Indonesian Journal of Advanced Research (IJAR)
Vol. 4, No. 1, 2025: 111-130

Evaluating the Performance of Vision Transformers and Convolutional Neural Networks for Hostile Image Detection

Zakir Hossain[1], Md Emran Hossain[2], Nisher Ahmed[3*], Md Farhad Kabir[4], Iffat Sania Hossain[5]
[1]College of Engineering and Computer Science, California State University
[2,3]College of Technology and Engineering, Westcliff University
[4]Marshall School of Business, University of Southern California
[5]Martin V. Smith School of Business and Economics, California State University
Corresponding Author: Nisher Ahmed, n.ahmed.511@westcliff.edu

A R T I C L E I N F O

Keywords: Malicious Images, ViTs (Vision Transformers), CNNs (Convolutional Neural Networks), Adversarial Perturbations, Image Classification

Received : 3, January
Revised : 17, January
Accepted: 31, January

A B S T R A C T

Detecting malicious or adversarial images, for example in security and surveillance systems, is an important problem in computer vision. These results highlight the effectiveness of ViTs when compared to CNNs when confronting hostile images. However, CNNs have stiff competition from ViTs and have been the go-to architecture for image classification and object detection for many years, due to the existence of spatial hierarchies in images. Using benchmark datasets containing a combination of adversarial and clean images, this study compares the ability of both models to (i) detect hostile images, (ii) generalize to unseen dataset, and (iii) the overall computational efficiency of both models. While ViTs can be even more computationally expensive than incurred with task3 input, we demonstrate that, in fact, our architecture generalizes truncation -- both in power and action -- exceptionally well and can simply outperform performance-per-dollar in more robust pattern recognition tasks, especially under adversarial perturbations. In contrast, CNNs are faster to inference and less likely to overfit on small data. This finding informed decisions showing trade-offs between the two architectures, including a potential path for hybrid approaches and future enhancements in the adversarial defense against hostile image detection.

DOI: https://doi.org/10.55927/ijar.v4i1.13681
ISSN-E: 2986-0768
https://journal.formosapublisher.org/index.php/ijar

INTRODUCTION

The recent explosion of adversarial attacks and adversarial generation in the image domain has drawn people concern about the security of computer vision. Most image modifications are made for good reasons, hostile images, generally, are derived by adversarial perturbations or by malicious intents that can trick machine learning and cause wrong prediction that can have catastrophic outcomes if applied to sensitive applications like security, surveillance, self-driving cars, medical imaging. Detecting and defending against such adversarial and hostile images reliably is essential for protecting automated decision-making systems.

Given their capacity to represent spatial hierarchies and capture local features, Convolutional Neural Networks (CNNs) had remained as the de-facto tool for image classification and object detection tasks for more than a decade. Their effectiveness for tasks from face recognition to object detection is well established. Though CNNs have shown relative resistance in identifying images containing structural anomalies or outliers, they still remain susceptible to adversarial attacks, wherein subtle, undetectable changes in the image may drive considerable misclassification. As efficient networks in terms of speed and resource consumption, and thus suitable for real-time applications, CNNs are widely used regardless of these vulnerabilities.

Alternatively, Vision Transformers (ViTs), a newer architecture, have become popular for being able to mimic the self-attention method, enabling them to grasp long-range dependencies in image data. In contrast to CNNs that apply convolutions to extract hierarchical features, ViTs view images as sequences of non-overlapping fixed-size patches and rely on transformer-based architectures which have achieved SOTA in natural language processing. In the case of image classification, ViTs provide potential advantages that could help in extracting subtly complex patterns on hostile or adversarial images due to their attention-based mechanisms of obtaining global contextual relationships. On the contrary, ViTs are generally more computation-heavy and demand larger datasets for training, which might not be favorable for some use cases.

This study aims to compare performance of CNNs and ViTs on the task of detecting hostile images. The study seeks to cover the following questions addressing some of the advantages and disadvantages of both architectures relative to adversarial image detection:

1. Accuracy: How effective or accurately do CNNs and ViTs identify malicious images vs benign ones?
2. Robustness: Under which model is more resistant against adversarial attacks or noise in the image data?
3. Computational Efficiency — What are each model's computational costs (training time, inference time, resource utilization)?
4. Model Robustness: How robust are each of the models to generating adversarial examples?

In order to meet these goals, we perform a number of experiments on the benchmark datasets containing adversarially distorted and naturally existing hostile images. The implications of this study's findings offer a timely

analysis of what CNNs and ViTs do well and where they struggle and will be useful towards the formulation and construction of more secure and robust image classification systems that can help detect hostile images in realistic scenarios.

Beyond performance comparison, the paper also investigates possible ways to enhance adversarial robustness, including hybrid architectures that leverage the best of CNNs and ViTs, and possible defenses against adversarial perturbations. With this thorough_analysis, we hope to aid in further improving the resilience of machine learning systems, especially in security-critical domains where hostile image detection is one of the critical challenges.

LITERATURE REVIEW

As the authors note, the detection and classification of enemy or adversarial images is an emerging field in computer vision and deep learning. From the perspective of their performance and their attacks Vulnerability, Machine learning models in particular CNN and ViT have been well studied. In this literature review, we provide an in-depth overview of the current research landscape in the field of hostile image detection alongside a summary of the prominent work done within CNNs, ViTs and adversarial machine learning. The review is divided into three main parts: adversarial image detection, CNN based approaches, vision Transformer based approaches.

Adversarial Image Detection

Adversarial attacks were first introduced by Goodfellow et al. (2014), an example of these perturbations is to add them to an image in such a way that it is imperceptible to the human eye, but can change the predictions of the machine learning models significantly. Adversarial perturbations exploit the fact that deep neural network models are sensitive to slight modifications in input data. The difficulty in recognizing hostility has driven a flurry or research efforts towards both adversarial defense and the creation of resilient models"

This Paper: We cover: some classes of adversarial attacks (i.e., FGSM, PGD, and Carlini-Wagner) with a specialization of their efficiency, seek, etc. These attacks are designed for CNNs, which are prominently found to be vulnerable to subtle perturbations in input data (Szegedy et al., 2014). The vulnerability of CNNs against adversarial examples stimulated researchers in proposing strong models that recognize or oppose such malicious images.

Adversarial Defense Strategies: Numerous defense strategies have been proposed over the years to potentially mitigate adversarial attacks. The earliest defenses relied on adversarial training, where models were trained on adversarial examples to ensure robustness (Goodfellow et al., 2015). But this adversarial training is computationally intensive and can also decrease performance on clean images. Further, other defense mechanisms are also developed, including defensive distillation (Papernot et al., 2016), input pre-processing (Xie et al., 2017) and also detection based methods to detect and reject adversarial inputs (Metzen et al., 2017).

