Journal of Data Science Methods and Applications
Vol. 01, No. 02, November 2025, pp. 59-65
E-ISSN: 3090-5605, DOI: 10.30873/jodmapps.v1i2.pp59-65

Prediksi Jenis Ancaman Siber Global Menggunakan Algoritma *Random Forest*

Yudista Nanda P*[1], Rionaldi Ali[2], Noki Aris[3], Agus Saputra[4], M. David Saputra[5]

[1,3,4,5]Program Studi Bisnis Digital, Institut Informatika Dan Bisnis Darmajaya
[2]Program Studi Teknik Informatika, Institut Informatika Dan Bisnis Darmajaya
email: [1]yudistananda84@gmail.com, [3]nokiaris2004gmail.com, [4]agussaputra040903@gmail.com, [5]md047300@gmail.com

Article History:
Received : 02-08-2025
Revised : 12-09-2025
Accepted : 29-11-2025
Online : 29-11-2025

A B S T R A K

Ancaman siber global terus meningkat seiring dengan transformasi digital yang meluas di berbagai sektor. Penelitian ini bertujuan untuk memprediksi jenis serangan siber berdasarkan data historis global dari tahun 2015 hingga 2024. Data diperoleh dari *dataset* Global Cybersecurity Threats yang mencakup informasi tentang negara, sektor terdampak, dan jenis serangan. Metode yang digunakan adalah *supervised learning* dengan algoritma *Random Forest*, yang dikenal efektif untuk klasifikasi dan analisis variabel kompleks. Hasil penelitian menunjukkan bahwa algoritma ini mampu mengidentifikasi pola serangan dengan akurasi tinggi dan membantu dalam deteksi dini ancaman. Penelitian ini diharapkan dapat berkontribusi pada pengembangan sistem keamanan siber berbasis data dan pemodelan prediktif. Ancaman siber global terus berkembang dengan kompleksitas yang semakin tinggi, seiring dengan meningkatnya ketergantungan masyarakat terhadap sistem digital. Penelitian ini bertujuan untuk menganalisis tren dan memprediksi jenis ancaman siber berdasarkan data historis dari tahun 2015 hingga 2024, yang diperoleh dari *dataset* Global Cybersecurity Threats. Metode yang digunakan adalah supervised learning dengan algoritma *Random Forest*, guna mengklasifikasikan jenis serangan dan memprediksi potensi kerugian finansial. Hasil analisis menunjukkan bahwa model dapat mengidentifikasi pola-pola penting yang dapat membantu organisasi dalam mitigasi risiko siber. Penelitian ini memberikan kontribusi dalam pengembangan sistem inteligensi ancaman siber berbasis data.

Kata kunci: Ancaman Siber, Prediksi Serangan, Pembelajaran Mesin, Random Forest, Keamanan Digital.

A B S T R A C T

Global cyber threats continue to increase along with the widespread digital transformation across various sectors. This study aims to predict the types of cyberattacks based on global historical data from 2015 to 2024. Data was obtained from the Global Cybersecurity Threats dataset, which includes information on countries, affected sectors, and types of attacks. The method used was supervised learning with the Random Forest algorithm, which is known to be effective for classifying and analyzing complex variables. The results show that this algorithm is capable of identifying attack patterns with high accuracy and assisting in early threat detection. This research is expected to contribute to the development of data-driven cybersecurity systems and predictive modeling. Global cyberthreats continue to grow in complexity, along with society's increasing reliance on digital systems. This study aims to analyze trends and predict the types of cyberthreats based on historical data from 2015 to 2024, obtained from the Global Cybersecurity Threats dataset. The method used was supervised learning with the Random Forest algorithm to classify attack types and predict potential financial losses. The analysis results show that the model can identify important patterns that can assist organizations in mitigating cyber risks. This research contributes to the development of data-driven cyber threat intelligence systems.

Keywords: *Cyber Threats, Attack Prediction, Machine Learning, Random Forest, Digital Security.*

1. INTRODUCTION

Keamanan siber telah menjadi perhatian utama di seluruh dunia karena semakin maraknya serangan digital yang menargetkan berbagai sektor strategis. Menurut laporan Check Point Research tahun 2023, rata-rata insiden serangan siber meningkat sebesar 38% secara global [1]. Berbagai jenis serangan seperti *ransomware*, *phishing*, dan *Distributed Denial of Service* (DDoS) telah menyebabkan kerugian finansial dan gangguan operasional yang serius pada instansi pemerintahan maupun perusahaan swasta [2], [3].

