International Journal of Sustainable Development Goals

www.ijsdg.org
eISSN: 3080-5023

Advancing Aviation Safety and Sustainable Infrastructure: High-Accuracy Detection and Classification of Foreign Object Debris Using Deep Learning Models

Yaseen Mushtaq[1], Wajid Ali[1*], Usman Ghani[1], Rahim Ullah Khan[1], Amal Kumar Adak[2]
[1] Department of Computer Science, Air University Islamabad, E-9, Islamabad, 44000, Pakistan
[2] Department of Mathematics, Ganesh Dutt College, Begusarai, Bihar 851101, India

ARTICLE INFO

Article history:
Received 23 April 2025
Received in revised form 24 May 2025
Accepted 25 May 2025
Available online 26 May 2025

Keywords:
Deep Learning; YOLOv8; Foreign Object Debris (FOD); Material Classification; CNN; Aviation Safety; Artificial Intelligence; Sustainable Development Goal 9; Infrastructure Innovation

ABSTRACT

Foreign Object Debris (FOD) presents a critical threat to aviation safety, with the potential to damage aircraft and jeopardize lives. This study explores the use of Deep Convolutional Neural Networks (DCNNs) for the precise detection and classification of FOD, aiming to transform existing prevention strategies. By employing models such as Xception and YOLOv8, the system achieved detection accuracies of up to 98% on diverse datasets. The integration of AI-based approaches significantly enhances operational efficiency, contributing directly to the United Nations Sustainable Development Goals (SDGs), particularly SDG 9: Industry, Innovation, and Infrastructure: Industry, Innovation, and Infrastructure, by promoting smart, safe, and sustainable aviation systems. The findings highlight the pivotal role of innovation in strengthening critical transportation infrastructure and ensuring resilient airport operations aligned with global development goals.

1. Introduction

Foreign Object Debris (FOD) refers to any unwanted materials on airport runways or taxiways that could cause harm to aircraft or passengers. Traditional FOD mitigation techniques often rely on manual inspections and mechanical equipment, which can be inefficient and error prone. To address this, we propose a deep learning-based solution using state-of-the-art detection models like Xception and YOLOv8 for real-time, high-accuracy identification of FOD. By embedding intelligent technologies into airport safety systems, this research directly supports Sustainable Development Goal 9 (SDG 9): Industry, Innovation, and Infrastructure, encouraging the development of resilient and technologically advanced aviation infrastructure. The adoption of AI in airport maintenance systems not only advances operational safety but also promotes sustainable innovation across the global transportation sector.

All of these systems operate on the basis of employing cameras to obtain the image needed for the identification of FOD, with a human expert performing the last verification [1], [2], [3]. These automated FOD detection devices are limited to a few airfields worldwide. This sparse deployment

* wajidali00258@gmail.com
https://www.doi.org/10.59543/ijsdg.v1i.14279

has a number of causes, the primary one being the final verification process, which has two unique drawbacks. It is first necessary to have someone who is well-trained and resilient, which requires the field authority to manage force outflow expenses. The other drawback is that human beings by nature tend to make mistakes, regardless of experience or mental toughness. Therefore, automatic FOD recognition with advanced and economical computer vision executions is a preferable outcome in such a script. Globally, deep literacy procedures are widely utilized in automatic recognition/discovery tasks, mostly in the medical field. Where object detection is the most powerful thing in computer vision, this field has a huge impact and can be investigated in many methods to produce an accurate FOD identification and bracket system. Conversely, material recognition is a very recent but rudimentary field in computer vision.

Metal Concrete Plastic

Figure 1 Real FOD dataset introduced by Zainab et al. [9].

Contrary to decades of research on object identification and field of material recognition is relatively new, and the application of deep neural networks for automatic point birth has demonstrated promise in overcoming the drawbacks of manual point birth. Deep neural network methods involve obtaining image features and are desirable because of their advanced perfection and rapid recognition. Material recognition is a relatively new way, in comparison to decades of work on object identification, and the use of deep neural networks for autonomous point birth has demonstrated promise in overcoming the drawbacks of manually performed point birth. Millions of images are used to train deep neural networks to achieve reasonable results in material recognition.

Using laborious and slow computational methods including BRDF [4], SIFT [5], Histograms of Gradient [6], interest points [7], and optic inflow [8]. We also proved that material detection can be very helpful in FOD detection as we not only just used material recognition but also used material detection in our work and results show that material detection is quite helpful in detection of FOD. Below, we summarize the major contributions of this work:

i. A robust feature-based approach is introduced to accurately classify the material categories of FOD items across all classes, achieving an average accuracy rate of 95%, not limited solely to the metal class.
ii. A detection of FOD using YoloV8 on real FOD dataset introduced by Zainab et al [9].
iii. FOD dataset of real FOD items and Chinese dataset is merged into single dataset and then performed material recognition on it which produced extraordinary results irrespective of the difference in shapes present in images.
iv. The theoretical framework supporting the refined classification algorithm is established to illustrate alignment between theory and the observed outcomes.

v. The proposed method has surpassed the previously used algorithm, achieving much better classification accuracy in the field of Foreign Object Debris (FOD) classification via image processing with DCNNs and also obtained much better accuracy on detection of FOD using YoloV8.

The remainder of the paper is structured as follows: Section 2 contains the literature survey, while Section 3 elaborates on the methodology employed for the research. In Section 4, we provide details about the newly merged dataset, and Section 5 give details of experiment and results. Finally, Section 6 concludes the research.

2. Literature Review

This section provides an in-depth examination of prior research concerning material recognition, Foreign Object Debris detection/recognition, and computer vision utilizing deep neural networks, all aimed at tackling the FOD classification challenge addressed in this paper.

2.1. FOD Classification and Detection

The literature review provides an in-depth exploration of the current landscape of Foreign Object Debris (FOD) detection methodologies, with a particular emphasis on advancements in image processing techniques and their application in enhancing aviation safety and operational efficiency. Traditional approaches to FOD detection, predominantly reliant on manual inspections or 'FOD walks,' are acknowledged for their inherent limitations, including susceptibility to human error, inefficiency, and vulnerability to weather-related challenges [10].

In response to these challenges, significant strides have been made towards the development and deployment of automated FOD detection systems. These systems leverage a diverse range of sensing modalities, encompassing radar, optical cameras, and advanced image processing techniques, to detect and classify FOD with increased accuracy and efficacy [10].