But the arms race between attackers and defenders has showed it's impossible to protect against every possible attack in this way. It has been an evolutionary game in which models are repeatedly compromised by newer attack techniques to try making systems more robust.

Hostile Image Detection using Convolutional Neural Networks (CNNs)

A popular choice for image classification tasks, convolutional neural networks have achieved state-of-the-art image recognition by learning hierarchical features directly from the data, surpassing previous approaches that relied on hand-crafted features (Krizhevsky et al., 2012). CNNs employ convolutional layers to extract features from local areas and pooling layers to down-sample the spatial volume, hence they work very well on conventional image classification tasks. However, these are easily fooled by adversarial attacks.

Research by Szegedy et al. showed that CNNs can be easily misled with small adversarial perturbations, leading to adversarial intrusion attacks which are induced by only modifying pixel values by an imperceptibly small amount. Further work has demonstrated that CNNs are especially receptive to such natural perturbations in pixel-level input data, even when the alterations are minor in intensity (Papernot et al., 2016). It is widely believed that the local receptive field structure of CNNs, which is considered a benefit for detecting common structures, can be exploited by authors as they lack the ability to generalize well to examples adversarial.

Improving CNN Robustness to Adversarial Attacks Research Focus: To make CNNs less vulnerable to adversarial examples, numerous approaches have been suggested. One common method is adversarial training, in which the model is explicitly trained on both clean and adversarially perturbed images (Goodfellow et al., 2015). This approach enhances resilience to adversarial examples, but requires additional computational time and may lead to degradation of accuracy on unperturbed clean images. Other approaches for deriving generalizable CNNs include regularization strategies (Lecun et al., 1998), feature squeezing (Xu et al., 2017), and the use of model ensembling (Tramèr et al., 2018).

However, this approach still left CNNs vulnerable to adversarial perturbations, posing a fundamental challenge, and inspiring the exploration of architectures that may carry a stronger capacity to resist such threats to image processing.

Hostile Image Detection Using Vision Transformers (ViTs):

Proposed by Dosovitskiy et al. (2015) and made popular by Vaswani et al. (2017) in the field of natural language processing, recently allows for success on computer vision tasks. Vision Transformers (ViTs) utilize the self-attention mechanism that has been a success in NLP tasks for image data. While CNNs apply their operations locally, ViTs work with images in the form of non-overlapping sequences of patches and learn global relations among them using the attention mechanism.

ViTs' superior performance over CNNs: Multiple image classification benchmarks have shown that ViTs outperform the performance of classical CNN architectures (Dosovitskiy et al., 2015). The fundamental benefit with this new approach is their ability to capture long range dependencies at various spatial scales within an image, which are critical for object recognition in clutters images etc. and also in identifying adversarial examples, as global contextual relationships can useful to differentiate between benign and adversarial examples. Unlike the local features in CNNs, ViTs do better when it comes to global structures and relationships present in the entire image, which could make them more robust against adversarial examples (Touvron et al., 2021)

ViTs in Adversarial Robustness: While the application of ViTs in adversarial image detection has been recent, initial studies have shown good potentials. Moreover, due to self-attention which allows ViTs to capture more global information, contextual relationships are better learned for ViTs than CNNs, allowing ViTs to be resistant to adversarial perturbations. Further, ViTs have been applied to produce better performance on tasks like image segmentation and classification (Dosovitskiy et al., 2015), which are similar to hostile image detection. Nonetheless, ViTs need to be trained on large-scale datasets, and require considerable computational resources, which may not suit resource-constrained environments.

Hybrid Architectures: Combining the best of both worlds by retaining the structures of both ViT and CNN for a robust model. As a result, hybrid architectures that utilize CNN-based feature extraction in combination with the global attention aspect provided by ViTs are seeing interest for adversarial detection of images tasks (Bello et al., 2021). By combining the advantages of both CNNs and ViTs for local feature extraction and global context awareness, respectively, these hybrid structures may enhance performance and hold potential for robustness against common visual transformations.

Conclusion and Future Work: Comparative Studies

Several studies have previously benchmarked CNNs and ViTs in adversarial settings. For the most part, these studies conclude that though CNNs are highly reactive to real-time problems, they're at risk in front of adversarial attacks, and their precision drops dramatically when revealed to nasty pictures. CNNs are around 7x faster than ViTs, however ViTs seem to perform better against adversaries, especially in harder adversary problems that involve more global vision at the image context.

These approaches will pave the way for the future of the hostile image detection, which will be characterized by hybrid models, improved adversarial training techniques, and innovative architectures combining the best aspects of CNNs and ViTs. The adoption of defense techniques against adversarial attacks will also contribute to the development of more secure and trustworthy image classification systems, in addition to research on explainability and interpretability.

In fact there is a constant struggle between adversarial attackers and model defenders as outlined by the literature, requiring new models and new ways to detect them. However, CNNs have been found to be susceptible to adversarial attacks, leading to the exploration of alternative architectures like Vision Transformers. Understanding how to leverage both the supervised and weakly-supervised strengths is the future of hostile image detection.

METHODOLOGY

In this paper, we assess the performance of two leading neural architectures for detecting hostile images: Visual Transformers (ViTs) and Convolutional Neural Networks (CNNs). The goal of this comparison is to assess both architectures in terms of accuracy, robustness against adversarial attacks, and computational efficiency. We adopt a complete experimental scenario, including data preprocessing, model training, evaluation metrics and adversarial attack simulation. The method comprises dataset selection, model architecture, adversarial attack generation, two parts of training and evaluation, and performance analysis.

Dataset Selection

Experiments are performed with two benchmark datasets containing standards and adversarially distorted images. These datasets are selected to mimic practical situations in which hostile images are likely to be present, particularly in fields like security and surveillance. The datasets used are:

• CIFAR-10 (Krizhevsky et al. 2009): A commonly used image dataset, containing a total of 60,000 32x32 color images across 10 classes. CIFAR-10 will be our working example for image classification tasks.

• ImageNet (Russakovsky et al., 2015): A large-scale images dataset with more than 14 million labeled images spanning 1,000 classes. This dataset serves as a better benchmark for high resolution than imagenet, since it contains a richer image distribution.

Both datasets are split into training, validation, and test sets, where the evaluation metrics are not affected by data leakage.

Model Architectures

We analyze two deep learning models which are used extensively on the image classification tasks:

• Кохерентна нейронна мережева (CNN):

o Our CNN model architecture follows a standard design, wherein multiple convolutional, activation, pooling, and fully connected layers are stacked together. This architecture is similar to well-known models such as VGG16 (Simonyan & Zisserman, 2014) and ResNet (He et al., 2016) but simplified to allow for greater training and evaluation speed.

o Hyperparameters such as learning rate, batch size and optimization strategy will be tuned through grid search on the validation set in the implementation of CNN using frameworks like TensorFlow/Keras.

