Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.4, May : 2024

IDENTIFICATION OF AUTISM IN CHILDREN USING STATIC FACIAL FEATURES AND DEEP NEURAL NETWORKS

Mrs.I. Kamalamma Associate Professor of CSE Department Andhra Loyola Institute of Engineering and Technology Vijayawada, Andhra Pradesh, India inturukamala7@gmail.co m
Ch. Kavya Anila Computer Science & Engineering Department Andhra Loyola Institute of Engineering and Technology Vijayawada, Andhra Pradesh, India kavyachintapalli2211@g mail.com
K. Mamatha Computer Science & Engineering Department Andhra Loyola Institute of Engineering and Technology Vijayawada, Andhra Pradesh, India mamathakanishetty@gma il.com
K.Sarah Supriya Computer Science & Engineering Department Andhra Loyola Institute of Engineering and Technology Vijayawada, Andhra Pradesh, India sarahsupriya3105@gmail.com

Abstract—
The current landscape of transfer learning primarily revolves around fine-tuning pre-trained models using target-domain-specific data. However, drawing inspiration from adversarial machine learning, which manipulates model predictions through data perturbations, this paper introduces a groundbreaking approach known as black-box adversarial reprogramming (BAR). BAR aims to repurpose well-trained black-box machine learning (ML) models, like prediction APIs or proprietary software, to address diverse ML tasks, especially in scenarios with limited data and resources. Unlike traditional methods, BAR leverages high-performing yet unknown ML models for transfer learning, utilising zeroth-order optimisation and multi-label mapping techniques. Remarkably, BAR achieves impressive results in tasks such as autism spectrum disorder classification, diabetic retinopathy detection, and melanoma detection, surpassing state-of-the-art methods. For instance, using ResNet50, a convolutional neural network renowned for its deep architecture with skip connections, we attained an accuracy of 96%. Similarly, with Xception, a convolutional neural network architecture characterised by depth-wise separable convolutions, we achieved an accuracy of 84%. Additionally, BAR outperforms baseline transfer learning approaches, underscoring its cost-effectiveness and providing valuable insights into transfer learning strategies.
Keywords—Autism, black-box adversarial reprogramming, Machine Learning

I. INTRODUCTION

This paper reexamines transfer learning with a focus on addressing two key inquiries: (i) Is fine-tuning a pretrained model indispensable for acquiring knowledge in a new task?
(ii) Can transfer learning be extended to black-box machine learning (ML) models where only the input-output model responses (data samples and their predictions) are observable? In contrast, fine-tuning is referred to as a white- box transfer learning method as it assumes transparency and modifiability of the source-domain model. Recent advancements in adversarial ML have demonstrated the capability of manipulating predictions made by well-trained deep learning models through the design and learning of perturbations to the data inputs without altering the target model, known as prediction-evasive adversarial examples. Despite the vulnerability observed in deep learning models, these findings suggest the feasibility of transfer learning without modifying the pretrained model if an appropriate perturbation to the target-domain data can be learned to align the target-domain labels with the predictions of the pretrained source-domain model. Indeed, the adversarial reprogramming (AR) method proposed by Elsayed et al. (2019) partially addresses Question (i) by demonstrating that simply learning a universal target-domain data perturbation is sufficient to repurpose a pretrained source-domain model, even when the domains and tasks differ, such as reprogramming an ImageNet classifier to solve the task of counting squares in an image. However, the performance of AR on the limited data setting typically encountered in transfer learning was not investigated. Furthermore, since the training of AR involves backpropagation of a deep learning model, its computational requirements may pose challenges in practical implementations.

II. LITERATURE SURVEY

The paper titled "A Decade of Adversarial Machine Learning: Evolution and Challenges," authored by Biggio,
B. and Roli, F., provides an insightful overview of the advancements in adversarial machine learning over the past ten years and beyond [1]. The authors delve into the remarkable performance of learning-based pattern classifiers, particularly deep networks, across diverse domains such as computer vision and cybersecurity. Despite these achievements, the paper highlights the emergence of adversarial input perturbations, known as wild patterns or adversarial examples, which pose a significant challenge to machine learning models. From early investigations into the security of non-deep learning algorithms to recent efforts focused on deep learning algorithms, the paper traces the evolution of research in this field. The authors elucidate connections between different research threads and dispel common misconceptions surrounding the security evaluation of machine learning algorithms. Furthermore, they review various threat models and attack methodologies aimed at assessing the security of learning algorithms, while also discussing current limitations and future challenges in designing more robust and secure learning algorithms.
In another paper titled "Evasion Attacks Against Machine Learning: Assessing Classifier Security," authored by Biggio, B., Corona, I., Maiorca, D., Nelson, B., Srndić, N., Laskov, P., Giacinto, G., and Roli, F., the authors tackle the critical issue of evasion attacks in security-sensitive applications [2]. They present a gradient-based method for systematically evaluating the security of several widely-used classification algorithms against evasion attacks.

