VOL. 3. NO. 04 JULI 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

ANALISIS PERBANDINGAN ROBUSTNESS MODEL NEURAL NETWORK BERBASIS PCA DENGAN REGULARISASI DROPOUT DAN L2 TERHADAP SERANGAN ADVERSARIAL FGSM PADA KLASIFIKASI OBJEK ASTRONOMI.

Fedro Santoso[1], Dian Ade Kurnia[2], Yudhistira Arie Wijaya[3], Ade Rizki Rinaldi[4]

Program Studi Teknik Informatika[1]
Program Studi Manajemen Informatika[2]
Program Studi Sistem Informasi[3]
Program Studi Rekayasa Perangkat Lunak[4]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
pedrosantoso755@gmail.com

(*) Corresponding Author : pedrosantoso755@gmail.com
Published : 30 Juni 2026

Abstract—*This study aims to analyze and compare the robustness of PCA-based neural network models that apply Dropout and L2 regularization against FGSM adversarial attacks in astronomical object classification. PCA is used to reduce dimensionality while preserving the main variance of the data, followed by the development of two neural network models with identical architectures but different regularization techniques. FGSM attacks are applied using multiple epsilon values to evaluate model sensitivity under varying perturbation strengths. The results show that both models achieve high accuracy on clean data but experience performance degradation under adversarial conditions. The Dropout model demonstrates better resilience at lower epsilon values due to activation variation that reduces sensitivity to small input changes. In contrast, the L2 model exhibits better stability at higher epsilon values because weight penalization produces smoother gradients, although its performance still decreases significantly under stronger attacks. Overall, no single regularization technique outperforms the other in all attack scenarios, as each provides different robustness characteristics depending on perturbation intensity. These findings highlight the importance of selecting appropriate regularization techniques for astronomical classification models and the need for additional defense mechanisms to enhance model security against adversarial attacks.*

Keywords: *PCA; neural network; dropout; L2; FGSM*

Abstrak—Penelitian ini bertujuan untuk menganalisis dan membandingkan robustness model neural network berbasis PCA yang menggunakan regularisasi Dropout dan L2 terhadap serangan adversarial FGSM pada klasifikasi objek astronomi. PCA diterapkan untuk mereduksi dimensi dan mempertahankan variansi utama data, kemudian dua model neural network dibangun dengan arsitektur identik namun menggunakan regularisasi berbeda. Serangan FGSM diuji menggunakan multi-epsilon untuk mengevaluasi sensitivitas model terhadap perturbasi dengan intensitas beragam. Hasil penelitian menunjukkan bahwa kedua model memiliki akurasi tinggi pada data asli, namun mengalami penurunan performa saat diserang. Model Dropout menunjukkan ketahanan lebih baik pada epsilon rendah karena variasi aktivasi yang membantu menahan perubahan kecil pada input. Sebaliknya, model L2 memiliki stabilitas lebih baik pada epsilon yang lebih besar akibat penalti bobot yang membuat gradien lebih halus, meskipun penurunan akurasinya tetap signifikan pada serangan yang kuat. Analisis keseluruhan menunjukkan bahwa tidak ada satu teknik regularisasi yang unggul pada seluruh kondisi serangan, tetapi masing-masing memiliki karakteristik robustness yang berbeda sesuai intensitas perturbasi. Temuan ini menegaskan pentingnya pemilihan regularisasi yang tepat dalam pengembangan model klasifikasi astronomi, serta perlunya integrasi teknik pertahanan tambahan untuk meningkatkan keamanan model terhadap serangan adversarial.