Radar-based detection systems, such as millimeter-wave radar, have long been established as a cornerstone of FOD detection, offering unparalleled precision and reliability, particularly in identifying larger objects on airport runways [11]. However, the high cost associated with radar technology has constrained its widespread adoption, prompting researchers to explore alternative sensing modalities, including optical camera systems, which offer a more cost-effective solution for enhanced FOD detection.

The integration of advanced image analysis techniques has significantly enhanced FOD detection capabilities, with researchers successfully incorporating sophisticated detection frameworks such as Faster R-CNN, Single Shot Detector, and Alex-Net to achieve remarkable improvements in detection accuracy [12, 13, 14]. These methods utilize pattern recognition and feature extraction for classification, with notable enhancements observed in model architecture such as DenseNet and VGG-16.

Moreover, the introduction of novel object detection frameworks such as DeepLabv3+ and YOLOv5 has further augmented FOD detection capabilities [16]. These frameworks incorporate advanced techniques like spatial pyramid pooling and attention mechanisms to optimize feature extraction and classification performance, resulting in heightened accuracy and speed.

Efforts to integrate multiple image processing techniq**ues** and curate comprehensive datasets have also been pivotal in advancing FOD detection accuracy [9]. Techniques such as combining multiple detection methods and data augmentation have proven instrumental in mitigating challenges associated with limited data availability and class imbalance, culminating in the development of more robust and generalizable detection systems [9].

In addition to radar and optical camera systems, hybrid approaches that combine radar with infrared sensors or electro-optical imagery have shown promising results in improving detection capabilities across diverse environmental conditions [16]. Commercially available FOD detection systems, including the Tarsier millimeter-wave radar and Hitachi's radar over fiber system, have garnered regulatory approval from authorities such as the Federal Aviation Administration (FAA), underscoring their efficacy in real-world applications.

Furthermore, the literature review underscores the significance of preprocessing techniques in optimizing image data for FOD detection. Grayscale conversion, histogram equalization, and Gaussian filtering are among the common preprocessing methods employed to enhance image quality and feature visibility, while feature extraction techniques such as Linear Discriminant Analysis and Gray-Level Co-occurrence Matrix contribute to precise classification. Statistical classification methods, including K-Nearest Neighbors and Random Forest, further improve detection accuracy, offering a versatile and effective framework for FOD detection tasks [17].

Moreover, the transformative potential of automated imaging systems in streamlining FOD detection and removal processes is highlighted, with the ultimate goal of minimizing disruption to airport operations and mitigating the risk of aircraft damage. However, significant challenges, including safe integration with existing air traffic, weather-related disruptions, and computational complexity, underscore the need for ongoing research and development in this domain.

2.2. Related Dataset

The available datasets in this context are CUReT [18], the KTH- TIPS [19] and FMD [20], and the raw items in environment Database MINC are the most common datasets extensively used for detection. For 61 texture images, the CUReT data accommodates illuminations and angles on a scale of 205 types. KTH- TIPS has a total of 11 material orders, with four samples for each order. In order to increase diversity, every sample is photographed in vibrant settings. The Accoutrements dataset have 10 material orders with 100 images of each order. About 3 million image patches of 23 material orders that have been adapted from ImageNet are available in MINC.

Even though all of these datasets have a great deal of diversity, they are inappropriate for identifying FOD waste because they were primarily collected indoors and do not correspond to the lighting or conditions associated with FOD waste emergence. The main challenge in finding delicate FOD is the absence of a suitable dataset that reflects real lighting conditions and various levels of rainfall. Taking this into account, Xu et al [12] introduced the FOD dataset having several classes. Although it is a huge contribution by them as availability of FOD datasets are very limited, still this data had some downsides like absence of real background, Placement of FOD at the center of image and utmost of the images comprise bulky objects such as bolts, wrenches, plastic bottles etc. Keeping this thing in mind S.M. Zainab [9] made another dataset they called real FOD particle (set up by safety labor force on a functional runway) in three basic classes were made in this dataset. The data consists of 2010 train, 336 validation, and 126 test images. Also, almost equal numbers of photos were shot for each script in the morning, evening and nighttime.

In response to the limitations identified in existing FOD datasets, particularly regarding unrealistic backgrounds, centralized placement of FOD, and predominance of bulky objects, a novel approach is devised. By amalgamating the datasets created by Xu et al. [12] and Zainab et al. [9], which respectively provided comprehensive class coverage and real-world FOD instances, a more diverse and representative dataset is established. This combined dataset addressed the shortcomings of individual datasets, offering a broader range of FOD scenarios captured across different times of day. Leveraging this enhanced dataset, an Xception model is trained, resulting in notably improved classification performance. This approach not only rectified the deficiencies of existing datasets but also demonstrated the efficacy of integrating diverse data sources to enhance model accuracy and robustness.

3. The Proposed Methodology

This section elaborates on how earlier studies implemented FOD material classification techniques and methods for recognition. Furthermore, it explains the proposed methodology concerning its development, implementation, and improvement in accuracy. It also provides insights into the composition of the newly merged dataset and details the process of object detection for FOD using advanced image processing techniques with YOLOv8.

3.1. Implementation of Inception V3

Zainab et al. [9] employed transfer learning, utilizing the Inception V3 pre-trained model, which underwent complete fine-tuning on the FOD dataset. Additionally, they utilized the AlexNet pre-trained on a Chinese dataset. The resulting accuracies were quite low on Chinese dataset but good enough on real FOD dataset having Focus on metal. In comparison to that our aim is to get better accuracy not just for metal but also for other classes as well and we succeed in that. The methodology adopted by Zainab et al. [9] is compared with the approach employed in this study as outlined below:

- In their study, Zainab et al. [9] opted for Inception V3 as their preferred neural network for FOD recognition, particularly focusing on material classification. While having focus on metal as they thought that metal is one that presents the most in FOD. Considering the complexity involved in material recognition, it is acknowledged that deeper neural networks tend to yield better recognition accuracy [18], assuming that factors such as appropriate dataset selection, training methodologies, and techniques to mitigate overfitting are properly addressed. In this paper we have used the Xception Model. In this study, Xception architecture is exclusively employed as the primary deep learning model for material recognition tasks. While previous research emphasized the effectiveness of models like InceptionV3 and ResNet, our investigation focused solely on Xception architecture. Through comprehensive experiments and implementation our study demonstrated the superior performance of the Xception model on both existing and newly developed FOD datasets.
- Zainab et al. [9] utilized pre-trained models, specifically AlexNet and Inception V3. They trained only the fully connected layers of these networks on both the Chinese dataset and the newly developed dataset in this study. Transfer learning strategy was employed in conjunction with training the last few layers, allowing the network to acquire complex features. On the other hand, in this research, we used Transfer learning and used pre trained Xception Model which have been trained on Chinese, newly developed dataset of real FOD by Zainab et al [9] and then at the end we merged both these datasets and again trained Xception model on it and the outcome is quite remarkable.
- The dataset presented by Xu et al. [12] includes training images featuring generic items within the irrelevant material category. In contrast, the dataset compiled by Zainab et al. [9] exclusively comprises FOD items retrieved from actual airfield runways and taxiways, ensuring a balanced representation across various times of the day to account for lighting variations. These items, primarily discovered by flight safety

personnel, predominantly consist of small objects such as nuts, bolts, screws etc. Furthermore, a novel dataset is formed by merging these two datasets to augment diversity and enhance outcomes.

3.2 Proposed Model

The fundamental principle underlying the concept of algorithm development is to leverage pre-trained networks trained on ImageNet, given their exceptional performance in object recognition tasks and also performed very well in the work of Zainab et al. [9] as well as Xu et al. [12] as they also used pre trained networks. Furthermore, these algorithms have been fine-tuned for recognition tasks, thus demonstrating improved performance when applied to datasets where FOD items are centrally positioned in the image with a runway background. The effectiveness of the networks is underscored by the outcomes, indicating enhanced performance with ImageNet, recognized as the largest dataset available for feature learning, thereby facilitating network with better training.

The Xception model, pre-trained on ImageNet, exhibits superior prediction accuracy compared to the results reported by Zainab et al. [9] in their study. Pre-processing techniques are employed to align input images with the size and format requirements of the respective model. In this study, three distinct optimizers, namely stochastic gradient descent, Adam, and RMS Prop, were utilized. It is observed that on the newly developed dataset developed by Zainab et al [9] and on the newly merged dataset, the Adam optimizer performed best for Xception model. The results are to be discussed in detail in later sections. Figure 2 shows the structure of the model.

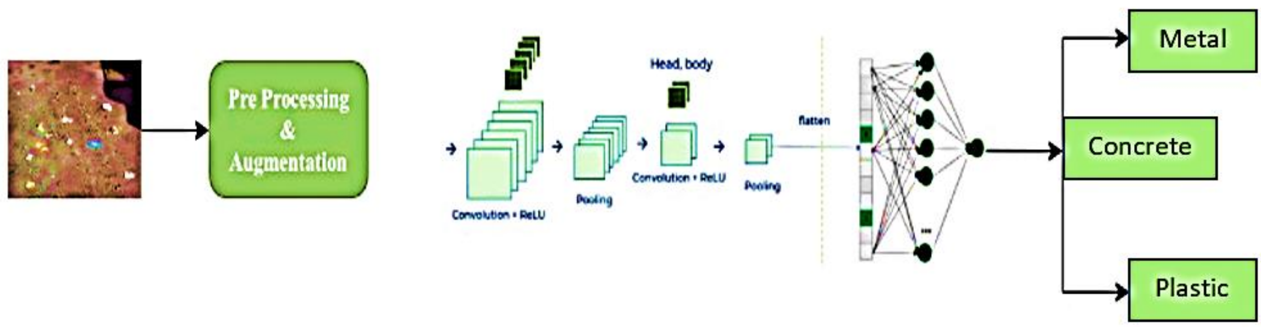

Figure 2. Proposed model for FOD classification.

3.3 Improved Recognition Accuracy

To enhance the overall test accuracy of the algorithm, Zainab et al. [9] primarily concentrates on augmenting the accuracy of metal recognition. Given the dataset's scale, comprising a few thousand training images, encountering the issue of overfitting was anticipated. Our focus is to enhance the accuracy not only on metal class but also on other classes as well, the following strategies were implemented, yielding significant effectiveness.

3.4 Selection of D-CNNs

Considering the prior research conducted by Zainab et al. [9], the Xception model has been chosen as the primary deep convolutional neural network (D-CNN) to enhance recognition accuracy across all classes. Among numerous available D-CNNs, Xception has demonstrated superior performance in terms of accuracy. Leveraging the pre-trained Xception model available within the Keras API applications, it was implemented using Keras. Comparative analysis reveals that the results obtained with Xception surpassed those of earlier network models. Even when combining the two datasets and training the Xception model on it the results are exceptionally good.

3.5 Selection of Pre-trained Weights

Leveraging pre-existing models is essential in image processing due to the challenges associated with collecting large datasets and the computational demands of developing models from scratch. Utilizing pre-trained models helps alleviate these burdens and facilitates faster results with higher accuracy. However, selecting a suitable model tailored to the task at hand is crucial. For instance, employing a model trained for image classification is essential for tasks related to visual data rather than one designed for text processing. ImageNet, renowned for its vast category coverage and optimization for image analysis, holds an advantage over datasets like MINC. While MINC may be compatible, ImageNet's extensive category range enables better generalization of features, particularly in tasks like FOD classification within material categories. Consequently, employing models trained on ImageNet yields superior performance compared to previous studies [12]. This approach ensures that deep feature extraction is optimized, allowing models to adapt effectively to specific datasets in image-based classification tasks.

3.6 Selection of Optimizer

In previous studies by Xu et al. [12], stochastic gradient descent was employed as the optimizer; nevertheless, Adaptive Moment Estimation (Adam) optimizer demonstrated superior performance. Similarly, in the study conducted by Zainab et al. [9], Adam optimizers were utilized. Hence, considering these findings, Adam optimizers were selected for training all models in this paper. The outcomes confirmed that Adam consistently achieved optimal convergence and yielded the highest accuracies across all models.

3.7 Methods to resolve Overfitting

Given the constrained size of the FOD dataset utilized in this study, the risk of limited adaptability remains a significant concern. This issue arises when a system becomes too specialized to the available data, leading to diminished performance when applied to new or varying conditions. This phenomenon can significantly impact the system's ability to function effectively beyond the initial dataset, reducing its overall reliability in real-world applications. The following techniques were used to improve adaptability and maintain accuracy across different scenarios:

1. Reduced Complexity
2. Early Stopping
3. Data Augmentation
4. Dropout

4. **Newly Merged Dataset**

An important milestone in our research was achieved by merging two distinct datasets—the Chinese FOD dataset [12] and the real FOD dataset [9]—to create a comprehensive and enriched dataset for analysis. This newly combined dataset was utilized to enhance classification accuracy, yielding exceptional results that highlight its robust ability to handle diverse data. The fusion of these datasets significantly expanded the diversity and volume of the data, enabling more effective pattern recognition and feature extraction from varied sources. This accomplishment underscores the improved adaptability of the approach, showcasing its capability to deliver high performance across diverse and complex data.