Salah satu jenis serangan yang paling berbahaya adalah *ransomware*, di mana penyerang mengenkripsi data korban dan meminta tebusan agar data tersebut dapat dipulihkan. Pencegahan *ransomware* dapat dilakukan dengan rutin melakukan backup data secara *offline* atau *cloud*, memperbarui perangkat lunak *anti-malware*, serta menghindari membuka lampiran atau tautan dari sumber yang tidak dikenal [4]. Selain itu, serangan phishing juga menjadi salah satu ancaman utama. Phishing dilakukan dengan menyamar sebagai pihak tepercaya untuk memperoleh informasi sensitif dari korban, seperti password atau data kartu kredit. Upaya pencegahan *phishing* meliputi pelatihan keamanan siber kepada pengguna, penerapan filter email canggih, dan penggunaan autentikasi dua faktor [5]. Serangan *Distributed Denial of Service* (DDoS) bertujuan melumpuhkan layanan dengan membanjiri jaringan menggunakan trafik palsu dalam jumlah besar. Pencegahan DDoS dilakukan melalui *firewall*, sistem deteksi intrusi, serta penggunaan layanan mitigasi DDoS berbasis *cloud* [6]. SQL Injection merupakan serangan yang mengeksploitasi celah pada input aplikasi untuk menjalankan perintah SQL berbahaya, dengan tujuan mencuri atau merusak data. Cara mengatasinya adalah dengan validasi input secara ketat, menggunakan *prepared statement*, dan memperkuat konfigurasi *database* [7]. Terakhir, malware merupakan perangkat lunak berbahaya yang dirancang untuk merusak sistem atau mencuri data. Pencegahan dilakukan dengan memperbarui sistem secara berkala, memasang *anti-malware* yang andal, dan berhati-hati dalam mengunduh *file* internet [8].

Prediksi ancaman siber saat ini banyak menggunakan algoritma *machine learning* karena mampu mengenali pola serangan dari data historis dan memprediksi kemungkinan serangan di masa depan [9]. Contoh algoritma yang umum digunakan adalah *Random Forest, Support Vector Machine (SVM), Naïve Bayes*, dan *K-Nearest Neighbors (KNN)* [10]. *Random Forest* bekerja dengan membangun banyak pohon keputusan dari subset data acak, kemudian menggabungkan hasilnya melalui voting sehingga tahan terhadap overfitting dan cocok untuk data besar dan kompleks [11]. SVM berfungsi dengan mencari garis pemisah terbaik (*hyperplane*) yang memaksimalkan margin antar kelas data, sehingga efektif untuk klasifikasi, misalnya deteksi *phishing* berbasis teks [12]. Naïve Bayes menggunakan pendekatan probabilistik dengan Teorema Bayes dan mengasumsikan fitur saling independen, sehingga cepat dan sering digunakan untuk filtering spam [13]. Sementara itu, KNN mengklasifikasikan data baru berdasarkan label mayoritas dari tetangga terdekatnya dan sering dipakai untuk mendeteksi anomali pada jaringan [14]. Dengan berbagai algoritma ini, sistem prediksi ancaman siber dapat membantu meningkatkan akurasi deteksi dan mendukung mitigasi risiko yang lebih efektif [15].

Penelitian ini bertujuan untuk memprediksi jenis ancaman siber berdasarkan data historis yang dikumpulkan dari berbagai insiden antara tahun 2015 hingga 2024, guna membangun model prediksi jenis ancaman siber untuk mendukung sistem deteksi dini serta memperkuat strategi mitigasi risiko di berbagai sektor strategis. Dengan prediksi yang akurat, organisasi dapat lebih cepat mengambil langkah antisipasi terhadap serangan siber, sehingga mengurangi potensi kerugian yang timbul [16]. Penggunaan algoritma machine learning, khususnya *Random Forest*, dipilih karena kemampuannya mengenali pola serangan dari data historis yang besar dan kompleks, memberikan akurasi tinggi, serta stabil terhadap variasi data dan masalah *overfitting* [17]. *Machine learning* juga mendukung otomatisasi proses analisis sehingga lebih efisien dibandingkan metode manual atau konvensional. Penelitian ini dilakukan untuk menjawab kekurangan penelitian sebelumnya yang pada umumnya hanya fokus pada satu jenis serangan atau skenario tertentu, sehingga kurang mampu memberikan gambaran menyeluruh tentang pola serangan siber secara global [18]. Selain itu, penelitian terdahulu juga cenderung belum memanfaatkan visualisasi data secara mendalam untuk membantu interpretasi pola serangan, dan sebagian besar model yang dikembangkan belum optimal saat diuji pada data yang tidak seimbang [19]. Oleh karena itu, penelitian ini berusaha menutup gap tersebut dengan membangun model prediktif yang lebih komprehensif dan dilengkapi analisis serta visualisasi yang mendukung sistem pendeteksian dini ancaman siber.