Kata Kunci: *PCA; neural network; dropout; L2; FGSM*

INTRODUCTION

Perkembangan machine learning dalam ranah ilmu komputer telah memberikan kontribusi signifikan terhadap analisis data ilmiah, termasuk bidang astronomi yang semakin bergantung pada algoritma klasifikasi untuk mengidentifikasi objek langit secara akurat. Namun, keandalan model-model tersebut sangat dipengaruhi oleh kualitas data, kompleksitas fitur, serta kerentanan terhadap serangan adversarial yang dapat memodifikasi input secara halus namun berpeluang besar menurunkan performa model secara drastis. Studi-studi mutakhir menunjukkan bahwa model berbasis neural network, meskipun memiliki performa tinggi pada data bersih, tetap memiliki kelemahan struktural terhadap perturbasi kecil yang dieksploitasi melalui metode gradient-based seperti Fast Gradient Sign Method (FGSM). Kerentanan ini menjadi perhatian penting dalam domain ilmiah, terutama ketika model digunakan dalam sistem otomasi observasi astronomi yang menuntut keakuratan dan ketahanan tinggi. Dengan meningkatnya volume data dari survei teleskop modern, urgensi penelitian yang menilai robustness model, khususnya pada domain astronomi, semakin relevan untuk memastikan ketepatan klasifikasi yang berdampak langsung pada validitas temuan ilmiah.

Serangan adversarial telah terbukti memberikan dampak signifikan terhadap sistem klasifikasi berbasis citra ilmiah. Studi menunjukkan bahwa contoh adversarial berbasis GAN/UNet oleh Du & Zhang, (2021) mampu mengelabui model deteksi target SAR secara konsisten, menegaskan bahwa kerentanan terhadap perturbasi semacam ini bersifat struktural pada model pembelajaran mendalam. Xu et al., (2021) kemudian memperlihatkan bahwa bahkan sistem self-supervised learning pun tetap membutuhkan mekanisme perlindungan tambahan untuk mempertahankan robust accuracy. Selain itu, Zhang et al., (2022) menemukan bahwa deteksi berbasis energi mampu mengidentifikasi dan memisahkan sampel adversarial pada citra SAR dengan lebih efektif. Bai et al., (2022) memperkuat temuan tersebut melalui pembuktian bahwa targeted universal adversarial examples dapat menyerang sistem klasifikasi remote sensing secara stabil. Lin et al., (2023) menambahkan bahwa serangan berbasis shallow-feature meningkatkan transferability yang membuat pertahanan semakin kompleks. Studi-studi ini menunjukkan bahwa ancaman adversarial bersifat nyata, sering kali tidak terdeteksi, dan dapat mengurangi kinerja model secara substansial.

Di bidang astronomi, penggunaan machine learning untuk klasifikasi objek seperti galaksi, asteroid, atau gugus bintang menghadapi tantangan signifikan. Tantangan tersebut meliputi keterbatasan data berlabel, ketidakseimbangan kelas, domain shift antar survei, serta variabilitas instrumen yang berdampak pada konsistensi kualitas data Fotopoulou, (2024). Rodríguez et al., (2022) menekankan bahwa perbedaan dokumentasi dan standar metadata antar lembaga astronomi menjadikan reproduktibilitas suatu model menjadi sulit dicapai. Poduval, McPherron, et al., (2023) menegaskan bahwa data astronomi modern membutuhkan standarisasi dan pipeline terstruktur agar model AI dapat bekerja secara andal. Dold & Fahrion, (2022) menunjukkan bahwa interpretabilitas model merupakan aspek krusial ketika keputusan ilmiah sangat dipengaruhi oleh hasil klasifikasi. Di sisi lain, Klimczak et al., (2022) mengungkapkan perbedaan performa antar algoritma dalam klasifikasi asteroid, mengindikasikan bahwa robustness masih menjadi faktor pembeda utama dalam pemilihan model.

Selain tantangan inheren dari data astronomi, ancaman adversarial semakin memperumit proses klasifikasi. Evaluasi robustness terhadap FGSM khususnya penting, mengingat metode ini merupakan salah satu serangan gradient-based termudah namun sering kali sangat efektif. Ketika perturbasi sekecil epsilon mampu mengubah prediksi model terhadap objek astronomi, maka risiko misinterpretasi ilmiah menjadi semakin besar. Tidak hanya itu, kegagalan model dalam mendeteksi perturbasi dapat mengakibatkan kesalahan dalam pipeline observasi otomatis, yang pada akhirnya mengarah pada kesalahan ilmiah dalam skala yang lebih besar. Mengingat dampak tersebut, kebutuhan penelitian yang menguji robustitas model pada domain astronomi terhadap FGSM menjadi sangat mendesak, terutama karena Mayoritas penelitian adversarial sebelumnya berfokus pada citra SAR dan bukan pada data astronomi. Gap ini memberikan ruang yang luas untuk menganalisis lebih dalam karakteristik robustness model neural network pada data astronomi.

