

JEIT

مجلة التقنيات الناشئة والمبتكرة

مجلة علمية محكّمة مفتوحة الوصول تُعنى بتطوير البحث والابتكار في الهندسة وعلوم الحاسب والتقنيات التطبيقية للصناعة 4.0 و5.0

المجلد 1، العدد 1

نوفمبر 2026

ردمد الإلكتروني	رقم الإيداع الإلكتروني	دورية الصدور	الموقع الإلكتروني	الناشر
3152-8651	1447/16829	ربع سنوية	journals.qu.edu.sa/jeit	جامعة القصيم، المملكة العربية السعودية



Inaugural Issue

JEIT

Journal of Emerging and Innovative Technologies

A peer-reviewed, open-access journal advancing research and innovation in engineering, computer science and applied technologies for Industry 4.0 and 5.0

Volume 1, Issue 1

November 2026

E-ISSN 3152-8651

Legal Deposit No. 1447/16829

Frequency Quarterly

Website journals.qu.edu.sa/jeit

Publisher Qassim University, Saudi Arabia

Journal of Emerging and Innovative Technologies

Volume 1, Issue 1
November 2026

About the Journal

The Journal of Emerging and Innovative Technologies (JEIT) is a peer-reviewed, open-access journal published by Qassim University. It is dedicated to advancing research and innovation in engineering, computer science and applied technologies that drive the transformation toward Industry 4.0 and 5.0, and aims to be a leading international platform that merges solid scientific research with applied innovation and bridges the gap between academia and industry. The journal encourages empirical, theoretical and applied studies.

Aims and Scope

JEIT publishes original research articles, review articles and surveys, technical notes and short communications, case studies, conceptual and theoretical papers, and letters to the editor in areas including:

- Artificial Intelligence and Machine Learning
- Internet of Things (IoT) and Cyber-Physical Systems
- Robotics and Automation
- Smart Manufacturing and Digital Transformation
- Renewable and Sustainable Energy Systems
- Advanced Engineering Applications and Computing Technologies
- Cybersecurity and Blockchain Technology
- Smart Cities and Infrastructure
- AI Ethics and Responsible Technology
- Industry 4.0 and 5.0

Peer Review and Publication Ethics

All manuscripts are screened for similarity with iThenticate and then undergo double-blind peer review by at least two independent experts. The journal follows the guidance of COPE, ICMJE and WAME. Submissions from members of the editorial board are handled by independent editors and reviewed externally.

Open Access and Licensing

JEIT is a fully open-access journal and charges no article processing or submission fees. Articles are published under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), and authors retain copyright of their work.

Indexing and Archiving

Google Scholar, Crossref (DOI registration) and Portico.

Publication Details

E-ISSN	3152-8651
Legal Deposit No.	1447/16829 (electronic)
Publisher	Qassim University, Saudi Arabia
Frequency	Four issues per year
This issue	Volume 1, Issue 1, November 2026 (published 05 November 2026)
Editor-in-Chief	Prof. Abdulaziz S. Alaboodi
Website	https://journals.qu.edu.sa/jeit/ · jeit@qu.edu.sa
Copyright	© 2026 Qassim University; articles © 2026 The Author(s).

Journal of Emerging and Innovative Technologies

Volume 1, Issue 1
November 2026

Contents

FRONT MATTER
Editorial Board v–vi
Journal of Emerging and Innovative Technologies

EDITORIAL
From the Chief Editor's Desk vii–viii
Abdulaziz S. Alaboodi

REVIEW ARTICLE
Artificial Intelligence and Machine Learning in Cyber Threat Detection for Modern Organizations: A Systematic Review of Challenges and Solutions 1–22
Abdul Wahid Khan, Bassam Saleh F Alghashim, Meshari Khalid N Alharbi, Tamim Sulaiman A Alghomiz, Faisal Fahad M Alshrahrani, Faisal Saud F Almsaad

REVIEW ARTICLE
Sodium-Ion Batteries, Direct Lithium Extraction, Bidirectional Grid Integration, and Advanced Nuclear Power: A Review of Their Transition to Commercial Deployment, 2020–2026 23–36
Faisal O. Mahroogi

RESEARCH ARTICLE
Vision-Based Self-Mapping for Autonomous Indoor Drones 37–49
Wurud Alrashidi, Jamaal Alotibi

RESEARCH ARTICLE
Evaluating Low-Light Image Enhancement for Object Detection 50–66
Abdulrahman Abdulaziz Almudaymigh, M. Shujah Islam Sameem

REVIEW ARTICLE
Digital Twin Technologies for Intelligent Monitoring, Predictive Maintenance, and Autonomous Decision Making in Industrial Cyber-Physical Systems: A Comprehensive Systematic Review 67–83
Idama Omokaro, Ekwemuka Jones, Afolabi Awodeyi, Omolegho Ibok, Obonyano Kingdom, Odesa Ogaga

RESEARCH ARTICLE
Machine Learning–Based Flow Stress Prediction and Processing Map Optimization of Al/SiCp Metal Matrix Composites 84–111
Hanafy M. Omar, Osama M. Irfan

Editorial Board

Editor-in-Chief

Prof. Abdulaziz S. Alaboodi
Department of Mechanical Engineering, College of Engineering, Qassim University, Saudi Arabia
alaboodi@qu.edu.sa

Editorial Board Members

Prof. Salman A. Al Qahtani
Computer Engineering
Saudi Arabia
alqahtani.s@gmail.com

Prof. Abdulrahman F. Al Motrif
Information Technology
King Saud University, Saudi Arabia
ALMOTRIF@ksu.edu.sa

Prof. Dina M. Hussein
Information Technology
Qassim University, Saudi Arabia
d.hussein@qu.edu.sa

Prof. Hany R. Ammar
Mechanical Engineering
Qassim University, Saudi Arabia
h.ammar@qu.edu.sa

Prof. Rehan Ullah Khan
Data Science & Machine Learning
Qassim University, Saudi Arabia
re.khan@qu.edu.sa

Dr. Ibrahim S. Al Saidan
Electrical Engineering
Qassim University, Saudi Arabia
ALSAIDAN@qu.edu.sa

Dr. Fawaz A. Al Harbi
Civil Engineering
Qassim University, Saudi Arabia
fawazalharbi@qu.edu.sa

Editorial Office

Publisher	Qassim University, Saudi Arabia
Editorial office email	jeit@qu.edu.sa
Website	https://journals.qu.edu.sa/jeit/
E-ISSN	3152-8651
Peer review	Double-blind, minimum of two independent reviewers
Licence	CC BY-NC-SA 4.0; no article processing charges

List of Reviewers

JEIT is indebted to its reviewers, who contribute toward improving the quality of the articles published in the journal.

Abdulaziz Alorf	Ali Qamar	Mohamed Arbi Khlifi
Rashid Nasimov	S. Sivasankaran	Khalil Khan
Awais Ahmed	Mohammed Almannaa	Waleed Shewakh
Muhammad Abdurrazaq	Wadood Wadood	Faisal Mahroogi
Irfan Khan	Fahad Almufadi	Mohamed Ibrahiem
Issah Ibrahim	Waleed Zeiada	Arshad Jamal
Mahmoud Khalil	Salem Khidir	Iskander Tlili
Abdullah Alqasir	Dina Mahmoud	Khalid Alsnaie
Muhammad Alam	Muhammad Awais	Hany Elghaish
Liaqat Khan	Dina Ali	Asmaa Alshargabi
Omar Smadi	Sherif Shokry	Afef Selmi

مجلة التقنيات
الناشئة والمبتكرة

المجلد 1، العدد 1
نوفمبر 2026

افتتاحية العدد

من مكتب رئيس التحرير

أ.د. عبدالعزيز سليمان العبودي، رئيس التحرير
قسم الهندسة الميكانيكية، كلية الهندسة، جامعة القصيم، بريدة، المملكة العربية السعودية · alaboodi@qu.edu.sa

تاريخ النشر: 5 نوفمبر 2026 · معرّف الوثيقة الرقمي (DOI): https://doi.org/10.71422/jeit.0001

يسعدني أن أقدّم العدد الأول من مجلة التقنيات الناشئة والمبتكرة التي تصدرها جامعة القصيم. ويمثّل هذا العدد الافتتاحي انطلاقة مجلة أُنشئت لتجمع بين البحث العلمي الرصين والابتكار التطبيقي، ولتكون جسرًا يصل الأوساط الأكاديمية بالصناعة. ويضم العدد ستة بحوث تعكس اتساع نطاق المجلة، من الذكاء الاصطناعي والأمن السيبراني إلى أنظمة الطاقة والروبوتات والتوائم الرقمية والمواد المتقدمة.

ويشكّل الذكاء الاصطناعي خيطًا مشتركًا في هذا العدد. إذ تجمع مراجعة منهجية شملت 45 دراسة حول الذكاء الاصطناعي وتعلم الآلة في اكتشاف التهديدات السيبرانية 271 تحديًا مُبلَّغًا عنه و252 حلًا مقترحًا في سبع عشرة فئة متكررة. وتبرز عوائق رئيسة مثل اكتشاف هجمات اليوم الصفري، وقابلية التوسع، ومعدلات الإنذار الكاذب، وقيود الموارد، وتُبيّن المراجعة كيف تختلف هذه الأولويات بين آسيا وأوروبا وأمريكا الشمالية والشرق الأوسط وأفريقيا.

وتتناول دراستان الرؤية الحاسوبية في ظروف الإضاءة الصعبة. فتقترح دراسة أصيلة حول الطائرات المسيّرة ذاتية القيادة في البيئات الداخلية إطارًا هجينًا تكيفيًا لتحسين الصور، تختار فيه شجرة قرار بين تصحيحات خفيفة وشبكة عصبية التفافية قبل وصول الإطارات إلى ORB-SLAM3، مما يحسّن دقة التوطين دون تعديل مسار SLAM. وتقيّم دراسة مصاحبة تحسين الصور منخفضة الإضاءة الموضوع قبل كاشف الأجسام YOLO11n، وتُظهر أن التدريب المتطابق مع المجال يستعيد معظم الخسارة الظاهرة، في حين لا يحقق أي أسلوب تحسين مكسبًا مستقرًا. وتبيّن الدراسة أن الجودة الإدراكية للصورة لا تغني عن التقييم على مستوى المهمة.

وتحظى قضية تحول الطاقة بمراجعة جاءت في وقتها تتناول بطاريات أيونات الصوديوم، والاستخلاص المباشر لليثيوم، والتكامل ثنائي الاتجاه مع الشبكة، والطاقة النووية المتقدمة. ويرى المؤلف أن هذه التقنيات بلغت نقاط تحول تجارية بين عامي 2024 و2026 بفعل الطلب وحجم التصنيع والسياسات، لا بفعل اختراقات مخبرية جديدة، وأنه ينبغي فهمها بوصفها أجزاءً من محفظة طاقة واحدة مترابطة.

ويحضر مجال الصناعة 4.0 و5.0 من خلال مراجعة منهجية شاملة تناولت 87 منشورًا حول تقنيات التوائم الرقمية في الأنظمة السيبرانية الفيزيائية الصناعية. وترسم المراجعة دور هذه التقنيات في المراقبة الذكية والصيانة التنبؤية واتخاذ القرار الذاتي، والعوائق التي لا تزال قائمة في مجالات الأمن وقابلية التشغيل البيني وقابلية التوسع.

وأخيرًا، تقارن دراسة أصيلة من جامعة القصيم أداء ثمانية نماذج لتعلم الآلة في التنبؤ بإجهاد السريان أثناء التشكيل الساخن لمركب ذي مصفوفة معدنية من نوع Al/SiCp. ويحقق التعزيز التدريجي والنموذج التجميعي المكدَّس أدنى أخطاء الاستيفاء، وتحمل الخصائص المبنية على معامل زينر-هولومون معظم الوزن التنبؤي، ويُستخدم نموذج بديل قائم على العملية الغاوسية لبناء خرائط معالجة تحدد نوافذ التشغيل الساخن الملائمة.

وتُظهر هذه البحوث مجتمعةً ما تمثله مجلة JEIT: بحث دقيق في منهجيته، صريح في بيان حدوده، ومتصل بمشكلات هندسية حقيقية. وأتوجه بالشكر إلى المؤلفين الذين وثقوا بمجلة جديدة فأودعوها أعمالهم، وإلى المحكّمين الذين بذلوا وقتهم وخبرتهم، وإلى أعضاء هيئة التحرير على التزامهم. كما أعرب عن خالص الامتنان لجامعة القصيم على دعمها الذي جعل إطلاق هذه المجلة ممكنًا. ونرحب بإسهامات الباحثين من أنحاء العالم في الأعداد القادمة.

قراءة ممتعة!

هذه مقالة مفتوحة الوصول تُوزَّع وفق شروط ترخيص المشاع الإبداعي CC BY-NC-SA 4.0 (نَسب المُصنَّف، غير تجاري، الترخيص بالمثل). © 2026 المؤلفون. نشرتها جامعة القصيم، المملكة العربية السعودية.

Journal of Emerging and Innovative Technologies

Volume 1, Issue 1
November 2026

Editorial

From the Chief Editor's Desk

Prof. Abdulaziz S. Alaboodi, Editor-in-Chief

Department of Mechanical Engineering, College of Engineering, Qassim University, Buraydah, Saudi Arabia · alaboodi@qu.edu.sa

Published: 05 November 2026 · DOI: https://doi.org/10.71422/jeit.0001

I am honored to present the first issue of the Journal of Emerging and Innovative Technologies, published by Qassim University. This inaugural issue marks the beginning of a journal created to bring rigorous scientific research and applied innovation together, and to serve as a bridge between academia and industry. It brings together six articles that reflect the breadth of our scope, from artificial intelligence and cybersecurity to energy systems, robotics, digital twins and advanced materials.

Artificial intelligence is a common thread in this issue. A systematic review of 45 studies on AI and machine learning for cyber threat detection consolidates 271 reported challenges and 252 proposed solutions into seventeen recurring categories. Zero-day attack detection, scalability, false alarm rates and resource constraints emerge as the leading obstacles, and the review shows how these priorities differ between Asia, Europe, North America and the Middle East and Africa.

Two contributions address computer vision under difficult lighting. An original study on autonomous indoor drones proposes an adaptive hybrid image enhancement framework, in which a decision tree selects between lightweight corrections and a convolutional neural network before frames reach ORB-SLAM3, improving localization accuracy without modifying the SLAM pipeline. A companion study evaluates low-light image enhancement placed in front of a YOLO11n object detector and shows that domain-matched training recovers most of the apparent loss, while no enhancement method delivers a stable gain. Perceptual image quality, it shows, is no substitute for task-level evaluation.

Energy transition is the subject of a timely review of sodium-ion batteries, direct lithium extraction, bidirectional grid integration and advanced nuclear power. The author argues that these technologies reached commercial inflection points between 2024 and 2026 because of demand, manufacturing scale and policy rather than new laboratory breakthroughs, and that they should be understood as parts of a single, interdependent energy portfolio.

Industry 4.0 and 5.0 are represented by a comprehensive systematic review of 87 publications on digital twin technologies in industrial cyber-physical systems. It maps their role in intelligent monitoring, predictive maintenance and autonomous decision-making and the barriers that remain in security, interoperability and scalability.

Finally, an original study from Qassim University benchmarks eight machine learning models for predicting the hot-deformation flow stress of an Al/SiCp metal matrix composite. Gradient boosting and a stacked ensemble achieve the lowest interpolation errors, features based on the Zener–Hollomon parameter carry most of the predictive weight, and a Gaussian-process surrogate is used to construct processing maps that identify favorable hot-working windows.

