Implementing Multi-Expert Fusion Model for Open Set Recognition

Brinda Ramesh Bhanderi
Electrical and Computer Engineering
University at Albany
Albany,USA
bhanderibrinda@gmail.com

***Abstract*—”Open Set Recognition (OSR) poses a significant challenge in machine learning, necessitating models that can robustly classify known categories while effectively identifying and rejecting unknowns. Traditional methods often struggle with the adaptability and accuracy required for dynamic environments where new, unseen categories frequently emerge. This project introduces the Multi-Expert Fusion Model (MEDAF), a novel approach employing multiple expert networks and a gating mechanism to enhance classification performance. MEDAF incorporates ResNet architectures with attention modules to focus on diverse features, significantly reducing feature redundancy through an innovative attention diversity loss. Implemented using the PyTorch framework and evaluated on CIFAR-10 and CIFAR-100 datasets, MEDAF demonstrates substantial improvements in accuracy, AUROC, and F1 scores compared to conventional models. The results highlight MEDAF’s capability to not only recognize known objects with high precision but also to effectively detect and categorize unknown objects, thereby increasing the reliability and applicability of OSR systems in real-world scenarios.”**

***Index Terms*—Open Set Recognition (OSR), Multi-Expert Diverse Attention Fusion (MEDAF), Adaptive Fusion, Machine Learning, Attention Modules, Diversity Loss, PyTorch.**

I. Introduction

In today’s digital era, systems are increasingly exposed to data from diverse and dynamic sources, necessitating advanced recognition algorithms capable of not only categorizing known data but also effectively handling unknown or novel inputs. This challenge is particularly pronounced in Open Set Recognition (OSR), where the objective extends beyond traditional closed set environments that only require classification among previously seen categories. OSR is critical for applications where encountering unseen data is expected, such as in autonomous systems, security surveillance, and content moderation platforms. These applications must not only identify and classify data into known categories but also detect and appropriately respond to new, unseen data types.

Among various approaches to OSR, the Multi-Expert Diverse Attention Fusion (MEDAF) model emerges as a compelling solution by effectively minimizing the open space risk—the likelihood of erroneously classifying unknown inputs as known categories. Originally developed for complex image recognition tasks, MEDAF leverages a system of multiple expert networks, each trained to focus on distinct aspects of the data through sophisticated attention mechanisms. This multi-expert architecture allows for a more granular and comprehensive analysis of data, enhancing the ability to distinguish between known and unknown inputs.

Applying the MEDAF model to general OSR tasks involves overcoming significant challenges, such as ensuring that the model can generalize well across different types of data beyond images, including textual, audio, and complex sensor data. The transition from specialized domains to a more generalized framework entails adapting the model to accommodate diverse data characteristics and integrating domain-specific knowledge to maintain high accuracy and reliability.

This paper explores the adaptation of the MEDAF model for broad OSR applications, detailing the necessary modifications to harness its potential across varied data environments. We discuss how the unique capabilities of MEDAF, such as its diverse expert integration and dynamic attention fusion, can be leveraged to enhance the robustness and adaptability of OSR systems. By expanding the scope of MEDAF beyond its initial image-centric applications, this work aims to establish a new standard in OSR, providing a robust framework for systems to accurately recognize known categories and effectively manage encounters with unknown data, thereby addressing one of the most pressing challenges in modern data processing environments.

A. Project Motivation

Traditional models in machine learning, including most deep learning approaches, are typically designed for closed set recognition and often struggle with the open set scenario. These models tend to either misclassify new, unseen inputs as one of the known classes or require extensive retraining with new data, which is not always feasible in real-time applications. For example, in security systems, failing to recognize an unauthorized entrant or misidentifying them could lead to breaches. Similarly, in autonomous driving, encountering an unrecognized obstacle or scenario could result in inadequate responses, posing safety risks.

The limitations of these traditional models stem largely from their design to maximize accuracy on a fixed set of known classes, often leading to overfitting on these classes and poor generalization to new data. Furthermore, generative models, which have been employed to enhance OSR capabilities by modeling the distribution of known classes and detecting anomalies, often suffer from high computational demands and instability in dynamic settings

B. Objective

The primary goal of this project is to significantly enhance the accuracy and adaptability of open set recognition systems through the implementation of the Multi-Expert Diverse Attention Fusion (MEDAF) model. MEDAF aims to leverage the diverse capabilities of multiple expert networks, each trained to specialize in different aspects of the data. By integrating these experts through a dynamic gating mechanism that assesses and weighs their outputs based on the presented data, the model seeks to achieve high accuracy in recognizing known classes and robust performance in identifying unknown classes. The objective is to create a scalable, efficient model that can adapt to new data without the need for constant retraining, thereby supporting real-time applications in varied and unpredictable environments.