Untuk mencapai tujuan penelitian ini, digunakan metode *supervised learning*, yaitu pendekatan pembelajaran mesin yang mengandalkan data berlabel untuk melatih model prediktif [20]. Algoritma yang dipilih adalah *Random Forest* karena kemampuannya dalam menangani klasifikasi dengan banyak fitur dan kestabilannya dalam menghadapi data yang tidak seimbang [21]. Melalui pendekatan ini, diharapkan dapat dibangun model yang mampu memprediksi jenis serangan berdasarkan pola historis, yang pada akhirnya dapat digunakan untuk mendukung sistem deteksi dini dan strategi mitigasi risiko [22].

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *supervised learning* berbasis algoritma *Random Forest*. Tujuan utama dari metode ini adalah untuk membangun model klasifikasi yang mampu memprediksi jenis serangan siber berdasarkan data historis. Langkah-langkah penelitian dilakukan secara sistematis mulai dari pengumpulan data, *preprocessing*, pelatihan model, hingga evaluasi kinerja model.

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari *dataset* 'Global Cybersecurity Threats 2015–2024' yang tersedia di *Kaggle*. *Dataset* ini mencakup atribut seperti tahun kejadian, negara asal serangan, sektor terdampak, estimasi kerugian, dan jenis serangan. *Dataset* ini terdiri dari 1.500 entri data dengan 7 atribut yang digunakan, yaitu sebagai berikut:

Tabel 1. *Dataset*

Atribut	Deskripsi
Year	Tahun terjadinya serangan siber
Country	Negara tempat serangan berlangsung
Sector	Sektor terdampak (misalnya pemerintahan, kesehatan, keuangan)
Attack Type	Jenis serangan siber (Ransomware, Phishing, DDoS, SQL Injection, Malware)
Impact	Dampak atau tingkat keparahan dari serangan
Loss Estimate	Estimasi nilai kerugian finansial dari serangan
Additional Notes	Informasi atau rincian lainnya terkait dengan serangan

Sumber: https://www.kaggle.com/datasets/atharvasoundankar/global-cybersecurity-threats-2015-2024

2.2. *Preprocessing Data*

Tahapan ini mencakup pembersihan data (menghapus missing values), transformasi data kategorikal menjadi numerik dengan teknik encoding, dan normalisasi jika diperlukan. Data kemudian dibagi menjadi dua bagian, yaitu data latih (training set) sebesar 70% dan data uji (testing set) sebesar 30%. Pada tahap preprocessing, fitur-fitur yang digunakan dari *dataset* Global Cybersecurity Threats 2015–2024 terdiri dari: Year, Country, Sector, Attack Type, dan Loss Estimate. Selama pemeriksaan awal dataset, ditemukan bahwa terdapat nilai-nilai yang hilang (missing values) terutama pada atribut 'Loss Estimate' dan 'Additional Notes'. Nilai-nilai yang hilang ini diisi dengan nilai rata-rata (imputation) untuk atribut numerik (Loss Estimate), sedangkan nilai-nilai yang hilang untuk atribut kategorikal diisi dengan nilai khusus 'Unknown'. Setelah itu, atribut kategorikal (Country, Sector, Attack Type) dikonversi ke bentuk numerik dengan metode One-Hot Encoding agar dapat digunakan oleh algoritma Random Forest yang membutuhkan nilai numerik sebagai input.