Together, these papers show what JEIT stands for: research that is methodologically careful, open about its limitations and connected to real engineering problems. I thank the authors who trusted a new journal with their work, the reviewers who gave their time and expertise, and the members of the editorial board for their commitment. I also gratefully acknowledge the support of Qassim University, which made the launch of this journal possible. We warmly invite researchers worldwide to contribute to the coming issues.

Happy reading!

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License. © 2026 The Author(s). Published by Qassim University, Saudi Arabia.

مقالة مراجعة

الذكاء الاصطناعي وتعلم الآلة في اكتشاف التهديدات السيبرانية في المؤسسات الحديثة: مراجعة منهجية للتحديات والحلول

عبد الواحد خان[1,*]، بسام صالح ف. الغشيم[1]، مشاري خالد ن. الحربي[1]، تميم سليمان أ. الغميز[1]، فيصل فهد م. الشهراني[1]، فيصل سعود ف. المسعد[1]

[1] قسم علوم الحاسب، كلية الهندسة وتقنية المعلومات، كليات عنيزة، القصيم، المملكة العربية السعودية

*المؤلف المراسل: wahidkn@gmail.com

الملخص

أصبحت القدرة على التكيف مع التهديدات باستخدام تقنيات الذكاء الاصطناعي وتعلم الآلة عاملًا رئيسًا في مكافحة التهديدات السيبرانية، إذ تقدم نهجًا تكيفيًا قائمًا على البيانات لاكتشاف التهديدات يتجاوز الأساليب المعتمدة على البصمات. غير أن التنامي المتسارع في البحث والتطوير في مجال كشف التسلل القائم على الذكاء الاصطناعي وتعلم الآلة صاحبه اضطراب في فهم التحديات التي تواجهها المؤسسات والحلول المتاحة لها لمعالجة تلك التحديات. تستند هذه الورقة إلى مراجعة منهجية للأدبيات (SLR) أُجريت وفق بروتوكول Kitchenham للمراجعة المنهجية واستراتيجية البحث PICOC، وإلى تحليل 45 دراسة محكّمة ومنشورة في تقارير صناعية تتعلق بتطبيق الذكاء الاصطناعي وتعلم الآلة في اكتشاف التهديدات السيبرانية. وإجمالًا، رصدت الدراسات المشمولة 271 تحديًا مختلفًا و252 حلًا، صُنّفت في سبع عشرة فئة متكررة من التحديات. وكانت أبرز التحديات أمام التبني الناجح: اكتشاف هجمات اليوم الصفري والهجمات الجديدة (36)، وقابلية التوسع والبيانات الضخمة (34)، وارتفاع معدلات الإيجابيات والسلبيات الكاذبة (33)، والقيود الحاسوبية وقيود الموارد (32)، وعدم توازن مجموعات البيانات (31)، تليها مجموعات البيانات المتقادمة، والخصوصية والامتثال التنظيمي، وقابلية التفسير، وفضاءات الخصائص عالية الأبعاد. كما كشفت المقارنة على مستوى القارات أن قابلية التوسع والاكتشاف القائم على إنترنت الأشياء (IoT) يمثلان التحدي الرئيس في آسيا، وأن الخصوصية وقابلية التفسير تمثلان التحدي الأول في أوروبا، وأن المتانة في مواجهة الهجمات العدائية وأمن المؤسسات يمثلان التحدي الرئيس في أمريكا الشمالية، وأن البنية التحتية الأساسية ومحدودية القدرات تمثلان التحدي الرئيس في الشرق الأوسط وأفريقيا. وتشير النتائج إلى توجه نحو مقاربات أكثر هجنة وتكيفًا ومراعاة للخصوصية في أمن الذكاء الاصطناعي، تدمج تقنيات التعلم التجميعي والتعلم العميق، والتعلم الاتحادي، والذكاء الاصطناعي القابل للتفسير، ومسارات التدريب المستمر. ويسلط هذا العمل البحثي الضوء على حدة التحديات التي ستواجهها المؤسسات مستقبلًا في توظيف الذكاء الاصطناعي وتعلم الآلة لكشف التسلل في الأمن السيبراني.

الكلمات المفتاحية: الذكاء الاصطناعي؛ تعلم الآلة؛ المراجعة المنهجية للأدبيات؛ تحديات الأمن السيبراني؛ الحلول؛ المؤسسات الحديثة.

جامعة القصيم
Qassim University

JEIT

مجلة التقنيات الناشئة والمبتكرة

المجلد 1، العدد 1
نوفمبر 2026

تاريخ الاستلام
12 سبتمبر 2026

تاريخ القبول
29 أكتوبر 2026

تاريخ النشر
5 نوفمبر 2026

معرّف الوثيقة الرقمي (DOI)
https://doi.org/10.71422/jeit.0002

الترخيص
CC BY-NC-SA 4.0

للاستشهاد بهذا البحث

عبد الواحد خان، بسام صالح ف. الغشيم، مشاري خالد ن. الحربي، تميم سليمان أ. الغميز، فيصل فهد م. الشهراني، فيصل سعود ف. المسعد. الذكاء الاصطناعي وتعلم الآلة في اكتشاف التهديدات السيبرانية في المؤسسات الحديثة: مراجعة منهجية للتحديات والحلول. مجلة التقنيات الناشئة والمبتكرة. 2026;1(1):1-22.

الموقع الإلكتروني للمجلة

© 2026 المؤلفون. نشرته جامعة القصيم بموجب ترخيص CC BY-NC-SA 4.0.

Qassim University جامعة القصيم

JEIT

Journal of Emerging and Innovative Technologies

Volume 1, Issue 1
November 2026

Received
12 September 2026

Accepted
29 October 2026

Published
05 November 2026

DOI
https://doi.org/10.71422/jeit.0002

License
CC BY-NC-SA 4.0

Cite this article

Khan AW, Alghashim BSF, Alharbi MKN, Alghomiz TSA, Alshrahrani FFM, Almsaad FSF. Artificial Intelligence and Machine Learning in Cyber Threat Detection for Modern Organizations: A Systematic Review of Challenges and Solutions. J Emerg Innov Technol. 2026;1(1):1–22.

Journal website

Review article

Artificial Intelligence and Machine Learning in Cyber Threat Detection for Modern Organizations: A Systematic Review of Challenges and Solutions

Abdul Wahid Khan[1,*], Bassam Saleh F Alghashim[1], Meshari Khalid N Alharbi[1], Tamim Sulaiman A Alghomiz[1], Faisal Fahad M Alshrahrani[1], Faisal Saud F Almsaad[1]

[1] Department of Computer Science, College of Engineering and Information Technology, Onaizah Colleges, Qassim, Saudi Arabia

*Corresponding author: wahidkn@gmail.com

Abstract

The ability to adapt to threats using AI and ML techniques has become a major factor in the fight against cyber threats, offering a data-driven, adaptive approach to threat detection that transcends signature-based methods. With the growing surge of research and development on AI/ML-based intrusion detection, however, there has been disarray in the perception of the challenges which organizations are facing and the solutions available to them to tackle the challenges. This paper is based on a systematic literature review (SLR) carried out following the Kitchenham SLR protocol and a PICOC search strategy, and the analysis of 45 peer-reviewed and industry-reported studies related to the application of AI/ML for cyber threat detection. In total, the corpus identified 271 different challenges and 252 solutions, which were grouped into seventeen recurring categories of challenges. The biggest challenges to successful adoption were zero day and novel attack detection (36), scalability and big data (34), high false positive/negative rates (33), computational and resource constraints (32), and dataset imbalance (31), followed by outdated datasets, privacy and regulatory compliance, explainability, and high dimensional feature spaces. The comparison on the continental level also identified the scalability and IoT-based detection as the main challenge in Asia, privacy and explainability as the primary challenge in Europe, adversarial robustness and enterprise security as the main challenge in North America and foundational infrastructure and capacity limitations as the main challenge in the Middle East/Africa. The findings indicate a trajectory towards more hybrid, adaptive, and privacy-conscious approaches to AI security that integrate ensemble and deep learning techniques, federated learning, explainable AI, and ongoing training pipelines. This research work highlights the severity of challenges faced by the future organisations in implementation of AI/ML for cybersecurity intrusion detection.

Keywords: Artificial Intelligence; Machine Learning; Systematic Literature Review; Cybersecurity Challenges; Solutions; Modern Organisations.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

1. Introduction

It is observed that cyber-attack chance for modern organisations going to expand due to vast interconnected system, cloud-based framework and huge number of data. The older approach of cybersecurity is highly ineffective against the emerging method like zero day attacks and advanced persistent threats [1, 2]. The cybersecurity needs the fresh approach of AI and ML concepts. The latest system needs to learn and adapt to new threat manners in real-time using AI and ML techniques [2]. These approaches are using constantly for various threats detection and security technologies but it also craft new hurdles like adversarial manipulation, algorithmic bias and low transparency [3, 4].

The confidentiality, integrity and availability of data are compromised due to internet of things (IoT), Industrial control systems (ICS) and digital twins that are interconnected. With the recent developments in deep learning, anomaly detection and automatic threat identification have gained significant boost [2] [6] however, explainability and trustworthiness are still essential requirements for the deployment of deep learning technologies in the real world [5] [7]. Explainable Artificial Intelligence (XAI) is therefore proving to be a critical enabler in enhancing transparency and compliance in AI-powered cybersecurity solutions [7].

The Intrusion Detection System is still the cornerstone of defense strategies and in modern systems, it is distinguished between normal and malicious behaviour as either misuse-based or anomaly-based detection; anomaly-based detection is especially important for detecting zero-day attacks [1], [6], [8]. Although the problems of spatiotemporal feature extraction and data imbalance are solved by deep learning approaches in complex network environments [9], the reliability of real-world applications is still limited because of the ageing datasets and the high rate of false-positive [1]. Adversarial machine learning (AML) against AI-based systems in critical infrastructure remains a problem to spur the creation of standardized taxonomies and mitigation strategies [3-4]. IDS solutions, built with the help of machine learning (ML), still fail to be generalized for industrial control systems (ICS) [10].

The organisations need to dig out some of the latest hurdles in advance as the regime is changed with the entry of AI and ML. The new challenges are quality of data, biasness of algorithm, lack of model explainability, old system integration, regulatory and ethical restrictions for these issues. This study focused to understand these issues in a general way and proposed the solutions for these challenges as identified from the literature to improve cyber security capability [11-13].

The current statistics of industry are enough to understand the enormousness of this obstacle. There are 69% of risks for small medium enterprises (SMBs) as they faced cyberattack in the last year. Moreover, 80% of SMBs not practicing Privileged Access Management solutions and 60% of enterprises constantly using cybersecurity methods like credential managers, two steps of factor authentication and training of cybersecurity awareness [14]. This concern gone high due to shift to remote and hybrid works: business report shows that 72% concerned about security risk; cybersecurity professional report that 80% security threats boost sine 2020 [14]. The trends reflect how critical it is to grasp the practicality of threat detection implementation via AI and ML models and who should do it.

This is a systematic literature review of 45 studies on the use of AI/ML in cyber threat detection that builds upon a previous structured literature protocol to produce a complete evidence synthesis. Specific goals are: (i) Identification of the key challenges that are technical, organizational and ethical in relation to threat detection using AI/ML; (ii) Collecting the solutions and best practices presented in the literature to address the challenges; (iii) Ranking the challenges by frequency of occurrence in the corpus to

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

distinguish between critical challenges and less critical challenges; and (iv) Exploring the variation of challenge profiles and research emphases across geographic regions. The rest of this paper is organised as follows. The conceptual background is provided in Section 2. Research Methodology is discussed in section 3. Results are reported in Section 4 as data extracted was in tabular format. Findings and implications are discussed in section 5. The paper is completed in section 6, and the limitations of the paper are discussed in section 7.

2. Background

2.1 Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) and Machine Learning (ML) are at the core of modern information systems through enabling automated decision making, pattern recognition and intelligent data analysis. Artificial Intelligence (AI) is the ability of a computer system to replicate human intelligence such as reasoning, learning and problem-solving. Machine Learning is one of the major branches of AI which deals with the system using historical data and learning from it without being explicitly programmed. It is usually categorized into supervised learning, unsupervised learning and reinforcement learning. Supervised learning is done on labeled data to classify and predict, whereas unsupervised learning is used to discover hidden patterns and anomalies in unlabeled data. This enables systems to learn to optimize their action through interaction with changing environments using reinforcement learning.

2.2 Cybersecurity and Cyber Threats

In cybersecurity the main focus is to protect computer systems, digital appliances from intruders and cyberattack. Cybersecurity became a pillar due to huge dependencies of organisations on digital framework. In today, cybersecurity infrastructure we encapsulate preventive, detective and responsive methods to overcome on all types of cyber-attacks [15]. A cyber-attack is an attempt to exploit a weakness in technology, computers, or networks that are digital in nature with the goal of causing harm to the victim, whether that's in the form of malware, phishing, ransomware, insider threats, denial-of-service attacks, or advanced persistent threats. The attackers keep changing their techniques aided by automation, social engineering and zero day exploits. In this scenario, static signature based defense is not enough.

2.3 Cyber Threat Detection and Intrusion Detection Systems

Cyber threat detection is the identification of unusual activities and security breaches in information systems. IDSs are widely deployed to monitor the network traffic and the system behavior, and can be broadly divided into two types: network-based (NIDS) and host-based (HIDS). Traditional detection methods are based on signature-based and rule-based methods. These approaches work well for known attacks but lack the power to deal with new and emerging threats. For this, machine learning detection models have been suggested by many to increase the adaptability, accuracy and real-time analysis [16].

2.4 AI and ML in Cyber Threat Detection

The security operations for cyber threat in today age are incomplete without use of AI and ML practices. To monitor the various weakness and apprehensive activity is possible by using ML-based model that make it possible in real time by meting out massive amounts of network traffic, system logs and user behavior patterns. The automatic danger classification, anomaly scoring, and predictive risk assessment, minimizing false positives and improving security procedures center situational

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

awareness can assist by AI-powered detection system [17]. Although the advantages are substantial, the procedural and executive trials cover data disparity, limited accessibility of labeled data, and the quality of the data, which can hamper the preparation and recital of AI detection systems. Alternative trial is system limpidity, as many ML models are deliberated “black boxes”, leading to lower assurance of the users and making it tough to meet controlling requirements, and adversarial attacks like data poisoning and ducking can affect the learning process and conciliation system sturdiness.

2.5 Emerging Solution Directions

Some practices that give the solution of these challenges suggested in the current studies. Explainable Artificial Intelligence (XAI) advances limpidity by giving an interpretable awareness on model behavior to help analyst decision-making and controlling audit. It enables collaborative model training while maintaining privacy and adherence to regulations between organisations, through federated learning. Furthermore, hybrid learning models, data balancing methods and blockchain-based security architectures enhance security and data integrity [18]. All these emerging areas point toward a suite of lightweight, ethical, and real-time AI-driven security solutions that balance performance, explainability and resilience. In the rest of this paper these themes are examined in a systematic manner, as they are also presented in the tabular format presented in the data extraction part.

