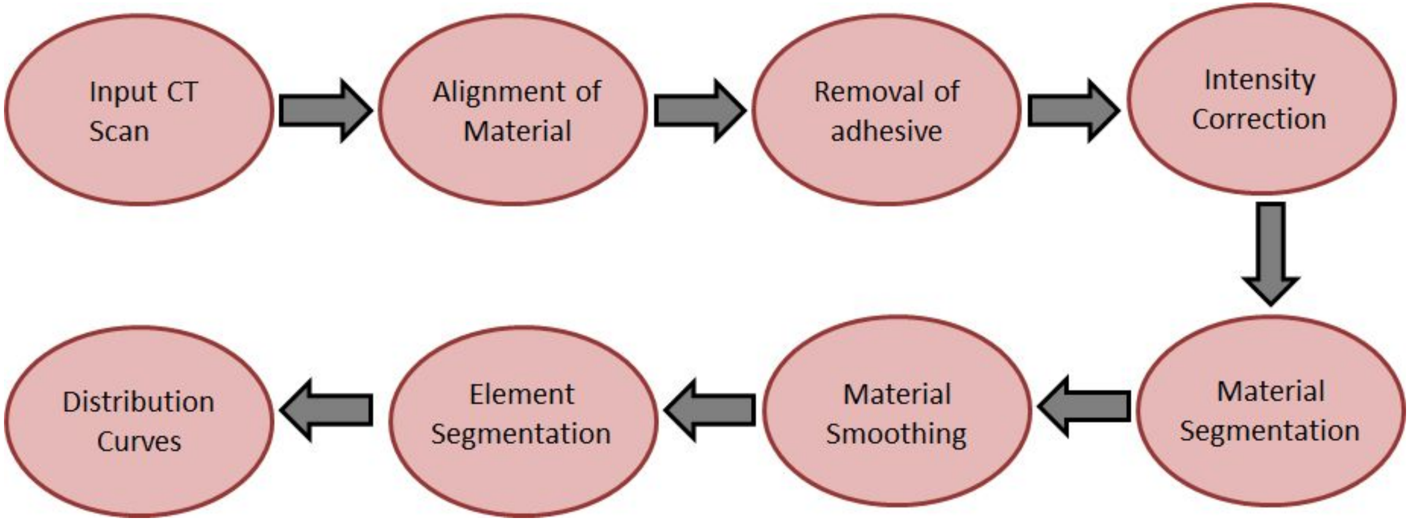
Figure 9: Sagittal view slice

Data Preprocessing

Data Preprocessing

Processing of scans

Following processing steps were performed on the GDL material CT scans to extract out the desired features before the application of any AI model:



These processing steps are discussed in detail in further slides.

Data Preprocessing

Alignment of Material

The scans had extra black area that needed to be cropped and since the scan region appeared to be tilted, we developed an algorithm to crop out the black part and rotate the required region to get straight scan as shown in the figure.

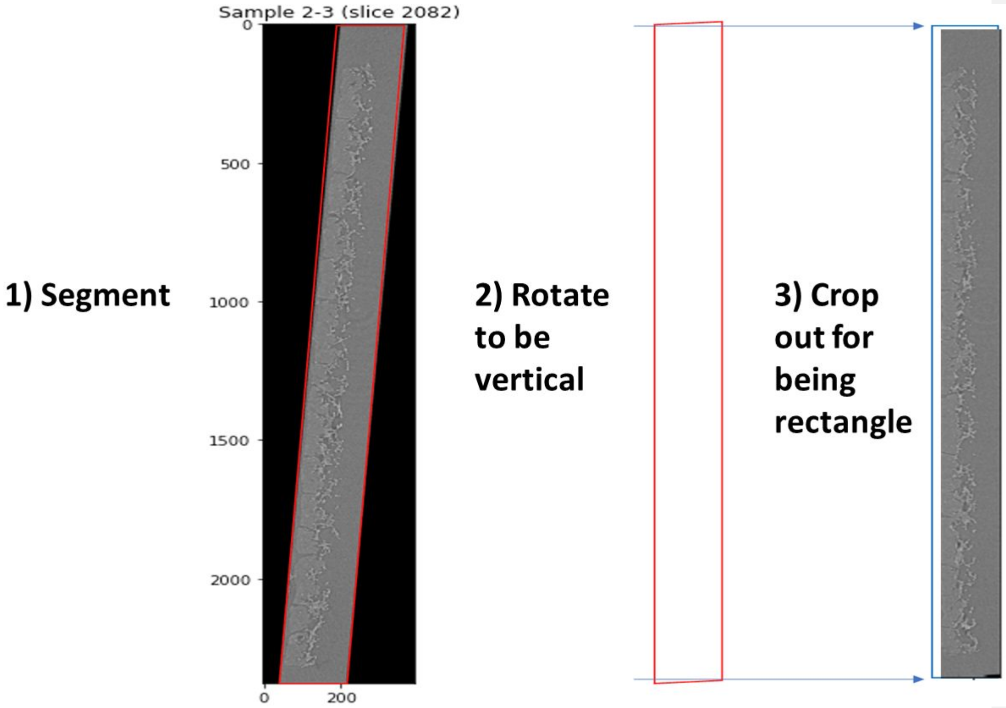


Figure 10: Removal of black region and straightening of scan

Data Preprocessing

Adhesive Removal

- We observed that some scans have adhesive portion visible as shown in fig 11.
- The intensity and morphology of adhesive is similar to that of material elements. Therefore it is not possible to remove the adhesive using image processing techniques.
- We cropped the scans in sagittal view to remove adhesive as shown in fig. 12.