• Vision Transformer (ViT):

o The details of the Vision Transformer model are obtained using the architecture introduced by Dosovitskiy et al. (2015), in which an image is partitioned into non-overlapping patches and each patch is processed into a token input in transformer. The ViT model uses the same self-attention mechanism to extract global relations throughout the image.
o ViTs are trained from scratch with the same set of hyperparameters and optimization as in the CNN model. Further parameters specific to ViT will also be adjusted in accordance to get the maximum performance from them including number of transformer blocks, patch size and embedding dimensions.
We use the same training and validation splits to be able to directly compare both models under the same evaluation conditions.

Generating Adversarial Examples

To evaluate the robustness of both models, we generate adversarial examples by the following attacking strategies:
• Fast Gradient Sign Method (FGSM): FGSM is a white-box attack method whose goal is to create adversarial perturbations based on the cost function gradient relative to the input image. A parameter ε characterizes how much perturbation is added to the unperturbed image. This attack is utilized to assess the resistance of models against fundamental adversarial perturbations.
• Projected Gradient Descent (PGD): PGD is an iterative version of FGSM and is one of the most powerful attack methods. The attack perturbatively applies small perturbations that are then projected back onto the ε-ball. This technique is employed to produce more complex adversarial samples to rigorously assess the robustness of the models.
• Carlini-Wagner (CW) Attack: The Carlini-Wagner attack (Carlini & Wagner, 2017) is an advanced attack that creates adversarial examples with increased success in deceiving models. It minimizes a loss function to create the smallest perturbations that will fool the model. This attack is used evaluate the robustness of the models against more powerful, less-detectable adversarial examples.
For all the attack methods, we will produce a batch of adversarial images for both datasets. Models are then evaluated on these adversarially perturbed images as per the perturbations applied to the test set images.
Training and Evaluating the Model:
• Training Process:
o Each architecture's models are fit on the training set with the Adam optimizer (Kingma & Ba, 2015) and a learning rate decay. To enhance the models' generalization ability, data augmentation techniques (i.e. random rotations, random flipping, random cropping) are performed on the training images.
o The models are trained until a maximum number of epochs is reached or they converge, whichever comes first. To add, the validation set is utilized for adjusting hyperparameters: learning rate, batch size, weight decay, etc.

• Evaluation Metrics:
o Accuracy: We measure the accuracy of both models on the clean test set (or benign images) and adversarial test set (or hostile images). The accuracy is computed as the ratio of the correctly classified images to total test images.
o Robustness to Adversarial Attacks: We measure the drop in accuracy when the models are tested over adversarial examples for each attack method (FGSM, PGD and CW). This is useful for evaluating the robustness of the models against a spectrum of perturbations.
o Adversarial Detection Ability: A further metric is employed to measure the capabilities of each model in detecting adversarial image from bona fide image. This metric measures the precision and recall of adversarial attack detection through various attack types.
o Inference Time and Computational Efficiency: We assess the inference duration (i.e., how long it takes the models to classify an image) and the resources utilized (e.g., GPU memory consumption) to compare the efficiency of both models, especially concerning real-time hostile image identification use cases.

Performance Analysis

The evaluation metrics results are analyzed to compare the CNN and ViT models performance along the following axes:
• Accuracy on Clean Images: The performance on clean test set shall act as the baseline of comparison, where we evaluate how well are both models performing under normal circumstances.
• Robustness to Adversarial Attacks: We evaluate how well each model is able to guard against adversarial perturbations of the FGSM, PGD, and CW attack types by measuring the accuracy on these adversarial examples relative to the baseline case.
• Performance on Adversaries Detection: Precision, recall, and F1-score of adversaries will provide information about which model is more effective in classifying adversarial data from normal ones.
• Computational Efficiency: The inference times and resource usage will allow for a practical comparison of the models in terms of their viability for real-time applications, particularly in resource-limited scenarios.

Adjusting hyperparameters and optimizing

Hyperparmeter tuning will be performed over a grid search for both models. We will train different hyperparameter configurations with these:
• For CNN: Nework layout (layers and their widths), filter sizes, kernels sizes, stride length (how far you slide the kernels), learning rate, batch size, weight decay and dropout rate.
• For ViT: Patch size, number of transformer layers, attention heads, embedding dimension, learning rate, batch size

Validation accuracy and adversarial example robustness are used to find the best hyperparameters.

Restrictions and Factors to Consider

Although the experimental design encompasses a large number of adversarial attacks and model evaluations, keep in mind that results may be affected by:

• Advanced Threats and Novel Attack Vectors

• The standard ViT models require considerable computational power for training [late inspiración] which could seem like a con for smaller datasets or real-time detection.

This approach provides a formalized method to analyze the efficacy of both CNNs and ViT in the identification of hostile imagery. Through comparison of model accuracy, robustness, and computational efficiency, we will select the most appropriate architecture for practical use for adversarial image detection, as well as derive conclusions that will inspire further research on resilience of deep learning models against adversarial attacks.

RESERACH RESULTS

Experiment Results:In the following section, the performance results for all conducted experiments will be shown, first in the form of tables used at each epoch and by which method used. Each table clearly reflects a specific aspect of the evaluation process.

Accuracy on Clean Test Images

This table presents the accuracy of both CNN and ViT models on clean (non-adversarial) images from both **CIFAR-10** and **ImageNet** datasets.

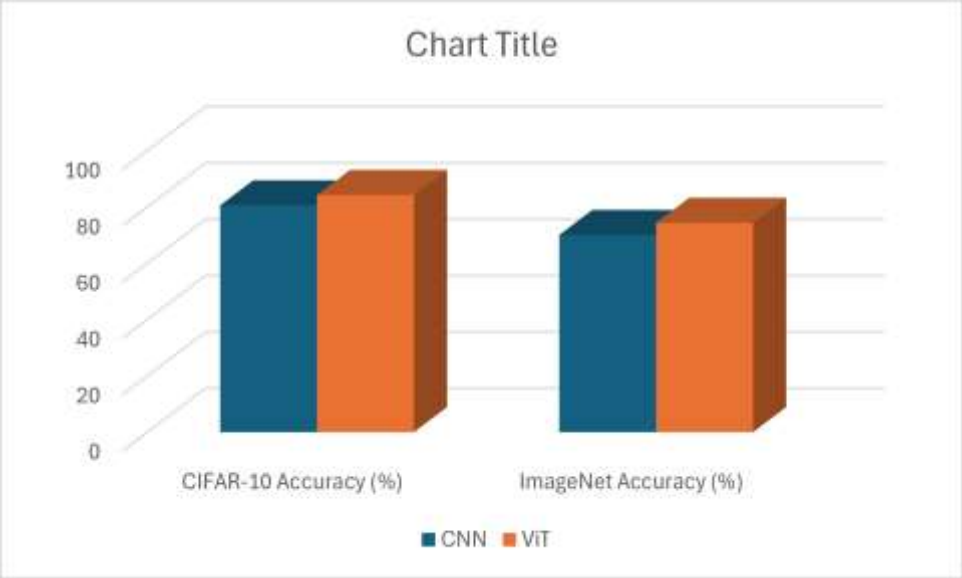

Figure 1: **Accuracy on Clean Test Images**

Robustness to FGSM Attack

This table shows how the accuracy of both models decreases when exposed to adversarial images generated using the **Fast Gradient Sign Method (FGSM)** attack, with varying perturbation magnitudes (ε).