The merged dataset contains a balanced distribution of images across key FOD categories, including metal, concrete, and plastic, as detailed in Table 1. With a total of 5,465 images allocated for training, validation, and testing, this dataset provides a robust platform for advancing FOD classification and detection research. The training set consists of 3,812 images, while the validation and test sets contain 1,123 and 530 images, respectively. This meticulous allocation ensures that the dataset supports comprehensive model evaluation and fine-tuning.

The successful integration of these datasets, coupled with the outstanding results achieved by the Xception model, reinforces the significance of this approach. By leveraging the strengths of both datasets, this research lays a strong foundation for future advancements in FOD detection. The merged dataset not only enhances model generalization capabilities but also serves as a valuable resource for developing and benchmarking innovative machine-learning techniques in aviation safety.

TABLE 1. Newly merged FOD dataset details.

Category	**No. of Images**		
	Training	Validation	Test
Metal	1193	332	188
Concrete	1300	396	171
Plastic	1320	395	171
Total	3812	1123	530

5. Experiments and results

This section outlines the outcomes derived from the predictions of the FOD classification algorithm, which are based on the methodologies delineated in Section 3.

5.1 Xception Algorithm

The Xception algorithm is implemented in three distinct approaches. Initially, the Xception model pre-trained on ImageNet undergoes fine-tuning using the Chinese FOD dataset. Subsequently, the same Xception model pre-trained on ImageNet is fine-tuned utilizing the newly developed FOD dataset introduced by Zainab et al. [9]. After that checking the diversity Xception pre-trained on ImageNet is fine-tuned on newly merged dataset.

5.2 Xception Algorithm on Chinese Dataset

We trained our model on a Chinese dataset and this dataset is also used by SM. Zainab et al [9] in their research as well. Our model's performance on the Chinese dataset surpassed the outcomes reported by SM. Zainab et al [9] and also by Xu et al [12]. This achievement underscores the efficiency of our approach, showcasing superior results in comparison to prior works. Table 2 gives us detailed comparative analysis.

Table 2. Results comparisons on Chinese dataset [12]

MODEL	**Accuracy**
Proposed Xception model	98%
AlexNet by Zainab et al [9]	81%
AlexNet by Xu et al [12]	67%

Figure 3 illustrates the F1-score, a metric that balances precision and recall, providing a comprehensive assessment of classification accuracy across all categories. Moreover, the individual graphs plotted for each class visually depict the patterns in precision, recall, and F1-score, enhancing the understanding of the system’s effectiveness across diverse categories. These findings underscore the dependability and efficiency of the proposed classification approach in accurately identifying metal, concrete, and plastic objects within the Chinese dataset, thereby establishing a strong foundation for its potential application in various real-world scenarios.

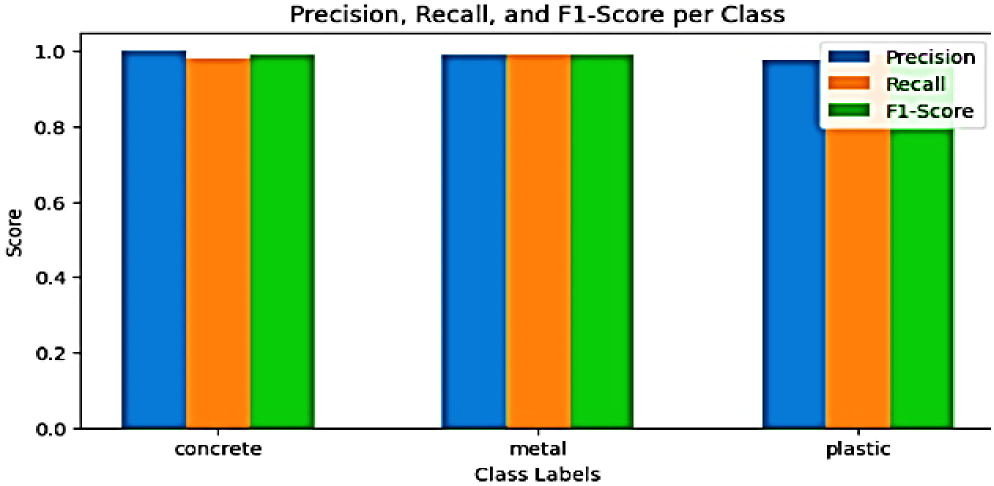

Figure 3. Graph showing, F1score, precision, recall on Chinese dataset [12].

The confusion matrix reflects on the overall performance of the model across different classes in Table 3. In the "concrete" class, out of 110 instances, 108 were accurately classified, while 2 were misclassified as "plastic," resulting in a precision of about 98%. For the "metal" class, 110 instances were correctly identified, with only 1 misclassification as "plastic," achieving a precision of approximately 99%. Similarly, in the "plastic" class, 123 instances were correctly classified, with 1 misclassified as "metal," resulting in a precision of around 99%. The average accuracy across all classes stands at approximately 98%.

TABLE 3.confusion matrix on Chinese dataset [12].

True Label \ Predicted Label	**Concrete**	**Metal**	**Plastic**
Concrete	108	0	2
Metal	0	110	1
Plastic	0	1	123

5.3 Xception Algorithm on real FOD Dataset

We trained our model on a real FOD dataset. This dataset is introduced by SM. Zainab et al [9] in their research as well. Our reported results on the new dataset surpassed the outcomes reported by SM. Zainab et al [9]. The details of the results are listed below in Table 4.

TABLE 4. Comparisons of Results on real FOD dataset [9].