2.3. Pelatihan Model

Pada tahap ini, model prediksi jenis ancaman siber dikembangkan dengan memanfaatkan algoritma Random Forest. Sebanyak 70% dari total *dataset* (±1.500 data) digunakan sebagai data pelatihan untuk memungkinkan model mempelajari pola dari atribut-atribut yang tersedia, seperti tahun kejadian, negara tempat serangan, sektor terdampak, jenis serangan, dan estimasi nilai kerugian. Hasil dari visualisasi *feature importance* menunjukkan bahwa atribut *Loss Estimate* dan *Year* memiliki pengaruh paling signifikan dalam memprediksi jenis serangan siber, dengan nilai masing-masing sekitar 0,5 dan 0,25. Atribut lainnya, seperti sektor dan negara, juga berkontribusi tetapi dengan tingkat yang lebih kecil (sekitar 0,03–0,04). Hal ini mengindikasikan bahwa nilai kerugian finansial dan tahun kejadian menjadi faktor kunci dalam pola klasifikasi ancaman siber**.**

2.4. EvaluasiModel

Kinerja model dievaluasi menggunakan metrik seperti *accuracy, precision, recall*, dan *F1-score*. Confusion matrix digunakan untuk mengukur tingkat keberhasilan klasifikasi. Evaluasi ini bertujuan untuk mengetahui seberapa baik model dapat memprediksi jenis serangan siber. Evaluasi model dilakukan dengan menggunakan 30% data uji dari total dataset. Beberapa metrik evaluasi yang digunakan adalah. *Accuracy***,** mengukur persentase prediksi yang benar dibandingkan seluruh prediksi. *Precision*, mengukur proporsi prediksi positif yang benar-benar positif. *Recall***,** mengukur seberapa baik model menemukan semua data positif. *F1-score***,** merupakan harmonisasi antara *precision* dan *recall*.

2.5. Visualisasi

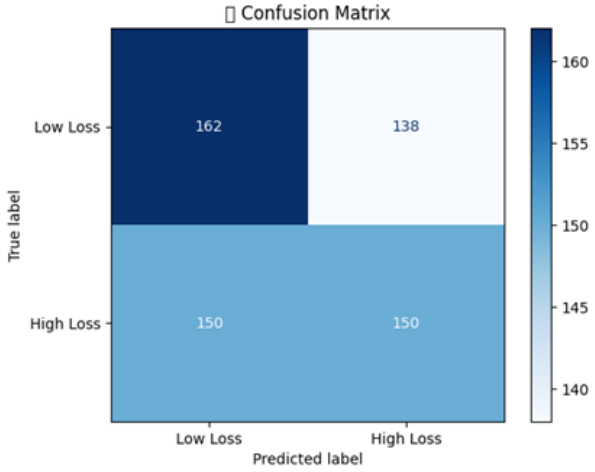
Untuk mendukung pemahaman hasil, dilakukan visualisasi seperti grafik distribusi serangan, *feature importance* dari *Random Forest*, dan hasil prediksi dalam bentuk *confusion matrix*. Hasil visualisasi data ini memberikan gambaran yang jelas terkait pola dan performa model dalam memprediksi jenis serangan

siber. Diagram pertama menunjukkan bahwa dari berbagai jenis serangan, Phishing dan Malware merupakan yang paling dominan dengan jumlah kasus tertinggi dibandingkan jenis lainnya. Diagram kedua memperlihatkan bahwa atribut Loss Estimate dan Year memiliki tingkat kontribusi terbesar dalam klasifikasi dengan Random Forest, masing-masing sebesar ±48% dan ±25%, sedangkan atribut lain hanya memberikan pengaruh kecil.