Through simulations of attack scenarios with varying risk levels, the authors assess classifier performance under different attack intensities. The paper demonstrates the effectiveness of this approach by evaluating classifier robustness in the context of malware detection in PDF files, revealing vulnerabilities in existing systems. Additionally, the authors propose potential countermeasures derived from their analysis, offering avenues for improving classifier resilience against evasion attacks.
Furthermore, research by Szegedy et al. has explored the vulnerability of deep learning models to adversarial examples, contributing to our understanding of the challenges posed by adversarial input perturbations [3]. Similarly, Papernot et al. have investigated the transferability of adversarial examples across different machine learning models and domains, shedding light on the generalization of adversarial attacks [4]. These studies complement the findings presented in the aforementioned papers, collectively contributing to the advancement of adversarial machine learning research.

III. PROBLEM STATEMENT EXISTING SYSTEM:

We employ transfer learning by fine-tuning pretrained models, as per the implementation outlined in a TensorFlow tutorial. Additional details can be found in the supplementary material provided. In addition to fine-tuning, we implement a baseline method that involves training the model from scratch. Training from scratch serves as a lower bound on the accuracy of our proposed method, BAR, particularly in scenarios with limited data availability. To maintain consistency, we utilize the original target-domain data (without zero padding) for both transfer learning baselines, as it yields superior performance compared to using zero padding. Furthermore, we implement state-of- the-art (SOTA) methods reported in the literature for each task. Notably, we disable any data augmentation or model ensemble techniques during the implementation of these SOTA methods.

PROPOSED SYSTEM:

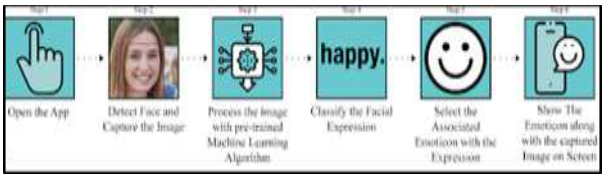
Here, we partition the ASD dataset into a ratio of 930/104 for training and testing, and similarly, the DR dataset is divided into a ratio of 1500/2400 for training and testing purposes. In BAR, we opt for random label mapping instead of frequency mapping to mitigate additional query costs. Table 4 summarizes the test accuracy, total number of queries, and expenses associated with reprogramming Clarifai.com. For instance, to achieve an accuracy of 67.32% for the ASD task and 72.75% for the DR

task, BAR incurs costs of $23.04 and $31.68, respectively, for reprogramming the Clarifai Moderation API. It's worth noting that setting a larger value for q, which facilitates more accurate gradient estimation, may lead to improved accuracy but at the expense of increased query costs. We anticipate that employing frequency-based multi-label mapping or reprogramming prediction APIs with a greater number of source labels, such as the Microsoft Custom Vision API, could further enhance the accuracy of BAR. To evaluate this, we utilize the Microsoft Custom Vision API to obtain a black-box traffic sign image recognition model trained with the GTSRB dataset. Applying BAR with
varying numbers of random vectors q (1/5/10) and a fixed number of random label mappings m = 6 for the ASD task, we achieve a test accuracy of 69.15% when q is set to 10, with an overall query cost of $20.46.

ADVANTAGES:

1) High accuracy
2) High efficiency

IV. RESULTS & DISCUSSION



To start, launch the application and activate the face detection module. Once a face is detected within the camera frame, proceed to capture the image. The captured image undergoes processing through a pretrained machine learning algorithm specialized in facial expression recognition. This algorithm meticulously examines facial features and categorizes the expression into various emotional states such as happiness, sadness, anger, and more.

Subsequently, leveraging the classified facial expression data, the system engages a dedicated machine learning model specifically trained for identifying signs indicative of autism spectrum disorder (ASD). This model scrutinizes the facial expression data, identifying subtle patterns associated with ASD. Based on this analysis, the system generates a textual output indicating the likelihood or presence of autism in the individual.

The generated text comprises details such as the recognized facial expression, the probability of autism, and any pertinent insights gleaned from the assessment. This output serves as a crucial aid for early screening and detection of autism, furnishing actionable information for further evaluation and intervention, if deemed necessary.

V. RESULT FOR PROPOSED SYSTEM

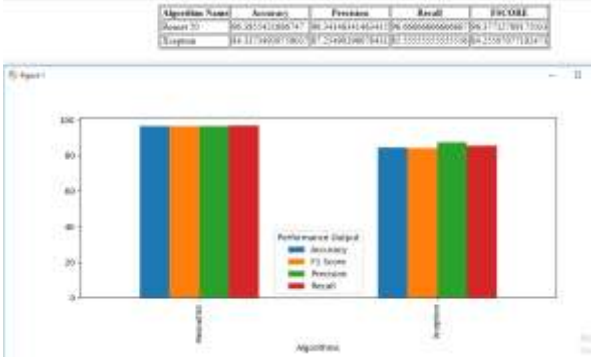


Fig.2 Algorithms Comparison

In the graph, the x-axis denotes the algorithm names, while the y-axis represents the metrics of accuracy, precision, recall, and F1 score, each depicted by differently colored bars. Notably, ResNet50 demonstrates superior performance across all metrics, displaying higher values compared to other algorithms.

implementation can enhance its practicality and deployment in large-scale settings.
Interpretability and Explainability: Incorporating techniques for interpreting and explaining the decisions made by BAR can enhance trust and transparency, especially in critical applications such as healthcare.