Model/Method	**Accuracy**

Proposed Xception model	96
Inception by Zainab et al [9]	92

Among the classification approaches analyzed, the Xception framework emerged as the most effective, achieving an impressive accuracy of 96%. Its strong results highlight the efficiency of its depth-wise separable layers, which excel at capturing intricate patterns within the dataset. Following closely behind, Inception V3 attained a commendable accuracy of 92%, utilizing its advanced feature extraction mechanisms. Despite its deeper structure designed to handle complex classification tasks, ResNet50 achieved an accuracy of 87%, slightly trailing Xception and Inception V3. This suggests that the dataset may not fully benefit from ResNet50's depth compared to the more optimized feature extraction techniques of Xception and Inception V3. Meanwhile, AlexNet, while innovative at its introduction, demonstrated the lowest accuracy at 82%. Its shallower structure relative to the other approaches likely limited its ability to capture detailed patterns within the dataset, resulting in comparatively lower accuracy.

In summary, the Xception model's exceptional accuracy highlights its effectiveness in capturing the complexities of the dataset, positioning it as the preferred choice among the models examined, followed closely by Inception V3, while ResNet50 and AlexNet offer competitive but comparatively lower performance.

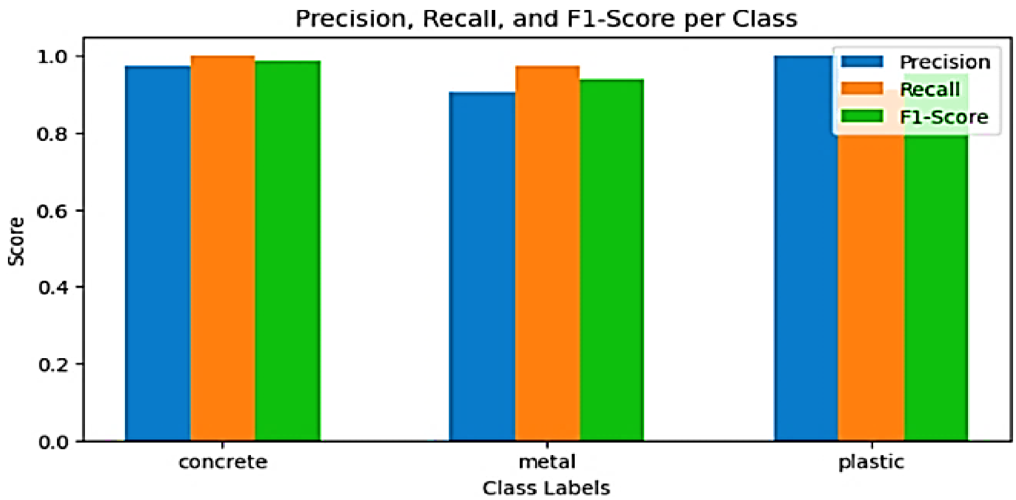

Figure 4. Graph showing, F1score, precision, recall on real FOD dataset [9].

The average precision, recall, and F1-score across all classes reached an impressive 96%. This demonstrates the system's strength in accurately categorizing objects into their appropriate classes. Precision evaluates classification accuracy in identifying relevant instances within a class, while recall assesses its ability to capture all pertinent instances. The F1 score, which combines precision and recall, provides an overall measure of classification effectiveness. Additionally, the graphs plotted for each class visually depict the trends in all three metrics, offering a comprehensive visual overview of classification accuracy across various categories. These results underscore the reliability and efficiency of the proposed classification approach in accurately identifying metal, concrete, and plastic objects, laying a strong foundation for their practical application in real-world scenarios.

TABLE 5. Confusion matrix on real FOD dataset [9].

Predicted Label	Concrete	Metal	Plastic
Concrete	40	0	0
Metal	1	39	0
Plastic	0	4	42

The provided confusion matrix reveals the classification performance of the model across three classes: "concrete," "metal," and "plastic" in Table 5. In the "concrete" class, the model accurately classified 40 out of 40 instances, resulting in a precision of 100%. For the "metal" class, 39 instances were correctly identified, with one misclassification as "plastic," yielding a precision of approximately 97.50%. In the "plastic" class, 42 instances were accurately classified, with four misclassified as "metal," achieving a precision of around 91.30%. The average accuracy across all classes stands at approximately 96.96%.

5.4 Xception Algorithm on Newly Merged Dataset

In the subsequent phases of our research, a notable advancement is achieved by merging two distinct datasets, followed by training the combined dataset using the Xception model. The outcome is highly promising, underscoring the model's robust ability to perform well on generalized data. This accomplishment signifies the model's adaptability to diverse datasets, affirming its capability to effectively learn and extract meaningful features across varied sources. The successful fusion of datasets, coupled with the stellar results attained, further reinforces the versatility and generalization of the Xception model in our research context.

TABLE 6. Confusion Matrix on Newly merged dataset

Predicted Label	Concrete	Metal	Plastic
Concrete	184	3	1
Metal	4	158	9
Plastic	2	9	160

The confusion matrix provides a detailed overview of the models' performance in Table 6. Across all the classes out of a total of 530 instances, 502 were correctly classified, resulting in an overall accuracy of 94.72%. Specifically, the model accurately classified 184 instances of concrete, with only 4 instances misclassified as metal and 2 as plastic. For the metal class, 158 instances were correctly identified, with 3 misclassifications as concrete and 9 as plastic. Similarly, in the plastic class, 160 instances were correctly classified, with 1 misclassification as concrete and 9 as metal. Overall, the model demonstrates a strong ability to differentiate between the three classes, with the majority of instances correctly classified.

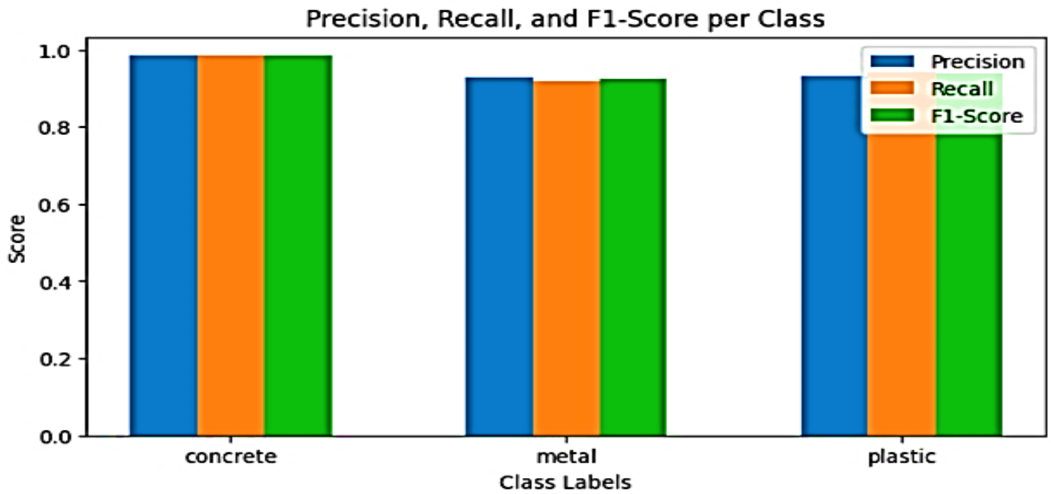

Figure 5. Graph showing, F1score, precision, recall on real Newly Merged dataset.

