Utilization of Machine Learning for Comprehensive Identification of Low Density Thin DebriSat Fragments

Bogachan Ondes[(1)], Norman Fitz-Coy[(2)]

[(1)] University of Florida, Department of Mechanical and Aerospace Engineering, 939 Center Dr, Gainesville, FL, USA
Email: bogachan.ondes@ufl.edu
[(2)] University of Florida, Department of Mechanical and Aerospace Engineering, 939 Center Dr, Gainesville, FL, USA
Email: nfc@ufl.edu

ABSTRACT

Orbital debris present a growing threat to operational satellites, necessitating accurate modelling of debris generation and evolution. In the early 1990s, NASA and the Department of Defense (DoD) conducted the SOCIT series of ground-based laboratory hypervelocity impact (LHVI) tests, providing critical data on debris characteristics. As satellite design evolved to include advanced materials such as carbon-fiber reinforced polymers and Kevlar, fragmentation behaviours changed significantly. To address these shifts, NASA, the DoD, The Aerospace Corporation, and the University of Florida initiated the DebriSat project in 2011, and subjected a modern LEO satellite test article to a LHVI at Arnold Engineering Development Complex in 2014. The resulting fragments were then transported to the University of Florida for comprehensive characterization of material, shape, mass, and dimensions, thereby delivering updated data to refine orbital debris modelling capabilities. Since 2022, machine learning (ML) capabilities are being developed to rapidly and systematically characterize materials of large volume of post-impact fragments generated during the hypervelocity tests.

1 BACKGROUND

The fragment material assessment in the DebriSat project relies primarily on non-destructive material characterization methods since the requirement that fragments remain undamaged during characterization. Due to large number of fragments, the principal technique involves measuring mass and dimensional data using the DebriSat mass and imaging systems. The UF team implemented several dimensional analysis instruments (2D and 3D fragment imagers) for material characterization, linking them through a database known as the Debris Categorization System (DCS) [1][2]. The DCS is an interactive database where students use an interface to record fragment dimensional data along with fragment images. Fragments are categorized by their shape, colour, size and materials. The size of the fragment is determined by its height, if a fragment has height over 3 mm, it is categorized as 3D fragment and imaged using the 3D imager. If the fragment's height is lesser than 3 mm, it is categorized as a 2D fragment (thin fragment) and imaged using the 2D imagers.

A preliminary survey indicated that the majority of the fragments were 2D fragments. Moreover, significant portion of the 2D fragments were found to be carbon-fiber reinforced polymer (CFRP). These CFRP fragments were found as needle-like or flat plate shapes, both of which exhibit a small height (Z_{DIM}). As shown in Fig. 1, 2D imaging systems were developed to efficiently determine the volumes of thin fragments. These systems capture both top and side views using a camera positioned above the fragment. The top-view pixel area is calculated by measuring the cross-sectional area from the image pixels, while a 45-deg mirror is used to provide a side view, enabling accurate height measurement (Z_{DIM}). [3]

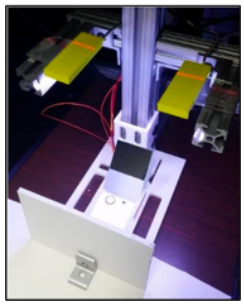

Figure 1. One of two 2D Imagers

The prismatic volume approach yields highly accurate volume estimates for flat, uniform shapes but proves inadequate for complex geometries, such as metal nuggets or entangled wires. Since density plays a crucial role in making decision during material characterization, it is very important to obtain an accurate measure of the volume in order to determine of an accurate density. In cases where volume measurements were known to be imprecise, operators relied on their pre-impact familiarity with the DebriSat test article to identify metals.

However, since char layers formed on most fragments during the hypervelocity impact testing, visual identifications at time are inaccurate. Consequently, density became the most dependable metric for classifying metals, yet volume inaccuracies introduced significant variability in density estimates. Although

predefined density ranges for each metal group helped guide decisions, fragments with density values near these boundaries remained challenging to classify.

2 CHARACHTERIZATION BIAS

In 2022, a labelling bias was identified within the DCS, wherein some stainless steel fragments had been misclassified as titanium. As shown in Fig. 2 and Fig, 3, the titanium fragments were found (mostly intact) after a search in the DCS data and the fragment repository.