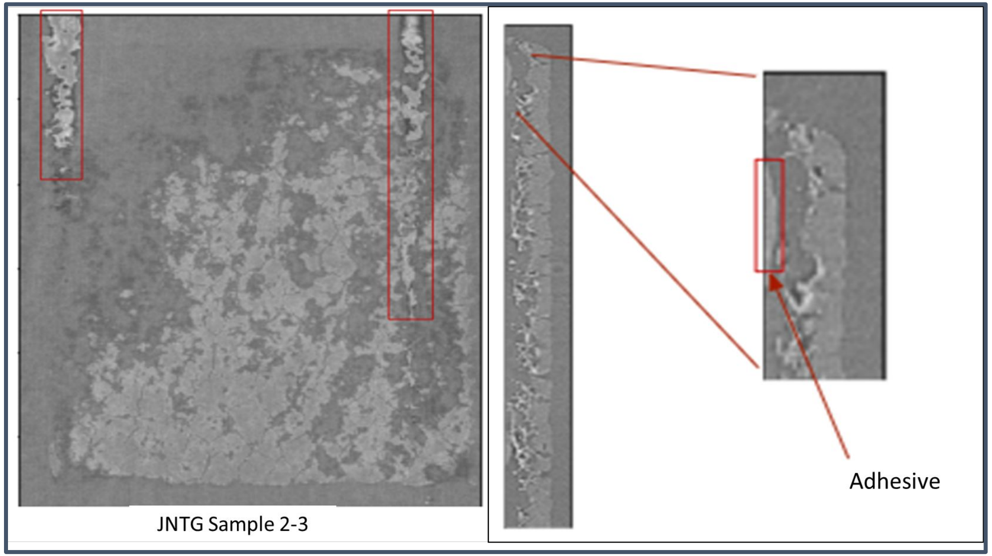


Figure 11: Adhesive in axial and sagittal view

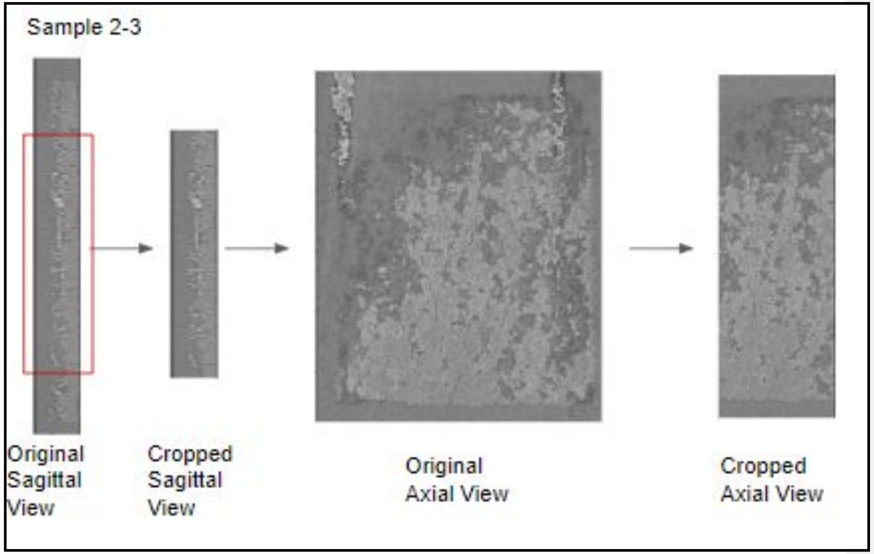
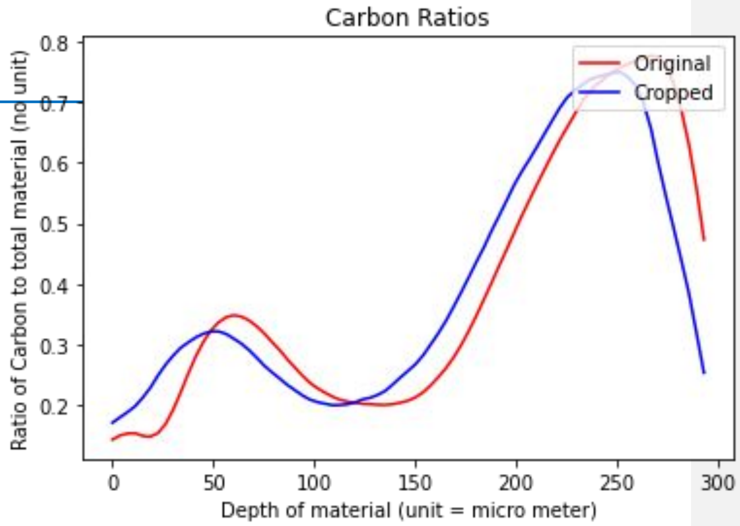


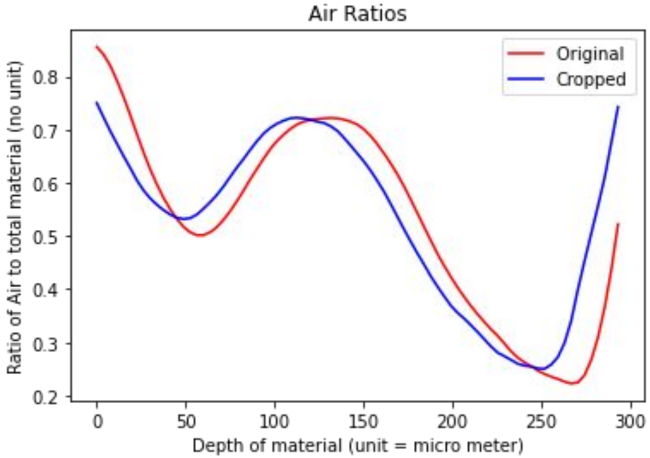
Figure 12: Cropping out the adhesive part

Data Preprocessing



Adhesive Removal

Comparison of distribution curves of elements show no significant difference after cropping suggesting that the adhesive part can be cropped out.
The distribution curves for Quarter view are shown here before and after cropping.



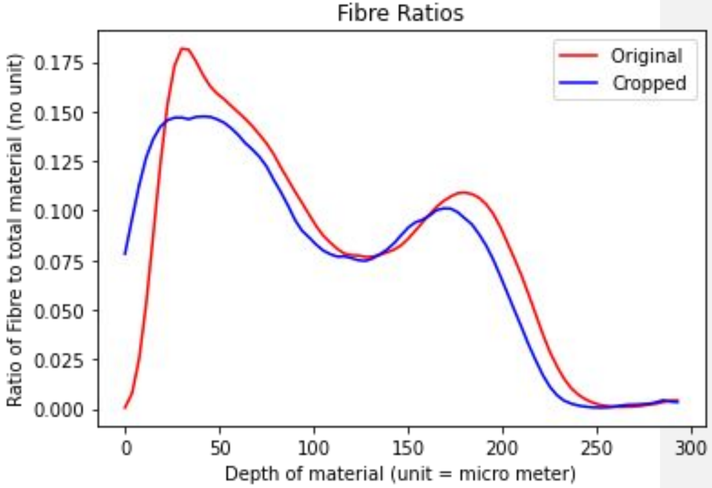


Figure 13: Distribution curves of elements before and after cropping

Data Preprocessing

Intensity Correction

We observed contrast variations in axial view (seen in fig. 16) and observed that this is because of intensity variations in sagittal slices (fig 14.) as can be seen in histograms (fig 15).