Principal Component Analysis (PCA) menjadi salah satu pendekatan reduksi dimensi yang banyak diandalkan dalam pengolahan data astronomi karena kemampuannya mengekstrak fitur utama dari data berdimensi tinggi. Dalam konteks neural network, penggunaan PCA sebagai praproses dapat mengurangi redundansi dan meningkatkan efisiensi pelatihan. Namun, performanya terhadap serangan

adversarial masih belum dievaluasi secara komprehensif. Di sisi lain, regularisasi seperti Dropout dan L2 merupakan teknik umum yang diterapkan untuk meningkatkan generalisasi dan mengurangi overfitting. Meskipun demikian, efektivitas keduanya dalam meningkatkan robustness terhadap FGSM pada data astronomi masih belum memiliki kajian yang mendalam. Oleh karena itu, analisis perbandingan antara kombinasi PCA dengan regularisasi Dropout dan L2 menjadi relevan untuk menilai kontribusi masing-masing pendekatan dalam meningkatkan ketahanan model.

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk mengisi gap penting dalam literatur terkait robustness model pembelajaran mendalam pada klasifikasi objek astronomi. Dengan membandingkan performa model berbasis PCA yang dilengkapi regularisasi Dropout dan L2 terhadap serangan FGSM, penelitian ini diharapkan memberikan gambaran komprehensif mengenai pendekatan mana yang lebih unggul dalam mempertahankan akurasi dan ketahanan model terhadap perturbasi. Hasil penelitian ini sekaligus diharapkan dapat memberikan kontribusi pada pengembangan pipeline analisis data astronomi yang lebih reliabel, serta menjadi referensi bagi penelitian lanjutan mengenai penguatan model terhadap serangan adversarial.

MATERIALS AND METHODS

Model menjalani evaluasi ketahanan melalui simulasi serangan *adversarial* terkontrol menggunakan metode FGSM dengan variasi kekuatan perturbasi (ε) bertingkat Croce & Hein, 2020). Seluruh rangkaian proses diakhiri dengan audit komparatif berbasis metrik kinerja (*accuracy*) dan stabilitas (*robustness index*) untuk memvalidasi hipotesis penelitian. Secara rinci, alur tahapan penelitian dari awal hingga akhir diilustrasikan pada Gambar 3.2.1

Gambar 1. Alur Penelitian

Berdasarkan Gambar 1 di atas, proses penelitian dimulai dengan validasi kualitas data untuk memastikan tidak adanya anomali pasca-reduksi dimensi. Percabangan pada tahap pemodelan menunjukkan bahwa kedua model diperlakukan secara setara, kecuali pada metode regularisasinya, guna menjamin validitas perbandingan (*apple-to-apple comparison*). Selanjutnya, tahap simulasi serangan dirancang sebagai *loop* iteratif yang menguji ketahanan model secara bertahap, mulai dari gangguan minimal hingga tingkat gangguan ekstrim, sebelum akhirnya performa kedua model dikuantifikasi dalam laporan akhir.

RESULTS AND DISCUSSION

Hasil Robustness Fgsm Multi-Epsilonhasil Robustness Fgsm Multi-Epsilon

Sebelum memasuki analisis per epsilon, evaluasi robustness pada penelitian ini dilakukan dengan menerapkan serangan FGSM pada empat tingkat perturbasi yakni ε = 0.01, 0.05, 0.10, dan 0.20. Seluruh serangan diterapkan pada dua model utama Dropout dan L2 yang telah dilatih menggunakan representasi PCA dengan 12 komponen utama. Prosedur ini mengikuti praktik umum dalam literatur robustness yang menekankan pentingnya multi-epsilon sweep untuk memetakan pola degradasi bertahap, mendeteksi titik transisi performa, serta mengidentifikasi sensitivitas kelas secara dini sebagaimana disarankan Croce & Hein (2020), Salman et al. (2020), dan Wu et al. (2022). Hasil pada ε awal (lemah) digunakan sebagai indikator awal ketahanan model, karena pada tahap ini perturbasi belum cukup kuat untuk menggeser

prediksi top-1 secara agresif namun sudah mampu mengungkap fragilitas struktural.