II. Literature Review

The exploration of Open Set Recognition (OSR) and data fusion in wireless communication has garnered significant attention in recent research, as evidenced by the contributions of various authors in the field. This review delves into the advancements and methodologies proposed in recent literature, highlighting their implications for the development of more adaptable and efficient AI and machine learning systems.

Wang et al. (2024) embark on an innovative journey to enhance OSR through "Exploring Diverse Representations for Open Set Recognition," leveraging varied data representations to improve the identification and classification of unknown classes. Their work underscores the crucial need for adaptable and robust AI systems capable of navigating the uncertainties of real-world applications [1] [4].

In a comprehensive survey by Geng et al. (2020), "Recent advances in open set recognition: A survey," the authors dissect the methodologies, challenges, and applications associated with OSR. Published in the esteemed IEEE Transactions on Pattern Analysis and Machine Intelligence, their analysis illuminates the path forward for OSR, advocating for methodologies that bolster the reliability and flexibility of AI models [5].

Complementing this exploration, Mahdavi and Carvalho (2021) present "A survey on open set recognition" at the IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering. Their review canvasses the landscape of OSR, emphasizing its pivotal role in enhancing AI systems' performance across varied domains [6].

Simultaneously, the realm of multimodal data fusion benefits from the insights of Gao et al. (2020), whose survey "A Survey on Deep Learning for Multimodal Data Fusion" in Neural Computation evaluates the integration of deep learning in processing and combining data from multiple sources. Their work heralds a promising direction for future research in multimodal data fusion [3].

Lastly, the innovative approach to data fusion in wireless sensor networks by Izadi et al. (2015) introduces a novel fuzzy-based method aimed at improving network efficiency and longevity. Published in Sensors, their methodology emphasizes the significance of distinguishing true data values and minimizing redundancy [2].

Together, these studies offer a panoramic view of the current state and prospects of OSR and data fusion, setting a vibrant stage for future investigations aimed at pushing the boundaries of AI and machine learning applications in the face of real-world challenges.

III. Methodology

The Multi-Expert Diverse Attention Fusion (MEDAF) model exemplifies a sophisticated approach to tackling the nuanced demands of open set recognition (OSR). In the realm of OSR, the primary challenge lies not only in accurately identifying instances from known categories but also in detecting and correctly labeling inputs from unseen, unknown categories. To adeptly manage this dual challenge, MEDAF integrates a multi-branch network architecture composed of several ResNet models, each serving as an independent expert. This design capitalizes on the strengths of the ResNet architecture, renowned for its deep learning efficiency and particularly adept at handling vanishing gradients—a common hurdle in deep neural networks—thanks to its innovative shortcut connections and batch normalization.

In MEDAF, each 'expert' is a ResNet model configured to independently analyze a unique segment of the input data. This setup allows the system to harness a diverse and rich array of feature sets from the same input, enabling a comprehensive understanding. The term "expert" underscores the model's capability to specialize in different facets of the data, with each expert developing unique insights that collectively enhance the system's diagnostic accuracy.

The process within each expert begins with initial convolutional layers and batch normalization that serve to standardize and stabilize the incoming data, facilitating a more reliable and uniform analysis across the network's depth. Following these initial layers, ResNet's Basic Blocks delve deeper into the data, extracting intricate features while pooling layers work to reduce dimensionality, refining the data into a manageable form for classification. This intricate processing ensures that each expert not only understands but also accentuates specific features of the data relevant to its specialized task.

Beyond the individual processing by each expert, MEDAF employs attention mechanisms that refine the feature maps generated by the experts. These modules strategically focus the model's computational power on the most informative elements of the data, enhancing the model's precision in pinpointing crucial features crucial for distinguishing between known and unknown inputs. The attention-refined feature maps are then concatenated to create a holistic set of features that represents a fusion of insights from all experts.