2.6 Related Work and Research Gap

Over the last 10 years, research on artificial intelligence and machine learning within the domain of cybersecurity has increased tremendously. Buczak and Guven [14] gave one of the first surveys on the use of data mining and ML techniques for intrusion detection. They talked about supervised and unsupervised methods of detecting anomalies in the network. Shone et al. in [16] tried to extend the above work by introducing a deep learning based intrusion detection system using stacked autoencoders with a higher classification accuracy than the traditional classifiers. Kim et al. [23] was interested in the idea of deep neural networks for the detection of complex attack patterns from massive network traffic. At the same time, several researchers have highlighted the shortcomings of this advancement, with Sommer and Paxson [15] stating that many models trained in the ML world fail to achieve good results in the real world because of unrealistic expectations regarding the quality and availability of the data, and Biggio and Roli [17] formalizing adversarial machine learning, where they showed how adversarial attacks can be based on manipulation of the training data and/or exploiting model weaknesses to evade detection. Recent research has been directed towards addressing the lack of transparency faced by ML based security systems, with Explainable AI (XAI) gaining significant focus. To encourage human trust and decision making in automated systems, Ribeiro et al. [18] suggested model-agnostic explanation techniques. Gartner and IBM Security industry reports also highlight the importance of pairing AI with human intelligence and organisational governance structures [19-22].

This collection of literature demonstrates the technical progress in AI/ML detection models, but there is limited literature that systematically brings together the organisational, ethical and operational issues and limitations. The majority of current research is oriented towards enhancing detection accuracy and explained by the lack of concern for explainability, regulations, workforce readiness or integration with existing security infrastructure. No single source that has gathered the frequency of challenges and solutions from a massive, recent and multi-regional collection of studies. This review aims to fill this gap by providing a framework for solutions and a systematic literature review protocol

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

with a thematic frequency analysis of 45 publications, including technical, organizational and governance approaches, further broken down by geographic region (Section 4.3).

3. Research Methodology

This study follows the systematic literature review (SLR) protocol proposed by Kitchenham [13], which structures an SLR into three phases: planning, conducting, and reporting, and ensures a transparent and reproducible research process. The protocol was developed with reference to established SLR practice in software and systems engineering [11], [12].

3.1 Research Objectives

The research methodology was designed to satisfy the following objectives:

- To identify the key technical, organizational, and ethical challenges in implementing AI and ML for threat detection in modern organizations.
- To analyze the most commonly proposed solutions and best practices reported in the literature.
- To evaluate the effectiveness of AI and ML approaches relative to traditional threat detection methods.
- To develop a structured, evidence-based framework for guiding organizations in adopting AI- and ML-based cybersecurity solutions.

3.2 Research Questions

RQ1. What are the primary challenges that modern organizations face when implementing AI- and ML-based threat detection systems?

RQ2. What technical and organizational solutions are most frequently recommended in the literature to address these challenges?

3.3 Search Strategy

Search strings were developed using the PICOC framework (Population, Intervention, Comparison, Outcome, Context), aligned with RQ1 and RQ2:

- Population (P): Modern organizations, enterprises, companies, IT environments, cloud and hybrid infrastructures.
- Intervention (I): Artificial Intelligence, Machine Learning, Deep Learning, AI-based security, ML-based threat detection.
- Comparison (C): Traditional detection systems, signature-based detection (optional).
- Outcome (O): Challenges, solutions, effectiveness, accuracy, scalability, explainability, adoption.
- Context (C): Cybersecurity, threat detection, intrusion detection systems, network security, cloud security.

3.3.1 Keyword Identification

Primary keywords and their synonyms were grouped into five categories to support systematic search-string construction:

- AI/ML terms: “Artificial Intelligence,” “Machine Learning,” “Deep Learning,” “AI-based,” “ML-based.”

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

- Threat detection terms: “threat detection,” “intrusion detection,” “anomaly detection,” “cyber attack detection.”
- Challenge terms: challenges, limitations, barriers, risks, issues, privacy, bias, adversarial.
- Solution terms: solutions, mitigation, best practices, strategies, framework, governance.
- Organizational context terms: organization*, enterprise*, company*, “modern organization*,” cloud, hybrid.

3.3.2 Boolean Search Strings

Three Boolean search strings were constructed: two aligned with the individual research questions and one combined query used for broader corpus retrieval.

Search String 1 — RQ1 (Challenges):

("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "AI-based" OR "ML-based") AND ("threat detection" OR "intrusion detection" OR "anomaly detection") AND (challenges OR limitations OR barriers OR risks OR issues OR adversarial OR privacy) AND (organization* OR enterprise* OR cloud OR hybrid)

Search String 2 — RQ2 (Solutions):

("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "AI-based" OR "ML-based") AND ("threat detection" OR "intrusion detection" OR "cyber attack detection") AND (solutions OR mitigation OR "best practices" OR strategies OR framework OR governance) AND (organization* OR enterprise* OR cloud OR hybrid)

Search String 3 — Combined (RQ1 + RQ2):

("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "AI-based" OR "ML-based") AND ("threat detection" OR "intrusion detection" OR "anomaly detection" OR "cyber attack detection") AND (challenges OR limitations OR barriers OR risks OR issues OR solutions OR mitigation OR "best practices" OR strategies OR framework OR governance) AND (organization* OR enterprise* OR cloud OR hybrid)

3.3.3 Database-Specific Adaptation

The search strings were adapted for database-specific syntax across four academic sources:

- IEEE Xplore: quotation marks applied, with search restricted to Abstract and Index Terms fields.
- Scopus / SpringerLink: TITLE-ABS-KEY search with wildcard (*) truncation.
- Google Scholar: simplified query forms used to accommodate query-length limits.

Example (Scopus format): TITLE-ABS-KEY(“Artificial Intelligence” OR “Machine Learning” OR “AI-based” AND “threat detection” OR “intrusion detection” AND challenges OR solutions AND organization).

3.3.4 Validation of Search Strings

A pilot search was conducted in IEEE Xplore and Google Scholar to verify that the search strings returned relevant and sufficient results. Strings were iteratively refined based on:

- Volume of results retrieved.
- Relevance of the top 20 articles returned.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

- Presence of peer-reviewed and high-impact sources.

3.4 Inclusion and Exclusion Criteria

Studies were encompassed if they:

IC-1: addressed the application of AI/ML in cybersecurity, explicitly threat detection;

IC-2: were peer-reviewed journal articles, conference papers, or credible industry/technical reports;

IC-3: were published in English;

IC-4: explicitly identified challenges and proposed solutions or best practices; and

IC-5: addressed organizations across varied geographical contexts.

Studies were excluded if they:

EC-1: did not directly address AI/ML applications in cybersecurity;

EC-2: were non-peer-reviewed sources such as blog posts or unrefereed white papers;

EC-3: duplicated findings already captured by another included study;

EC-4: discussed generic cybersecurity measures unrelated to AI/ML; or

EC-5: lacked sufficient detail on challenges and solutions.

3.5 Study Selection, Quality Assessment, and Data Extraction

After the title/abstract screening and full-text assessment of inclusion and exclusion criteria listed above, 45 studies were selected for the synthesis. Using a structured data extraction form, each study was systematically documented for: study identification (authors, year of publication), origin of the study (country/region), publication venue, challenges identified to address AI/ML in cybersecurity, proposed solutions/good practices, results and conclusions. Iterative thematic coding of extracted challenges and solutions led to a common set of challenges subsumed into the following categories: challenges were then tabulated by frequency of each challenge category across the 45 studies, providing a severity rank. Challenge categories with a frequency of occurrence between 22 and 36 (appeared in about 50% or more of the reviewed studies) were considered critical and categories with a lower frequency were considered less critical but still relevant. A continent-wise comparison was also conducted to compare the variation of challenges and solution emphases across the continents of Asia, Europe, North America and Middle East and Africa. The resulting synthesis is reproduced in the same table format as was used for extraction and analysis in Section 4.

4. Results

The 45 studies in this review reported a total of 271 problems and 252 solutions for the use of AI/ML in cyber threat detection. These extracted items were consolidated with a thematic approach to obtain seventeen recurring challenge categories. The consolidated challenge–solution mapping is reproduced in Table 1 in the same form as it appeared in the data extraction process: each row represents a consolidated challenge and its description, the consolidated solutions proposed in the literature, the identifiers of the papers that contributed to the consolidated challenges, and the count of how many studies the consolidated challenge/solution occurred in (Section 4.1, landscape orientation). As per the review protocol, those categories that are seen to occur between 22 and 36 times are considered as the

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

critical category; the other categories are considered less critical but relevant in specific organizations. Table 2 shows the challenge/solution profile of every individual paper and Table 3 shows the continent-wise comparison of the challenges, root causes and solution emphasis of the papers.

4.1 Consolidated Challenge–Solution Table (n = 45 papers)

Table 1. Consolidated challenges, descriptions, solutions, contributing paper IDs, and frequency

S. No	Consolidated Challenge	Description	Consolidated Solutions	Paper IDs	Freq (45)
1	**Zero-Day & Novel Attack Detection**	Difficulty in detecting previously unknown, polymorphic, stealthy, and emerging attacks because conventional signature-based/ML approaches mainly depend on known attack patterns.	The use of Hybrid IDS and Open Set Recognition can improve the detection of unknown and previously unseen attacks. The assumption of Deep Learning techniques such as CNN, LSTM, and DAE can advance the identification of complex, polymorphic, and stealthy attacks. The presentation of unsupervised anomaly detection can relief detects novel attacks without reliant entirely on formerly known attack signatures. The execution of adaptive retraining can assist the IDS to continuously adapt to evolving attack patterns and shrink the impact of zero-day attacks.	02, 04, 05, 07, 10, 12, 13, 14, 16, 17, 18, 19, 20, 21–32, 33, 34, 35, 36, 37, 40, 42, 43, 44, 45	**36**
2	**High False Positives / False Negatives**	Incorrect classification of legitimate activities as attacks or actual attacks as normal, resulting in alert fatigue, healthcare risks, and excessive workload for ICS operators.	The use of Collective ML techniques can advance classification reliability and shrink incorrect attack classifications. The application of hyper parameter modification can optimize model performance and help achieve a better balance between false positives and false negatives. The unification of context-aware detection can improve alert assessment by considering the operational context of detected activities. The implementation of human-in-the-loop validation can allow security experts to verify critical alerts and reduce the consequences of incorrect automated decisions.	02, 04, 05, 09, 10, 11, 12, 13, 16, 18, 19, 20, 21–32, 33, 34, 35, 41, 42, 45	**33**

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

S. No	Consolidated Challenge	Description	Consolidated Solutions	Paper IDs	Freq (45)
3	**Dataset Imbalance & Minority Attack Detection**	Rare User-to-Root (U2R), Remote-to-Local (R2L), insider-attack, and healthcare anomaly classes are poorly represented in datasets, causing ML models to favor majority classes.	The use of SMOTE and SMOTEENN can increase the representation of minority attack classes and improve their detection. The application of GAN-based data generation can provide additional representative samples for rare attack categories. The use of One-Class ML can support the detection of rare or insufficiently represented attack classes. The adoption of balanced sampling can reduce model bias toward majority attack classes and improve minority attack detection.	04, 05, 07, 09, 10, 11, 12, 14, 15, 18, 19, 21–32, 33, 34, 35, 36, 37, 40, 41, 43, 45	**31**
4	**Outdated / Unrealistic Datasets**	Heavy dependence on older datasets such as KDD99 and NSL-KDD, together with limited representation of realistic IoT/ICS environments, restricts reliability and generalizability of IDS models.	The adoption of modern datasets such as UNSW-NB15, CICIDS2017, and TON-IoT can reduce dependence on outdated datasets and provide more realistic attack scenarios. The implementation of systematic dataset cleansing and preprocessing can improve data quality and the reliability of experimental results. The use of cross-dataset validation can provide a more realistic assessment of model generalizability and reduce dependence on a single dataset.	02, 04, 05, 10, 11, 16, 19, 21–32, 34, 37, 40, 43, 44, 45	**27**
5	**Computational & Resource Constraints**	IoT, IoMT, and ICS devices often have limited computational resources, while sophisticated Deep Learning models may require substantial GPU/TPU resources.	The use of lightweight CNN1D models can reduce computational requirements for resource-constrained IoT, IoMT, and ICS environments. The adoption of Edge/Fog-based IDS architectures can move detection closer to the data source and reduce dependence on resource-intensive centralized processing. The application of model pruning can reduce unnecessary model parameters and computational overhead. The use of efficient feature selection can reduce the amount of data that needs to be processed while maintaining relevant detection	02, 04, 05, 07, 10, 13, 14, 16, 18, 19, 21–32, 33, 34, 36, 37, 40, 43, 45	**32**

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

S. No	Consolidated Challenge	Description	Consolidated Solutions	Paper IDs	Freq (45)
			information.		
6	**Scalability & Big Data Handling**	The large volume of data generated by Cloud, Fog, Edge, CI/CD, IoMT, and smart-healthcare environments creates challenges in processing and monitoring security events at scale.	The adoption of distributed IDS architectures can distribute security monitoring and processing across multiple computing resources. The integration of Cloud–Fog–Edge architectures can distribute workloads according to the capabilities of different computing layers. The use of Federated Learning can support collaborative model training across distributed environments without requiring all data to be centrally collected. The integration of SDN/NFV technologies can provide flexible and scalable security functions for large-scale environments.	02, 05, 07, 11, 13, 14, 16, 17, 18, 20, 21–32, 33, 35, 36, 37, 40, 42, 43, 45	**34**
7	**Explainability & Trust (Black-Box AI)**	The lack of transparency in AI/ML-based decisions creates difficulties in regulatory compliance, operator acceptance, and understanding why a model generates a particular security alert.	The use of Explainable AI (XAI) techniques such as SHAP and LIME can make AI-based IDS decisions easier for operators and stakeholders to understand. The establishment of transparent AI governance can improve trust and address regulatory and organizational concerns associated with black-box AI. The application of feature contribution analysis can help security operators understand which features contributed to specific detection decisions.	01, 02, 07, 11, 12, 13, 14, 19, 20, 21, 23, 28, 29, 30, 31, 34, 35, 38, 39, 41, 42, 43	**22**
8	**Adversarial ML Vulnerability**	Machine Learning-based IDS models may be manipulated through evasion, poisoning, or black-box attacks, potentially causing security models to produce incorrect results.	The use of adversarial training can improve model resistance against adversarial evasion attacks. The development of robust ML pipelines can reduce vulnerabilities throughout the machine-learning-based detection process. The implementation of a secure ML lifecycle can incorporate security controls throughout model development, deployment, monitoring, and updating.	01, 02, 03, 07, 09, 10, 11, 12, 14, 15, 17, 23, 28, 29, 30, 34, 37, 40, 42	**19**
9	**Feature Engineering & High**	Very large feature spaces, ranging from approximately 80–20,000 features, create	The application of PCA can reduce feature dimensionality while retaining important	04, 05, 07, 09, 10, 14, 16, 19, 22, 23, 24, 25, 26, 27, 33,	**22**