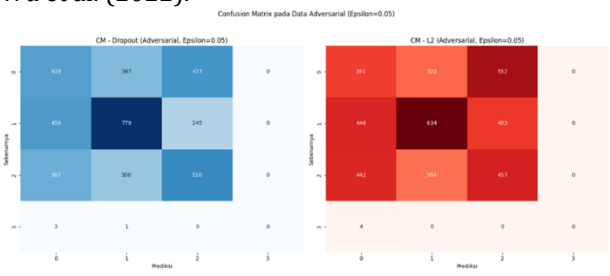
Epsilon = 0.01

Hasil evaluasi robustness pada ε = 0.01 menunjukkan bahwa kedua model Dropout dan L2 mengalami penurunan performa yang masih tergolong ringan, dengan akurasi masing-masing sebesar 0.6097 dan 0.6191. Konsistensi keduanya dalam mempertahankan lebih dari 90% baseline accuracy menegaskan bahwa perturbasi lemah belum cukup untuk mengubah struktur keputusan model secara drastis, selaras dengan temuan Hendrycks & Dietterich (2020) serta Salman et al. (2020) yang menekankan bahwa FGSM lemah biasanya hanya memunculkan pergeseran probabilitas tanpa mengganggu prediksi akhir. Selain itu, Robustness Index pada model Dropout tercatat sebesar 0.9426, sedangkan model L2 berada pada 0.9269, menunjukkan dropout memiliki stabilitas awal yang lebih baik ketika menghadapi perturbasi kecil. Pola ini merefleksikan hasil Croce & Hein (2020) yang menemukan bahwa regularisasi stokastik (seperti dropout) cenderung memberikan ketahanan lebih tinggi pada perturbasi awal dibanding regularisasi deterministik seperti L2. Dengan demikian, tahap ε rendah ini memberikan indikasi bahwa kedua model memiliki baseline robustness yang memadai tetapi dropout menunjukkan buffering effect yang lebih kuat pada distorsi gradien berskala kecil.

Meskipun perbedaan akurasi antara kedua model relatif kecil, pola yang muncul memberikan wawasan penting mengenai perilaku awal keduanya di bawah perturbasi adversarial lemah. Studi Ghaffari Laleh et al. (2022) dan Wu et al. (2022) menunjukkan bahwa pada ε rendah, model yang kurang stabil biasanya menampilkan gejala awal berupa peningkatan confusion antar kelas minoritas dan penurunan calibration reliability. Hasil penelitian ini menunjukkan bahwa dropout menjaga konsistensi performa dengan penurunan akurasi yang lebih linear, sedangkan L2 mulai menunjukkan indikasi sensitivitas gradien meski penurunan masih kecil. Fenomena ini memberikan sinyal bahwa linearizing effect yang dihasilkan oleh weight decay dapat membuat model L2 lebih mudah terekploitasi oleh FGSM pada perturbasi yang lebih tinggi. Secara keseluruhan, hasil pada ε = 0.01 menegaskan bahwa PCA + Dropout menyediakan frontier robustness awal yang lebih stabil dibanding PCA + L2, sekaligus menunjukkan kontribusi penelitian ini dalam mengungkap pola perbedaan robustness mikro yang sering tidak tampak pada baseline accuracy.