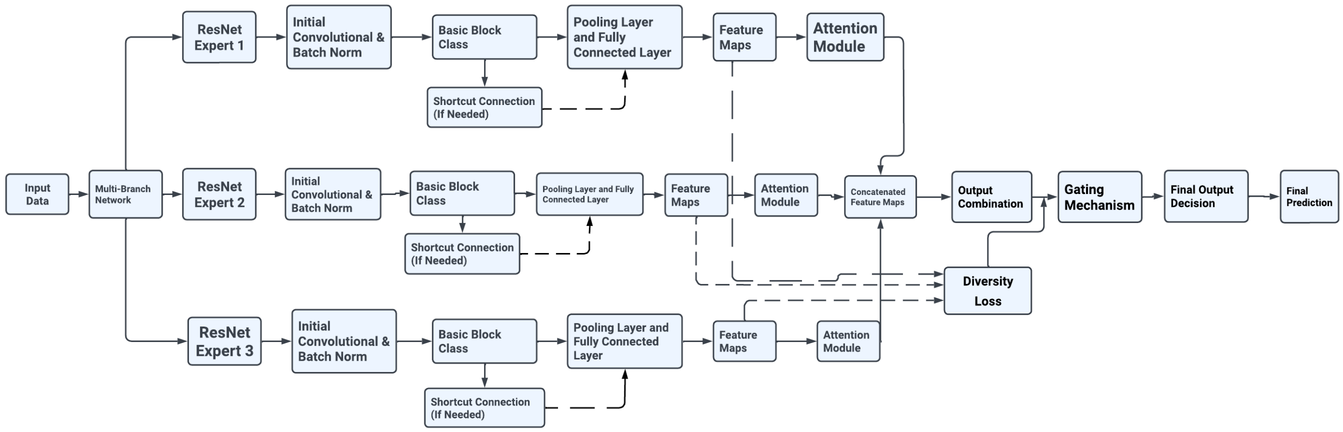

Fig. 1. Architecture of MEDAF

The synthesized expert outputs are adeptly woven together through a dynamic gating mechanism. This mechanism assesses and strategically weighs the contributions of each expert according to their relevance to the current input, ensuring the decision-making process is both adaptive and contextually aware. This capability is critical in environments where data variability is high, and the system must be adept at recognizing and adjusting to new patterns. The final output from the gating mechanism determines the classification of the input—whether it fits a known category or should be designated as unknown.

To preserve the uniqueness and integrity of the insights from each expert, the MEDAF model incorporates a diversity loss function. This function is designed to minimize redundancy across the attention maps of different experts, encouraging each to explore and emphasize distinct aspects of the data. Such diversity is not merely beneficial but essential for robust recognition of known classes and effective detection of novel categories, significantly boosting the model's ability to generalize across varied scenarios.

In essence, the MEDAF model's methodology is crafted to enhance the classification accuracy and adaptability required for effective open set recognition. By leveraging a team of expert models, each fortified with targeted attention mechanisms and collectively integrated through an intelligent gating system, MEDAF addresses the core challenges of OSR with notable adeptness. This strategic ensemble of experts ensures that the system not only excels at recognizing familiar entities but is also well-equipped to handle the unexpected, marking a significant leap forward in machine learning applications.

A. Technical Details and Mathematical Formulations

1) Prediction and Uncertainty Estimation: The MEDAF model assesses the probability distribution over classes to predict whether a sample belongs to any known class. If the probability of the unknown class U is below a threshold τ, the sample is classified as known. The probability that the model assigns the sample to any known class can be formulated as:

$$p(\hat{y}_u = U|z_u) = \max_{k \in K} p(\hat{y}_u = k|z_u), \tag{1}$$

where $\hat{y}_u$ is the predicted label determined by:

$$\hat{y}_u = \begin{cases} k & \text{if } p(\hat{y}_u = U|z) \geq \tau, \\ U & \text{if } p(\hat{y}_u = U|z) < \tau. \end{cases} \tag{2}$$

[1]

2) Diversity Regularization and Cross-Entropy Loss: The diversity loss L_d is designed to minimize the cosine similarity between feature maps of different experts, encouraging each expert to capture unique aspects of the input data. This is crucial for preventing the model from focusing redundantly on similar features across different experts, thus enhancing its capability to generalize well to new, unseen data. For experts i and j with feature representations M_y^i and M_y^j, the diversity loss is defined as:

$$L_d = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{M_y^i \cdot M_y^j}{||M_y^i||_2 \cdot ||M_y^j||_2}, \tag{3}$$

[1] where $||\cdot||_2$ denotes the magnitude, promoting distinctiveness in the learned features by different experts.