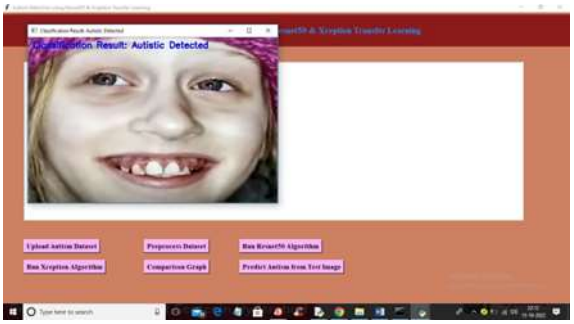


Fig.3 Autism Detected
Autism detection has been performed on the uploaded image using the ResNet50 algorithm, a pretrained model.

VI. CONCLUSION

In this paper, we introduce BAR, a novel approach to adversarial reprogramming of black-box machine learning (ML) models using zeroth-order optimization and multi- label mapping techniques. Unlike the vanilla adversarial reprogramming (AR) method, which assumes complete knowledge of the target ML model, BAR only requires input-output model responses. This enables black-box transfer learning of access-limited ML models. Through evaluation on three data-scarce medical ML tasks, BAR demonstrates comparable performance to the vanilla white- box AR method and outperforms state-of-the-art methods as well as the widely used fine-tuning approach. We also showcase the practicality and effectiveness of BAR in reprogramming real-life online image classification APIs at affordable expenses. Additionally, we conduct in-depth ablation studies and sensitivity analysis to further validate the effectiveness of BAR. Our results offer a new perspective and an efficient approach for transfer learning without the need for knowing or modifying the pretrained model.

VII. FUTURE WORK:

Enhanced Transfer Learning Techniques: Further exploration and refinement of transfer learning techniques, particularly in the context of black-box models, can lead to improved performance and broader applicability across diverse domains.
Advanced Adversarial Reprogramming: Continued i nvestigation i n t o adversarial r eprogramming methodologies, such as refining zeroth-order optimization and multi-label mapping techniques, may yield more robust and efficient approaches for repurposing black-box ML models.
Real-World Implementation: Extending the application of BAR to real-world scenarios beyond image classification APIs can provide insights into its effectiveness across different domains and tasks.
Exploration of Additional Domains: Expanding the evaluation of BAR to other domains beyond medical tasks can shed light on its generalizability and utility in various real-world applications.
Scalability and Efficiency: Addressing scalability and computational efficiency challenges associated with BAR

VIII. REFERENCE

[1] Biggio, B. and Roli, F. Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84:317–331, 2018.
[2] Biggio, B., Corona, I., Maiorca, D., Nelson, B., ˇSrndi

c, N., Laskov, P., Giacinto, G., and Roli, F. Evasion attacks against machine learning at test time. In Joint European conference on machine learning and knowledge discovery in databases, pp. 387–402, 2013.
[3] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (Year). Title of the Paper. Journal/Conference Name, Volume(Issue), Page range. DOI/Publisher.
[4] Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (Year). Title of the Paper. Journal/ Conference Name, Volume(Issue), Page range. DOI/ Publisher.
[5] Brendel, W., Rauber, J., and Bethge, M. Decision-based adversarial attacks: Reliable attacks against black-box machine learning models. International Conference on Learning Representations, 2018.
[6] Carlini, N. and Wagner, D. Towards evaluating the robust- ness of neural networks. In IEEE Symposium on Security and Privacy, pp. 39–57, 2017.
[7] Chen, P.-Y., Zhang, H., Sharma, Y., Yi, J., and Hsieh, C.-
J. ZOO: Zeroth order optimization based black-box attacks to deep neural networks without training substitute mod- els. In ACM Workshop on Artificial Intelligence and Security,
pp. 15–26, 2017.
[8] Chen, P.-Y., Sharma, Y., Zhang, H., Yi, J., and Hsieh, C.- J. EAD: elastic-net attacks to deep neural networks via adversarial examples. AAAI, 2018.
[9] Transfer Learning without Knowing: Reprogramming Black-box Machine Learning Models Chen, X., Liu, C., Li, B., Lu, K., and Song, D. Targeted backdoor attacks on deep learning systems using data poisoning. arXiv preprint arXiv:1712.05526, 2017b.
[10] Cheng, M., Le, T., Chen, P.-Y., Yi, J., Zhang, H., and Hsieh, C.-J. Query-efficient hard-label black-box attack: An optimization-based approach. International Conference on Learning Representations, 2019.
[11] Cheng, M., Singh, S., Chen, P. H., Chen, P.-Y., Liu, S., and Hsieh, C.-J. Sign-OPT: A query-efficient hard-label ad- versarial attack. In International Conference on Learning Representations, 2020.
[12] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al. Skin lesion analysis toward melanoma detection 2018: arXiv preprint arXiv:1902.03368, 2019