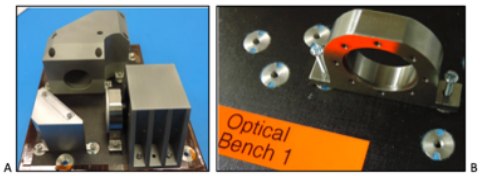

Figure 2. Titanium Camera Mounts (Before Impact)

Figure 3. Titanium Fragments (After Impact)

Further investigation revealed that the root cause of this issue stemmed from the approach used to compute volumes for thin fragments. As seen in Eq. 1 and Fig. 4, the prismatic volume calculation was leading to overestimation of the fragment volumes when the fragment was not perfectly prismatic. The labelling bias was also found in aluminium labelled fragments; many SS fragments volumes were overestimated and were labelled as AL in the DCS due to their low density value.

$$V_{prismatic} = \text{pxa} * \text{height}$$

$$V_{prismatic} \geq V_{true} \tag{1}$$

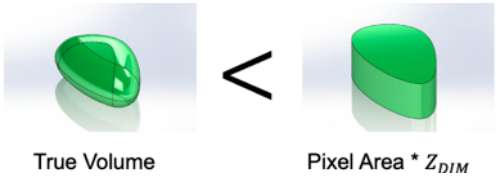

Figure 4. Prismatic Volume Comparison

To compensate this volume overestimation and resolve the labelling bias, an algorithm was developed using an ellipsoidal volume of the fragments. This algorithm was labelled the Dynamic Volume Algorithm (DVOL), and it was used to relabel (previously biased) stainless steel fragments. The DVOL has shown to successfully place stainless steel fragments into the high density region where stainless steel fragments would reside [4]. The selection bias is desired to be avoided for future characterizations; thus, machine learning (ML) was utilized as a proactive measure. The ML model would predict the stainless steel fragments and would prevent students from classifying thin stainless steel fragments as titanium or aluminium.

3 BIAS MINIMIZATION USING ML

The DVOL has minimized the selection bias in the thin SS fragments dataset as it was able to obtain true stainless steel labels. Thus, it was decided to train a ML model that can identify stainless steel fragments from the not yet categorized thin metal fragment dataset. This binary classification task was accomplished using Logistic Regression. The logistic regression model (LRM) was trained using a set of numerical features, including mass, dimensions, and density estimates, to improve the classification accuracy of stainless steel fragments.

3.1 Training the Logistic Regression Model

Logistic Regression algorithm is widely recognized as a simple yet powerful classification algorithm that excels in both interpretability and computational efficiency. Its relatively small number of parameters reduces the risk of overfitting compared to more complex methods, while still delivering robust performance in many real-world applications [4]. The key advantage of the LRM is its ability to find and label SS fragments before any biased decisions are made.

The dataset used for training was pre-processed to remove inconsistencies, and feature selection was conducted to ensure that only the most relevant attributes contributed to the classification process. To optimize the model, hyperparameter tuning was performed by adjusting the regularization strength to prevent overfitting while maintaining generalizability. The model was trained using a stratified dataset to account for material characterization imbalances, ensuring that the decision boundary was not skewed toward the more prevalent material categories. Cross-validation was employed to assess model robustness, and performance metrics such as accuracy, precision, recall, and F1-score were analysed to validate its effectiveness. The final version of the LRM was able to obtain over 95% accuracy for predicting SS in the training and the test datasets [4].

Once trained, the LRM was applied to the remaining thin metal fragments in the dataset. The predicted classifications were then compared against known samples to assess model confidence and refine decision

thresholds. Overall, this process effectively reduced the reliance on manual identification and mitigated SS selection biases associated with previous volume-based classifications.

3.2 LRM Limitations

The results demonstrated that the LRM could successfully differentiate stainless steel fragments, providing a scalable and systematic approach to minimizing selection bias in the DebriSat dataset. However, the LRM success did not transfer to predicting aluminium. This was revealed as the not yet categorized metal dataset was found to have fragment types that are metallic by visual inspection yet has very low density. One of the most common examples to these types of fragments is multi-layer insulation (MLI) fragments. As seen in Fig. 5, it visually similar to a metallic foil, yet its density is significantly low to be a metal fragment ($0.00038\ \mathrm{g/mm^3}$).