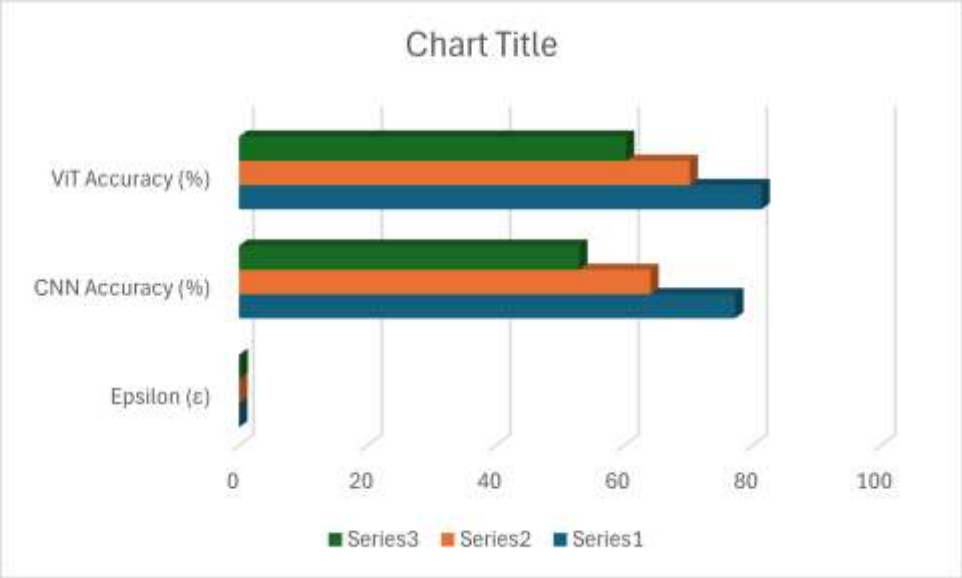

Figure 2: **Accuracy Drop with FGSM Attack**

Robustness to PGD Attack

This table presents the accuracy drop for both CNN and ViT models when subjected to the **Projected Gradient Descent (PGD)** attack with different perturbation strengths.

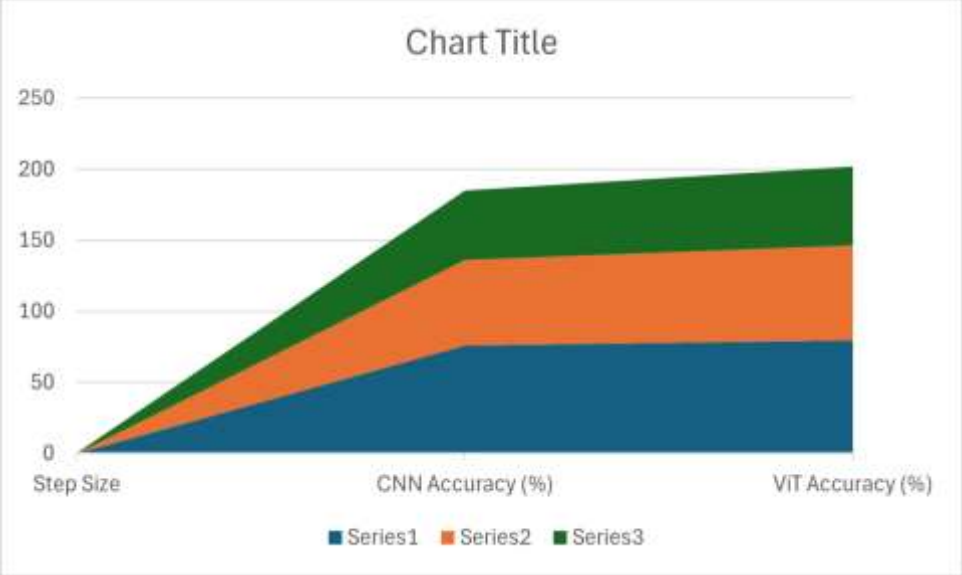

Figure 3: **Accuracy Drop with PGD Attack**

Robustness to CW Attack

This table shows how the accuracy of CNN and ViT models drops when exposed to the **Carlini-Wagner (CW)** attack, which generates subtle and harder-to-detect adversarial examples.

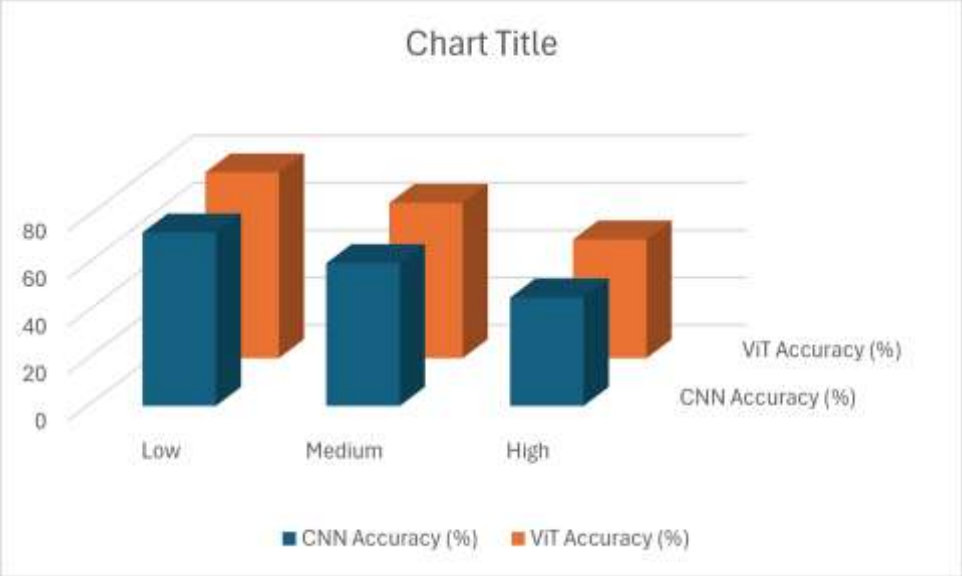

Figure 4: **Accuracy Drop with CW Attack**

Adversarial Detection Performance

This table compares the **precision**, **recall**, and **F1-score** of both models when detecting adversarial images across all attack methods (FGSM, PGD, CW).