3. HASIL DAN PEMBAHASAN

3.1. Confusion Matrix

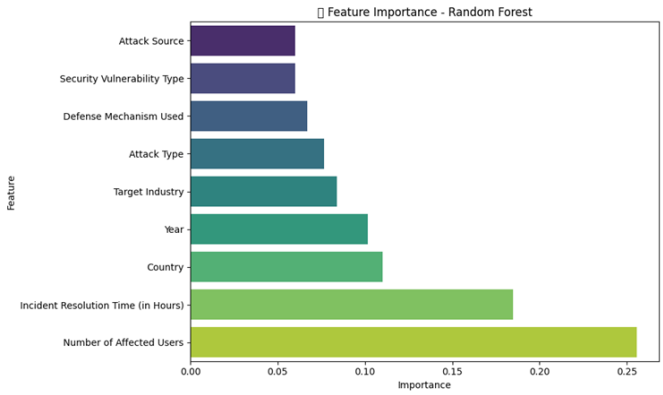
Evaluasi awal performa model klasifikasi dilakukan dengan meninjau Confusion Matrix (Gambar 1). Matriks ini menginformasikan jumlah prediksi benar dan salah untuk kedua kelas target, yaitu "Low Loss" dan "High Loss". Model *Random Forest* berhasil mengklasifikasikan 162 dari 300 kasus "Low Loss" dengan benar dan 150 dari 300 kasus "High Loss" dengan benar. Namun, terdapat 138 False Positive dan 150 False Negative, menunjukkan bahwa model masih kesulitan membedakan insiden kerugian tinggi dan rendah secara konsisten. Tingginya tingkat kesalahan ini dapat disebabkan oleh fitur yang kurang informatif, distribusi kelas yang seimbang, maupun pola non-linier yang belum terpetakan dengan baik oleh model.



Gambar 1. Confusion Matrix Klasifikasi Tingkat Kerugian Serangan Siber

3.2. Feature Importance

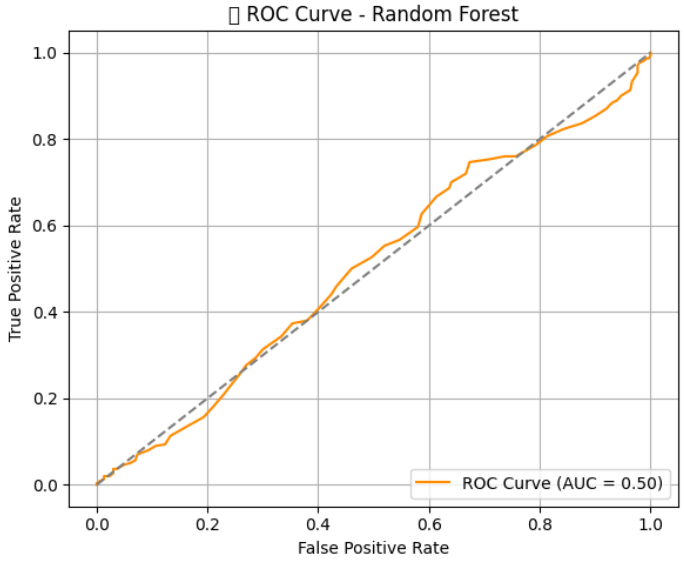
Analisis selanjutnya menyoroti kontribusi masing-masing fitur terhadap akurasi model melalui visualisasi *Feature Importance* (Gambar 2). Hasil menunjukkan bahwa *Number of Affected Users* dan *Incident Resolution Time (in Hours)* memiliki pengaruh paling besar terhadap prediksi, sedangkan fitur seperti *Attack Source* dan *Security Vulnerability Type* memberikan kontribusi relatif kecil. Informasi ini penting untuk proses *feature selection* di iterasi berikutnya, di mana fitur dengan kepentingan rendah dapat dipertimbangkan untuk dihapus atau digabungkan dengan fitur lain yang lebih relevan.



Gambar 2. Visualisasi *Feature Importance* dari model *Random Forest*

3.3. ROC Curve

Untuk mengukur kemampuan diskriminatif model, digunakan kurva *ROC (Receiver Operating Characteristic)* yang menggambarkan hubungan antara *True Positive Rate (TPR)* dan *False Positive Rate (FPR)* pada berbagai *threshold* klasifikasi. Berdasarkan Gambar 3, diperoleh nilai AUC sebesar 0,50, yang setara dengan tebakan acak. Hal ini menegaskan bahwa meski model telah memanfaatkan fitur-fitur penting, performa keseluruhan masih sangat lemah dan memerlukan perbaikan lebih lanjut melalui *hyperparameter tuning*, peningkatan kualitas data, atau penerapan teknik ensemble tambahan.



Gambar 3. ROC Curve model Random Forest dengan nilai AUC = 0,50.