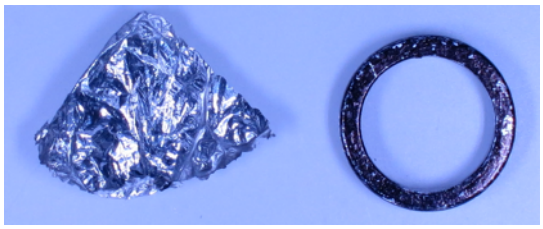

Figure 5. Example MLI Fragment

The LRM utilized is inherently a binary classifier**.** While it can be extended to multiclass classification using one-vs-rest or SoftMax regression (multinomial logistic regression)**, t**hese approaches often struggle when the dataset contains multiple classes with overlapping distributions [5]. The classification task in the dataset involves multiple material labels, which require more complex decision boundaries that LRM may not effectively model.

LRM assumes that data can be separated using a linear decision boundary**,** which works well when the relationship between input variables and class probabilities is approximately linear [5]. However, materials such as AL, MLI, and CFRP do not exhibit clearly linear separability in the DCS numerical feature space. For instance, MLI fragments are extremely low-density, but their densities may overlap numerically with plastic fragments. Similarly, AL and Glass fragments have densities that are numerically equivalent and thus require more complex feature relationships for differentiation. LRM is sensitive to feature overlap, where different materials share similar numerical characteristics (e.g., mass, density, dimensions).

4 UTILIZATION OF DECISION TREES FOR MULTIPLE LABELS

The transition from using the LRM as a classification algorithm to the Decision Tree model (DTM) in material classification significantly improved the ability to distinguish between different fragment types in the dataset. LRM's inherent limitations, particularly its reliance on a linear decision boundary, made it unsuitable for capturing the complex relationships among fragment properties such as mass, density, and shape. The DTM provided a more flexible and effective approach by enabling hierarchical decision-making, better handling class imbalances, and offering robustness against measurement noise [6].

A major advantage of the DTM is the ability to handle nonlinear relationships in the dataset. Unlike LRM, which assumes a linear separation between classes, DTM partition the feature space iteratively based on optimal threshold values. This was particularly beneficial for classifying materials that exhibited overlapping densities but had distinct geometric characteristics. By allowing splits based on features where dimensional data is compared to each other for better understanding of the fragments shape. These features, such as xyRatio ($\frac{X_{DIM}}{Y_{DIM}}$), PXA ratios ($\frac{X_{DIM}}{PXA}, \frac{Y_{DIM}}{PXA}$), and ellipsoid area ($\frac{\pi * X_{DIM} * Y_{DIM}}{4}$), help the DTM to capture patterns that were not apparent in a linear framework. This capability proved especially effective in differentiating materials such as MLI (fragments appear more of a square, xyRatio is close to 1) and CFRP (mostly slender fragments, xyRatio is significantly greater than 1), where shape variations were crucial in classification.

Another key improvement introduced by the DTM was their ability to capture feature interactions dynamically. Unlike LRM, which assigns a fixed weight to each feature across all data points, the DTM allow the importance of a feature to change based on previous splits in the decision path. This adaptive process enabled the model to refine its classification strategy based on multiple criteria. For example, in distinguishing stainless steel from aluminium, the DTM initially considered density as a primary factor but further refined classifications using secondary attributes such as cross-sectional area and geometric ratios. This flexibility ensured that materials with complex fragmentation patterns were more accurately categorized.

Class imbalance, which posed a significant challenge when using the LRM as a classification algorithm, was also effectively managed by the DTM approach. The dataset contained underrepresented materials such as MLI and CFRP which LRM tended to misclassify due to its bias toward majority classes. The DTM mitigated this issue by adjusting their splitting criteria to ensure that minority classes were adequately represented in the classification process. Additionally, the use of stratified cross-validation ensured that all material types contributed proportionally to the training process,

reducing the likelihood of underrepresented materials being misclassified. As a result, The DTM provided a more balanced classification framework that accurately identified rare material fragments.