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

S. No	Consolidated Challenge	Description	Consolidated Solutions	Paper IDs	Freq (45)
	Dimensionality	significant complexity in IoMT and CI/CD environments and can increase processing requirements.	information. The use of Autoencoders can provide a more efficient representation of high-dimensional feature spaces. The application of Extra Trees and RFE can identify relevant features and remove unnecessary or redundant features. The use of correlation filtering can eliminate highly correlated features and reduce feature complexity.	34, 36, 37, 38, 41, 43, 45	
10	**Integration & Deployment Complexity**	Integrating ML-based security solutions with CI/CD pipelines, Operational Technology (OT), and legacy SIEM platforms can be technically complex.	The integration of ML models into CI/CD pipelines can streamline the deployment and management of ML-based security solutions. The adoption of modular API-based frameworks can facilitate integration with existing OT environments and legacy SIEM systems. The use of modular architectures can allow individual ML components to be modified or upgraded without requiring major changes to the complete security infrastructure.	11, 12, 13, 14, 18, 20, 21, 28, 29, 30, 31, 32, 33, 35, 41, 42	**16**
11	**Real-Time & Low-Latency Requirements**	IoMT and ICS environments require rapid detection and response, while edge devices have limited processing capabilities and communication delays can affect timely detection.	The adoption of Edge-based detection can process security events closer to the data source and reduce communication and processing delays. The use of Online Learning can enable models to learn from newly arriving data without relying exclusively on periodic offline retraining. The implementation of stream processing can support near-real-time analysis of continuously generated healthcare IoMT and ICS data.	07, 10, 13, 14, 18, 20, 23, 24, 27, 28, 30, 31, 32, 33, 35, 38, 41, 45	**18**
12	**Concept Drift & Model Adaptation**	Continuous changes in threat behavior and system characteristics can cause trained ML models to degrade over time and become less effective against new patterns.	The implementation of continuous retraining can keep detection models updated as threats and system behavior evolve. The use of Online Learning can enable models to adapt to newly emerging patterns in incoming data. The adoption of adaptive	12, 14, 17, 18, 20, 27, 29, 30, 31, 32, 33, 35, 42, 44, 45	**15**

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

S. No	Consolidated Challenge	Description	Consolidated Solutions	Paper IDs	Freq (45)
			frameworks can help address model degradation caused by changes in the threat environment.		
13	**Privacy, Legal & Ethical Compliance**	The use and sharing of sensitive IoMT and multi-facility data creates privacy, legal, regulatory, and ethical concerns, including requirements associated with the General Data Protection Regulation (GDPR).	The use of Federated Learning can support collaborative model training while reducing the need to centrally share sensitive data. The establishment of AI governance frameworks can provide systematic mechanisms for addressing privacy, legal, and ethical requirements. The implementation of secure aggregation can strengthen the protection of information exchanged during federated model training. The adoption of these privacy-preserving approaches can help address concerns associated with GDPR, IoMT privacy, and multi-facility data sharing.	01, 02, 03, 07, 09, 11, 13, 14, 15, 17, 18, 19, 20, 21, 24, 28, 29, 30, 31, 39, 40, 42, 45	**23**
14	**Encrypted Traffic Detection Limitations**	Encryption technologies such as TLS/HTTPS restrict access to packet contents, reducing the visibility available to traditional traffic-based detection mechanisms.	The use of behavioral anomaly detection can identify suspicious activities based on traffic behavior even when packet contents are encrypted. The adoption of hybrid feature models can combine multiple observable characteristics of encrypted traffic to improve detection despite limited TLS/HTTPS visibility.	23, 29, 43	**3**
15	**Organizational & Workforce Readiness**	Shortages of appropriate technical skills and the complexity of AI governance can make organizations less prepared to deploy, operate, and manage AI-based cybersecurity solutions.	The provision of workforce training can reduce the skills shortage associated with AI-based cybersecurity technologies. The establishment of AI governance mechanisms can improve organizational management of AI-based cybersecurity solutions. The maintenance of comprehensive lifecycle documentation can improve the management, maintenance, and accountability of AI-based security solutions.	15, 17, 20, 21, 28, 30, 31, 35, 42	**9**
16	**Multi-Layer / Heterogeneous Architecture**	IoMT and IoT systems consist of multiple architectural layers and	The adoption of a layer-aware IDS taxonomy can organize detection mechanisms	13, 16, 45	**3**

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

S. No	Consolidated Challenge	Description	Consolidated Solutions	Paper IDs	Freq (45)
	Complexity	diverse communication protocols, making it difficult to design a unified detection mechanism for the entire environment.	according to the different layers of IoMT/IoT architectures. The implementation of protocol-aware detection can account for differences among diverse IoT communication protocols. The use of layer- and protocol-specific detection mechanisms can reduce the complexity associated with heterogeneous and multi-layer environments.		
17	**Predictive Maintenance & RUL Estimation Integration**	Industrial systems require capabilities beyond anomaly detection, including prediction of equipment degradation and estimation of Remaining Useful Life (RUL).	The integration of Hybrid ML + RUL models can extend security analysis beyond anomaly detection toward equipment degradation prediction. The application of degradation scoring can quantify the severity and progression of equipment degradation and support predictive maintenance and RUL estimation.	39	**1**

4.2 Paper-Level Challenge and Solution Summary

Table 2. Challenges identified and solutions proposed, by individual paper (AP-01 to AP-45)

Paper ID	Challenges Identified	Solutions Proposed
AP-01	Adversarial ML attacks; Explainability gap; Data leakage	Adversarial training; Secure ML lifecycle; XAI
AP-02	Zero-day detection; False positives; Dataset imbalance; Privacy; Scalability	Hybrid IDS; Ensemble ML; SMOTE; Federated learning
AP-03	ML supply chain risks; Governance gap	Secure ML governance; Risk framework
AP-04	Dataset imbalance; High dimensionality; Outdated datasets	SMOTE; PCA; Modern datasets
AP-05	Feature redundancy; Dataset realism gap; Computational cost	Feature selection; Lightweight DL; Dataset modernization
AP-06	High false alarm rate	Ensemble optimization
AP-07	IoT resource limits; Real-time constraints; Drift	Edge IDS; Lightweight models; Online learning
AP-08	Big data traffic; Scalability	Distributed IDS; Cloud deployment
AP-09	Evaluation inconsistency; Dataset bias	Standard metrics; Benchmarking
AP-10	Concept drift; Zero-day gap	Continuous retraining; Hybrid detection
AP-11	Insider threat complexity	Behavioral analytics (UBA)
AP-12	Poor generalization; Zero-day; Drift	Cross-dataset validation; Adaptive learning
AP-13	IoT multi-layer complexity; Privacy	Layer-aware IDS; Federated learning
AP-14	Distributed privacy risk; Dataset imbalance	Federated learning; SMOTE

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Paper ID	Challenges Identified	Solutions Proposed
AP-15	Organizational resistance; Skills shortage	AI governance; Workforce training
AP-16	IoT protocol heterogeneity	Protocol-aware detection
AP-17	Evolving threat landscape; Insider risk	Adaptive ML; Continuous retraining
AP-18	Large-scale traffic; Real-time need	Stream processing; Distributed ML
AP-19	Privacy compliance; False alarms	Differential privacy; Ensemble ML
AP-20	Alert fatigue; Detection-response gap	Automated remediation; Intelligent prioritization
AP-21	Dataset realism limitation	Modern datasets
AP-22	Feature complexity; High dimensionality	PCA; Autoencoders
AP-23	Encrypted traffic limitation	Behavioral traffic analysis
AP-24	Healthcare data imbalance; Latency	SMOTE; Edge deployment
AP-25	Rare attack under-detection	Hybrid DL models
AP-26	Resource constraints	Lightweight ML
AP-27	Real-time detection need	Online learning
AP-28	IoT privacy concerns	Federated learning
AP-29	Enterprise scalability; Drift	Cloud IDS; Continuous learning
AP-30	Adaptive detection requirement	ML lifecycle management
AP-31	Integration complexity	Modular API framework
AP-32	Dataset bias; Imbalance	Balanced resampling
AP-33	CI/CD pipeline vulnerabilities; High false positives; Imbalance; Scalability	CNN-LSTM; SMOTEENN; ML-integrated CI/CD monitoring
AP-34	Zero-day detection; Overfitting; Imbalance; Adversarial risk	Deep Neural Network; Regularization; SMOTE; Adversarial robustness
AP-35	ICS dataset quality; Drift; Explainability; Scalability	Semi-supervised ML; XAI; Structured ML lifecycle
AP-36	Severe imbalance; High dimensionality; Rare attacks; Overfitting	SMOTEENN; DAE + Isolation Forest hybrid
AP-37	Zero-day; False alarms; Outdated datasets; Computational cost	Hybrid IDS; Modern datasets; GPU optimization
AP-38	High-dimensional sensor data; Unlabeled anomalies	PCA + DBSCAN; Human-in-loop validation
AP-39	Minority fault detection; RUL estimation; Privacy	Deep Autoencoder; RUL model; Federated learning
AP-40	IoT resource constraint; Adversarial vulnerability; Imbalance	Lightweight CNN; Federated learning; Hybrid DL
AP-41	Imbalanced anomaly rate; Interpretability; Feature complexity	RF/SVM; SHAP; Dimensionality reduction
AP-42	Alert fatigue; Governance gap; Adversarial ML; Privacy	Automated response; AI governance; Robust ML
AP-43	Outdated datasets; Imbalance; Overfitting; Encrypted traffic	Modern datasets; Ensemble ML; XAI
AP-44	Closed-set limitation; Zero-day gap; High dimensionality	Open Set Recognition; Clustering; Adaptive updating
AP-45	IoMT resource constraint; Heterogeneity; Imbalance; Latency; Privacy	Cloud-Fog-Edge IDS; SMOTE; Federated learning

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

4.3 Continent-Wise Comparison

Table 3. Continent-wise comparison of dominant challenges, root causes, and strategic solutions

Continent	Dominant Challenges	Root Causes / Driving Factors	Strategic Solutions Emphasized	Research Emphasis Trend
Asia	Zero-day attack detection Dataset imbalance High false positives Scalability in IoT/smart systems Computational constraints	Rapid digital transformation Large-scale IoT deployment Limited high-quality labeled datasets Resource-constrained environments Uneven regulatory maturity	Hybrid IDS (Signature + Anomaly) SMOTE / GAN-based balancing Lightweight DL models (CNN1D, LSTM) Cloud–Fog–Edge architecture Adaptive retraining	Performance-oriented and scalability-focused research
Europe	Privacy & GDPR compliance Explainability (XAI) Ethical AI governance Cross-border data-sharing limitations Federated learning complexity	Strict regulatory frameworks (GDPR, AI Act) Strong emphasis on ethical AI Data-sovereignty laws Public-sector cybersecurity integration	Federated learning Differential privacy SHAP / LIME explainability Secure ML lifecycle governance Risk-based AI frameworks	Trust, transparency, and compliance-driven research
North America	Adversarial ML attacks ML supply-chain vulnerabilities Insider threats Large-scale enterprise integration Zero-trust adaptation	Advanced enterprise AI deployment National security focus Mature cloud ecosystems Active adversarial research community	Adversarial training Secure ML pipelines Zero-trust AI frameworks Continuous retraining AI-SOAR integration	Security hardening and adversarial robustness focus
Middle East & Africa	Infrastructure readiness gaps Workforce skill shortages Governance maturity issues Resource constraints Dataset scarcity	Rapid but recent digital transformation Limited local AI research ecosystem Dependency on imported AI tools / talent shortage	Capacity-building programs Government AI strategy alignment Cloud-based scalable IDS Regional dataset repositories	Organizational and capacity-development emphasis
Cross-Continental (Global Trend)	Zero-day detection False positives Dataset imbalance Scalability Concept drift	Evolving cyber threats Massive heterogeneous data Dynamic attack techniques AI lifecycle vulnerabilities	Hybrid IDS architectures Ensemble & Deep Learning Federated learning Continuous adaptive retraining Explainable AI integration	Convergence toward hybrid, adaptive, and privacy-aware AI security

As shown in Table 1, zero-day and novel attack detection (n = 36), scalability and big-data handling (n = 34), high false positive/negative rates (n = 33), computational and resource constraints (n = 32), and dataset imbalance (n = 31) are the five most frequently reported challenges, followed by outdated/unrealistic datasets (n = 27), privacy/legal/ethical compliance (n = 23), and explainability and feature-engineering/high-dimensionality (n = 22 each). Table 2 reveals that the majority of individual papers report multiple, co-occurring challenges—for instance, zero-day detection is often co-occurring with dataset imbalance and outdated datasets, and privacy is often co-occurring with scalability in IoMT and federated deployments—and therefore the challenges are not independent. Table 3 reveals that while all regions have an interest in hybrid, adaptive, and privacy-aware solution architectures, the relative importance and root causes of specific challenges vary significantly by region: Asia is

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

dominated by scalability and IoT-driven detection challenges, Europe by privacy and explainability, North America by adversarial robustness and enterprise integration, and the Middle East and Africa by foundational infrastructure and capacity constraints.

5. Discussion

It has been found that the challenges to utilizing AI/ML for cyber threat detection are not just technical but involve data, computational, organizational and regulatory dimensions. The most common challenges mentioned in Table 1 are mostly related to zero-day detection, scalability, and false-positive/false-negative rates, indicating that existing AI/ML-based IDSs are still primarily limited by their ability to be generalizable beyond known patterns of attacks and by their scale requirements which are needed for the fast growing industrial, IoT and cloud infrastructures. This aligns with the conceptual background presented in Section 2 which determined anomaly-based detection as the most important method for detecting new attacks, but that it is also highly dependent on the quality and representativeness of training data, as indicated by the high rates of dataset imbalance (n = 31) and outdated/unrealistic datasets (n = 27) as critical challenges.

The fact that computational and resource constraints (n = 32) and scalability (n = 34) were also highly cited indicates a practical conflict in deploying AI for security: the more powerful deep learning algorithms tend to be the less practical in IoT, IoMT, and ICS contexts where threat detection is most critical. This tension is not being resolved by giving up on deep learning, but by re-architecturing the location and method of inference, as solutions converge towards lightweight models, edge based inference, and tiered edge–cloud architectures (Table 1).

The presence of explainability and trust (n = 22) and privacy, legal and ethical compliance (n = 23) as critical challenges highlights the need for technical accuracy not to be the only factor for organizational adoption. In structured diligences like healthcare and critical infrastructure, security analysts and regulators need verdicts that are interpretable and auditable. This is stable with the regional analysis in Table 3, as explainability, privacy and governance are seen as separate from detection performance, whereas studies in Asia have a focus on detection performance and scalability in resource constrained environments, and North American studies have a focus on adversarial robustness in well-established enterprise security operations, typically aligned with Zero Trust principles of continuous verification and least-privilege access [21]. These disparities indicate that there is no one-size-fits-all solution for the ‘right’ AI/ML security architecture for an organisation, as it will depend on the regulatory landscape, infrastructure readiness and threat profile.

These findings have some practical implications for modern organisations. First, it is important to recognize that there are many examples of resilience, scalability, and minimizing false alarms in the literature and that these attributes should be considered first order design requirements, not second order.First, it is important to note that resilience, scalability, and minimizing false alarms are common in the literature and should be considered first-order design attributes, not second-order. Second, it is important to consider dataset governance, namely using modern and representative datasets and systematic class balancing methods, as an early part of system development and not an afterthought. Third, explainability and privacy-preserving mechanisms (such as XAI layers, federated learning, differential privacy) should be incorporated into system architecture early in the design process, especially for organisations in regulated industries and/or across jurisdictions. Lastly, when organizations have limited local datasets, a shortage of local workers, etc., they might be better served by

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

capacity-building and cloud-based scalable IDS solutions than with the most intense models that are published in the literature.

The evidence suggests that the field is moving toward a wider consideration of AI/ML intrusion detection systems beyond accuracy measures — from what is now a multidimensional view of 'trustworthy' cybersecurity AI, as embodied by the architectures of proposed solutions that are now hybrid, adaptive and privacy-compliant (Tables 1-3).