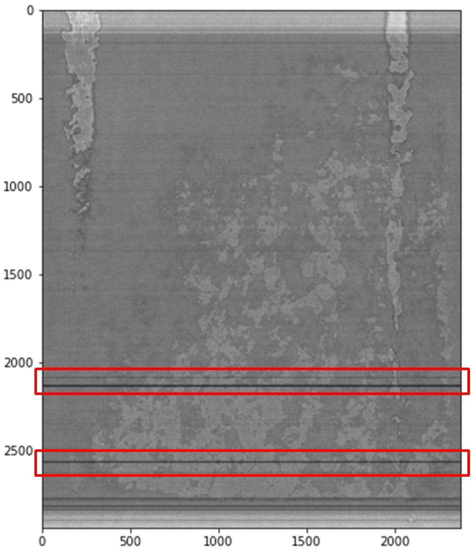


Figure 16: Dark lines in axial view

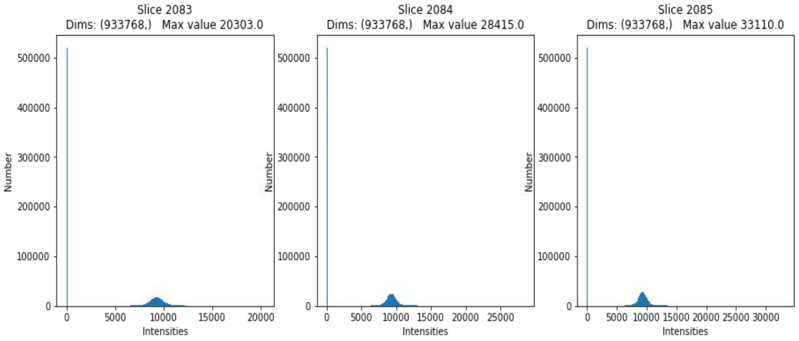


Figure 15: Respective histograms of slices in sagittal view

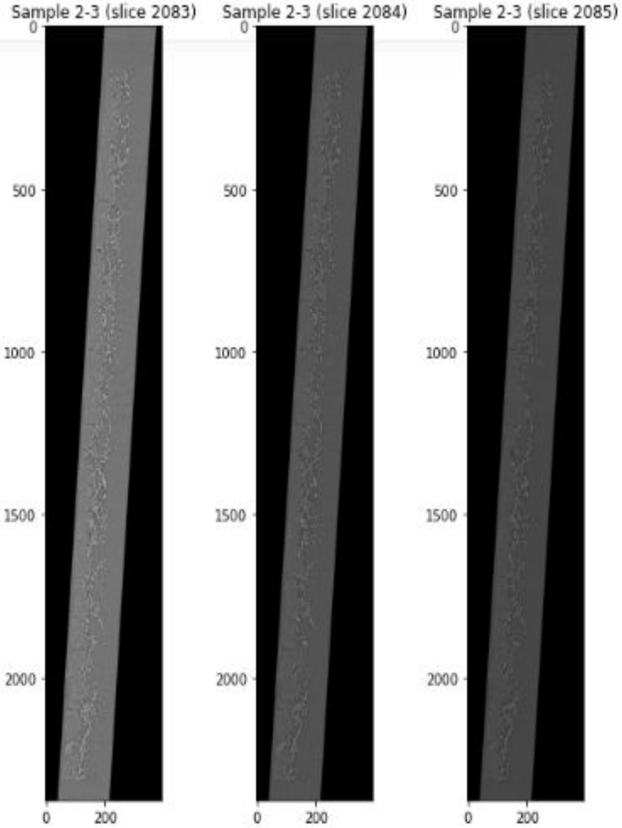
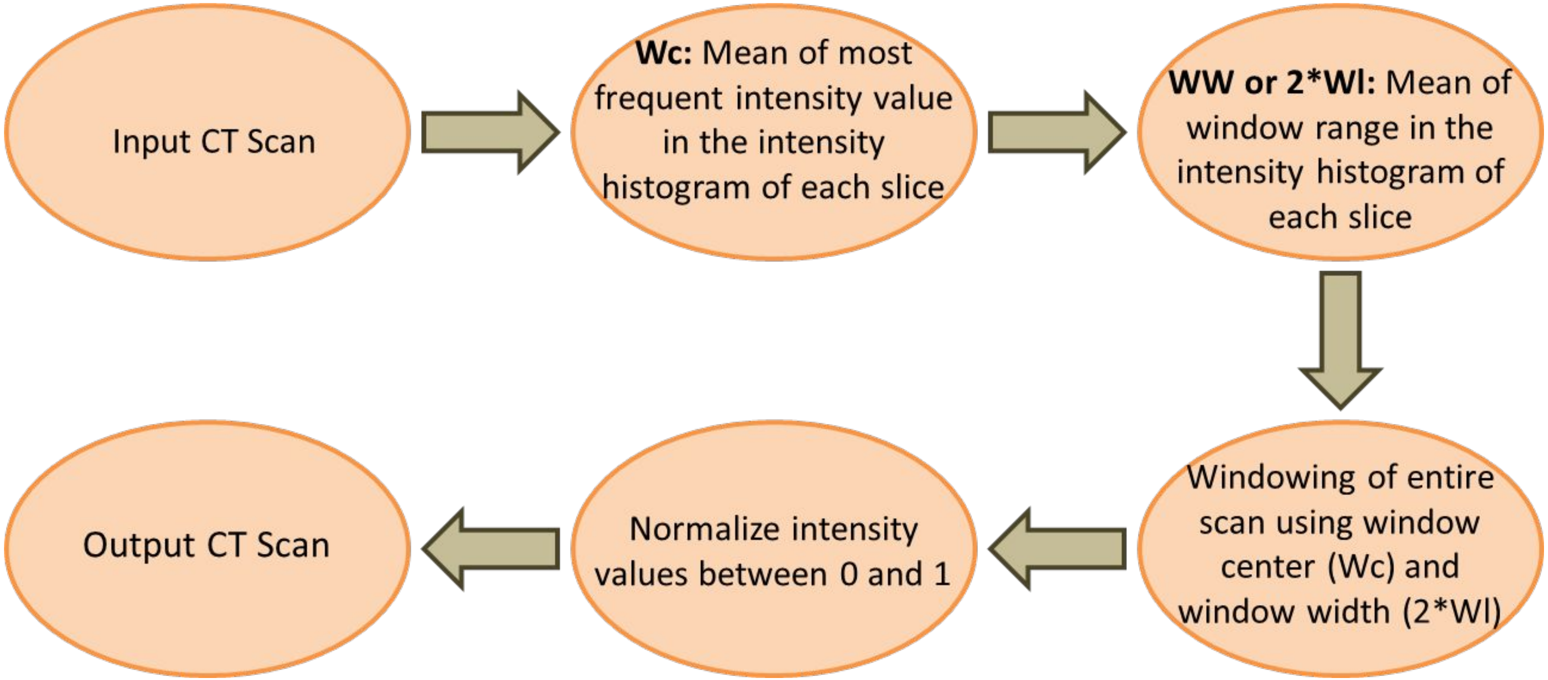


Figure 14: Contrast variations in sagittal view

Data Preprocessing

Intensity Correction

We developed an algorithm as shown:



Data Preprocessing

Intensity Correction

- For each slice the value of Wc and Wl is computed as shown in fig. 17
- Mean of Wc and 2*Wl is used as window centre and window width respectively for windowing.
- The Output can be seen in axial view in fig. 18