Beyond classification performance, The DTM also demonstrated greater robustness to outliers and noise in the dataset. Due to the nature of hypervelocity impact fragmentation, certain materials exhibited extreme density values or irregular shapes that traditional classification models struggled to handle. LRM's reliance on a continuous optimization process made it highly sensitive to such anomalies, leading to misclassifications. The DTM, by contrast, utilized threshold-based splits that minimized the influence of outliers. This resulted in a more stable classification system that was less prone to errors caused by measurement variations. Additionally, fragments with unexpected density values, which previously led to mislabelling, were reassigned based on a combination of multiple criteria rather than a single density threshold.

An important benefit of the DTM approach was its interpretability. Unlike LRM, which provides only global feature weights, The DTM allow for a detailed analysis of feature importance scores, offering insights into the most relevant characteristics for classification. The ability to visualize decision paths provided further transparency, allowing researchers to validate classification logic and refine the dataset accordingly. This analysis revealed that, while density remained an important feature, additional attributes such as shape descriptors and cross-sectional area ratios played a crucial role in improving classification accuracy. By highlighting these key predictors, The DTM enabled a more informed and systematic classification process.

The shift from using LRM as a classification algorithm to The DTM resulted in significant advancements in material classification by overcoming issues related to linearity, feature interactions, class imbalances, and outliers. The ability to model complex relationships, dynamically adjust classification criteria, and remain robust against noise made The DTM a superior choice for fragment identification. While challenges remained, particularly in distinguishing aluminium from glass, The DTM established a stronger foundation for classifying lower density fragments. Their hierarchical decision-making process and ability to incorporate multiple feature interactions provided a more reliable and interpretable approach, ultimately improving classification accuracy and dataset integrity.

4.1 Training Dataset Construction

The selection of fragment classes for training the DTM is based on the brief observations of these (visually classified as metal) specific materials—AL, MLI, SS, CFRP, Glass, and copper (CU)—that are present in the not yet categorized dataset. The goal of this manual selection process is to ensure that the model is trained on the same material categories expected to appear in the not yet categorized set, improving its ability to classify new fragments effectively. Moreover, any existing biases in the labelling of materials other than stainless steel could propagate through the DTM classification, affecting its overall accuracy. To mitigate this risk, the manual construction of the training dataset serves as a proactive measure to prevent the reinforcement of such biases, ensuring a more balanced and representative classification process.

The not yet categorized metal dataset was seen to consist of different sizes of classes. A specific case in this dataset is CU, which, despite being a metal, is typically found in a wire-like shape. This distinct geometric form makes CU easier to characterize as metal compared to other materials. It is encountered in high quantities in the not yet categorized metal dataset and usually appears in tiny wire shapes. Thus, the manually curated dataset has reflected this domain knowledge by providing more samples of CU to ensure a good representative mix of fragment types that would be encountered by the model while making predictions. A specific case in this dataset is CU, which, despite being a metal, is typically found in a wire-like shape. This distinct geometric form makes CU easier to classify compared to other metals. However, due to the presence of char layers or colour inconsistencies resulting from the hypervelocity impact, CU fragments may not always be immediately identifiable based on visual inspection alone. As a result, CU still appears in the not yet categorized set before it is properly labelled, highlighting the need for a classification model that accounts for both numerical attributes and potential visual ambiguities.

This method provided a practical balance between ensuring that the DTM encountered all relevant material types while also allowing it to generalize effectively to the not yet categorized dataset. The somewhat random nature of sample selection may introduce slight biases but given the Decision Tree's ability to handle complex feature relationships and class separations, this approach still allowed the model to learn meaningful classification rules.

4.2 Feature Engineering for Decision Tree Algorithm

The DCS data underwent pre-processing following the same methodology used for the LRM ensuring consistency in feature extraction and transformation across both models. Python's NumPy and Pandas libraries were utilized extensively to manage numerical data, perform transformations, and engineer new features to improve classification accuracy. The primary objective of this stage was to refine the dataset in a way that maximized the discriminative power of key

attributes while maintaining compatibility with the model's processing pipeline.