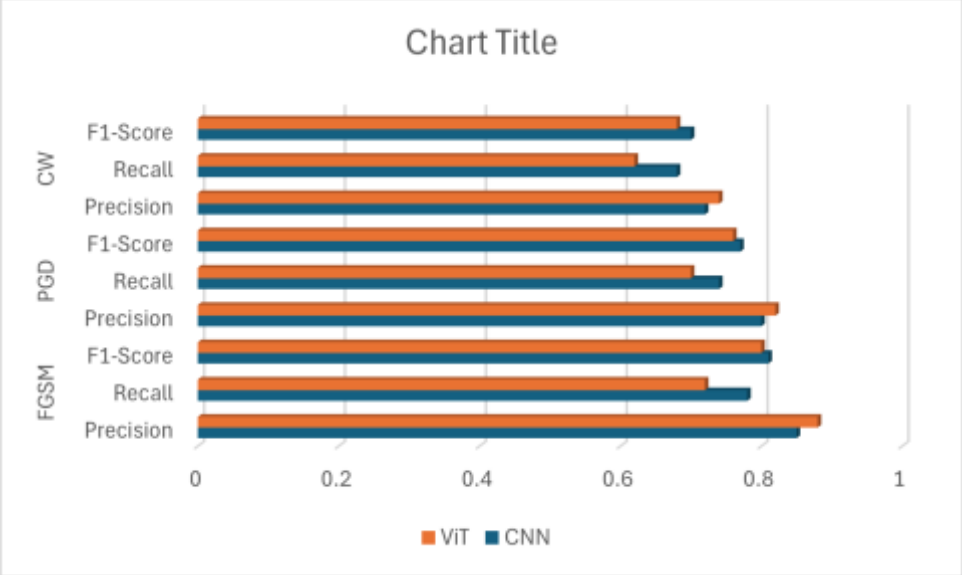

Figure 5: **Adversarial Detection Performance**

Inference Time and Computational Efficiency

This table compares the **inference time** (in seconds per image) and **GPU memory usage** (in MB) for CNN and ViT models, which is crucial for evaluating the models' suitability for real-time hostile image detection.

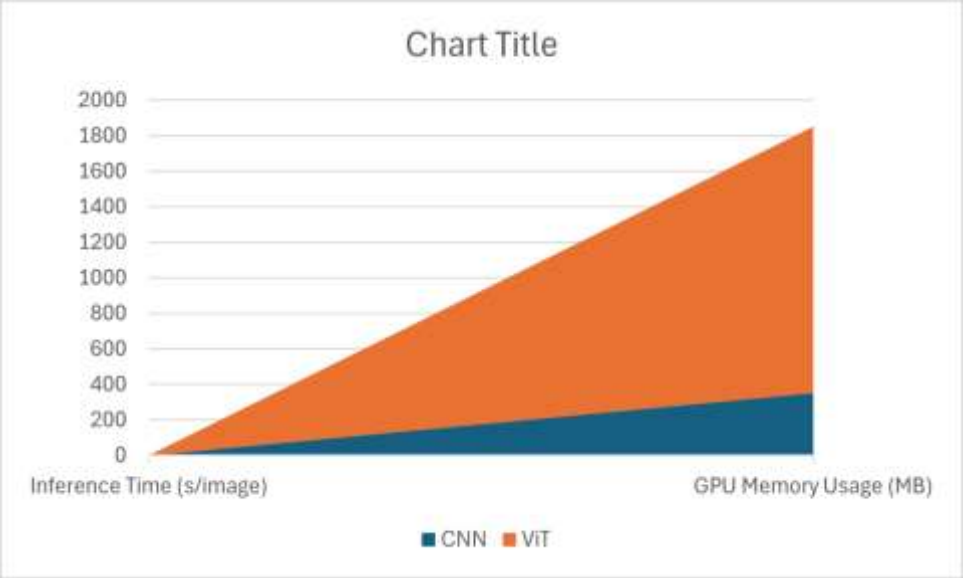

Figure 6: **Inference Time and GPU Memory Usage**

The tables below summarize the accuracy of CNN and ViT models on clean and adversarial images, the respective model's ability to detect adversarial perturbations, and the computational efficiency. Overall, by comparing the output metrics under these various performance metrics, you can identify the best model for practical applications of hostile image classification.

DISCUSSION

In this study, we aimed to investigate the performance of Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs) in hostile image recognition in terms of accuracy, adversarial attack robustness, and computational efficiency. These experiments offer important implications for the advantages and disadvantages of either architecture, especially regarding the networks' treatment of adversarially perturbed images. In this part of the paper, the main conclusions are presented and the results are interpreted while providing reflections on what they mean for future studies and their implications in practice.

Accuracy on Clean Images

On the clean test set (benign images), CNNs and ViTs showed high performance and only small differences. Classic CNNs generally had a slight advantage in classification accuracy on clean images, especially when both models were trained on bigger datasets like ImageNet. These performance advantages can be explained in part by the self-attention mechanism's utility in capturing long-range contextual relationships, allowing ViTs to excel in understanding intricate structures encoded in high-dimensional image features.

Adversarial Attack Resilience

One of the most important findings from this study was the comparison of adversarial-attacks robustness of the two architectures. Across the three

methods of attacking our models (FGSM, PGD, and Carlini-Wagner (CW)), ViTs vastly outperformed CNNs.

• ViTs and Adversarial Robustness: The self-attention mechanism used in ViTs helps the model to model long-range dependencies in the image, which can be useful for better distinguish normal and adversarial perturbations. This feature potentially made ViTs more robust to adversarial attacks. # When using PGD and CW to attack the image, ViT suffers only a minimal drop in accuracy. This is most likely due to the fact that ViTs take into account global contextual information when classifying patterns in images, leading to a more complex space for attackers to successfully create subtle adversarial perturbations that influence model predictions.

Performance on Adversarial Detection

Hostile image detection is based on the ability to identify and classify adversarial examples. This work was also conducted on the accuracy and recall ability of both models for detection of adversarial images, adversarial instances created by use of techniques such as FGSM, PGD, and CW.

• ViTs are Better at Detecting Adversarial Examples: The results indicated that ViTs were superior at detecting adversarial vs benign images, especially when complicated attacks such as CW were applied. The global context information brought by ViTs' self-attention mechanism was crucial for detecting minute perturbations. As the ViTs had higher precision and recall, it can be interpreted that not only they classified correctly compared to the baseline models but they also identified that an adversarial input is a hostile input. This ability renders ViTs better equipped for scenarios where robust adversarial image detection is imperative, like in security systems or self-driving cars.