The average precision, recall, and F1-score across all classes reached an impressive 94%. This highlights the robustness of the model in accurately classifying objects into their respective categories, indicating its effectiveness in handling the intricacies of the merged dataset. The precision metric evaluates the model's accuracy in correctly identifying pertinent instances within individual classes, whereas recall assesses its capability to capture all relevant instances. The F1-score, which harmonizes precision and recall, offers a holistic assessment of the model's overall performance in classifying objects across all categories. Furthermore, the graphs plotted for each class visually depict the patterns in precision, recall, and F1-score, thereby strengthening the depiction of the model's superior performance across diverse categories. These results underscore the reliability and effectiveness of the proposed classification model in accurately identifying metal, concrete, and plastic objects within the merged dataset, laying a solid foundation for its practical application in various real-world scenarios.

TABLE 7. Results on Newly merged dataset

Model	F1Score	Recall	Precision
The proposed Xception model	95	94	94

5.5 *Yolo V8 on real FOD dataset*

YOLOv8, the eighth iteration of the YOLO (You Only Look Once) algorithm presented by Ultralytics, introduces a significant leap in capabilities compared to its predecessors. Unlike its predecessors, YOLOv8 provides versatile support for various computer vision tasks, encompassing object detection, image classification, instance segmentation, and pose estimation. This versatility is further enhanced by the availability of five different model variants, ranging from nano to extra-large, catering to the different computational resources and accuracy requirements. A key advantage of YOLOv8 is its inclusion in the Ultralytics package, which also supports YOLOv3 and v5; however, YOLOv8 stands out as the sole version capable of handling all four computer vision tasks.

The architecture of YOLOv8 comprises a backbone and head, with the backbone utilizing cross-stage partial (CSP) blocks reminiscent of YOLOv5. These CSP blocks offer several advantages, including strengthening the learning capability of lightweight models, removing computational bottlenecks for faster inference, and reducing RAM utilization, enabling deployment on edge devices like Nvidia Jetson. YOLOv8, along with its predecessor YOLOv5, has gained popularity for its accuracy and real-time performance in object detection tasks, largely attributed to the efficiency of the CSP backbone. Comparisons between YOLOv8 and YOLOv5 repositories likely highlight differences in features, performance, and supported tasks, further underlining the advancements brought forth by YOLOv8 in the field of computer vision.

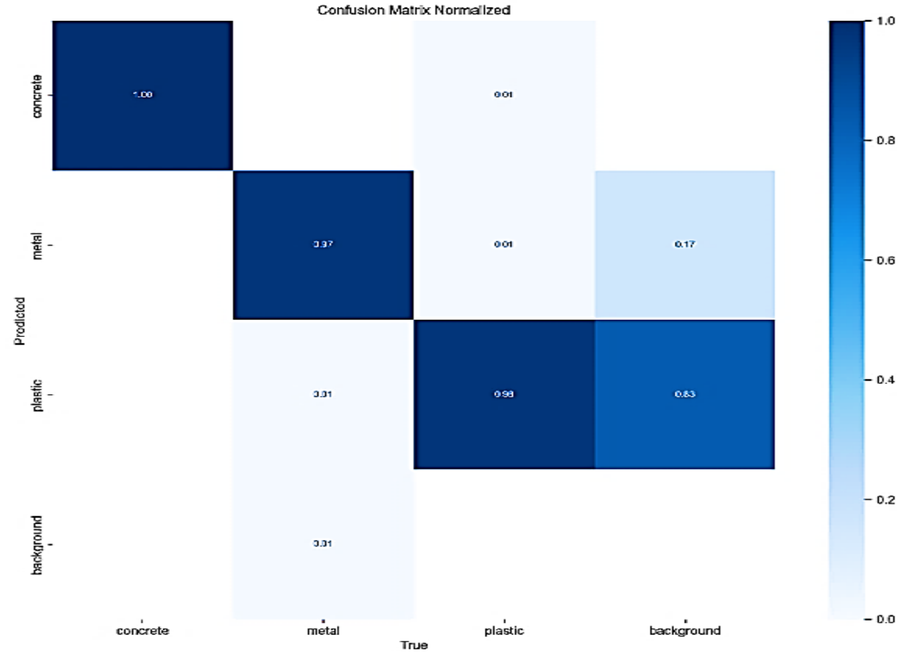

Figure 6. Graph showing, Confusion Matrix on real FOD dataset by Zainab et al [9]

Figure 6. Shows YOLO v8 results on my dataset featuring three classes' metal, concrete, and plastic. We obtained precision-recall curves that provide valuable insights into the model's performance. The curves showcase precision on the y-axis and recall on the x-axis, with the plotted line starting from a precision of 1.0 and gradually declining to approximately 0.98, covering a vast area under the curve.

This decline in precision with increasing recall indicates the model's ability to effectively balance precision and recall across various thresholds. Impressively, the achieved accuracies for metal, concrete, and plastic classes stand at 0.99, 0.98, and 0.98, respectively, underscoring the model's high accuracy in classifying objects within each category. Additionally, "mAP at 0.5" is included in the graph, denoting the mean average precision (mAP) at an intersection over union (IoU) threshold of 0.5. This metric serves as a comprehensive measure of the model's performance across all classes, further validating its effectiveness in accurately detecting and localizing objects. Overall, the precision-recall curves and associated accuracy metrics offer valuable insights into the robustness and reliability of the YOLO v8 model in accurately identifying metal, concrete, and plastic objects in the dataset. Figure 7, 8 shows the results and precision recall curve on.