There were two additional features that were not present in LRMs feature set, the X_{DIM} and Y_{DIM} ratios to PXA, provide information about the slenderness of the fragment, an important geometric property that could influence classification. The rationale behind including these ratios was that certain low density materials tend to fragment into distinct shapes (e.g. CFRP needles), and their aspect ratios could serve as useful indicators for distinguishing between different material classes. However, these features were excluded from the LRM's feature set because this relation is found to be nonlinear and it was unnecessary as it was not helpful in LRMs classification capabilities. To ensure uniformity in data representation, categorical and numerical features were transformed using the same pipeline as the LRM. OneHotEncoder was applied to categorical variables, converting them into numerical format suitable for machine learning models while preserving the distinctiveness of each category [4]. For numerical attributes, MinMaxScaler was used to scale the data, normalizing feature values within a specific range to prevent any single attribute from disproportionately influencing the model's decision-making process [4]. This transformation ensured that all input variables were appropriately scaled and encoded, allowing the Decision Tree classifier to efficiently learn patterns and relationships from the dataset while effectively incorporating the newly introduced slenderness-based features.

4.3 Hyperparameter Tuning

A systematic exploration of the model's hyperparameters was performed using grid search with cross validation from scikit-learn [4]. The grid search process involved evaluating multiple configurations to identify the most effective decision tree structure. The criterion for measuring impurity was examined by comparing both the gini and entropy functions to determine which provided better classification performance. As an illustration, first 3 levels of the DTM are shown in Fig. 6. The model's complexity was controlled by testing maximum depth values ranging from 1 to 9, ensuring that the tree was neither too shallow nor excessively deep, which could lead to overfitting. Additionally, different thresholds were explored for the minimum number of samples required to split an internal node and to form a leaf,