3.4 Implikasi dan Langkah Selanjutnya

Berdasarkan hasil evaluasi di atas, beberapa strategi perbaikan dapat diusulkan:

1. *Hyperparameter Tuning*: Melakukan *grid search* atau *random search* untuk menemukan kombinasi parameter yang optimal.
2. *Feature Engineering*: Mencoba teknik transformasi atau pembuatan fitur baru untuk menangkap pola non-linier.
3. Penambahan Data: Mengumpulkan lebih banyak sampel atau mengatasi ketidakseimbangan kelas dengan teknik *sampling*.
4. *Ensemble Model*: Menggabungkan beberapa algoritma atau iterasi model untuk meningkatkan kestabilan dan akurasi.

Implementasi strategi-strategi ini diharapkan dapat meningkatkan kemampuan model dalam mengklasifikasikan insiden kerugian tinggi dan rendah dengan lebih robust dan konsisten.

4. KESIMPULAN

Hasil analisis terhadap performa model *Random Forest* dalam mengklasifikasikan tingkat kerugian akibat serangan siber menunjukkan bahwa model belum mampu memberikan hasil yang memuaskan. Meskipun terdapat prediksi yang benar pada masing-masing kelas, jumlah kesalahan klasifikasi masih cukup signifikan. Hal ini mencerminkan bahwa model mengalami kesulitan dalam membedakan antara kategori kerugian tinggi dan rendah secara akurat.

Evaluasi menggunakan kurva ROC memperkuat kelemahan model, di mana bentuk kurva yang mendekati garis diagonal mencerminkan kemampuan diskriminatif yang sangat rendah. Artinya, model tidak mampu secara konsisten membedakan antara dua kelas target dan cenderung melakukan prediksi secara acak.

Selain itu, distribusi probabilitas hasil prediksi memperlihatkan sebaran yang saling tumpang tindih antara dua kategori kerugian. Hal ini menunjukkan bahwa model belum memiliki tingkat kepercayaan yang cukup dalam proses klasifikasi, karena nilai probabilitas tersebar merata tanpa kecenderungan yang jelas terhadap salah satu kelas.

Secara keseluruhan, model *Random Forest* dalam penelitian ini belum dapat diandalkan untuk klasifikasi tingkat kerugian serangan siber. Oleh karena itu, diperlukan pengembangan lebih lanjut, seperti pemilihan fitur yang lebih relevan, penyeimbangan data antar kelas, penyetelan parameter model secara optimal, serta eksplorasi terhadap algoritma pembelajaran mesin lainnya yang lebih adaptif terhadap karakteristik data.

DAFTAR PUSTAKA

[1] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[2] G. Nguyen *et al.*, "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 77–124, 2019, doi: 10.1007/s10462-018-09679-z.

[3] N. Al-lQubaydhi *et al.*, "Deep learning for unmanned aerial vehicles detection: A review," *Computer Science Review*, vol. 51, p. 100614, Feb. 2024, doi: 10.1016/j.cosrev.2023.100614.

[4] K. Sivaraman, R. M. V. Krishnan, B. Sundarraj, and S. Sri Gowthem, "Network failure detection and diagnosis by analyzing syslog and SNS data: Applying big data analysis to network operations," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9 Special Issue 3, pp. 883–887, 2019, doi: 10.35940/ijitee.I3187.0789S319.

[5] J. Chen, J. Zhu, B. Feng, S. Xie, H. Yang, and F. Nie, "A novel method for optimizing spectral rotation embedding K-means with coordinate descent," *Information Sciences*, vol. 612, pp. 1095–1110, Oct. 2022, doi: 10.1016/j.ins.2022.09.011.

[6] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE Access*, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[7] R. Ameri, C.-C. Hsu, and S. S. Band, "A systematic review of deep learning approaches for surface defect detection in industrial applications," *Engineering Applications of Artificial Intelligence*, vol. 130, p. 107717, Apr. 2024, doi: 10.1016/j.engappai.2023.107717.

[8] S. Kumar and M. Singh, "Big data analytics for healthcare industry: Impact, applications, and tools," *Big Data Mining and Analytics*, vol. 2, no. 1, pp. 48–57, 2019, doi: 10.26599/BDMA.2018.9020031.

[9] L. M. Ang, K. P. Seng, G. K. Ijemaru, and A. M. Zungeru, "Deployment of IoV for Smart Cities: Applications, Architecture, and Challenges," *IEEE Access*, vol. 7, pp. 6473–6492, 2019, doi: 10.1109/ACCESS.2018.2887076.