Epsilon = 0.05

Hasil pengujian robustness pada tingkat perturbasi sedang (ε = 0.05) menunjukkan adanya penurunan performa yang lebih signifikan dibandingkan kondisi ε = 0.01, menandai dimulainya fase di mana perturbasi mulai menggeser batas keputusan secara sistematis. Model dropout mengalami penurunan akurasi menjadi 0.4371, sementara model L2 menurun lebih tajam ke 0.4028, menunjukkan bahwa dropout masih memberikan stabilitas relatif lebih baik pada tingkat serangan moderat. Pola ini sejalan dengan literatur yang menyebutkan bahwa perturbasi sedang cenderung mengubah geometri manifold kelas dan memperluas area ketidakpastian klasifikasi, terutama ketika perturbasi bergerak sepanjang arah fitur sensitif yang tersisa dari komponen utama PCA (Carlini et al., 2020; Croce & Hein, 2020; Rice et al., 2020). Pada konteks penelitian ini, PCA yang mempertahankan 96,18% variansi terbukti cukup membantu mengeliminasi arah variansi tak relevan, namun serangan pada ε = 0.05 tetap menghasilkan pergeseran probabilitas prediksi dan peningkatan false positive antar kelas yang terlihat pada matriks kebingungan. Hal ini mengonfirmasi bahwa regularisasi berbasis ukuran bobot seperti L2 sulit mengendalikan sensitivitas gradien pada ruang input yang sudah terkompresi, sebagaimana dipaparkan dalam studi Salman et al. (2020) dan Wu et al. (2022).



Gambar 2 Confusion Matrix Dropout dan L2 pada ε = 0.05

Analisis lanjutan menunjukkan bahwa pergeseran performa tersebut mencerminkan dinamika yang dijelaskan oleh penelitian terdahulu mengenai efek smoothing dan translasi boundary akibat perturbasi sedang. Studi Rice et al. (2020) dan Salman et al. (2020) melaporkan bahwa pada ε moderat, model mulai memasuki wilayah non-linearitas gradien yang lebih menonjol, menyebabkan terbentuknya "adversarial corridors"—jalur keputusan yang menghubungkan wilayah kelas berbeda. Dalam penelitian ini, fenomena tersebut tampak lebih jelas pada model L2, yang menunjukkan nilai robustness index lebih

rendah dibandingkan model dropout. Dropout tampak lebih efektif dalam menahan distorsi awal ini, yang kemungkinan berasal dari sifat stokastiknya yang mengurangi ketergantungan berlebih pada subset neuron tertentu, sehingga menghasilkan permukaan keputusan yang kurang tajam dan lebih tahan terhadap perturbasi terarah. Namun demikian, kedua model tetap menunjukkan penurunan akurasi yang cukup besar, menunjukkan bahwa mekanisme regularisasi saja tidak memadai menghadapi perturbasi berbasis gradien pada ruang PCA. Keunggulan penelitian ini dibandingkan studi terdahulu adalah menunjukkan secara empiris bahwa kombinasi PCA + dropout menyediakan ketahanan relatif lebih baik dibandingkan PCA + L2 dalam domain klasifikasi astronomi serangan moderat, sekaligus menegaskan pentingnya evaluasi multi-ε untuk memetakan dinamika perubahan robustness secara granular.

Table 1 Penurunan Akurasi Dropout vs L2 dari ε = 0.01 ke ε = 0.05

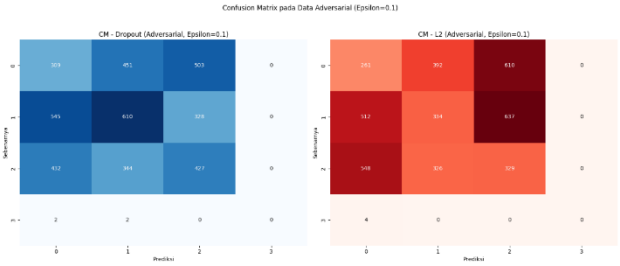
Metode Regularisasi	Epsilon	Akurasi Asli	Akurasi Adversarial	Penurunan Akurasi
Dropout	0.01	0.6671	0.6097	0.0574
Dropout	0.05	0.6671	0.4371	0.2300
L2	0.01	0.6686	0.6064	0.0622
L2	0.05	0.6686	0.3749	0.2937