6. Conclusion

In this study, a systematic literature review of 45 papers about the use of AI and ML in cyber threat detection for modern organizations was carried out, which resulted in a total of 271 reported challenges and 252 proposed solutions and were reduced to seventeen recurring themes (Table 1). The most significant challenges identified were zero day / novel attacks, scalability, handling large amounts of data, high false positive / negative rates, computation / resource limitations, imbalance in data sets, outdated datasets, privacy / regulatory compliance, explainability, and high dimensional feature space, with the latter being additional issues that still need to be addressed. As illustrated in the paper-level summary (Table 2), these challenges are not isolated – they generally occur in combination within any given study – and the continent-wise comparison (Table 3) provides insight into the systematic differences among research priorities across each continent as a function of varying regulatory environments, infrastructure maturity, and threat landscape, even as the field as a whole evolves toward hybrid, adaptive and privacy-aware AI security architectures that combine ensemble learning, federated learning, explainable AI, and ongoing adaptive retraining.

This synthesized evidence base serves as a practical benchmark for organizations developing or testing AI-based intrusion detection systems, while also providing a structured reference point for ongoing research exploring how to improve the gap between experiment and deployment to provide a more reliable, trusted experience. Further research should be conducted to integrate this synthesis with a quantitative meta-analysis of the detection performance reported in the literature, and to validate the proposed consolidated framework in operational environments, and to monitor the change in the frequency of challenges over time, as adversarial skills and regulatory requirements continue to shift.

7. Limitations

- The review is limited to peer-reviewed studies in English and a few credible technical reports in English; other relevant English-language or non-English literature (grey literature) may not have been included.
- Challenge and solution frequencies (Table 1) are indicative of the themes reported on in the literature, rather than of their empirically validated severity or impact in the real world; frequency does not reflect real-world risk but rather real-world research attention.
- Although the studies were systematically selected, the study is not a meta-analysis of quantitative detection performance metrics (e.g., accuracy, F1 score) in the different studies, but rather a mapping/synthesis study.
- Some studies (Table 3) were attributed to a country/continent based on the affiliation of the authors or the journal in which the article was published, but did not explicitly specify a geographic scope, which may contribute to imprecision in the regional analysis.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

- Several entries in the appendix-A do not include full bibliographic information in the original source records (e.g., volume number) and these are indicated in the appendixes, not estimated.
- The selection of published challenges and solutions might have been affected by publication bias, common to all SLRs.

References

1. A. K. B. Arnob, R. R. Chowdhury, N. A. Chaiti, S. Saha, and A. Roy, "A comprehensive systematic review of intrusion detection systems: Emerging techniques, challenges, and future research directions," Journal of Edge Computing, vol. 4, no. 1, pp. 73–104, 2025.
2. A. Hozouri, A. Mirzaei, and M. Effatparvar, "A comprehensive survey on intrusion detection systems with advances in machine learning, deep learning and emerging cybersecurity challenges," Discover Artificial Intelligence, vol. 5, art. 314, 2025.
3. A. Alotaibi and M. A. Rassam, "Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense," Future Internet, vol. 15, no. 2, art. 62, 2023.
4. A. Vassilev et al., "Adversarial machine learning: A taxonomy and terminology of attacks and mitigations (NIST AI 100-2e2025)," National Institute of Standards and Technology, 2025.
5. I. H. Sarker et al., "Explainable AI for cybersecurity automation, intelligence and trustworthiness in digital twin: Methods, taxonomy, challenges and prospects," ICT Express, vol. 10, pp. 935–958, 2024.
6. B. I. Farhan and A. D. Jasim, "Survey of intrusion detection using deep learning in the Internet of Things," Iraqi Journal for Computer Science and Mathematics, vol. 3, no. 1, pp. 81–91, 2022.
7. S. Oriaro and S. Mishra, "Improving cybersecurity through explainable artificial intelligence: A systematic literature review," Issues in Information Systems, vol. 26, no. 3, pp. 387–400, 2025.
8. E. E. Abdallah, W. Eleisah, and A. F. Otoom, "Intrusion detection systems using supervised machine learning techniques: A survey," Procedia Computer Science, vol. 201, pp. 205–212, 2022.
9. Y. Zhang, R. C. Muniyandi, and F. Qamar, "A review of deep learning applications in intrusion detection systems: Overcoming challenges in spatiotemporal feature extraction and data imbalance," Preprints, 2024.
10. A. Pinto et al., "Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure," Sensors, vol. 23, no. 5, art. 2415, 2023.
11. Y. Muhammad et al., "An improved framework for requirement implementation in the context of global software development: A systematic literature review protocol," International Journal of Database Theory and Application, vol. 9, no. 6, pp. 161–170, 2016.
12. A. Philip et al., "A systematic literature review of software requirements prioritization research," Information and Software Technology, vol. 56, pp. 568–585, 2014.
13. B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," EBSE Technical Report, 2007.
14. A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.
15. R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in Proc. IEEE Symposium on Security and Privacy, 2010.
16. N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 41–50, 2018.
17. B. Biggio and F. Roli, "Wild patterns: Ten years after the rise of adversarial machine learning," Pattern Recognition, vol. 84, pp. 317–331, 2018.
18. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
19. Gartner, "Top security and risk management trends," Gartner Research, 2022.
20. IBM Security, "AI in cybersecurity report," IBM Corporation, 2022.
21. S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero trust architecture (NIST Special Publication 800-207)," National Institute of Standards and Technology, 2020.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

22. L. Cadieux, "20 shocking cybercrime statistics: 2024 edition," The Devolutions Blog, Jul. 2024. [Online]. Available: https://blog.devolutions.net/2024/07/20-shocking-cybercrime-statistics2024-edition/.

23. J. Kim et al., "Deep neural networks for complex attack pattern detection in large-scale network traffic," 2021. (Bibliographic details limited in source record.)

Appendix A. List of Reviewed Papers (n = 45)

The Table below lists the 45 studies synthesized in this review, identified by internal paper ID (AP-01 to AP-45), along with title, authors, volume, publication year, and journal/venue. Volume information was not consistently available in the source bibliographic records and is marked "N/A" where absent; authorship is listed as "Multiple authors" where individual author names were not extracted in the underlying data-extraction records.

Paper ID	Title	Authors	Vol.	Year	Journal / Venue
AP-01	Adversarial machine learning in intrusion detection systems (dataset-specific study)	Multiple authors	N/A	2022	IEEE Access
AP-02	Hybrid machine learning-based intrusion detection system study	Multiple authors	N/A	2021	Computers & Security
AP-03	Machine learning lifecycle security risks	Multiple authors	N/A	2022	ACM CCS Workshop
AP-04	Feature optimization in intrusion detection systems	Multiple authors	N/A	2020	Future Generation Computer Systems
AP-05	Deep learning for intrusion detection system evaluation	Multiple authors	N/A	2021	Journal of Network & Computer Applications
AP-06	False alarm reduction in intrusion detection systems	Multiple authors	N/A	2020	Security and Communication Networks
AP-07	IoT-based machine learning intrusion detection framework	Multiple authors	N/A	2022	IEEE Internet of Things Journal
AP-08	Big data intrusion detection	Multiple authors	N/A	2021	IEEE Transactions on Big Data
AP-09	Benchmarking machine learning-based intrusion detection systems	Multiple authors	N/A	2021	Computers & Security
AP-10	Adaptive intrusion detection systems	Multiple authors	N/A	2022	Applied Sciences
AP-11	Insider threat detection review	Multiple authors	N/A	2020	ACM Computing Surveys
AP-12	Cross-dataset validation in intrusion detection systems	Multiple authors	N/A	2021	Information Sciences
AP-13	IoT layered security framework	Multiple authors	N/A	2022	IEEE Access
AP-14	Federated learning for intrusion detection systems	Multiple authors	N/A	2023	IEEE Transactions on Network Science
AP-15	AI governance in cybersecurity	Multiple authors	N/A	2023	Government Information Quarterly
AP-16	Protocol-aware intrusion detection systems for IoT	Multiple authors	N/A	2022	Sensors
AP-17	Adaptive threat detection	Multiple authors	N/A	2023	ACM Digital Threats: Research and Practice
AP-18	Distributed machine learning-based intrusion detection systems	Multiple authors	N/A	2022	Future Internet
AP-19	Privacy-preserving machine learning intrusion detection systems	Multiple authors	N/A	2023	IEEE Access

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Paper ID	Title	Authors	Vol.	Year	Journal / Venue
AP-20	AI-driven incident response	Multiple authors	N/A	2022	Computers & Security
AP-21	Modern dataset evaluation study	Multiple authors	N/A	2023	Electronics
AP-22	Dimensionality reduction in intrusion detection systems	Multiple authors	N/A	2023	Expert Systems with Applications
AP-23	Encrypted traffic detection	Multiple authors	N/A	2023	IEEE Transactions on Information Forensics and Security
AP-24	Healthcare intrusion detection system framework	Multiple authors	N/A	2023	Journal of Medical Systems
AP-25	Rare attack detection study	Multiple authors	N/A	2023	Security and Communication Networks
AP-26	Lightweight machine learning for intrusion detection systems	Multiple authors	N/A	2023	Sensors
AP-27	Online learning intrusion detection framework	Multiple authors	N/A	2023	IEEE Access
AP-28	Federated IoT security study	Multiple authors	N/A	2023	IEEE Internet of Things Journal
AP-29	Enterprise cloud intrusion detection system	Multiple authors	N/A	2023	Future Generation Computer Systems
AP-30	Continuous learning security models	Multiple authors	N/A	2023	Information Systems
AP-31	API-based intrusion detection system integration	Multiple authors	N/A	2023	Software: Practice and Experience
AP-32	Dataset bias mitigation in intrusion detection systems	Multiple authors	N/A	2023	Journal of Big Data
AP-33	Advancing software security and reliability in cloud platforms through AI-based anomaly detection	Multiple authors	N/A	2024	ACM Cloud Computing Security Workshop (CCSW)
AP-34	Network-based intrusion detection using deep learning technique	Multiple authors	N/A	2025	Scientific Reports
AP-35	Deployment guidelines for industry: machine learning-based intrusion detection systems	PETRAS Research Team	N/A	2023	PETRAS National Centre of Excellence (UKRI) Report
AP-36	NetDAIL: optimized deep learning hybrid model	Multiple authors	N/A	2025	International Journal of Advanced Computer Science and Applications (IJACSA)
AP-37	Network intrusion detection system: a systematic study of machine learning and deep learning approaches	Zeeshan Ahmad et al.	N/A	2021	Transactions on Emerging Telecommunications Technologies
AP-38	Machine learning-based anomaly detection for ITER Tokamak systems	Multiple authors	N/A	2025	TechRxiv
AP-39	Machine learning framework for cryogenic systems	Multiple authors	N/A	2025	IOP Conference Series
AP-40	Deep learning for cyber threat detection in IoT networks	Multiple authors	N/A	2024	Internet of Things & Cyber-Physical Systems
AP-41	Machine learning-based network anomaly detection	Multiple authors	N/A	2024	AI (MDPI)
AP-42	Applying AI in cybersecurity to enhance threat detection	Multiple authors	N/A	2024	Computer Science & IT Research Journal

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Paper ID	Title	Authors	Vol.	Year	Journal / Venue
AP-43	Systematic literature study of machine learning-based intrusion detection systems	Multiple authors	N/A	2025	Journal of Big Data
AP-44	Adaptable deep learning-based intrusion detection system for zero-day attacks	Multiple authors	N/A	2021	arXiv (cs.CR)
AP-45	AI for IoMT security: intrusion detection systems, attacks, and architectures	Multiple authors	N/A	2023	Internet of Things (Elsevier)

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

مقالة مراجعة

بطاريات أيونات الصوديوم، والاستخلاص المباشر لليثيوم، والتكامل ثنائي الاتجاه مع الشبكة، والطاقة النووية المتقدمة: مراجعة لانتقالها إلى النشر التجاري، 2020-2026

فيصل ع. محروقي[1,*]

[1] قسم الهندسة الميكانيكية، كلية الهندسة، الجامعة الإسلامية بالمدينة المنورة، المدينة المنورة، المملكة العربية السعودية

*المؤلف المراسل: f.mahroogi@iu.edu.sa

الملخص

تتساءل هذه المراجعة عن السبب الذي جعل أربع تقنيات كانت ممكنة تقنيًا قبل عقد أو أكثر، وهي بطاريات أيونات الصوديوم، والاستخلاص المباشر لليثيوم، والتكامل ثنائي الاتجاه مع الشبكة، والطاقة النووية المتقدمة، تبلغ جميعها نقاط تحول تجارية في الفترة نفسها، من 2024 إلى 2026. وتختبر المراجعة أطروحة مفادها أن هذه التحولات كانت مدفوعة في المقام الأول بتغيرات في سياق الطلب وحجم التصنيع والدعم السياسي، لا باختراقات تقنية منفصلة، وأن التقنيات الأربع أجزاء مترابطة من محفظة طاقة واحدة لا خيارات مستقلة. وتجمع هذه المراجعة السردية بين الأدبيات المحكّمة وتقارير الهيئات (ومنها تقارير الوكالة الدولية للطاقة والوكالة الدولية للطاقة الذرية) وإفصاحات الصناعة المنشورة حتى منتصف عام 2026. ولكل تقنية، تُفحص الأدلة في ضوء ثلاثة أسئلة: متى أُثبتت القدرة التقنية الأساسية أول مرة، وما الذي تغير بين 2020 و2026 فأتاح النشر التجاري، وما العوائق التي لا تزال قائمة. ثم تُحلَّل التفاعلات بين المجالات الأربعة. فقد دخلت بطاريات أيونات الصوديوم مرحلة الإنتاج الضخم في عام 2025، وباتت تقدم بديلًا أقل تكلفة وأوفر موارد لكيمياء الليثيوم في تخزين الشبكة والمركبات الكهربائية الاقتصادية، وإن ظلت كثافة طاقتها دون كثافة خلايا فوسفات الحديد والليثيوم. ويستخلص الاستخلاص المباشر الليثيوم من المحاليل الملحية في غضون ساعات بدلًا من أشهر، ببصمة أصغر على الأرض والمياه، وقد فتح المحاليل الملحية الحرارية الجوفية ومحاليل حقول النفط بوصفها موارد جديدة. ويحوّل الشحن ثنائي الاتجاه بطاريات المركبات والمنازل غير المستغلة إلى أصول موزعة للشبكة عبر بنية من أربع طبقات تشمل البطاريات، وتحويل القدرة، والتجميع، والأسواق. أما الطاقة النووية المتقدمة، بما فيها المفاعلات المعيارية الصغيرة والوقود المقاوم للحوادث، فقد اجتذبت استثمارات متجددة مدفوعة أساسًا بطلب مراكز البيانات على الكهرباء. وفي الحالات الأربع جميعها، يسبق العلمُ الممكِّن عامَ 2015، في حين ظهرت المحفزات التجارية بعد عام 2020. وتدعم الأدلة هذه الأطروحة؛ إذ حدد سياق الطلب وحجم التصنيع والسياسات توقيت التسويق التجاري لا الاختراقات المخبرية، وتتفاعل التقنيات الأربع عبر متغيرات مشتركة مثل ندرة الليثيوم ومتطلبات التخزين الثابت والحاجة إلى قدرة مضمونة. ونتيجة لذلك، تغدو منظومات الطاقة أكثر توزعًا وأكثر تنوعًا في موادها وأكثر استجابة لطلب تعيد الحوسبة تشكيله، وينبغي لتوقعات التقنية أن تتابع الأسواق والمؤسسات بقدر لا يقل عن متابعتها للأداء المخبري.