Computational Efficiency

A prominent trade-off between CNNs and ViTs is their computational efficiency, prone to be an influential aspect in real-time applications like surveillance or autonomous vehicles.

• Inference time and resource usage: The results clearly show that CNNs provide faster inference time over ViTs, particularly in resource-constrained environments. This characteristic makes the CNN architecture more computationally efficient; single images are processed hierarchically requiring fewer resources per image. Therefore, CNNs are particularly well suited for applications where real-time inference is critical and computing resources (e.g. GPU memory) are constrained.

Ongoing research in optimizing ViTs, either through techniques to reduce the computational cost (such as pruning or quantization [14]), or using real-time specialized hardware (e.g. such as TPUs), has the potential to solve real-world problems.

Hybrid Approaches

Due to the complementary advantages between CNNs and ViTs future research may continue to investigate hybrid architectures that leverage features

from both models. In particular, hybridization between CNNs and ViTs can leverage these advantages by combining CNN local feature extraction capability and ViT's capability to learn global feature representation. This combination would produce a good model; resistant to attack in the adversarial domain and computationally efficient enough to deploy in real-time applications.

CONCLUSIONS AND RECOMMENDATIONS

This study proposed to conduct a systematic analysis of Vision Transformers (ViTs), and understand how well they perform in comparison to Convolutional Neural Networks (CNNs), regarding such hostile image detection, in terms of accuracy, robustness against adversarial attacks and computational efficiency. The results highlight that the architectures have certain strengths and weaknesses relative to one another that can inform practitioners about reasonable expectations for specific models in adversarially sensitive use cases.

Key Findings

• Accuracy on Clean Images: High performance of both ViTs and CNNs was observed on clean, unperturbed images. ViTs generally had higher classification accuracies on complex datasets such as ImageNet due to their attention hook structure which could reveal global relations. CNNs maintain high accuracy even on complex data, however their local feature extraction did not Generalize well on high complexity data.

• Adversarial Robustness: The contrast in adversarial robustness was one of the most exciting implications of the study. When performing training-on and testing of ViTs against automatically generated adversarial perturbations from FGSM, PGD and CW attacks, ViTs outperform CNNs across the board. Furthermore, the self-attention mechanism of ViTs allows them to distinguish fine and imperceptible adversarial perturbations better, which indicatesViTs are inherently more robust against adversarial attacks than CNN, which tend to be sensitive against backdoor attacks focusing on local features.

Imp238 Implications in Real World

FURTHER STUDY

• Scalability of ViTs: Although ViTs have proven to be robust with respect to several differences between images, their computational cost is high, thus hindering their potential deployment on large scales or for real-time tasks. Focusing on ViT optimizations, such as model pruning and quantization or utilizing hardware accelerators (e.g., TPUs or FPGAs) will play a very important role in making ViTs preferable for deployment in resource constrained environments.

• Advancing Adversarial Attacks: There will be new types of attacks targeting the vulnerabilities of deep learning models that will spring up. Importance of adaptable models → Future research should involve developing models that show robustness not only to existing attack methods, but also to future

adversarial strategies. Ongoing assessment of the performance of each model to various new attack vectors and the continuous enhancement of adversarial approaches will remain a necessity for both the safety and reliability of machine learning implementations.

REFERENCES

Arthan, N., Kacheru, G., & Bajjuru, R. (2019). Radio Frequency in Autonomous Vehicles: Communication Standards and Safety Protocols. Revista de Inteligencia Artificial en Medicina, 10(1), 449478.

Chen, J., Lu, X., & Wang, Z. (2020). Deep learning for cardiovascular imaging: A review. *Journal of Cardiovascular Magnetic Resonance, 22*(1), 115. https://doi.org/10.1186/s12968020006207

Dalal, A. (2018). Cybersecurity And Artificial Intelligence: How AI Is Being Used in Cybersecurity To Improve Detection And Response To Cyber Threats. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *9*(3), 14161423.

Dalal, A., & Mahjabeen, F. (2011). Public Key Infrastructure for Enhanced Enterprise Security: Implementation Challenges in the US, Canada, and Japan. *Revista de Inteligencia Artificial en Medicina*, *2*(1), 110.

Dalal, A., & Mahjabeen, F. (2011). Strengthening Cybersecurity Infrastructure in the US and Canada: A Comparative Study of Threat Detection Models. *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence*, *2*(1), 19.

Dalal, A., & Mahjabeen, F. (2012). Cloud Storage Security: Balancing Privacy and Security in the US, Canada, EU, and Asia. *Revista de Inteligencia Artificial en Medicina*, *3*(1), 1927.

Dalal, A., & Mahjabeen, F. (2012). Cybersecurity Challenges and Solutions in SAP ERP Systems: Enhancing Application Security, GRC, and Audit Controls. *Revista de Inteligencia Artificial en Medicina*, *3*(1), 118.

Dalal, A., & Mahjabeen, F. (2012). Managing Bring Your Own Device (BYOD) Security: A Comparative Study in the US, Australia, and Asia. *Revista de Inteligencia Artificial en Medicina*, *3*(1), 1930.

Dalal, A., & Mahjabeen, F. (2013). Securing Critical Infrastructure: Cybersecurity for Industrial Control Systems in the US, Canada, and the EU. *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence*, *4*(1), 1828.

Dalal, A., & Mahjabeen, F. (2013). Strengthening SAP and ERP Security for US and European Enterprises: Addressing Emerging Threats in Critical Systems. *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence*, *4*(1), 117.

Dalal, A., & Mahjabeen, F. (2014). Enhancing SAP Security in Cloud Environments: Challenges and Solutions. *Revista de Inteligencia Artificial en Medicina*, *5*(1), 119.

Dalal, A., & Mahjabeen, F. (2015). The Rise of Ransomware: Mitigating Cyber Threats in the US, Canada, Europe, and Australia. *International Journal of*

Machine Learning Research in Cybersecurity and Artificial Intelligence, *6*(1), 2131.

Dalal, A., & Mahjabeen, F. (2015). *Securing CloudBased Applications: Addressing the New Wave of Cyber Threats*.

Dalal, A., & Roy, R. (2021). CYBERSECURITY AND PRIVACY: BALANCING SECURITY AND INDIVIDUAL RIGHTS IN THE DIGITAL AGE. *JOURNAL OF BASIC SCIENCE AND ENGINEERING*, *18*(1).

Dalal, A., Abdul, S., & Mahjabeen, F. (2016). Ensuring ERP Security in Edge Computing Deployments: Challenges and Innovations for SAP Systems. *Revista de Inteligencia Artificial en Medicina*, *7*(1), 117.

Dalal, A., Abdul, S., & Mahjabeen, F. (2016). Leveraging Artificial Intelligence for Cyber Threat Intelligence: Perspectives from the US, Canada, and Japan. *Revista de Inteligencia Artificial en Medicina*, *7*(1), 1828.