[10] B. P. L. Lau *et al.*, "A survey of data fusion in smart city applications," *Information Fusion*, vol. 52, no. January, pp. 357–374, 2019, doi: 10.1016/j.inffus.2019.05.004.

[11] Y. Wu *et al.*, "Large scale incremental learning," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 374–382, 2019, doi: 10.1109/CVPR.2019.00046.

[12] A. Mosavi, S. Shamshirband, E. Salwana, K. wing Chau, and J. H. M. Tah, "Prediction of multi-inputs bubble column reactor using a novel hybrid model of computational fluid dynamics and machine learning," *Engineering Applications of Computational Fluid Mechanics*, vol. 13, no. 1, pp. 482–492, 2019, doi: 10.1080/19942060.2019.1613448.

[13] V. Palanisamy and R. Thirunavukarasu, "Implications of big data analytics in developing healthcare frameworks – A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 31, no. 4, pp. 415–425, 2019, doi: 10.1016/j.jksuci.2017.12.007.

[14] J. Sadowski, "When data is capital: Datafication, accumulation, and extraction," *Big Data and Society*, vol. 6, no. 1, pp. 1–12, 2019, doi: 10.1177/2053951718820549.

[15] J. R. Saura, B. R. Herraez, and A. Reyes-Menendez, "Comparing a traditional approach for financial brand communication analysis with a big data analytics technique," *IEEE Access*, vol. 7, pp. 37100–37108, 2019, doi: 10.1109/ACCESS.2019.2905301.

[16] X. Deng *et al.*, "Estimation of photosynthetic parameters from hyperspectral images using optimal deep learning architecture," *Computers and Electronics in Agriculture*, vol. 216, p. 108540, Jan. 2024, doi: 10.1016/j.compag.2023.108540.

[17] H. Aizenstein, R. C. Moore, I. Vahia, and A. Ciarleglio, "Deep Learning and Geriatric Mental Health," *The American Journal of Geriatric Psychiatry*, Dec. 2023, doi: 10.1016/j.jagp.2023.11.008.

[18] J. L. Ávila-Jiménez, V. Cantón-Habas, M. del P. Carrera-González, M. Rich-Ruiz, and S. Ventura, "A deep learning model for Alzheimer's disease diagnosis based on patient clinical records,"

Computers in Biology and Medicine, vol. 169, p. 107814, Feb. 2024, doi: 10.1016/j.compbiomed.2023.107814.

[19] Y. Yu, M. Li, L. Liu, Y. Li, and J. Wang, "Clinical big data and deep learning: Applications, challenges, and future outlooks," *Big Data Mining and Analytics*, vol. 2, no. 4, pp. 288–305, 2019, doi: 10.26599/BDMA.2019.9020007.

[20] K. Bayoudh, "A survey of multimodal hybrid deep learning for computer vision: Architectures, applications, trends, and challenges," *Information Fusion*, p. 102217, Dec. 2023, doi: 10.1016/j.inffus.2023.102217.

[21] A. Abernathy and M. E. Celebi, "The incremental online k-means clustering algorithm and its application to color quantization," *Expert Systems with Applications*, vol. 207, p. 117927, Nov. 2022, doi: 10.1016/j.eswa.2022.117927.

[22] B. Jo and S.-J. Lee, "A study on the positioning of fine scintillation pixels in a positron emission tomography detector through deep learning of simulation data," *Nuclear Engineering and Technology*, Dec. 2023, doi: 10.1016/j.net.2023.12.028.

[23] S.-J. Byun, S. Cho, and D.-H. Kim, "Can a machine learn from behavioral biases? Evidence from stock return predictability of deep learning models," *Journal of Behavioral and Experimental Finance*, vol. 41, p. 100881, Mar. 2024, doi: 10.1016/j.jbef.2023.100881.

[24] A. Bouteska, M. Z. Abedin, P. Hajek, and K. Yuan, "Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods," *International Review of Financial Analysis*, vol. 92, p. 103055, Mar. 2024, doi: 10.1016/j.irfa.2023.103055.

[25] X. Deng *et al.*, "Estimation of photosynthetic parameters from hyperspectral images using optimal deep learning architecture," *Computers and Electronics in Agriculture*, vol. 216, p. 108540, Jan. 2024, doi: 10.1016/j.compag.2023.108540.