The cross-entropy loss L_{ce} quantifies the classification error, using the softmax output for each class k:

$$L_{ce} = -\sum_k y_k \cdot \log(\text{Softmax}(l)^k), \tag{4}$$

[1] where y_k represents the true label in one-hot encoded form, and $\text{Softmax}(l)^k$ is the softmax probability of class k.

The overall loss function L combines the global cross-entropy loss L_{ce}^g, individual cross-entropy losses L_{ce}^i for each expert, and the diversity loss L_d, adjusted by scaling factors β_1 and β_2:

$$L = L_{ce}^g + \sum_{i=1}^{N} \beta_1 \cdot L_{ce}^i + \beta_2 \cdot L_d. \tag{5}$$

[1]

IV. Implementation of the MEDAF Model

Data Preparation and Augmentation Techniques: The foundation of any effective machine learning model is robust data preparation. For the MEDAF model, this involves utilizing the CIFAR-10 dataset for training and validation purposes, while the CIFAR-100 dataset is employed to simulate unknown class scenarios during testing. Both datasets comprise small 32x32 pixel images categorized into various classes, with CIFAR-100 treated uniformly as an 'unknown' class to challenge the model's open set recognition capabilities.

To enhance the model's generalization ability and mitigate overfitting, a series of data augmentation techniques are applied. These include random horizontal flipping, which helps the model learn orientation invariance, and random rotation by 10 degrees to introduce variability in the image presentation. Color jittering adjusts the brightness, contrast, saturation, and hue to make the model robust against color variations in images. Additionally, random cropping simulates different image positioning and partial occlusions. All images are normalized using predefined mean and standard deviation values to ensure consistent scaling, which is crucial for maintaining stable training dynamics. These preprocessing steps are efficiently implemented using PyTorch's transforms module.

Configuration of Neural Networks: At the core of the MEDAF model is a sophisticated architecture based on multiple ResNet networks, each acting as an independent expert. These networks are carefully designed with convolutional layers, batch normalization, ReLU activations, pooling, and fully connected layers, with shortcut connections integral to each. These residual links are particularly crucial as they help alleviate the vanishing gradient problem, enabling the effective training of deeper network architectures. The configuration of each expert network—including the number of layers, size of the fully connected layers, and the overall architecture—is tailored based on empirical evaluations and computational efficiency to optimize performance.

Software Tools, Libraries, and Hardware:

The implementation leverages PyTorch for its dynamic computational graph capabilities, which are essential for building complex models that involve conditional operations. The torchvision library is utilized for accessing pre-processed datasets and implementing predefined data transformations. For performance evaluation, additional tools from scikit-learn are used to compute metrics such as ROC AUC and F1 scores, essential for assessing the model's ability to discriminate between known and unknown classes effectively. Visualization and numerical operations are handled using matplotlib and numpy, respectively.

Given the computational demands of training the MEDAF model, especially due to its multi-expert system processing large volumes of data, CUDA-enabled GPUs are employed. These GPUs significantly accelerate the training process by facilitating faster matrix operations and parallel computations, making them indispensable for this project.

Together, these steps ensure the MEDAF model is not only theoretically robust but also practically effective, capable of achieving high accuracy and adaptability in open set recognition tasks across diverse and dynamic input scenarios.

V. Evaluation and Results

Datasets Used: The evaluation of the MEDAF model relies on two distinct datasets to rigorously assess its capabilities in both closed and open set recognition contexts: CIFAR-10: This dataset serves as the "closed set" for training and validation. It includes 60,000 images across 10 classes (50,000 for training and 10,000 for validation), featuring objects such as cars, animals, and ships. The well-defined, limited scope of CIFAR-10 makes it ideal for training the model to recognize and classify these known categories with high accuracy. CIFAR-100: Used as the "open set" for testing, CIFAR-100 contains 100 classes, significantly more than CIFAR-10, and none of the CIFAR-10 classes overlap with CIFAR-100 classes. This dataset tests the model's ability to correctly identify when an input does not belong to any of the categories it was trained on, effectively simulating the model's response to new, unseen categories by treating all CIFAR-100 classes as unknowns. Evaluation Metrics:

To comprehensively evaluate the performance of the MEDAF model, the following metrics are utilized: Accuracy: Measures the proportion of predictions that the model got right, including both true positives and true negatives. High accuracy on the CIFAR-10 test set indicates strong recognition capabilities for known classes, while accuracy on the CIFAR-100 test set measures how well the model avoids false positives from unknown classes. AUROC (Area Under the Receiver Operating Characteristic Curve): This metric is particularly useful in evaluating how well the model can distinguish between classes that it was trained on (known) and new, unseen classes (unknown). A higher AUROC value indicates that the model effectively discriminates between known and unknown classes, making it a critical measure for open set recognition. F1 Score: Harmonizes precision and recall in a single metric by taking their harmonic mean. This is crucial in scenarios where class imbalance might exist (e.g., fewer unknown samples in CIFAR-100 compared to known samples in CIFAR-10), ensuring that the model does not favor the majority class. Presentation of Results:

The results are presented through a series of graphs and visual examples to illustrate the model's performance dynamics over training epochs and its ability to handle both known and unknown classes:

Graphical Visualizations: The training process is documented through graphs showing the decline in loss over epochs and the improvement in validation metrics, indicating the model's learning efficiency. Separate charts for each metric within the MEDAF model highlight specific contributions and performance variations among the different network components(Fig.2).

Performance Metrics: Additional graph details the Accuracy, AUROC, and F1 Score for each expert on test data from CIFAR-100, providing insights into how each part of the model responds to unknown data (Fig.3).