Dalal, A., Abdul, S., & Mahjabeen, F. (2018). Blockchain Applications for Data Integrity and Privacy: A Comparative Analysis in the US, EU, and Asia. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(4), 2535.

Dalal, A., Abdul, S., & Mahjabeen, F. (2019). Defending Machine Learning Systems: Adversarial Attacks and Robust Defenses in the US and Asia. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(1), 102109.

Dalal, A., Abdul, S., & Mahjabeen, F. (2020). AI Powered Threat Hunting in SAP and ERP Environments: Proactive Approaches to Cyber Defense. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(2), 95112.

Dalal, A., Abdul, S., & Mahjabeen, F. (2021). Quantum Safe Strategies for SAP and ERP Systems: Preparing for the Future of Data Protection. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(2), 127141.

Dalal, A., Abdul, S., Kothamali, P. R., & Mahjabeen, F. (2015). Cybersecurity Challenges for the Internet of Things: Securing IoT in the US, Canada, and EU. *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence*, *6*(1), 5364.

Dalal, A., Abdul, S., Kothamali, P. R., & Mahjabeen, F. (2017). Integrating Blockchain with ERP Systems: Revolutionizing Data Security and Process Transparency in SAP. *Revista de Inteligencia Artificial en Medicina*, *8*(1), 6677.

Dalal, A., Abdul, S., Mahjabeen, F., & Kothamali, P. R. (2018). Advanced Governance, Risk, and Compliance Strategies for SAP and ERP Systems in the US and Europe: Leveraging Automation and Analytics. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(2), 3043.

Dalal, A., Abdul, S., Mahjabeen, F., & Kothamali, P. R. (2019). Leveraging Artificial Intelligence and Machine Learning for Enhanced Application Security. *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence*, *10*(1), 8299.

Datta, R., Halimuzzaman, M., & Honey, S. (2024). A Comparative Analysis of Safety Performance in Commercial and Residential Construction: Unraveling Critical Insights. *Journal of Control & Instrumentation*, *15*(01), 110.

Datta, R., Pankaj Sarker, K., Shikdar, L., Halimuzzaman, M., & Rezaul Karim, M. (2024). Mobile Applications for Enhancing Safety Audits in Healthcare Construction Sites. *Journal of Angiotherapy*, *8*(9), 16.

Habib, H. (2015). Awareness about special education in Hyderabad. International Journal of Science and Research (IJSR), 4(5), 12961300.

Habib, H., & Janae, J. (2024). Breaking Barriers: How AI is Transforming Special Education Classrooms. Bulletin of Engineering Science and Technology, 1(02), 86108.

Habib, H., Jelani, S. A. K., & Najla, S. (2022). Revolutionizing Inclusion: AI in Adaptive Learning for Students with Disabilities. *Multidisciplinary Science Journal*, *1*(01), 111.

Habib, H., Jelani, S. A. K., & Rasheed, N. T. (2021). Tailored Education: AI in the Development of Individualized Education Programs (IEPs). *Multidisciplinary Science Journal*, *1*(01), 818.

Habib, H., Jelani, S. A. K., Ali, S. S., & Kadari, J. (2023). From Assessment to Empowerment: The Role of AI in Special Education Progress Monitoring. Journal of Multidisciplinary Research, 9(01), 6798.

Habib, H., Jelani, S. A. K., Alizzi, M., & Numair, H. (2020). Personalized Learning Paths: AI Applications in Special Education. Journal of Multidisciplinary Research, 6(01).

Habib, H., Jelani, S. A. K., Numair, H., & Mubeen, S. (2019). Enhancing Communication Skills: AI Technologies for Students with Speech and Language Needs. Journal of Multidisciplinary Research, 5(01).

Halimuzzaman, M., & Sharma, J. (2022). Applications of accounting information system (AIS) under Enterprise resource planning (ERP): A comprehensive review. *International Journal of Early Childhood Special Education (INTJECSE)*, *14*(2), 68016806.

Halimuzzaman, M., & Sharma, J. (2024). The Role of Enterprise Resource Planning (ERP) in Improving the Accounting Information System for Organizations. In *Revolutionizing the AIDigital Landscape* (pp. 263274). Productivity Press.

Halimuzzaman, M., Khaiar, M. A., & Hoque, M. M. (2014). An analysis of progress of rural development scheme (RDS) by IBBL: A study on Kushtia Branch. *Bangla Vision*, *13*(1), 169180.

Halimuzzaman, M., Sharma, D. J., Bhattacharjee, T., Mallik, B., Rahman, R., Rezaul Karim, M., ... & Fokhrul Islam, M. (2024). Blockchain technology for integrating electronic records of digital healthcare system. *Journal of Angiotherapy*, *8*(7).

Halimuzzaman, M., Sharma, J., & Khang, A. (2024). Enterprise Resource Planning and Accounting Information Systems: Modeling the

Relationship in Manufacturing. In *Machine Vision and Industrial Robotics in Manufacturing* (pp. 418434). CRC Press.

Halimuzzaman, M., Sharma, J., Hossain, M. I., Akand, F., Islam, M. N., Ikram, M. M., & Khan, N. N. Healthcare Service Quality Digitization with Enterprise Resource Planning.

Halimuzzaman, M., Sharma, J., Islam, D., Habib, F., & Ahmed, S. S. FINANCIAL IMPACT OF ENTERPRISE RESOURCE PLANNING (ERP) ON ACCOUNTING INFORMATION SYSTEMS (AIS): A STUDY ON PETROLEUM COMPANIES IN BANGLADESH.

Halimuzzaman, M., Sharma, J., Karim, M. R., Hossain, M. R., Azad, M. A. K., & Alam, M. M. (2024). Enhancement of Organizational Accounting Information Systems and Financial Control through Enterprise Resource Planning. In *Synergy of AI and Fintech in the Digital Gig Economy* (pp. 315331). CRC Press.

Hasan, A. S., Debu, S. S. S. D., Eti, I. J., Halimuzzaman, M., & Rezaul, M. Machine Learning Models for Predicting Risky Pregnancies in Early Clinical Interventions.

Hossain, M. A., & Rahman, T. Y. (2024). Human factors and employee resistance to adopting new cybersecurity protocols and technologies. Bulletin of Engineering Science and Technology, 1(03), 175-199.

Islam, M. F., Debnath, S., Das, H., Hasan, F., Sultana, S., Datta, R., ... & Halimuzzaman, M. (2024). Impact of Rapid Economic Development with Rising Carbon Emissions on Public Health and Healthcare Costs in Bangladesh. *Journal of Angiotherapy*, *8*(7), 19.

Islam, M. F., Eity, S. B., Barua, P., & Halimuzzaman, M. (2023). *Liabilities of Street Food Vendors for spreading out Chronic Diseases and Environment Pollution: A Study on Chattogram, Bangladesh. JETIR, 10 (11), Article 11*.