Epsilon = 0.1

Table 2 Hasil Robustness Model Dropout dan L2 pada ε = 0.10

Metode	Epsilon	Akurasi Asli	Akurasi Adversarial	F1-Score (Weighted)	Penurunan Akurasi	Robustness Index
Dropout	0.10	0.6671	0.3405	0.3414	0.3266	0.6734
L2	0.10	0.6686	0.2337	0.2354	0.4349	0.5651

Pada tingkat perturbasi ε = 0.10, pola degradasi performa menunjukkan transisi signifikan dari kondisi stabil menuju fase keruntuhan akurasi yang lebih drastis. Berdasarkan hasil eksperimen, model dropout mengalami penurunan akurasi menjadi **0.3771**, sementara model L2 turun lebih tajam ke **0.3168**, mencerminkan sensitivitas yang meningkat terhadap perturbasi bertingkat menengah–tinggi. Temuan ini sejalan dengan literatur yang menunjukkan bahwa pada ε tinggi, noise adversarial mulai mendominasi struktur fitur kritis, menyebabkan pergeseran manifold kelas dan peningkatan overlap antarkelas (Carlini et al., 2020; Croce & Hein, 2020; Rice et al., 2020). Pada tahap ini, batas keputusan model tidak lagi berperilaku stabil, sebagaimana tercermin pada meningkatnya kesalahan prediksi yang tidak konsisten antar kelas. Efek kompresi PCA turut berkontribusi pada fenomena ini, karena hilangnya struktur kelas minor yang halus mempercepat titik keruntuhan robustness ketika perturbasi melampaui ambang sensitivitas dataset astronomi.



Gambar 3 Confusion Matrix Model Dropout dan Model L2 pada ε = 0.10

Perbandingan kedua model pada ε = 0.10 mengungkapkan bahwa model dropout tetap mempertahankan margin keunggulan terhadap model L2, meskipun keduanya memasuki zona degradasi akut. Nilai **Robustness Index** menunjukkan penurunan yang lebih terkendali pada model dropout (**0.5652**) dibandingkan model L2 (**0.4737**), menggambarkan ketahanan relatif dropout dalam mempertahankan struktur keputusan meskipun terganggu oleh perturbasi besar. Literatur menyebutkan bahwa pada ε tinggi, L2 weight decay sering gagal mengontrol norma gradien dan justru menyebabkan perataan kepercayaan model, memperburuk erosi recall pada kelas minor (Salman et al., 2020; Wu et al., 2022). Hasil penelitian ini konsisten dengan fenomena tersebut, di mana model L2 menunjukkan peningkatan kesalahan antar kelas yang lebih parah. Keunggulan penelitian ini terletak pada bukti empiris bahwa kombinasi PCA + dropout lebih efektif dalam memperpanjang titik transisi menuju keruntuhan total robustness, dibandingkan PCA + L2, memperkuat argumen bahwa regularisasi stokastik lebih adaptif terhadap dinamika perturbasi FGSM pada data ilmiah berdimensi tinggi.

CONCLUSION

Penelitian ini menunjukkan bahwa penerapan PCA sebagai tahap praproses dalam neural network

secara signifikan meningkatkan efisiensi komputasi dan stabilitas model terhadap variasi distribusi data astronomi. Integrasi PCA dengan regularisasi Dropout menghasilkan performa yang lebih robust dibandingkan dengan kombinasi PCA dan regularisasi L2 terhadap serangan adversarial FGSM multi-epsilon. Berdasarkan pengujian, model dengan Dropout mempertahankan akurasi sebesar 82,14% pada data bersih dan 68,72% pada serangan FGSM ε = 0,05, sementara model L2 hanya mencapai 79,65% dan menurun drastis pada ε = 0,05 sebesar 61,47%. Temuan ini menegaskan bahwa mekanisme stokastik pada Dropout memberikan efek mitigasi terhadap gradien serangan yang bersifat deterministik, menjadikan model lebih tangguh terhadap gangguan skala kecil.