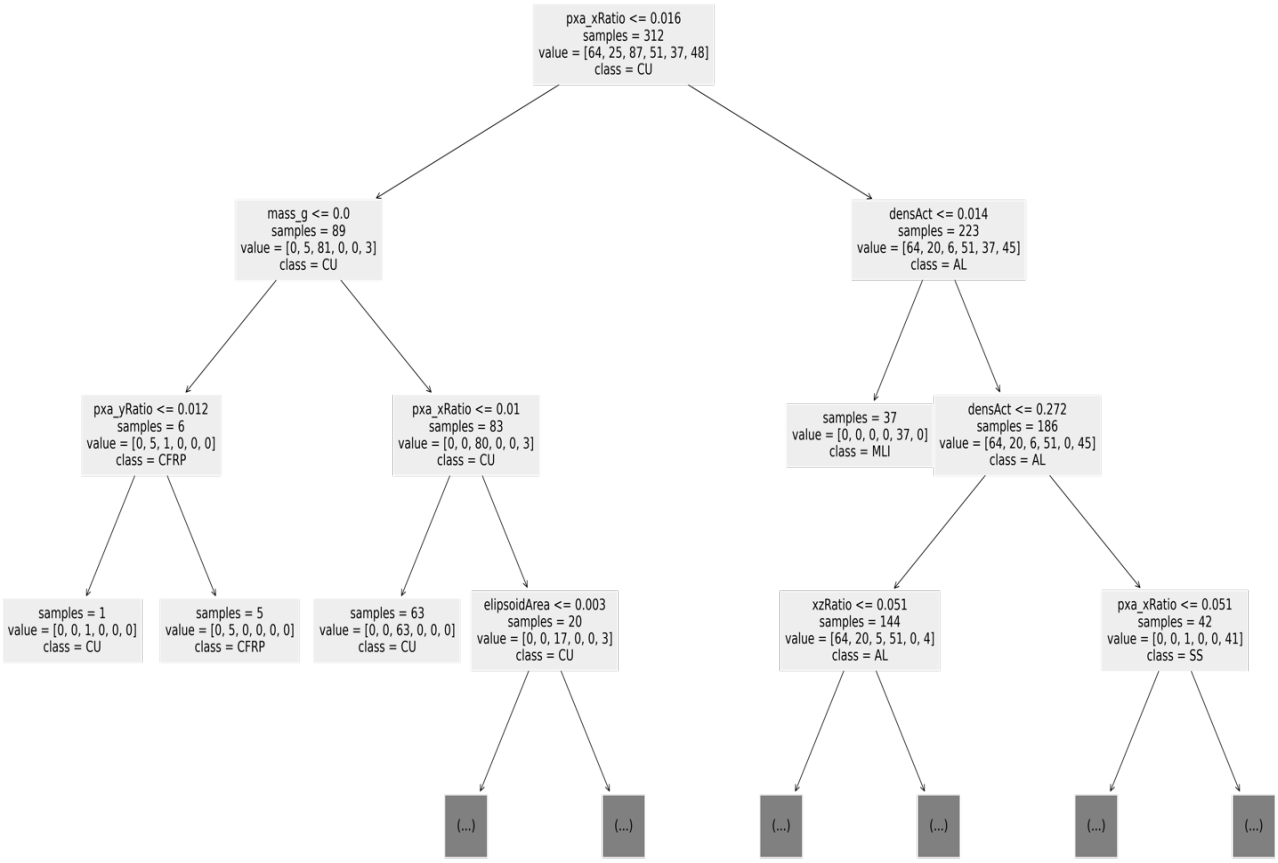

Figure 6: The first three levels of the Decision Tree

allowing for a balance between generalization and model specificity. Through this optimization process, by maximizing models overall, the most effective configuration was determined to utilize the entropy criterion, with a maximum depth of 7 and a minimum of 5 samples for splitting nodes. This relatively shallow tree structure was chosen to prevent overfitting while still preserving the ability to model essential nonlinear relationships within the dataset.

4.4 Decision Tree Model Results

As seen from Tab. 1 and Tab. 2, the performance of the DTM demonstrated strong classification capabilities for most material types, particularly for distinct classes such as SS, CU, CFRP, and MLI. However, the model encountered some challenges in differentiating between AL and glass, which contributed to a noticeable drop in overall accuracy on the test dataset. The DTM achieved an overall accuracy of 82% on the test set, which is significantly worse than the training set. The primary reason for this decline is misclassification which stemmed from glass and AL differentiation. The precision for glass (0.52) confirms that it is frequently mislabelled, reinforcing that its numerical features overlap significantly with AL. With AL having a recall of 0.72, the model often fails to capture all its instances correctly, further contributing to classification errors. The strong precision for copper suggests its unique wire shape helps separation, but its recall of 0.80 indicates some misclassification before final labelling.

Table 1: Decision Tree Performance on Training Set

Labels	Precision	Recall	F1-Score	Support
AL	0.95	0.96	0.95	64
CFRP	1.00	0.92	0.96	25
CU	1.00	0.95	0.98	87
GLASS	0.85	1.00	0.92	51
MLI	1.00	1.00	1.00	37
SS	0.96	0.98	0.97	48
Accuracy Scores			0.95	312
Macro Average	0.96	0.96	0.96	312
Weighted Average	0.96	0.96	0.96	312

Table 2. Decision Tree Performance on Test Set

Labels	Precision	Recall	F1-Score	Support
AL	0.84	0.72	0.78	65
CFRP	0.90	0.79	0.84	24
CU	0.99	0.80	0.88	88
GLASS	0.52	0.84	0.65	51
MLI	0.95	1.00	0.97	37
SS	0.91	0.83	0.87	48
Accuracy Scores			0.82	313
Macro Average	0.85	0.83	0.83	313
Weighted Average	0.86	0.82	0.83	313

The primary reason for this classification difficulty stems from the highly similar numerical properties of AL and glass fragments, particularly in terms of density. Since density is a key distinguishing feature in the dataset, the overlap between these two materials creates an inherent limitation in the model's ability to establish a clear decision boundary. Unlike other material types that exhibit distinct ratios or unique geometric features that reveals with dimensional comparison features, AL and glass often share nearly identical measurements, making them indistinguishable using numerical attributes alone.