الكلمات المفتاحية: بطاريات أيونات الصوديوم؛ الاستخلاص المباشر لليثيوم؛ ربط المركبات بالشبكة؛ المفاعلات المعيارية الصغيرة؛ تخزين الطاقة؛ مرونة الشبكة.

© 2026 المؤلفون. نشرته جامعة القصيم بموجب ترخيص CC BY-NC-SA 4.0.

جامعة القصيم Qassim University

JEIT

مجلة التقنيات الناشئة والمبتكرة

المجلد 1، العدد 1

نوفمبر 2026

تاريخ الاستلام

15 سبتمبر 2026

تاريخ القبول

25 سبتمبر 2026

تاريخ النشر

5 نوفمبر 2026

معرّف الوثيقة الرقمي (DOI)

https://doi.org/10.71422/jeit.0003

الترخيص

CC BY-NC-SA 4.0

للاستشهاد بهذا البحث

فيصل ع. محروقي. بطاريات أيونات الصوديوم، والاستخلاص المباشر لليثيوم، والتكامل ثنائي الاتجاه مع الشبكة، والطاقة النووية المتقدمة: مراجعة لانتقالها إلى النشر التجاري، 2020-2026. مجلة التقنيات الناشئة والمبتكرة. 2026;1(1):23-36.

الموقع الإلكتروني للمجلة

Qassim University جامعة القصيم

JEIT

Journal of Emerging and Innovative Technologies

Volume 1, Issue 1
November 2026

Received
15 September 2026

Accepted
25 September 2026

Published
05 November 2026

DOI
https://doi.org/10.71422/jeit.0003

License
CC BY-NC-SA 4.0

Cite this article

Mahroogi FO. Sodium-Ion Batteries, Direct Lithium Extraction, Bidirectional Grid Integration, and Advanced Nuclear Power: A Review of Their Transition to Commercial Deployment, 2020–2026. J Emerg Innov Technol. 2026;1(1):23–36.

Journal website

Review article

Sodium-Ion Batteries, Direct Lithium Extraction, Bidirectional Grid Integration, and Advanced Nuclear Power: A Review of Their Transition to Commercial Deployment, 2020–2026

Faisal O. Mahroogi[1,*]

[1] Department of Mechanical Engineering, Faculty of Engineering, Islamic University of Madinah, Madinah, Saudi Arabia

*Corresponding author: f.mahroogi@iu.edu.sa

Abstract

This review asks why four technologies that were technically feasible a decade or more ago, namely sodium-ion batteries, direct lithium extraction, bidirectional grid integration, and advanced nuclear power, all reached commercial inflection points in the same period, 2024 to 2026. It tests the thesis that these transitions were driven primarily by changes in demand context, manufacturing scale, and policy support rather than by discrete technical breakthroughs, and that the four technologies are interdependent parts of a single energy portfolio rather than independent options. This narrative review synthesizes peer-reviewed literature, agency reports (including those of the International Energy Agency and the International Atomic Energy Agency), and industry disclosures published up to mid-2026. For each technology, the evidence is examined against three questions: when the core technical capability was first demonstrated, what changed between 2020 and 2026 to enable commercial deployment, and which barriers remain. The interactions among the four fields are then analyzed. Sodium-ion batteries entered mass production in 2025 and now offer a lower-cost, more resource-abundant alternative to lithium chemistries for grid storage and entry-level electric vehicles, although their energy density still lags that of lithium iron phosphate cells. Direct lithium extraction recovers lithium from brines within hours rather than months, with a smaller land and water footprint, and has opened geothermal and oilfield brines as new resources. Bidirectional charging is converting idle vehicle and home batteries into distributed grid assets through a four-layer architecture spanning batteries, power conversion, aggregation, and markets. Advanced nuclear power, including small modular reactors and accident-tolerant fuels, has attracted renewed investment driven mainly by data center electricity demand. In all four cases, the enabling science predates 2015, whereas the commercial triggers emerged after 2020. The evidence supports the thesis. Commercial timing was set by demand context, manufacturing scale, and policy rather than by laboratory breakthroughs, and the four technologies interact through shared variables such as lithium scarcity, stationary storage requirements, and the need for firm capacity. Energy systems are consequently becoming more distributed, more materially diverse, and more responsive to demand reshaped by computing, and technology forecasts should track markets and institutions at least as closely as laboratory performance.

Keywords: Sodium-ion batteries; direct lithium extraction; vehicle-to-grid; small modular reactors; energy storage; grid flexibility.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

1. Introduction

Electricity systems built during the twentieth century assumed steady demand growth and centralized supply. Both assumptions are now under strain. Electricity demand is rising again in advanced economies after nearly two decades of stagnation, driven largely by data centers, electrified transport, and industrial electrification [1].

At the same time, supply is shifting toward variable renewable sources that require storage and flexibility to remain reliable [2, 3]. This review examines four technologies that independent assessments published in 2026 identified as having reached commercial inflection points between 2024 and 2026.

MIT Technology Review placed sodium-ion batteries on its list of breakthrough technologies for 2026, citing their reliance on abundant materials and their readiness to power grids and affordable electric vehicles [4]. The World Economic Forum highlighted direct lithium extraction and everything-to-grid technology in its annual survey of emerging technologies, noting that both have left the laboratory and are already in commercial use [5]. Advanced nuclear power features in several recent outlooks, propelled by power purchase agreements signed by technology companies seeking firm, clean electricity [6].

These technologies are usually reviewed separately, and most existing reviews describe their technical progress without asking why commercialization occurred when it did. The coincidence in timing is notable, because the core science of each technology was established well before 2015, as Section 6 documents.

This review therefore addresses the following research question: why did sodium-ion batteries, direct lithium extraction, bidirectional grid integration, and advanced nuclear power reach commercial deployment in the same narrow window of 2024 to 2026? It tests two linked propositions. The first is that the timing of each transition was determined mainly by changes in demand context, manufacturing scale, and policy support rather than by a discrete technical breakthrough. The second is that the four technologies are interdependent, so that progress or failure in one alters the case for the others; they are therefore better evaluated as parts of a single portfolio than as competing options.

The temporal stance of the paper is as follows. Developments up to mid-2026 are treated as the recent past and the present and are described on the basis of published evidence; forward-looking statements are confined to the signposts identified in Section 7. The review is narrative rather than systematic. It draws on peer-reviewed literature, agency reports, and industry disclosures, and for each technology it examines when the core capability was first demonstrated, what changed between 2020 and 2026, and which barriers remain. Section 2 covers sodium-ion batteries, Section 3 direct lithium extraction, Section 4 bidirectional grid integration, and Section 5 advanced nuclear power. Section 6 evaluates the two propositions against the combined evidence, and Section 7 presents the conclusions and an outlook.

2. Sodium-Ion Batteries

2.1 Technical principles

Sodium-ion batteries operate on the same “rocking-chair” principle as lithium-ion cells. During charging, sodium ions leave a layered oxide or polyanionic cathode, travel through the electrolyte, and

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

insert into the anode; during discharging, the ions return to the cathode and release energy through an external circuit [7]. The architecture of a typical cell is shown in Figure 1. The chemistry differs from that of lithium-ion cells in several important ways. Sodium ions are larger and heavier than lithium ions, which lowers gravimetric energy density. Graphite does not intercalate sodium effectively, so most designs use hard carbon anodes derived from biomass or petroleum precursors [8, 9]. On the cathode side, layered transition metal oxides, Prussian blue analogues, and polyanionic compounds such as sodium vanadium phosphate compete for dominance, each with distinct trade-offs in energy, cost, and cycle life [10, 11].

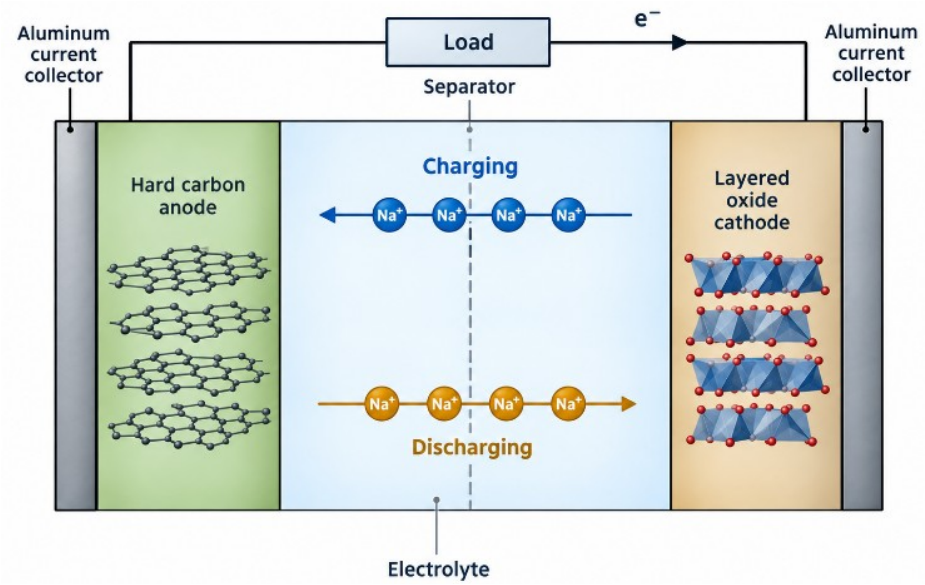


Figure 1. Schematic of a sodium-ion battery cell showing the hard carbon anode, layered oxide cathode, electrolyte, and movement of sodium ions during charging and discharging, with aluminum current collectors on both electrodes.

Two structural advantages follow from this chemistry. First, sodium-ion cells can use aluminum current collectors on both electrodes, because sodium does not alloy with aluminum at low potentials. This removes copper from the bill of materials and allows cells to be transported and stored at zero volts, which improves safety [12]. Second, sodium raw materials such as sodium carbonate are geographically widespread and inexpensive. A cost and resource analysis by Vaalma et al. [13] concluded that sodium chemistries reduce exposure to lithium, cobalt, and copper price volatility, even when cell-level costs remain similar during early production [14].

2.2 Performance status

Commercial sodium-ion cells announced between 2023 and 2025 reached energy densities of 160 to 175 Wh/kg, and laboratory cells have exceeded 200 Wh/kg [11, 12, 15]. These values remain below those of mainstream lithium iron phosphate (LFP) cells, at approximately 180 to 210 Wh/kg, and well below those of nickel-rich chemistries [16, 17]. The gap matters less for stationary storage, where footprint and cost per cycle dominate, than for long-range vehicles. Sodium-ion cells also perform well at low temperatures: manufacturers report usable capacity retention at −30 to −40 °C, a range in which LFP cells lose substantial capacity [15, 18]. Cycle life claims for recent commercial cells reach 8,000 to 10,000 cycles, which suits daily cycling in grid applications [19]. Table 1 compares representative characteristics of sodium-ion and lithium-ion chemistries, with the source of each row given beneath

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

the table. The values are indicative rather than exact, because both fields are evolving rapidly and specifications vary by manufacturer.

Table 1. Comparison of representative sodium-ion and lithium-ion battery characteristics.

Characteristic	Sodium-ion	Lithium iron phosphate	Nickel-rich lithium-ion
Gravimetric energy density (Wh/kg)	140 to 200	180 to 210	250 to 300
Low-temperature performance	Strong to −40 °C	Weak below −10 °C	Moderate
Cycle life (full cycles)	5,000 to 10,000	4,000 to 8,000	1,000 to 2,000
Anode current collector	Aluminum	Copper	Copper
Critical mineral exposure	Low	Lithium	Lithium, nickel, cobalt
Safe transport at 0 V	Yes	No	No
Primary near-term markets	Grid storage, entry-level EVs, starter batteries	EVs, grid storage	Premium EVs, electronics

Sources: energy density from Schmuch et al. [16], Abraham [17], Rudola et al. [12], and CATL [15]; low-temperature performance from Jaguemont et al. [18], Hwang et al. [10], and CATL [15]; cycle life from Preger et al. [20] and Hirsh et al. [19]; current collectors and zero-volt transport from Rudola et al. [12]; critical mineral exposure from Vaalma et al. [13] and Schmuch et al. [16]; near-term markets from Abraham [17] and CATL [15]. Values are indicative ranges for commercial cells. EV, electric vehicle.

2.3 Commercial deployment

The decisive shift came when large manufacturers committed production capacity. CATL, the world's largest battery producer, unveiled its second-generation sodium-ion product line under the Naxtra brand in April 2025, with mass production scheduled from December 2025 for passenger vehicle batteries and heavy-truck start-stop batteries [15]. BYD and several Chinese grid storage integrators followed with sodium-ion energy storage stations at the 100 MWh scale. Chinese public investment supported this scale-up, and analysts expect sodium-ion capacity to grow fastest in markets where lithium supply security is a policy concern [4]. Outside China, start-ups in the United States and Europe, including Natron Energy and Northvolt, pursued Prussian blue and layered oxide designs for industrial power and grid niches. Both companies had entered bankruptcy or ceased operations by late 2025, which shows how strongly commercial outcomes depended on access to capital and manufacturing scale rather than on cell chemistry alone.

Deployment has followed price dynamics. When lithium carbonate prices spiked in 2022, sodium-ion cells appeared immediately competitive. When lithium prices collapsed in 2023 and 2024, the cost case narrowed, and advocates of sodium-ion technology shifted their emphasis to supply diversification, safety, and cold-climate performance [21]. The long-term case rests on learning curves. Sodium-ion production borrows most of its equipment and processes from lithium-ion lines, so factories can be converted quickly, and costs can fall with cumulative production, as they did for lithium-ion cells [22, 23].

2.4 Barriers

Three barriers deserve attention. First, limited energy density restricts penetration into mid-range and long-range vehicles, so sodium-ion cells are likely to coexist with lithium chemistries rather than

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

replace them [17]. Second, hard carbon anode supply chains remain immature, and precursor quality strongly affects first-cycle efficiency [9]. Third, the field must avoid repeating the early safety complacency of the lithium-ion industry. Sodium-ion cells are more tolerant of abuse in many tests, yet large installations still require rigorous thermal management and fire codes.

3. Direct Lithium Extraction

3.1 The problem with incumbent methods

Even as sodium-ion technology diversifies storage chemistry, lithium demand continues to grow with electric vehicle production [21]. Most brine-derived lithium is currently produced from evaporation ponds in the South American Altiplano. Brine is pumped from underground aquifers into shallow ponds and concentrated by solar evaporation over 12 to 24 months. The method has low operating costs but is slow, land-intensive, and water-intensive, and it typically recovers only 40 to 60 percent of the lithium in the brine [24, 25]. Hard-rock mining of spodumene, the other main route, involves high energy use and processing emissions.

3.2 How direct lithium extraction works

Direct lithium extraction (DLE) replaces evaporation with selective chemical separation. Brine passes through an engineered medium that captures lithium ions while rejecting sodium, magnesium, calcium, and other competing species. Lithium is then stripped from the medium with water or a mild acid, producing a concentrated lithium solution ready for conversion to carbonate or hydroxide. The spent brine can be reinjected into the aquifer, which reduces net water extraction [26]. A generic process flow is shown in Figure 2.