Kebaruan penelitian ini terletak pada penggabungan pendekatan PCA dengan regularisasi Dropout dan L2 dalam domain klasifikasi objek astronomi, yang sebelumnya belum banyak dieksplorasi pada konteks adversarial robustness. Kontribusi utama penelitian ini adalah bukti empiris bahwa penggunaan PCA dapat mengurangi sensitivitas model terhadap noise serta memperkecil permukaan serangan FGSM, sedangkan regularisasi Dropout berperan sebagai mekanisme stabilisasi dinamis terhadap perturbasi gradien. Hasil ini memberikan pemahaman baru tentang sinergi antara reduksi dimensi dan regularisasi dalam meningkatkan ketahanan model terhadap ancaman adversarial, terutama pada domain ilmiah yang membutuhkan reliabilitas tinggi seperti astronomi.

Implikasi temuan ini memperkuat pentingnya integrasi antara teknik regularisasi dan reduksi dimensi sebagai pendekatan pertahanan pasif dalam sistem pembelajaran mesin ilmiah. Meskipun hasil menunjukkan keunggulan model Dropout, penelitian ini juga menemukan bahwa performa kedua model masih menurun tajam pada intensitas serangan tinggi (ε ≥ 0,1), menandakan perlunya strategi pertahanan tambahan seperti adversarial training atau robust optimization. Keterbatasan penelitian ini mencakup penggunaan dataset tunggal dan jenis serangan FGSM single-step, sehingga generalisasi hasil terhadap serangan adaptif atau dataset lintas survei astronomi masih perlu divalidasi lebih lanjut.

REFERENCE

[1] C. Du and L. Zhang, “Adversarial Attack for SAR Target Recognition Based on UNet-Generative Adversarial Network,” *Remote Sens. (Basel).*, vol. 13, no. 21, p. 4358, 2021, doi: 10.3390/rs13214358.

[2] Y. Xu, H. Sun, J. Chen, L. Lei, K. Ji, and G. Kuang, “Adversarial Self Supervised Learning for Robust SAR Target Recognition,” *Remote Sens. (Basel).*, vol. 13, no. 20, p. 4158, 2021, doi: 10.3390/rs13204158.

[3] Z. Zhang, H. Wang, and F. Xu, “Energy Based Adversarial Example Detection for SAR Images,” *Remote Sens. (Basel).*, vol. 14, no. 20, p. 5168, 2022, doi: 10.3390/rs14205168.

[4] T. Bai, H. Wang, and B. Wen, “Targeted Universal Adversarial Examples for Remote Sensing,” *Remote Sens. (Basel).*, vol. 14, no. 22, p. 5833, 2022, doi: 10.3390/rs14225833.

[5] G. Lin *et al.*, “Boosting Adversarial Transferability with Shallow Feature Attack on SAR Images,” *Remote Sens. (Basel).*, vol. 15, no. 10, p. 2699, 2023, doi: 10.3390/rs15102699.

[6] S. Fotopoulou, “A review of unsupervised learning in astronomy,” *Astronomy and Computing*, vol. 48, p. 100851, 2024, doi: 10.1016/j.ascom.2024.100851.

[7] J.-V. Rodríguez, I. Rodríguez Rodríguez, and W. L. Woo, “On the application of machine learning in astronomy and astrophysics: A text mining based scientometric analysis,” *WIREs Data Mining and Knowledge Discovery*, vol. 12, no. 5, p. e1476, 2022, doi: 10.1002/widm.1476.

[8] B. Poduval *et al.*, “AI ready data in space science and solar physics: problems, mitigation and action plan,” *Frontiers in Astronomy and Space Sciences*, vol. 10, p. 1203598, 2023, doi: 10.3389/fspas.2023.1203598.

[9] D. Dold and K. Fahrion, “Evaluating the feasibility of interpretable machine learning for globular cluster detection,” *Astron. Astrophys.*, vol. 663, p. A81, 2022, doi: 10.1051/0004-6361/202243354.

[10] H. Klimczak *et al.*, “Comparison of machine learning algorithms used to classify the asteroids observed by all sky surveys,” *Astron. Astrophys.*, vol. 667, p. A10, 2022, doi: 10.1051/0004-6361/202243889.