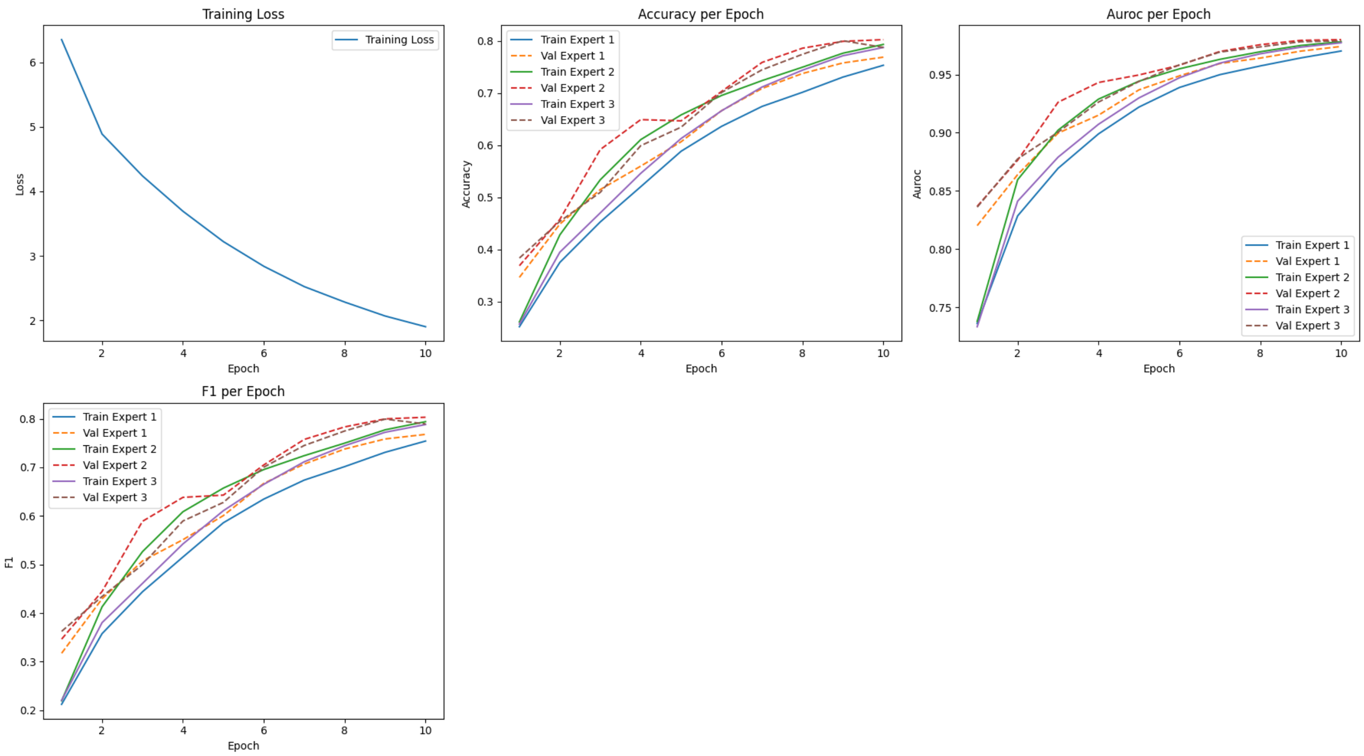

Fig. 2. Expert-wise Metrics for while Traning for OSR on Closed SET

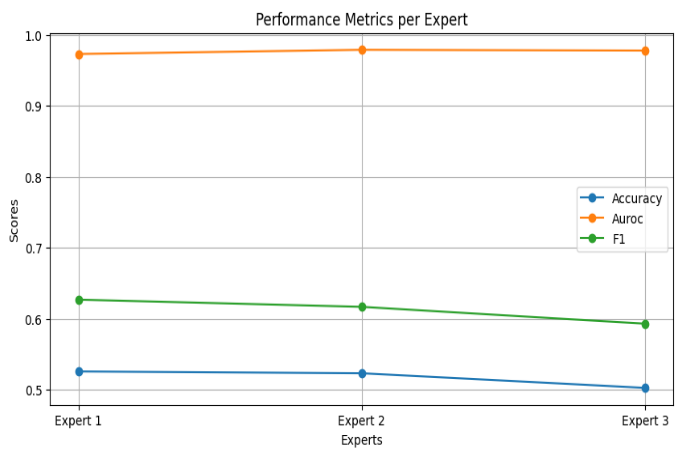

Fig. 3. Expert-Wise metrics of OSR

Example Predictions: Visual representations of the model's predictions showcase its practical effectiveness(Fig.4). These include images from both datasets where the model correctly identifies known classes and accurately flags images from CIFAR-100 as unknown. This direct visualization helps in assessing the qualitative aspects of model performance.

VI. Future Work

Exploration of Deeper and More Complex Network Architectures: Future research should delve into the exploration of deeper neural network architectures to further enhance the MEDAF model's capacity to capture intricate patterns and representations within both known and unknown classes. Investigating more complex architectures, such as those incorporating attention mechanisms or graph neural networks, could potentially unlock deeper insights into data representations, leading to improved performance in open-set recognition tasks.

Investigation of Alternative Fusion Mechanisms: Alternative fusion mechanisms for combining expert advice should be considered to refine decision-making processes and enhance the model's adaptability to diverse data landscapes. Exploring ensemble methods, Bayesian frameworks, or attention-based fusion techniques could offer new avenues for improving the MEDAF model's ability to effectively integrate diverse expert opinions and make accurate predictions in complex open-set recognition scenarios.

Adaptation of the Model for Real-Time Processing and Edge Deployment: Adapting the MEDAF model for real-time processing and deployment on edge devices represents a crucial area for future research. Optimizing the model's computational efficiency and memory footprint to ensure seamless integration into edge computing ecosystems is essential. This involves exploring algorithmic optimizations, model compression techniques, and hardware acceleration strategies to enable efficient deployment of the MEDAF model on resource-constrained edge devices. Additionally, addressing challenges related to latency, energy consumption, and scalability is paramount to realizing the potential of the MEDAF model in real-time applications across diverse domains, including IoT, autonomous