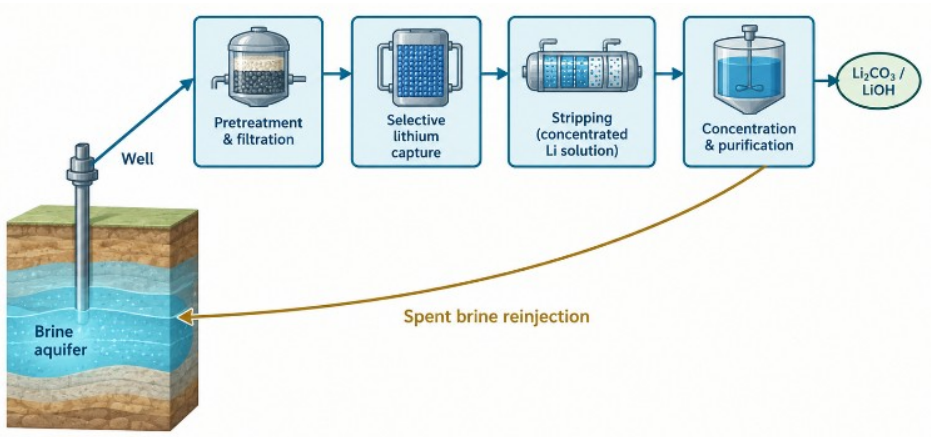


Figure 2. Generic process flow for direct lithium extraction, from brine intake through selective capture, stripping, concentration, and conversion to lithium carbonate or hydroxide, to reinjection of spent brine.

Several separation mechanisms compete. Adsorption uses aluminum-based sorbents that host lithium in their layered structure; it is the most commercially mature route, with decades of operation at plants in Argentina and China [27]. Ion exchange uses ceramic or polymeric materials with lithium-selective lattice sites and offers higher selectivity at the cost of more demanding regeneration chemistry. Solvent extraction and membrane processes, including nanofiltration and electrodialysis, are advancing in pilot programs [26, 28]. Across all routes, the headline advantages are recovery rates above 80 percent and processing times measured in hours rather than months.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

3.3 Deployment and new resources

DLE changes the map of viable lithium resources. Brines that are too dilute, too rich in magnesium, or located in climates unsuitable for evaporation become accessible. The most closely watched examples are geothermal brines at the Salton Sea in California, where operators plan to co-produce lithium and geothermal electricity [29], and oilfield brines [30] in the Smackover Formation across Arkansas and Texas, where ExxonMobil and Standard Lithium hold projects. Chinese producers have deployed adsorption-based DLE at scale on the Qinghai Plateau, and Eramet commissioned a commercial DLE plant in Argentina in 2024. The World Economic Forum included DLE in its 2026 list of emerging technologies on the grounds that it extracts essential raw materials with less environmental impact and is ready to scale [5].

3.4 Open questions

The technology still faces genuine uncertainties. Sorbent lifetime under harsh brine chemistry determines operating cost, and few public datasets cover multiyear performance. Freshwater consumption during sorbent washing can be significant in arid basins, which complicates the environmental case unless closed-loop water systems are used [25]. Reinjection of spent brine must be managed carefully to avoid diluting the resource. Finally, DLE projects have repeatedly slipped their schedules, and the sector still needs several full-scale plants operating profitably outside China to prove the model across diverse brine chemistries.

4. Bidirectional Grid Integration

4.1 From vehicle-to-grid to everything-to-grid

The concept of using electric vehicle batteries as grid resources dates back at least to the foundational analysis of Kempton and Tomić [31], who showed that vehicles are parked for more than 90 percent of the day and that their combined power capacity could rival national generation fleets. Twenty years later, the enabling conditions are in place: electric vehicle fleets are large, bidirectional chargers are commercially available, and grid operators face a growing need for flexible capacity [32, 33]. The framing has also widened. Everything-to-grid (X2G) includes home batteries, heat pump thermal storage, and commercial equipment alongside vehicles. The World Economic Forum summarizes the core idea as follows: electric vehicles and home batteries sit idle for much of the day, and everything-to-grid technology puts their stored energy to work when the grid needs it most [5].

4.2 System architecture

A bidirectional system has four layers, as illustrated in Figure 3. The first layer is the battery asset itself, whether in a vehicle or mounted on a garage wall. The second is the power conversion hardware, either an onboard bidirectional inverter or a DC charger that manages the power flow [34]. The third is the aggregation platform, that is, software that pools thousands of small assets into a virtual power plant that can bid into electricity markets [35]. The fourth is the market and regulatory interface that compensates participants for energy, capacity, or ancillary services. Vehicle-to-home operation, in which a car powers a house during an outage, requires only the first two layers and has become a mainstream selling point on models such as the Ford F-150 Lightning. Full vehicle-to-grid (V2G) participation requires all four layers and depends on interconnection rules that vary widely across jurisdictions.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

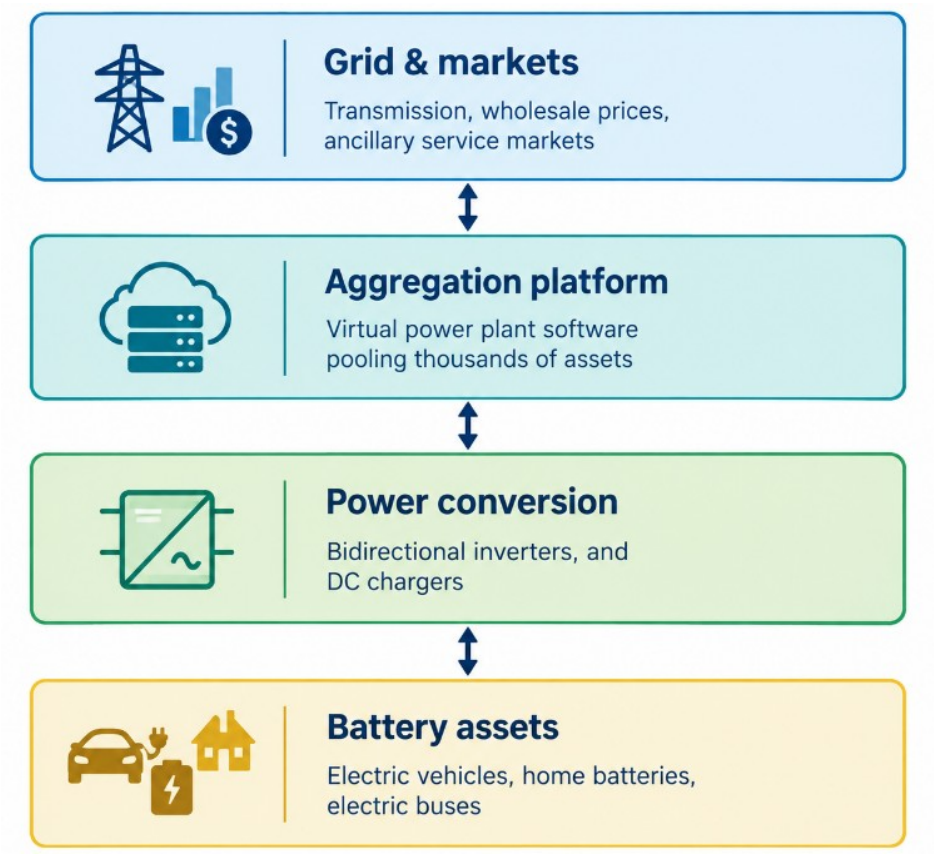


Figure 3. Four-layer architecture of a bidirectional everything-to-grid system, spanning battery assets, power conversion hardware, aggregation software, and market interfaces.

4.3 Value and evidence

The value streams are diverse. Frequency regulation pays for fast response and was the focus of early pilots [32]. Capacity payments reward availability during peak hours, and energy arbitrage exploits the price spread between inexpensive overnight charging and expensive evening discharge. Utility pilots in California, the Netherlands, Denmark, the United Kingdom, and Japan have demonstrated each of these streams. School buses have proven especially attractive, because their large batteries and predictable summer idle periods match grid needs. On the home battery side, virtual power plants operated by Tesla, Sunrun, and several utilities dispatched hundreds of megawatts during heat waves between 2023 and 2025, providing evidence that aggregated behind-the-meter storage performs when called upon.

4.4 Barriers

Battery degradation worries vehicle owners. Laboratory testing has shown that additional bidirectional cycling can accelerate capacity loss [36], but the authors of the leading studies subsequently concluded that smart management of state of charge and depth of discharge can limit this wear and, in some cases, extend battery life [37]. Warranty terms have begun to adapt, with some manufacturers explicitly covering approved bidirectional use. Interconnection and metering rules remain the largest practical obstacle, since many utilities still treat a discharging vehicle as an unfamiliar generator. Social dimensions matter as well: Sovacool et al. [38] argued that user trust, clarity of compensation, and control preferences will determine participation rates as much as hardware will. Standardization is improving through the ISO 15118 communication protocols, which allow the vehicle and charger to negotiate power flows securely [39].

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

5. Advanced Nuclear Power

5.1 Demand pull from computing

Nuclear power supplies steady, low-carbon electricity, and its economics improve when buyers value firm supply around the clock [3]. The surge in data center construction for artificial intelligence has created exactly such buyers. Technology companies signed a wave of nuclear procurement agreements between 2024 and 2026, including agreements to restart Three Mile Island Unit 1 for Microsoft, to purchase output from planned small modular reactors for Google, and to support new capacity at existing plants for Amazon. Global electricity demand from data centers is projected to more than double by 2030, reaching a level comparable to the current electricity consumption of Japan [6]. This demand pull, more than any technical breakthrough, explains the renewed investor interest in nuclear energy.

5.2 Small modular reactors

Small modular reactors (SMRs) are designed with electrical outputs below approximately 300 MW and are built largely in factories rather than on site. Their intended advantages are serial production, shorter construction schedules, lower capital at risk per project, and flexibility to be sited near industrial loads [40, 41]. Design families include light-water SMRs, such as those developed by NuScale and GE Hitachi; high-temperature gas-cooled reactors, such as X-energy's Xe-100; and sodium-cooled fast reactors, such as TerraPower's Natrium plant under construction in Wyoming. China connected the HTR-PM high-temperature gas-cooled reactor to the grid in 2021 and is building the Linglong One light-water SMR, giving it the most operational experience to date [42, 43]. Figure 4 places SMR capacity in the context of projected data center demand growth.

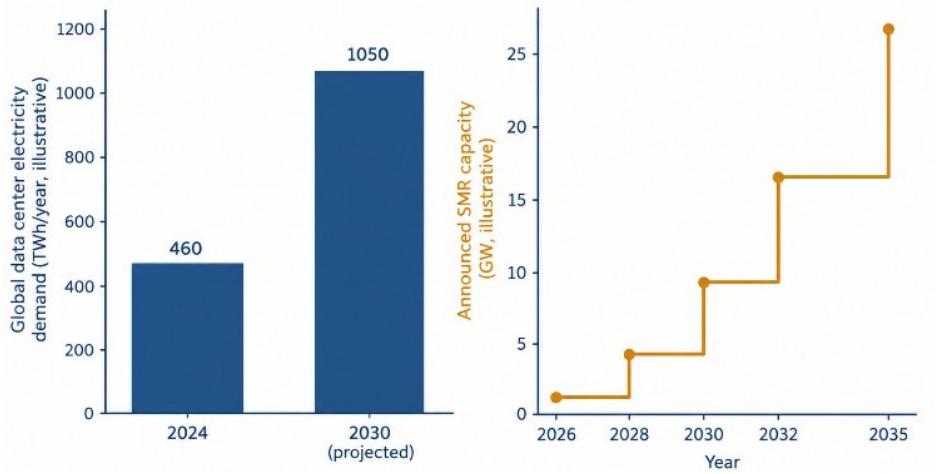


Figure 4. Projected growth of global data center electricity demand to 2030 alongside announced small modular reactor capacity, illustrating the demand pull behind advanced nuclear investment (illustrative values).

The economic case remains unproven at scale. NuScale's first planned project in Utah, the Carbon Free Power Project, was terminated in 2023 after it failed to attract sufficient utility subscription following a sharp rise in its estimated power price [44]. Advocates respond that first-of-a-kind units always carry cost premiums and that learning across identical factory-built units is the central premise of the model [45]. Construction programs scheduled for 2026 to 2030 in the United States, Canada, China, and the United Kingdom will test that claim directly.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

5.3 Safer fuel

Fuel innovation is proceeding in parallel with reactor design. Conventional reactors use uranium dioxide pellets in zirconium alloy cladding. Under severe accident conditions, the cladding can oxidize in steam and generate hydrogen, which contributed to the explosions at Fukushima Daiichi. Accident-tolerant fuels address this pathway with coated claddings, doped pellets, and alternative materials that resist oxidation and retain fission products at higher temperatures [46, 47]. TRISO fuel takes the concept further by encapsulating each uranium kernel in layers of carbon and silicon carbide, creating particles that act as their own containment and remain intact well beyond normal accident temperatures [48]. Several SMR designs pair TRISO fuel with passive cooling, so that decay heat is removed by conduction and natural circulation without operator action or external power.

5.4 Constraints

Advanced nuclear power faces constraints that engineering alone cannot remove. Many designs require high-assay low-enriched uranium (HALEU), which is enriched to between 5 and 20 percent uranium-235; Russia dominated its commercial supply until recently, prompting new Western investment in enrichment [49]. Licensing processes remain lengthy, although regulators in the United States and Canada have created dedicated pathways for advanced designs. Public acceptance varies by country, and waste management policy remains unresolved in most markets [50]. Cost, however, is the decisive variable. Nuclear power does not need to beat solar power on price per megawatt-hour; it needs to beat the combination of renewables, storage, and transmission for the specific service of firm, carbon-free power delivered continuously [3].

6. Discussion: Evaluating the Two Propositions

Consider first the proposition of interdependence. Although the four technologies are usually treated separately, their interactions are where the most important dynamics appear. Sodium-ion batteries and direct lithium extraction pull in opposite directions on the same variable, namely lithium scarcity. If DLE succeeds broadly, abundant and inexpensive lithium weakens the resource-security argument for sodium-ion cells; if DLE disappoints, sodium-ion adoption accelerates. Both outcomes reduce the risk of storage bottlenecks, which explains why investors fund both simultaneously.

Bidirectional integration changes the arithmetic of how much stationary storage the grid must build. A national fleet of electric vehicles with even partial V2G participation represents terawatt-hours of distributed capacity that already exists for another purpose [32]. Every megawatt-hour supplied from a parked car is a megawatt-hour of grid battery capacity that does not need to be manufactured, which in turn feeds back into demand projections for both lithium and sodium chemistries.

Advanced nuclear power interacts with all three. Data center demand growth is the strongest force pushing utilities toward new firm capacity, and the same growth raises peak stresses that storage and V2G can relieve [6]. High-temperature reactors could also supply industrial heat for materials processing, including battery precursor manufacturing and brine treatment. The evidence therefore supports the proposition of interdependence: energy systems in 2026 are best understood not as a contest between generation and storage but as a portfolio problem, in which flexibility, firmness, and material supply must be optimized together [3].