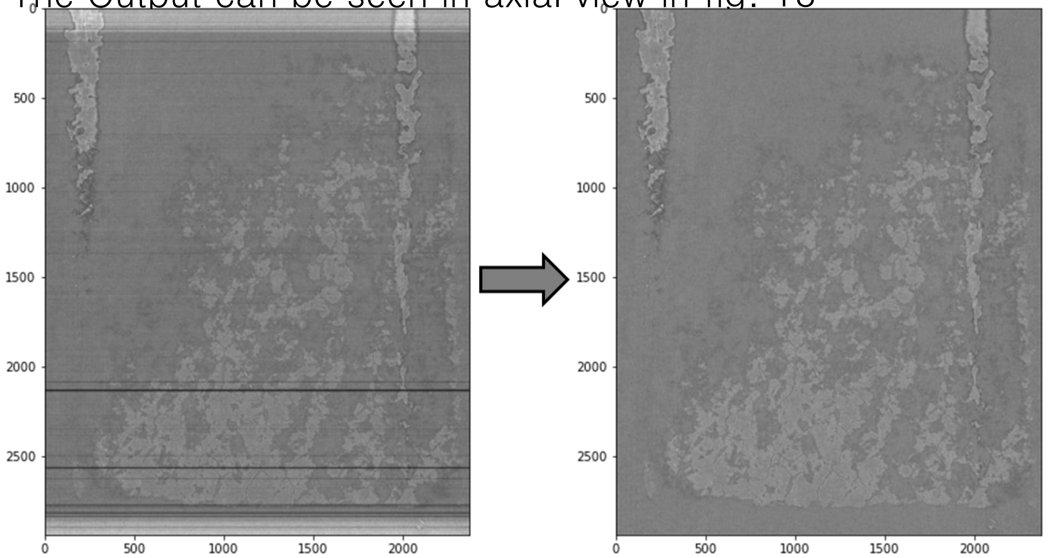


Figure 18: Output after removal of black lines

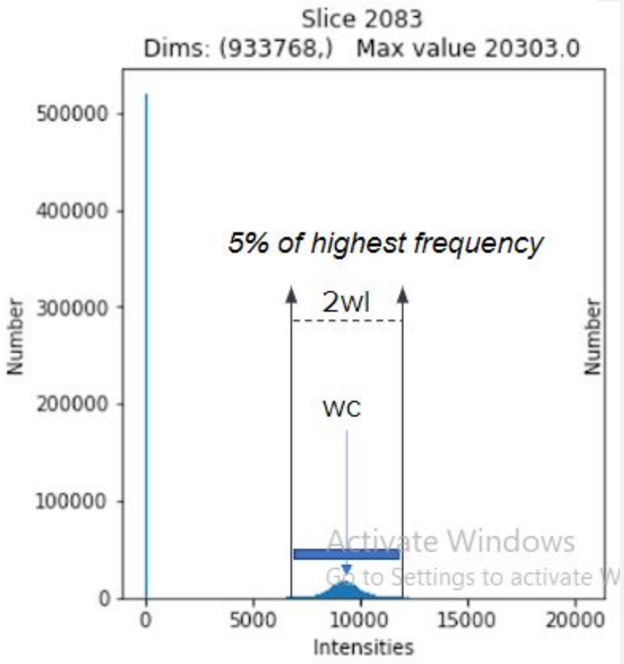


Figure 17: wc and wl computation

Data Preprocessing

Material Segmentation

In order to segment the materials in all new scan, We propose an algorithm as shown below:

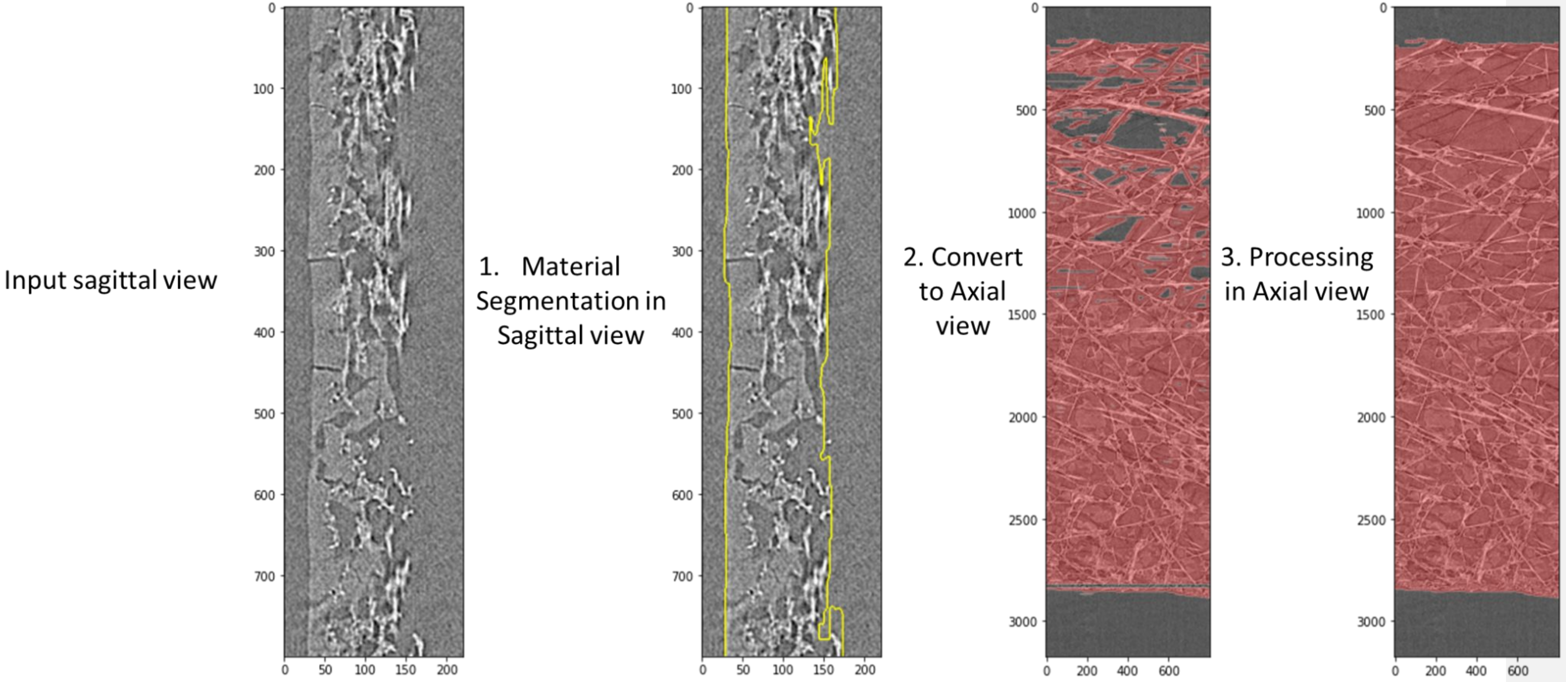


Figure 19: Algorithm for Material Segmentation

Data Preprocessing

Alignment/Smoothing of Material

GDL materials are thin sheets and the surface of a GDL sheet is similar to that of shown in fig. 20. To straighten the material as shown in fig. 22, we developed an algorithm that uses the material segmentation and traverses through the entire surface of material to perform alignment as shown in fig. 21.