This issue is further exacerbated by the fragment characterization process, where visual cues—such as transparency, edge sharpness, and surface texture—are not captured in the numerical dataset. The decision making process relies solely on measurements such as mass, volume, and derived density metrics, which, while effective for certain materials, are insufficient to accurately distinguishing between AL and glass. Glass fragments exhibit physical properties that make them visually distinct, including their sharp edges, brittle fracture patterns, and transparency, characteristics that are easily recognizable to a human observer. However, since these attributes are not represented (as textual or numerical descriptions) in the DCS feature columns, the DTM has no access to the information that would allow it to make such distinctions. Instead, it relies solely on numerical attributes that fail to provide adequate separation between AL and glass, both of which can have similar densities and geometric ratios in the dataset.

As a result, the Decision Tree frequently misclassifies these two materials, introducing systematic errors in test predictions and lowering overall classification performance. This limitation underscores a fundamental

shortcoming of purely numerical classification systems, highlighting the necessity of incorporating image-based artificial intelligence (AI) for more accurate material identification. Students, when classifying fragments, inherently rely on visual inspection, using edge patterns, reflections, and overall structural characteristics to differentiate between AL and glass.

Since this level of perception is absent from the current numerical dataset, an AI model capable of processing images would be essential for bridging the gap. By integrating computer vision techniques, such as convolutional neural networks (CNNs), the classification system could leverage visual features that are imperceptible to numerical descriptors alone, improving accuracy where traditional DTMs fall short. Thus, while numerical classifiers provide an efficient baseline, they are inherently limited when tasked with differentiating materials that exhibit overlapping physical properties, making image-based classification a necessary next step for improving fragment identification.

5 FUTURE WORK

Building upon these observations, further improvements to the classification framework must address the fundamental limitations of numerical-based material identification. Due to the persistent classification inconsistency between AL and glass, the overall accuracy of the model on the test dataset remains notably lower than on the training set. While the Decision Tree effectively learns patterns for most materials, its inability to reliably separate AL from glass leads to frequent misclassifications, significantly impacting test performance. These results reinforce the conclusion that the model is constrained by the available numerical features, necessitating alternative approaches such as image-based analysis to resolve material ambiguities [8].

Despite this limitation, the DTM remains an effective tool for categorizing most materials, particularly those with distinct density distributions or geometric properties. However, cases where numerical similarities prevent clear differentiation demonstrate that relying solely on numerical data is insufficient. Addressing these challenges in future work will involve developing hybrid classification methods that integrate both numerical descriptors and image-based AI techniques. By incorporating machine vision models alongside the DTM, classification accuracy can be improved, particularly for materials where numerical attributes alone do not provide enough separation [9].

6 REFERENCES

1. NASA (n.d.). *DebriSat*. Orbital Debris Program Office. Available at: https://www.orbitaldebris.jsc.nasa.gov/measurements/debrisat.html (Accessed: 18 March 2025).
2. Kleespies, J. (2018). *Design, Development, and Implementation of the DebriSat Debris Categorization System*. Master's Thesis. Department of Electrical and Computer Engineering, University of Florida, Gainesville.
3. Toledo, R., Shiotani, B., and Fitz-Coy, N. (2019). *Imaging Systems Utilized in the DebriSat Fragment Size Characterization Process*. First International Orbital Debris Conference, Sugar Land, TX.
4. Ondes, B. (2023). *The Minimization of the Selection Bias for Thin DebriSat Fragments with Machine Learning*. Master's Thesis. University of Florida. Available at: https://ufdcimages.uflib.ufl.edu/UF/E0/06/11/86/00001/Ondes_B.pdf (Accessed: 18 March 2025).
5. Scikit-learn Developers. (2024). *Logistic regression: Linear models for classification*. Available at: https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (Accessed: 18 March 2025).
6. Scikit-learn Developers. (2024). *Decision Trees: Understanding classification and regression trees*. Available at: https://scikit-learn.org/stable/modules/tree.html (Accessed: 18 March 2025).
7. Scikit-learn Developers. (2024)*: Exhaustive grid search over specified parameter values*. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (Accessed: 18 March 2025). (Accessed: 18 March 2025).
8. LeCun, Y., Bengio, Y., and Hinton, G. (2015). *Deep learning*. Nature, 521(7553), 436-444.
9. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks*. Advances in Neural Information Processing Systems, 25, 1097