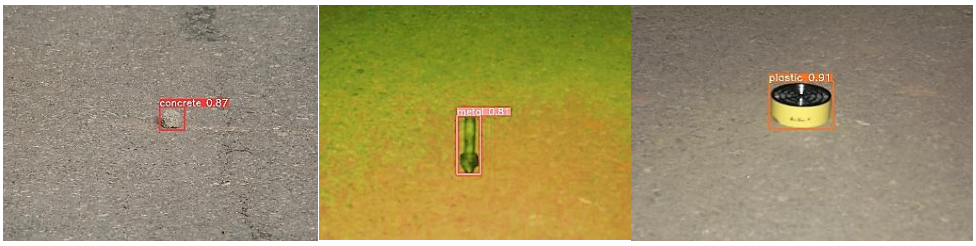

Figure 7. Shows results of Yolo v8 on real FOD dataset by Zainab et al [9]

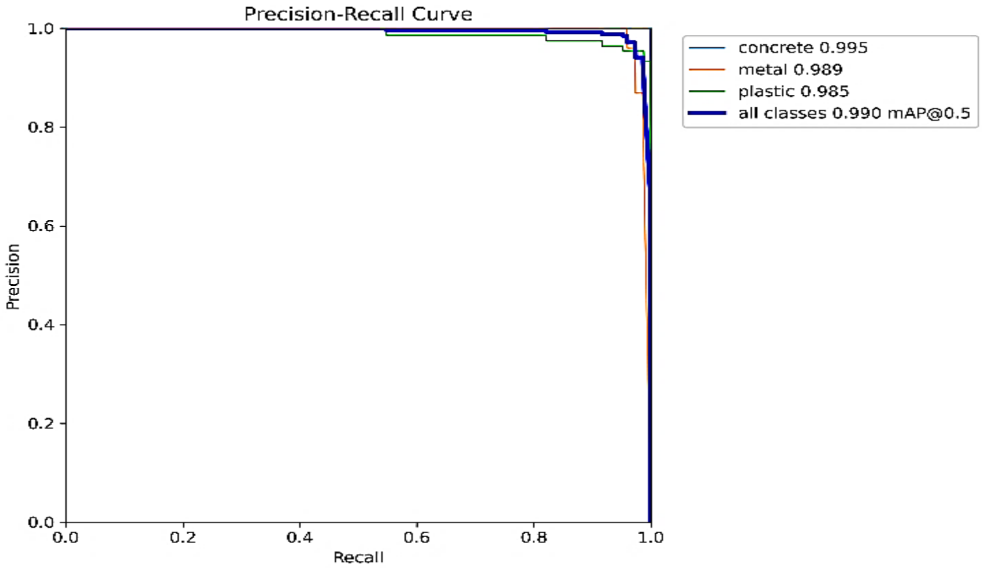

Figure 8 . Graph showing Precision Recall on real FOD dataset [9].

6 Conclusion

This research highlights the remarkable potential of using advanced image analysis techniques for Foreign Object Debris (FOD) detection and classification, paving the way for safer and more efficient aviation operations. By utilizing pre-trained classification methods and an enriched, merged dataset, we demonstrated that achieving consistently high accuracy across diverse FOD classes is possible. The merged dataset, with its larger and more varied collection of images, proved to be a game-changer, helping us achieve results that outperform previous benchmarks. Among all the classification techniques tested, the Xception framework stood out, delivering the highest accuracy across all categories. Additionally, YOLOv8 showed exceptional performance in detecting even small objects on real FOD data, addressing one of the critical challenges in debris detection.

Looking ahead, there are exciting opportunities to build on this work. One promising direction is to apply enhanced object detection techniques to the merged dataset, which, with its larger pool of instances, could further improve classification, accuracy and reliability. Another important step is creating a new dataset featuring multiple objects in a single image, mimicking real-world scenarios where airport runways often have multiple types of debris present simultaneously. This would allow detection systems to tackle more complex classification challenges and bring us closer to practical, on-the-ground implementations.

By addressing these future directions, we can unlock even greater potential in FOD detection systems, making aviation safer and more reliable. This research is a step forward in leveraging the power of automated detection technologies to address real-world challenges, and the possibilities for further innovation are limitless. Moreover, we will develop some more combined deep learning models for multicriteria decision making for fuzzy logic and stress analysis by using [21-27].

Conflict of Interest:

The authors declare that they have no known competing financial

CRediT authorship contribution statement

Yaseen Mushtaq: Validation, Software, Investigation, Data curation. Wajid Ali: Visualization, Validation, Supervision, Investigation, Data curation, Conceptualization. Amal Kumar Adak: Writing – original draft, Methodology, Formal analysis, Conceptualization, Usman: Writing – original draft, Methodology. Rahim Ullah: Writing – original draft, Methodology, Formal analysis

Funding information

This research was not supported by any specific grant from public, commercial, or non-profit funding organizations.

Ethics statement

This study complies with ethical standards, involving no human or animal subjects. All data were sourced ethically or generated by the authors.

References

1. Bhosale, Yogesh H., and K. Sridhar Patnaik. "Application of deep learning techniques in diagnosis of covid-19 (coronavirus): a systematic review." Neural processing letters 55, no. 3 (2023): 3551-3603.https://doi.org/10.1007/s11063-022-11023-0 .
2. Bhosale, Yogesh H., and K. Sridhar Patnaik. "IoT deployable lightweight deep learning application for COVID-19 detection with lung diseases using RaspberryPi." In 2022 International conference on IoT and blockchain technology (ICIBT), pp. 1-6. IEEE, 2022. 10.1109/ICIBT52874.2022.9807725.
3. Bhosale, Yogesh H., Shrinivas Zanwar, Zakee Ahmed, Mahendra Nakrani, Devendra Bhuyar, and Ulhas Shinde. "Deep convolutional neural network based Covid-19 classification from radiology X-Ray images for IoT enabled devices." In 2022 8th international conference on advanced computing and communication systems (ICACCS), vol. 1, pp. 1398-1402. IEEE, 2022. 10.1109/ICACCS54159.2022.9785113.
4. Jehle, Markus, Christoph Sommer, and Bernd Jähne. "Learning of optimal illumination for material classification." In Joint Pattern Recognition Symposium, pp. 563-572.Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. https://doi.org/10.1007/978-3-642-15986-2_57.
5. Hayman, Eric, Barbara Caputo, Mario Fritz, and Jan-Olof Eklundh. "On the significance of real-world conditions for material classification." In Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV 8, pp. 253-266. Springer Berlin Heidelberg, 2004. https://doi.org/10.1007/978-3-540-24673-2_21.
6. Liu, Ce, Lavanya Sharan, Edward H. Adelson, and Ruth Rosenholtz. "Exploring features in a bayesian framework for material recognition." In 2010 ieee computer society conference on computer vision and pattern recognition, pp. 239-246. IEEE, 2010. 10.1109/CVPR.2010.5540207.
7. Koenderink, J. J., A. J. Van Doorn, K. J. Dana, and S. Nayar. "Bidirectional reflection distribution function of thoroughly pitted surfaces." *International Journal of Computer Vision* 31, no. 2 (1999): 129–144. https://doi.org/10.1023/A:1008061730969.