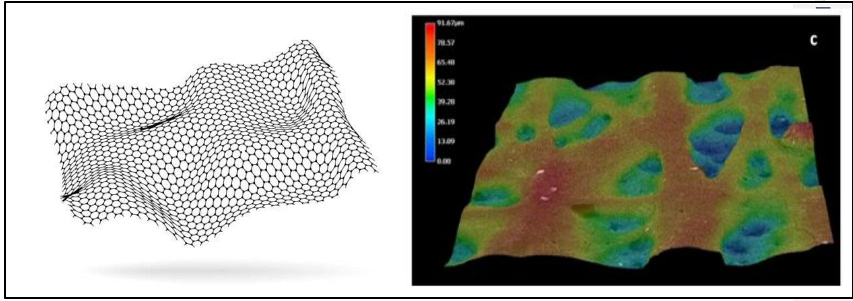


Figure 20: Visual representation to understand the wavy surface of GDL material

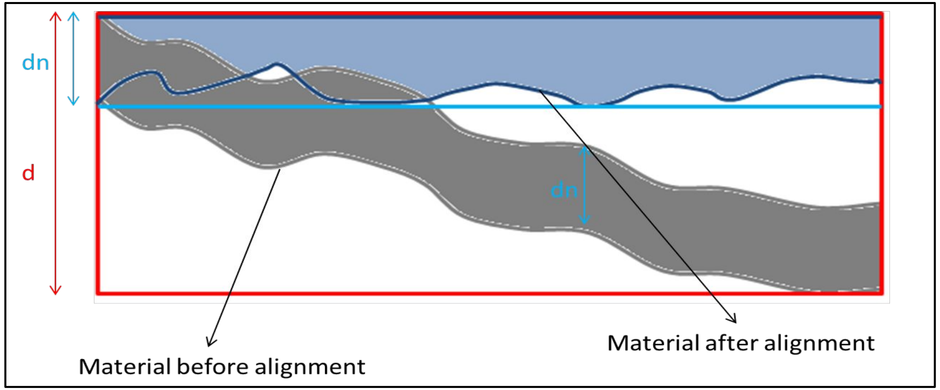


Figure 22: Straightening of GDL material

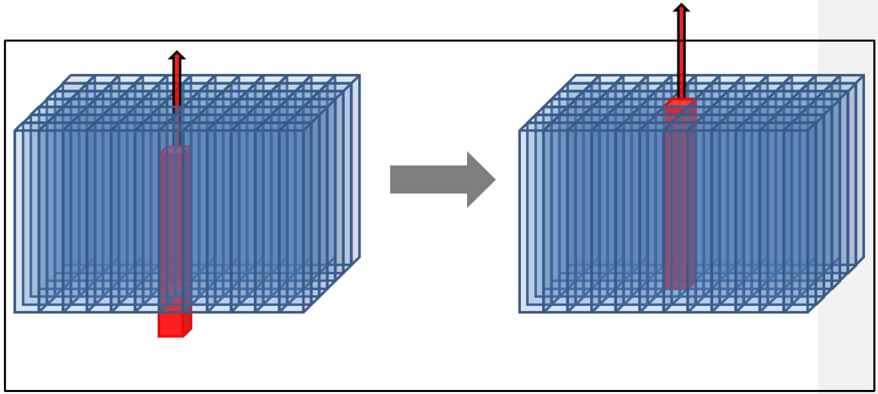


Figure 21: Basic working of alignment algorithm

Data Preprocessing

Alignment/Smoothing of Material

Output after smoothing can be seen as:

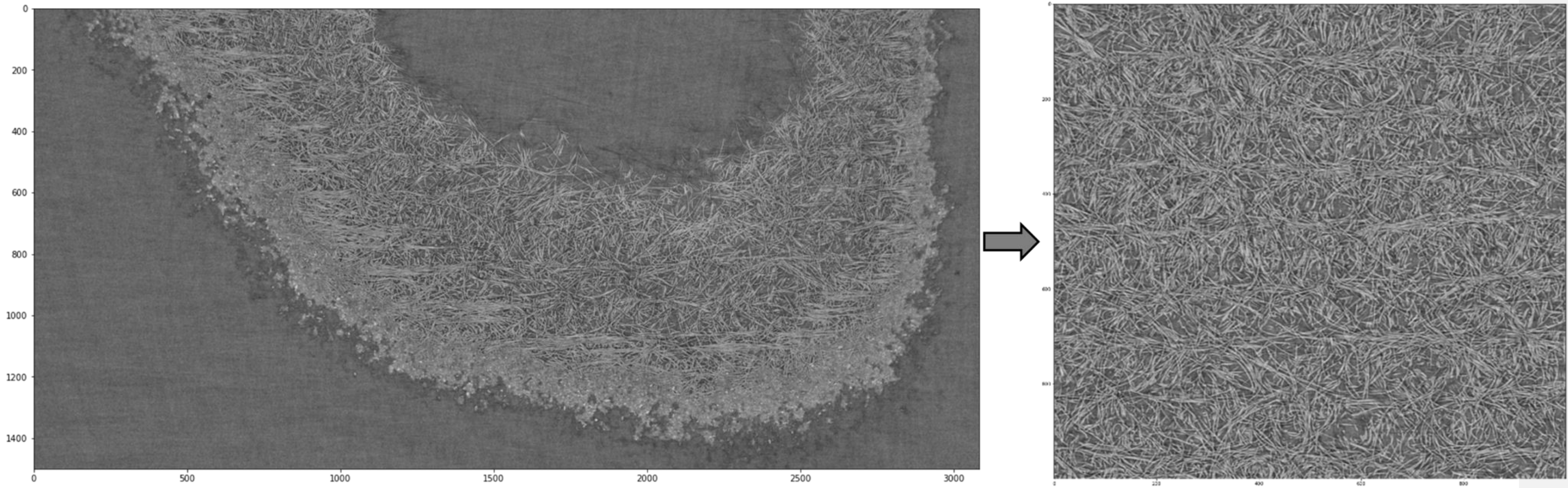


Figure 23: Slice in axial view before and after alignment (Slice taken from FRB sample provided for data analysis phase 3)

Data Preprocessing

Alignment/Smoothing of Material

Distribution curves before and after smoothing can be seen as:

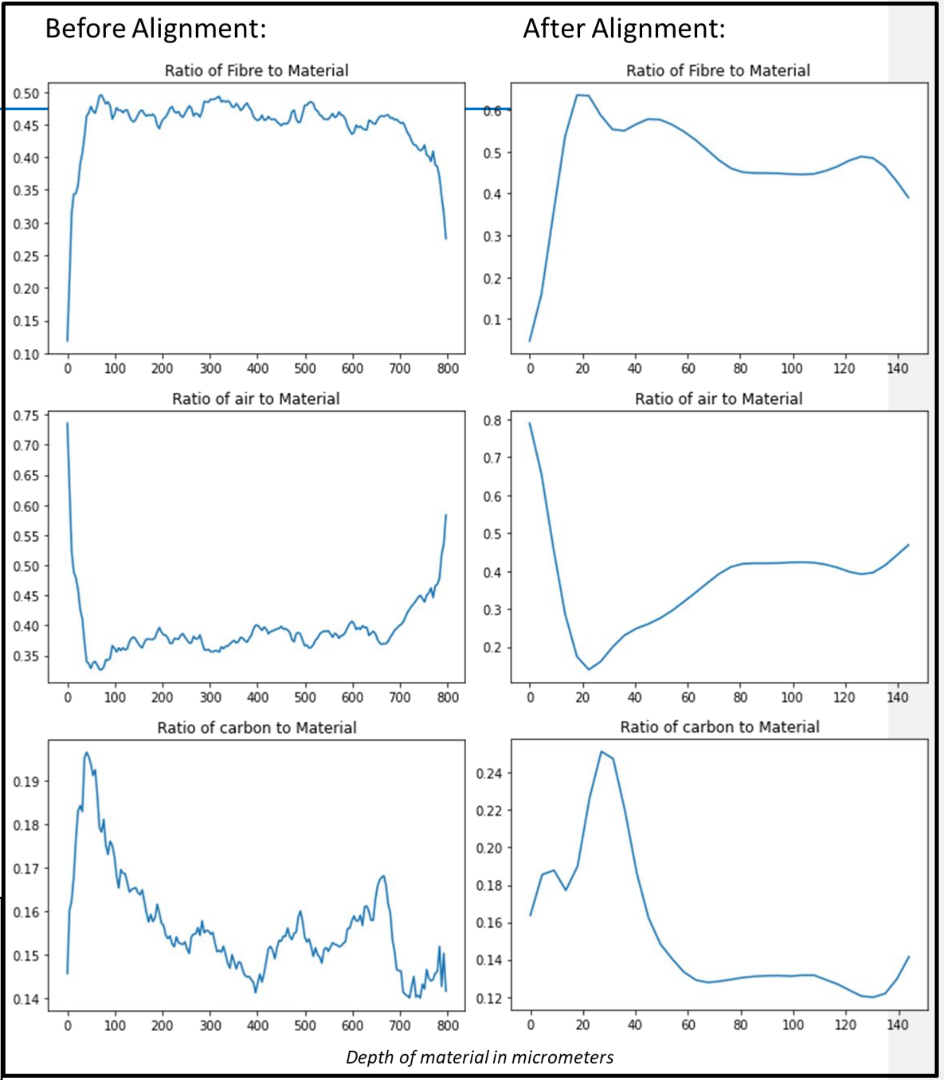


Figure 24: Distribution curves before and after alignment (FRB sample provided for data analysis phase 3)

Data Preprocessing

Element Segmentation and Distribution curves

- Since there is a significant difference in the intensities of elements, all three elements (fiber, air and carbon) can be segmented out by applying intensity based thresholding.
- Since the scans are already normalised between 0 and 1 after windowing for contrast correction, we set absolute threshold values as following to extract the respective element out based on visual analysis.
  - Fiber —> greater than 0.725
  - Air —> less than 0.54
  - Carbon —> between 0.54 and 0.725

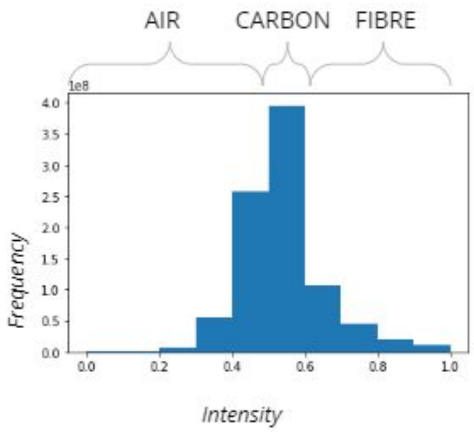


Figure 26: Intensity histogram showing thresholds

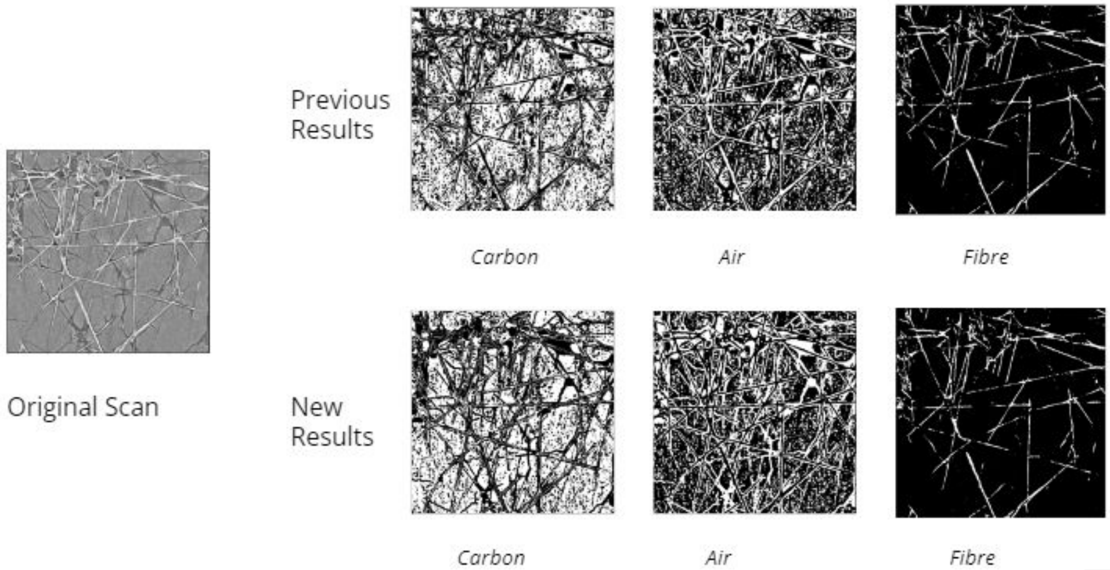


Figure 25: visual comparison of quarter view

Data Preprocessing

Element Segmentation and Distribution curves

Distribution curves are computed by taking ratio of each element to the material in each slice.
We compared new algorithm results with the previous results of Quarter view from phase 1 of Data analysis.

- Computed the distribution curves and ratios of elements before and after applying alignment/smoothing.
- The folders for distribution curves of all samples can be found [here.]()
- The element ratios and the ground truth provided by hyundai motors can be found [here.]()

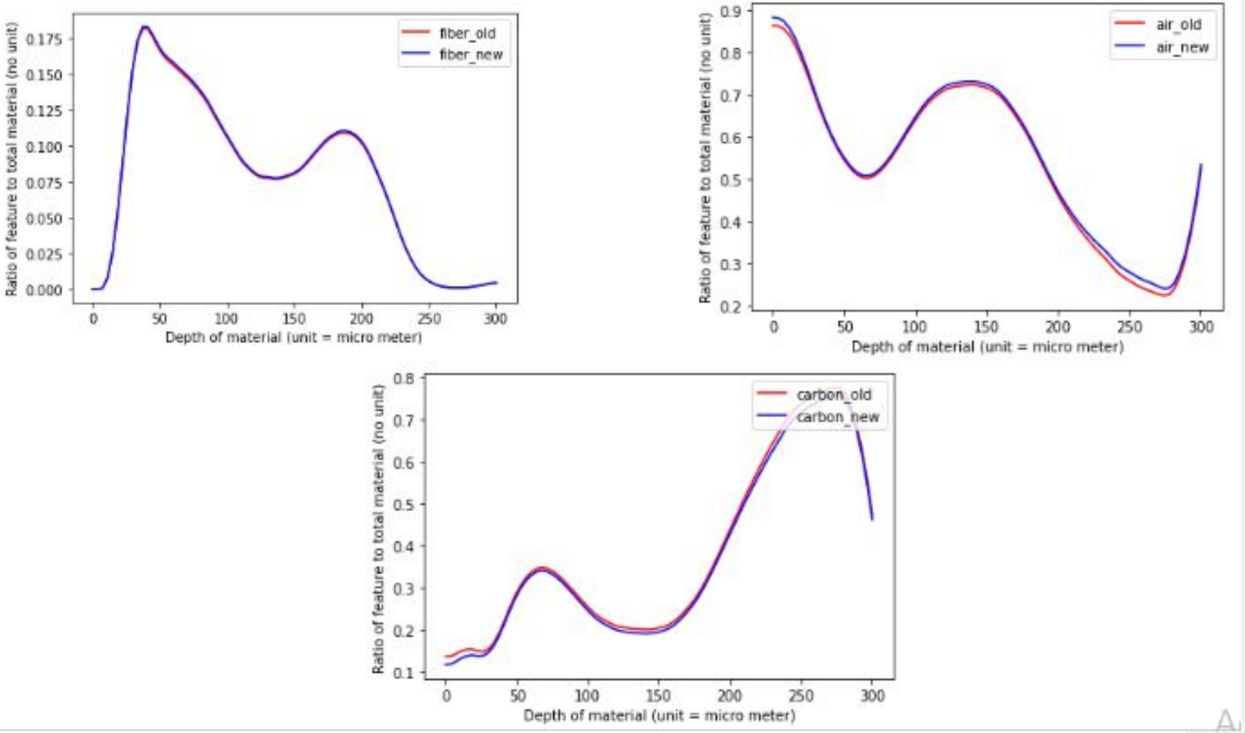


Figure 27: Distribution curves comparison of quarter view

Feature Engineering

Feature Engineering

Similarity analysis

We thoroughly analyzed scans in several ways to better understand the data and shortlist good features.