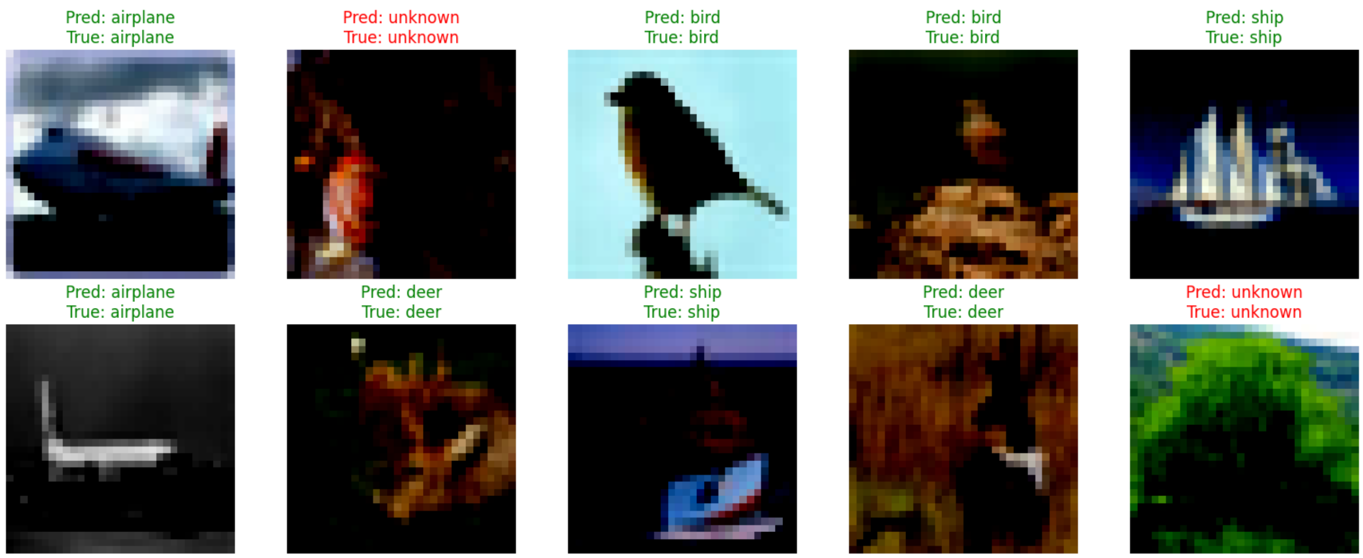

Fig. 4. Visualisation

systems, and security.

VII. Conclusion

The MEDAF model demonstrates significant advancements in open set recognition, effectively distinguishing between known and unknown classes. With robust performance metrics including an accuracy of 0.5173, AUROC of 0.9765, and F1 Score of 0.6123, the model showcases its reliability across various network components. Future research should explore deeper network architectures and alternative fusion mechanisms to further enhance its capabilities. Additionally, optimizing the model for real-time processing and deployment on edge devices holds promise for practical applications in IoT, autonomous systems, and security. Overall, MEDAF represents a significant leap forward in addressing the challenges of open set recognition, offering accuracy, adaptability, and scalability for real-world machine learning applications.

References

[1] Wang, Y., Mu, J., Zhu, P. and Hu, Q., 2024. Exploring Diverse Representations for Open Set Recognition. arXiv preprint arXiv:2401.06521

[2] Izadi D, Abawajy JH, Ghanavati S, Herawan T. A Data Fusion Method in Wireless Sensor Networks. Sensors. 2015; 15(2):2964-2979.

[3] Jing Gao, Peng Li, Zhikui Chen, Jianing Zhang; A Survey on Deep Learning for Multimodal Data Fusion. Neural Comput 2020; 32 (5): 829–864.

[4] Wang, Yu, Junxian Mu, Pengfei Zhu, and Qinghua Hu. "Exploring Diverse Representations for Open Set Recognition." arXiv preprint arXiv:2401.06521 (2024).

[5] Geng, Chuanxing, Sheng-jun Huang, and Songcan Chen. "Recent advances in open set recognition: A survey." IEEE transactions on pattern analysis and machine intelligence 43.10 (2020): 3614-3631.

[6] Mahdavi, Atefeh, and Marco Carvalho. "A survey on open set recognition." 2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). IEEE, 2021.

[7] Bendale, A.; and Boult, T. E. 2016. Towards Open Set Deep Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1563–1572.

[8] Scheirer, W. J.; de Rezende Rocha, A.; Sapkota, A.; and Boult, T. E. 2013. Toward Open Set Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(7): 1757–1772.

[9] Guo, Y.; Camporese, G.; Yang, W.; Sperduti, A.; and Ballan, L. 2021. Conditional Variational Capsule Network for Open Set Recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 103–111.

[10] He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 770–778.

[11] Huang, H.; Wang, Y.; Hu, Q.; and Cheng, M.-M. 2023. Class-Specific Semantic Reconstruction for Open Set Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 4214–4228

[12] Liu, C.; Yang, C.; Qin, H.-B.; Zhu, X.; Liu, C.-L.; and Yin, X.-C. 2023. Towards open-set text recognition via label-toprototype learning. Pattern Recognition, 134: 109109.