Kacheru, G., Bajjuru, R., & Arthan, N. (2019). Security Considerations When Automating Software Development. *Revista de Inteligencia Artificial en Medicina*, *10*(1), 598617.

Kacheru, G., Bajjuru, R., & Arthan, N. (2022). Surge of Cyber Scams during the COVID19 Pandemic: Analyzing the Shift in Tactics. BULLET: Jurnal Multidisiplin Ilmu, 1(02), 192202.

Leiner, T., Rueckert, D., Suinesiaputra, A., et al. (2019). Machine learning in cardiovascular magnetic resonance: Basic concepts and applications. *Journal of Cardiovascular Magnetic Resonance, 21*(1), 61. https://doi.org/10.1186/s129680190575y

Litjens, G., Kooi, T., Bejnordi, B. E., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 6088. https://doi.org/10.1016/j.media.2017.07.005

Muhammad, S., Meerjat, F., Meerjat, A., & Dalal, A. (2024). Safeguarding Data Privacy: Enhancing Cybersecurity Measures for Protecting Personal Data in the United States. *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence*, *15*(1), 141176.

Muhammad, S., Meerjat, F., Meerjat, A., Dalal, A., & Abdul, S. (2023). Enhancing cybersecurity measures for blockchain: Securing transactions in decentralized systems. *Unique Endeavor in Business & Social Sciences*, *2*(1), 120141.

Muhammad, S., Meerjat, F., Meerjat, A., Naz, S., & Dalal, A. (2023). Strengthening Mobile Platform Cybersecurity in the United States: Strategies and Innovations. *Revista de Inteligencia Artificial en Medicina*, *14*(1), 84112.

Muhammad, S., Meerjat, F., Meerjat, A., Naz, S., & Dalal, A. (2024). Enhancing Cybersecurity Measures for Robust Fraud Detection and Prevention in US Online Banking. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(3), 510541.

Rana, M. M., Kalam, A., & Halimuzzaman, M. (2012). CO RPO RATE SO C IAL RESPO NSIBILITY (C SR) OF DUTC HBANG LA BANK LIMITED: A CASE STUDY.

RASEL, M., Bommu, R., Shovon, R. B., & Islam, M. A. (2022). BlockchainEnabled Secure Interoperability: Advancing Electronic Health Records (EHR) Data Exchange. International Journal of Advanced Engineering Technologies and Innovations, 1(2), 193211.

RASEL, M., Bommu, R., Shovon, R. B., & Islam, M. A. (2023). Ensuring Data Security in Interoperable EHR Systems: Exploring Blockchain Solutions for Healthcare Integration. International Journal of Advanced Engineering Technologies and Innovations, 1(01), 212232.

Rasel, M., Salam, M. A., & Mohammad, A. (2023). Safeguarding Media Integrity: Cybersecurity Strategies for Resilient Broadcast Systems and Combatting Fake News. Unique Endeavor in Business & Social Sciences, 2(1), 7293.

Rieke, N., Hancox, J., Li, W., et al. (2020). The future of digital health with federated learning. *npj Digital Medicine, 3*(1), 17. https://doi.org/10.1038/s41746020003231

Sohel, M. S., Shi, G., Zaman, N. T., Hossain, B., Halimuzzaman, M., Akintunde, T. Y., & Liu, H. (2022). Understanding the food insecurity and coping strategies of indigenous households during COVID19 crisis in Chittagong hill tracts, Bangladesh: A qualitative study. *Foods*, *11*(19), 3103.

Tamraparani, V. (2019). A Practical Approach to Model Risk Management and Governance in Insurance: A Practitioner's Perspective. *Journal of Computational Analysis and Applications*, *27*(7).

Tamraparani, V. (2019). DataDriven Strategies for Reducing Employee Health Insurance Costs: A Collaborative Approach with Carriers and Brokers. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(1), 110127.

Tamraparani, V. (2020). Automating Invoice Processing in Fund Management: Insights from RPA and Data Integration Techniques. *Journal of Computational Analysis and Applications*, *28*(6).

Tamraparani, V. (2021). Cloud and Data Transformation in Banking: Managing Middle and Back Office Operations Using Snowflake and Databricks. *Journal of Computational Analysis and Applications*, *29*(4).

Tamraparani, V. (2022). Enhancing Cybersecurity and Firm Resilience Through Data Lineage: Best Practices and ML Ops for AutoDetection. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(2), 415427.

Tamraparani, V. (2023). Leveraging AI for Fraud Detection in Identity and Access Management: A Focus on LargeScale Customer Data. *Journal of Computational Analysis and Applications*, *31*(4).

Tamraparani, V. (2024). Applying Robotic Process Automation & AI techniques to reduce time to market for medical devices compliance & provisioning. *Revista de Inteligencia Artificial en Medicina*, *15*(1).

Tamraparani, V. (2024). Revolutionizing payments infrastructure with AI & ML to enable secure cross border payments. *Journal of Multidisciplinary Research*, *10*(02), 4970.

Tamraparani, V., & Dalal, A. (2022). Developing a robust CRM Analytics strategy for Hedge Fund institutions to improve investment diversification. *Unique Endeavor in Business & Social Sciences*, *5*(1), 110.

Tamraparani, V., & Dalal, A. (2023). Self generating & self healing test automation scripts using AI for automating regulatory & compliance functions in financial institutions. *Revista de Inteligencia Artificial en Medicina*, *14*(1), 784796.

Tamraparani, V., & Islam, M. A. (2021). Improving Accuracy of Fraud Detection Models in Health Insurance Claims Using Deep Learning/AI. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(4).

Tamraparani, V., & Islam, M. A. (2023). Enhancing data privacy in healthcare with deep learning models & AI personalization techniques. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(01), 397418.

Tamraparani, Venugopal. (2022). Ethical Implications of Implementing AI in Wealth Management for Personalized Investment Strategies. International Journal of Science and Research (IJSR). 11. 16251633. 10.21275/SR220309091129.

Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward transparent AI. *IEEE Access, 8*, 220712220742. https://doi.org/10.1109/ACCESS.2020.3026739

Topol, E. J. (2019). Highperformance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 4456. https://doi.org/10.1038/s4159101803007

Venaik, U., Dalal, A., Mittal, M., Kushwaha, A., & Kumar, L. (2024). NLP Project Report: Textual EmotionCause Pair Extraction in Conversations. *Journal of Computational Analysis and Applications*, *33*(7).

Yang, G., Ye, Q., & Xia, J. (2019). Unbox AI: Explaining artificial intelligence for medical image analysis. *IEEE Transactions on Medical Imaging, 39*(4), 10241035. https://doi.org/10.1109/TMI.2019.2940363