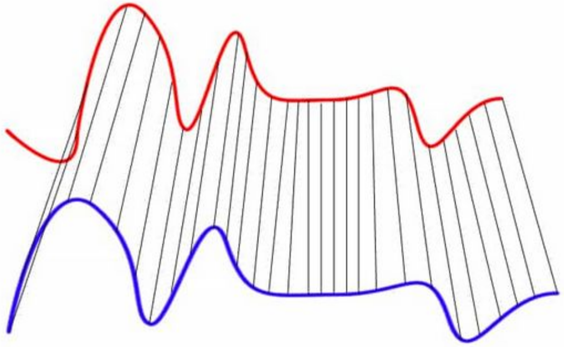
First, we analyzed the **similarities of distribution curves** using DTW metric in different capacities.
DTW (Dynamic Time Warping) compares the ups and downs of the graphs regardless of their lengths. It calculates the distance between two arrays or time series with different lengths.

This is important because almost all scans have different thicknesses. So this metric can be used robustly for pattern matching of our graphs as shown in the figure.

In this analysis we compared similarities of:

1. Samples of the same materials
2. Materials of the same company
3. Materials with similar ground truth values

For material to material comparison, we took the mean of material samples distribution curves and then computed DTW score. We define high similarity as above 70%. **Materials with similarity greater than 70% are considered to be similar**



Dynamic Time Warping Matching

[Link to detailed document]

Similarity analysis

Findings:

From our analysis, we observed that:

1) Generally, there is high similarity between samples of the same material for fibre, air and carbon.
2) Within the same company, Intra-material and inter-material similarities lie close to each other.
3) *Materials from different companies lying close in ground truth do not have high similarity.* Their element distributions are quite dissimilar. This shows that different companies have achieved the same properties, resistivity and permeability, using different compositions.

Conclusion:

Based on our findings, we concluded that:

1) As two dissimilar materials, belonging from different companies, can have equal or close physical property values, AI should be developed to **classify material according to companies first.** Two dissimilar distribution curves and similar ground truth can confuse the AI model.
2) Similarity between the materials of same company is high, therefore, it will be easy for the AI

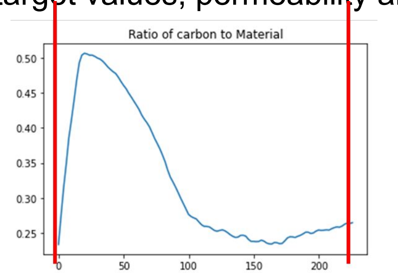
Link to detailed document

Correlation analysis

Analysis:

- Correlation coefficient values ranging from *0.50 to 0.75 or -0.50 to -0.75* indicate moderate to good correlation, and values from *0.75 to 1 or from -0.75 to -1* point to very good to excellent correlation between the variables.
- Positive correlation means direct relationship between the quantities and negative correlation means indirect relationship.
- We calculated the linear correlation between Area under the curve (AUC) and target values, permeability and resistance.

Correlation	Auc carbon	Auc fiber	Auc air
Permeability	-0.87389	0.084192	-0.73986
Resistance	0.897741	0.077085	0.757318
Auc carbon	1	0.090744	0.841605
Auc fiber	0.090744	1	0.069002



Findings:

There is a strong inverse relationship between carbon and permeability. We also found that carbon and resistance have a strong direct relationship. Even though air has a high correlation with the properties as well but it also has a high correlation with carbon.

Conclusion:

Based on the analysis, it can be concluded that AUC of carbon is a good feature to use for AI model training.

AI development

healthhub

Technique 1: Regression after classification

Based on the findings of Similarity analysis, Instead of training on the entire dataset, we used the following approach:

- classify the scans according to their companies (total 5 classes: JNTG An, JNTG Ca, FRB, SGL Ca, SGL An):
  - **Model:** Random Forest Classifier
  - **Features used:** Carbon, Fiber and air distribution graphs
- predict material's properties (using regression)
  - **Model:** Random Forest Regressor
  - **Features used:** Carbon, Fiber and air distribution graphs

We performed the following 3 tests to check the robustness of the model.

- Testing on 1 sample of each material from each company
- Testing on all samples of 2 whole materials from each company
- Testing on all samples of any one company.

Tests	Mean Percentage Error	
	Resistance	Permeability
1 sample from all companies (3 fold cross validation average)	2.77%	8.01%
2 materials from all companies (4 fold cross validation average)	4.7%	17.34%
All samples from one company (5 fold cross validation average)	25.01%	39.18%

Figure 28: Performance of model

Technique 2: Regression model for entire data

Based on the findings of Correlation analysis, we Trained 1 regression model for entire data:

- **Feature used:** Area under the curve (AUC) of Carbon distribution
- **Model for Resistance:** SVR (Support Vector regressor)
- **Model for Permeability:** Random Forest Regressor

Tests	Mean Percentage Error	
	Resistance	Permeability
1 sample from all companies (3 fold cross validation average)	7.14%	18.92%
2 materials from all companies (4 fold cross validation average)	7.60%	18.28%
All samples from one company (5 fold cross validation average)	11.75%	45.09%

Figure 29: Performance of model

Technique 3: Separate Regression model for each company

Based on the Previous two experiments, we observed that regression models work better when trained on each company separately instead of entire data.
Since classification through AI also introduces some error, We developed the solution to get company name as input from the user and train separate regression model or each company.

- **Features used:** Distribution curves of fiber, air and carbon
- **Model:** Random Forest Regressor

This technique displayed the best performance, among all.
But it obviously **cannot be used for unseen data** from some manufacturer other than the 5 provided at the moment.

Therefore, in the **final solution**:

- If user provides the manufacturer name (1 of the 5 manufacturers currently available), Technique 3 will be used for prediction of results
- If the user does not provide the manufacturer name, Technique 2 (1 regressor for entire data) will be used.