8. P. C. Ng and S. Henikoff, sift: Predicting amino acid changes that affect protein function,’ Nucl. Acids Res., vol. 31, no. 13, pp. 3812–3814, 2003.10.1093/nar/gkg509.
9. Zainab, Syeda Mahrukh, Khurram Khan, Adnan Fazil, and Muhammad Zakwan. "Foreign Object Debris (FOD) Classification Through Material Recognition Using Deep Convolutional Neural Network with Focus on Metal." IEEE Access 11 (2023): 10925-10934. 10.1109/ACCESS.2023.3239424.
10. Ghahremani, Morteza, Yitian Zhao, Bernard Tiddeman, and Yonghuai Liu. "Interwoven texture-based description of interest points in images." Pattern Recognition 113 (2021): 107821. https://doi.org/10.1016/j.patcog.2021.107821.
11. Qin, Fei, Xiangxi Bu, Yunlong Liu, Xingdong Liang, and Jihao Xin. "Foreign object debris automatic target detection for millimeter-wave surveillance radar." Sensors 21, no. 11 (2021): 3853. https://doi.org/10.3390/s21113853.
12. Xu, Haoyu, Zhenqi Han, Songlin Feng, Han Zhou, and Yuchun Fang. "Foreign object debris material recognition based on convolutional neural networks." EURASIP Journal on Image and Video Processing (2018).1-10. https://doi.org/10.1186/s13640-018-0261-2.
13. Yuan, Zhong-Da, Jia-Qing Lii, Zhii-Nan Qiuu, and Yong Zhang. "Research on FOD detection system of airport runways based on artificial intelligence." In Journal of Physics: Conference Series, vol. 1635, no. 1, p. 012065. IOP Publishing, 2020. 10.1088/1742-6596/1635/1/012065.
14. Gao, Qiang, Ruifeng Hong, Yutong Chen, and Jiaxing Lei. "Research on the detection algorithm of foreign object debris and small targets in airport runway based on ssd." In The 2nd International Conference on Computing and Data Science, pp. 1-4. 2021. 10.1145/3448734.3450862.
15. Wang, Chien-Yao, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. "CSPNet: A new backbone that can enhance learning capability of CNN." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp. 390-391. 2020. https://doi.org/10.48550/arXiv.1911.11929.
16. Ravi, Radhika, Darcy Bullock, and Ayman Habib. "Pavement distress and debris detection using a mobile mapping system with 2d profiler lidar." Transportation research record 2675, no. 9 2021: 428-438. https://doi.org/10.1177/03611981211002529.
17. Mozaffari, J., Amirkhani, A. & Shokouhi, S.B. A survey on deep learning models for detection of COVID-19. *Neural Comput & Applic* **35**, 16945–16973 (2023). https://doi.org/10.1007/s00521-023-08683-x.
18. Benedykciuk, E., Denkowski, M. & Dmitruk, K. Material classification in X-ray images based on multi-scale CNN. *SIViP* **15**, 1285–1293 (2021). https://doi.org/10.1007/s11760-021-01859-9.
19. Özdem, S., Orak, İ.M. A novel method based on deep learning algorithms for material deformation rate detection. *J Intell Manuf* (2024). https://doi.org/10.1007/s10845-024-02409-z.
20. Ullah, Rahim, Muhammad Saeed, Wajid Ali, Junaid Nazar, and Fakhra Nazar. "A Cooperative Heterogeneous Multi-Agent System Leveraging Deep Reinforcement Learning." *Knowledge and Decision Systems with Applications* 1 (2025): 112-124.
21. Ali, Wajid, Tanzeela Shaheen, Iftikhar Ul Haq, and Mohammad Mehedi Hassan. "A Novel Generalization of Hesitant Fuzzy Model with Application in Sustainable Supply Chain Optimization." *Transactions on Fuzzy Sets and Systems* 7, no. 1 (2025): 159.

22. Ali, Wajid, Tanzeela Shaheen, Iftikhar Ul Haq, Hamza Toor, Faraz Akram, Harish Garg, Md Zia Uddin, and Mohammad Mehedi Hassan. "Aczel-Alsina-based aggregation operators for intuitionistic hesitant fuzzy set environment and their application to multiple attribute decision-making process." (2023).
23. Ali, Wajid, Tanzeela Shaheen, Hamza Ghazanfar Toor, Tmader Alballa, Alhanouf Alburaikan, and Hamiden Abd El-Wahed Khalifa. "An improved intuitionistic fuzzy decision-theoretic rough set model and its application." *Axioms* 12, no. 11 (2023): 1003.
24. Ul Haq, Iftikhar, Tanzeela Shaheen, Wajid Ali, and Tapan Senapati. "A Novel SIR Approach to Closeness Coefficient-Based MAGDM Problems Using Pythagorean Fuzzy Aczel–Alsina Aggregation Operators for Investment Policy." *Discrete Dynamics in Nature and Society* 2022, no. 1 (2022): 5172679.
25. Qayyum, Afshan, Noreen Mushtaq, Tanzeela Shaheen, Wajid Ali, and Florentin Smarandache. "Some Similarity Measures on Generalized Interval-Valued Intuitionistic Fuzzy Soft Expert Sets and Their Applications in Medical Diagnosis." *Neutrosophic Systems with Applications* 23 (2024): 47-67.
26. Badi, Ibrahim, Mouhamed Bayane Bouraima, and Babatounde Ifred Paterne Zonon. "A Multi-Criteria Decision-Making Model for Prioritizing the Causes of Inflation in Libya Using the Best-Worst Method and Analytic Hierarchy Process." *Knowledge and Decision Systems with Applications* 1 (2025): 165-176.
27. KAYIRAN, Hüseyin Fırat. "Machine Learning-Based Investigation of Stress in Carbon Fiber Rotating Cylinders." *Knowledge and Decision Systems with Applications* 1 (2025): 1-10.