Tests	Mean Percentage Error	
	Resistance	Permeability
1 sample from all companies (3 fold cross validation average)	2.59%	7.69%
2 materials from all companies (4 fold cross validation average)	3.55%	10.13%
All samples from one company (5 fold cross validation average)	—	—

Figure 30: Performance of model

App Development

App Development

We developed an app using the following packages:

- Tkinter for 2D view
- PyQt and QVTK for 3D view

The app comprises of 3 screens:

- Screen 1 —> Analysis Board
- Screen 2 —> 2D visualisation and display of values from quantitative analysis and AI property predictions
- Screen 3 —> 3D visualisation

Link to the user guide : Link

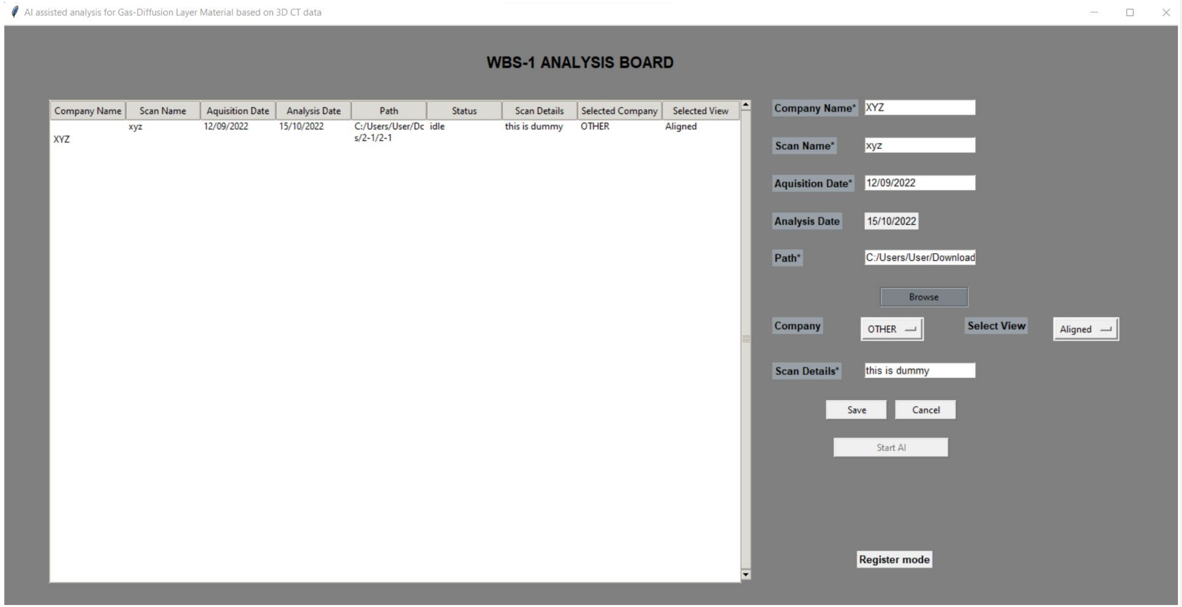


Figure 31: Front End Screen 1

App Development

Link to the demo for understanding the working of app can be found here.

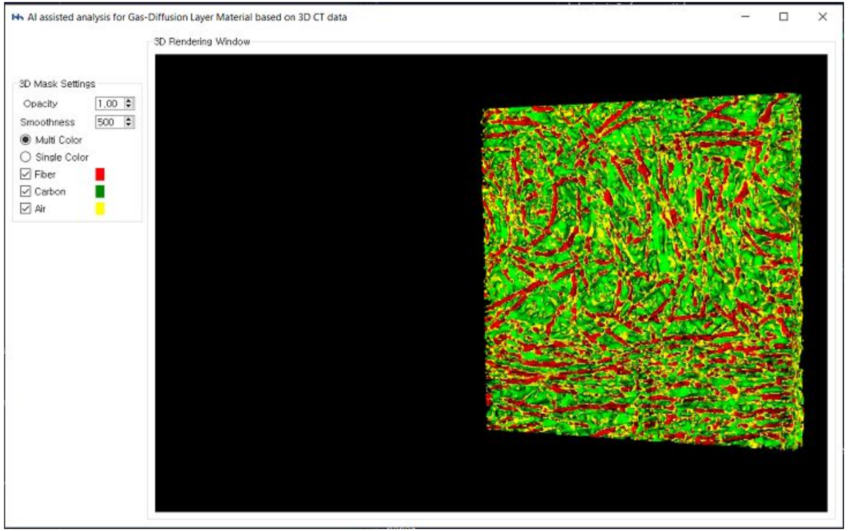


Figure 32: Front End Screen 3

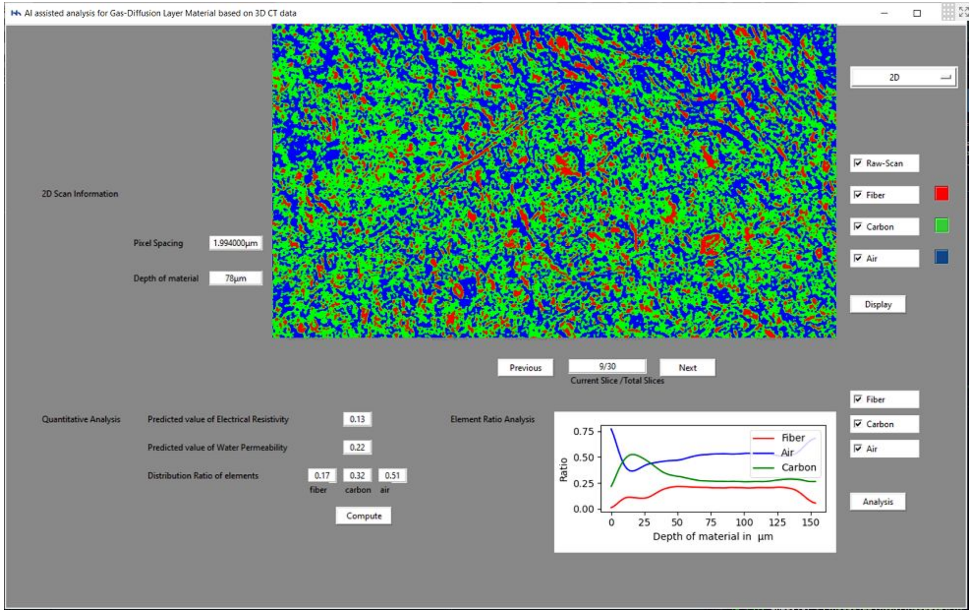


Figure 33: Front End Screen 2

Improvement of AI

Improvement of AI

The AI solution can further be improved as:

- Provided data was limited, as only 8 materials were provided for each company and ground truth values of properties were not diverse enough to train the AI models. Therefore, the performance of AI can be improved by providing more diverse data for training which will enable the AIs to explore all kinds of variation in the material and their corresponding properties.
- During the analysis, it was observed that the chemical properties of materials have a great effect on the physical properties but this information was not provided for each material used. We may exploit that information while training our AI models to improve the prediction and develop a robust model which can be applied to any unseen samples.
- We can also focus on predicting the possible composition of each element (ratio and structure of fiber, carbon and air) while manufacturing to obtain the product (GDL) having desired physical properties.
- We are currently predicting the value of two physical properties (Resistance and permeability). In the future, we can develop a solution to predict the values of other properties (like pore size distribution in 3D and fiber diameter, limiting current etc.) as well.

Conclusion

Conclusion

We have successfully developed AI for prediction of physical properties.
GUI has also been developed for displaying predicted physical properties of Gas-Diffusion Layer Material, visualisation of scan element segmentation results in both 2D and 3D along with the element ratios and distributions curves.

THANKS!

Any Questions?