The other proposition concerns the drivers of timing, and Table 2 summarizes the relevant evidence. In each case, the core technical capability was demonstrated well before 2015: sodium intercalation

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

chemistry was studied in the 1980s [51], adsorption-based DLE has operated in Argentina since the 1990s [27], V2G economics were formalized in 2005 [31], and SMR concepts date to the earliest years of civil nuclear power [40]. What changed between 2020 and 2026 was the demand context, manufacturing scale, and policy support. The contrasting fortunes of sodium-ion producers reinforce this reading: the same chemistry succeeded commercially where it was backed by established gigawatt-hour-scale manufacturing and failed where start-ups could not finance scale-up. This pattern is consistent with the historical record of energy technology diffusion, in which market formation and institutional support have repeatedly determined the pace of adoption [52]. The evidence therefore supports the timing proposition, with one qualification: incremental engineering, such as higher sodium-ion energy density and more selective DLE sorbents, was necessary for commercialization, even though it was not sufficient to trigger it. Forecasting technology impact consequently requires watching markets and institutions at least as closely as laboratories.

Table 2. Evidence on the timing of commercial inflection for the four technologies.

Technology	Core capability first shown	Main triggers, 2020–2026	Main remaining barriers
Sodium-ion batteries	1980s (intercalation chemistry)	2022 lithium price spike; conversion of gigawatt-hour lithium-ion lines; Chinese industrial policy	Energy density; hard carbon supply
Direct lithium extraction	1990s (commercial adsorption plants)	Rising EV lithium demand; supply-security policy; geothermal and oilfield resources	Sorbent lifetime; freshwater use; project delays
Bidirectional grid integration	2005 (V2G economics)	Large EV fleets; commercial bidirectional chargers; ISO 15118-20	Interconnection rules; user trust
Advanced nuclear power	1950s–1960s (small reactors; TRISO fuel)	Data center demand for firm power; corporate procurement	Cost; HALEU supply; licensing

Sources: Delmas [51], Rudola et al. [12], and CATL [15] for sodium-ion batteries; Liu et al. [27], Warren [29], and International Energy Agency [21] for direct lithium extraction; Kempton and Tomić [31] and International Organization for Standardization [39] for bidirectional integration; Ingersoll [40], Demkowicz et al. [48], and International Energy Agency [6] for advanced nuclear power. EV, electric vehicle; HALEU, high-assay low-enriched uranium; V2G, vehicle-to-grid.

7. Conclusion and Outlook

This review asked why sodium-ion batteries, direct lithium extraction, bidirectional grid integration, and advanced nuclear power all reached commercial deployment between 2024 and 2026. The evidence supports two conclusions. First, the timing was set less by technical breakthroughs than by shifts in demand context, manufacturing scale, and policy support, since each technology was technically plausible a decade or more earlier. Second, the four technologies are interdependent: DLE and sodium-ion chemistry act on the same lithium-scarcity variable, bidirectional integration reduces the amount of stationary storage that must be manufactured, and data center demand links firm nuclear capacity to the flexibility provided by storage. Together, they are producing energy systems that are more distributed, more materially diverse, and more responsive to a demand curve increasingly shaped

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

by computing. None of the four technologies is free of unresolved barriers, from anode supply chains and sorbent lifetimes to interconnection rules and enrichment supply.

Several signposts will indicate the direction of travel between 2026 and 2030. For sodium-ion batteries, the key indicators are cell prices falling below those of LFP cells on a sustained basis and the first gigawatt-hour-scale deployments outside China. For DLE, the test is whether commercial plants achieve nameplate recovery rates on geothermal and oilfield brines, particularly at the Salton Sea and in the Smackover Formation. For bidirectional integration, the indicators are interconnection reform and whether automakers standardize bidirectional capability across mass-market models rather than premium trims. For advanced nuclear power, completion of the Natrium and Xe-100 projects near their revised budgets would validate the factory-build thesis, whereas further large overruns would push buyers back toward renewables paired with storage. Because the analysis indicates that markets and institutions govern timing, these signposts are deliberately commercial and regulatory rather than purely technical.

Acknowledgment

The author thanks the editorial team of the Journal of Emerging and Innovative Technologies for the opportunity to submit this review. No external funding was received for this work, and the author declares no conflicts of interest.

References

1. International Energy Agency. (2024). Electricity 2024: Analysis and forecast to 2026. IEA.
2. Kittner, N., Lill, F., & Kammen, D. M. (2017). Energy storage deployment and innovation for the clean energy transition. Nature Energy, 2(9), 17125. https://doi.org/10.1038/nenergy.2017.125
3. Sepulveda, N. A., Jenkins, J. D., de Sisternes, F. J., & Lester, R. K. (2018). The role of firm low-carbon electricity resources in deep decarbonization of power generation. Joule, 2(11), 2403–2420. https://doi.org/10.1016/j.joule.2018.08.006
4. MIT Technology Review. (2026, January 12). 10 breakthrough technologies 2026. https://www.technologyreview.com/2026/01/12/1130697/10-breakthrough-technologies-2026/
5. World Economic Forum. (2026, June 23). Top 10 emerging technologies of 2026 [Insight report, in collaboration with Frontiers]. https://reports.weforum.org/docs/WEF_Top_10_Emerging_Technologies_Report_2026.pdf
6. International Energy Agency. (2025). Energy and AI. IEA.
7. Yabuuchi, N., Kubota, K., Dahbi, M., & Komaba, S. (2014). Research development on sodium-ion batteries. Chemical Reviews, 114(23), 11636–11682. https://doi.org/10.1021/cr500192f
8. Slater, M. D., Kim, D., Lee, E., & Johnson, C. S. (2013). Sodium-ion batteries. Advanced Functional Materials, 23(8), 947–958. https://doi.org/10.1002/adfm.201200691
9. Irisarri, E., Ponrouch, A., & Palacín, M. R. (2015). Review—Hard carbon negative electrode materials for sodium-ion batteries. Journal of The Electrochemical Society, 162(14), A2476–A2482. https://doi.org/10.1149/2.0091514jes
10. Hwang, J.-Y., Myung, S.-T., & Sun, Y.-K. (2017). Sodium-ion batteries: Present and future. Chemical Society Reviews, 46(12), 3529–3614. https://doi.org/10.1039/C6CS00776G
11. Tapia-Ruiz, N., Armstrong, A. R., Alptekin, H., Amores, M., Au, H., Barker, J., et al. (2021). 2021 roadmap for sodium-ion batteries. Journal of Physics: Energy, 3(3), 031503. https://doi.org/10.1088/2515-7655/ac01ef
12. Rudola, A., Rennie, A. J. R., Heap, R., Meysami, S. S., Lowbridge, A., Mazzali, F., Sayers, R., Wright, C. J., & Barker, J. (2021). Commercialisation of high energy density sodium-ion batteries: Faradion's journey and outlook. Journal of Materials Chemistry A, 9(13), 8279–8302. https://doi.org/10.1039/D1TA00376C
13. Vaalma, C., Buchholz, D., Weil, M., & Passerini, S. (2018). A cost and resource analysis of sodium-ion batteries. Nature Reviews Materials, 3, 18013. https://doi.org/10.1038/natrevmats.2018.13

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

14. Peters, J. F., Peña Cruz, A., & Weil, M. (2019). Exploring the economic potential of sodium-ion batteries. Batteries, 5(1), 10. https://doi.org/10.3390/batteries5010010
15. CATL. (2025, April 21). Naxtra battery breakthrough & dual-power architecture: CATL pioneers the multi-power era [Press release]. Contemporary Amperex Technology Co., Limited; PR Newswire.
16. Schmuch, R., Wagner, R., Hörpel, G., Placke, T., & Winter, M. (2018). Performance and cost of materials for lithium-based rechargeable automotive batteries. Nature Energy, 3(4), 267–278. https://doi.org/10.1038/s41560-018-0107-2
17. Abraham, K. M. (2020). How comparable are sodium-ion batteries to lithium-ion counterparts? ACS Energy Letters, 5(11), 3544–3547. https://doi.org/10.1021/acsenergylett.0c02181
18. Jaguemont, J., Boulon, L., & Dubé, Y. (2016). A comprehensive review of lithium-ion batteries used in hybrid and electric vehicles at cold temperatures. Applied Energy, 164, 99–114. https://doi.org/10.1016/j.apenergy.2015.11.034
19. Hirsh, H. S., Li, Y., Tan, D. H. S., Zhang, M., Zhao, E., & Meng, Y. S. (2020). Sodium-ion batteries paving the way for grid energy storage. Advanced Energy Materials, 10(32), 2001274. https://doi.org/10.1002/aenm.202001274
20. Preger, Y., Barkholtz, H. M., Fresquez, A., Campbell, D. L., Juba, B. W., Romàn-Kustas, J., Ferreira, S. R., & Chalamala, B. (2020). Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. Journal of The Electrochemical Society, 167(12), 120532. https://doi.org/10.1149/1945-7111/abae37
21. International Energy Agency. (2024). Batteries and secure energy transitions. IEA.
22. Schmidt, O., Hawkes, A., Gambhir, A., & Staffell, I. (2017). The future cost of electrical energy storage based on experience rates. Nature Energy, 2(8), 17110. https://doi.org/10.1038/nenergy.2017.110
23. Ziegler, M. S., & Trancik, J. E. (2021). Re-examining rates of lithium-ion battery technology improvement and cost decline. Energy & Environmental Science, 14(4), 1635–1651. https://doi.org/10.1039/D0EE02681F
24. Flexer, V., Baspineiro, C. F., & Galli, C. I. (2018). Lithium recovery from brines: A vital raw material for green energies with a potential environmental impact in its mining and processing. Science of the Total Environment, 639, 1188–1204. https://doi.org/10.1016/j.scitotenv.2018.05.223
25. Vera, M. L., Torres, W. R., Galli, C. I., Chagnes, A., & Flexer, V. (2023). Environmental impact of direct lithium extraction from brines. Nature Reviews Earth & Environment, 4(3), 149–165. https://doi.org/10.1038/s43017-022-00387-5
26. Farahbakhsh, J., Arshadi, F., Mofidi, Z., Mohseni-Dargah, M., Kök, C., Assefi, M., Soozanipour, A., Zargar, M., Asadnia, M., Boroumand, Y., Presser, V., & Razmjou, A. (2024). Direct lithium extraction: A new paradigm for lithium production and resource utilization. Desalination, 575, 117249. https://doi.org/10.1016/j.desal.2023.117249
27. Liu, G., Zhao, Z., & Ghahreman, A. (2019). Novel approaches for lithium extraction from salt-lake brines: A review. Hydrometallurgy, 187, 81–100. https://doi.org/10.1016/j.hydromet.2019.05.005
28. Stringfellow, W. T., & Dobson, P. F. (2021). Technology for the recovery of lithium from geothermal brines. Energies, 14(20), 6805. https://doi.org/10.3390/en14206805
29. Warren, I. (2021). Techno-economic analysis of lithium extraction from geothermal brines (Technical Report NREL/TP-5700-79178). National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy21osti/79178.pdf
30. Kumar, A., Fukuda, H., Hatton, T. A., & Lienhard, J. H. (2019). Lithium recovery from oil and gas produced water: A need for a growing energy industry. ACS Energy Letters, 4(6), 1471–1474. https://doi.org/10.1021/acsenergylett.9b00779
31. Kempton, W., & Tomić, J. (2005). Vehicle-to-grid power fundamentals: Calculating capacity and net revenue. Journal of Power Sources, 144(1), 268–279. https://doi.org/10.1016/j.jpowsour.2004.12.025
32. Kempton, W., & Tomić, J. (2005). Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy. Journal of Power Sources, 144(1), 280–294. https://doi.org/10.1016/j.jpowsour.2004.12.022
33. Tan, K. M., Ramachandaramurthy, V. K., & Yong, J. Y. (2016). Integration of electric vehicles in smart grid: A review on vehicle to grid technologies and optimization techniques. Renewable and Sustainable Energy Reviews, 53, 720–732. https://doi.org/10.1016/j.rser.2015.09.012
34. Yilmaz, M., & Krein, P. T. (2013). Review of the impact of vehicle-to-grid technologies on distribution systems and utility interfaces. IEEE Transactions on Power Electronics, 28(12), 5673–5689. https://doi.org/10.1109/TPEL.2012.2227500

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

35. Pudjianto, D., Ramsay, C., & Strbac, G. (2007). Virtual power plant and system integration of distributed energy resources. IET Renewable Power Generation, 1(1), 10–16. https://doi.org/10.1049/iet-rpg:20060023

36. Dubarry, M., Devie, A., & McKenzie, K. (2017). Durability and reliability of electric vehicle batteries under electric utility grid operations: Bidirectional charging impact analysis. Journal of Power Sources, 358, 39–49. https://doi.org/10.1016/j.jpowsour.2017.05.015

37. Uddin, K., Dubarry, M., & Glick, M. B. (2018). The viability of vehicle-to-grid operations from a battery technology and policy perspective. Energy Policy, 113, 342–347. https://doi.org/10.1016/j.enpol.2017.11.015

38. Sovacool, B. K., Noel, L., Axsen, J., & Kempton, W. (2018). The neglected social dimensions to a vehicle-to-grid (V2G) transition: A critical and systematic review. Environmental Research Letters, 13(1), 013001. https://doi.org/10.1088/1748-9326/aa9c6d

39. International Organization for Standardization. (2022). Road vehicles — Vehicle to grid communication interface — Part 20: 2nd generation network layer and application layer requirements (ISO Standard No. 15118-20:2022). ISO.

40. Ingersoll, D. T. (2009). Deliberately small reactors and the second nuclear era. Progress in Nuclear Energy, 51(4–5), 589–603. https://doi.org/10.1016/j.pnucene.2009.01.003

41. Locatelli, G., Bingham, C., & Mancini, M. (2014). Small modular reactors: A comprehensive overview of their economics and strategic aspects. Progress in Nuclear Energy, 73, 75–85. https://doi.org/10.1016/j.pnucene.2014.01.010

42. Zhang, Z., Dong, Y., Li, F., Zhang, Z., Wang, H., Huang, X., Li, H., Liu, B., Wu, X., Wang, H., Diao, X., Zhang, H., & Wang, J. (2016). The Shandong Shidao Bay 200 MWe high-temperature gas-cooled reactor pebble-bed module (HTR-PM) demonstration power plant: An engineering and technological innovation. Engineering, 2(1), 112–118. https://doi.org/10.1016/J.ENG.2016.01.020

43. International Atomic Energy Agency. (2022). Advances in small modular reactor technology developments: 2022 edition. IAEA.

44. NuScale Power. (2023, November 8). Utah Associated Municipal Power Systems (UAMPS) and NuScale Power agree to terminate the Carbon Free Power Project (CFPP) [Press release]. Filed as Exhibit 99.1 to NuScale Power Corporation Form 8-K, U.S. Securities and Exchange Commission.

45. Mignacca, B., & Locatelli, G. (2020). Economics and finance of small modular reactors: A systematic review and research agenda. Renewable and Sustainable Energy Reviews, 118, 109519. https://doi.org/10.1016/j.rser.2019.109519

46. Zinkle, S. J., Terrani, K. A., Gehin, J. C., Ott, L. J., & Snead, L. L. (2014). Accident tolerant fuels for LWRs: A perspective. Journal of Nuclear Materials, 448(1–3), 374–379. https://doi.org/10.1016/j.jnucmat.2013.12.005

47. Terrani, K. A. (2018). Accident tolerant fuel cladding development: Promise, status, and challenges. Journal of Nuclear Materials, 501, 13–30. https://doi.org/10.1016/j.jnucmat.2017.12.043

48. Demkowicz, P. A., Liu, B., & Hunn, J. D. (2019). Coated particle fuel: Historical perspectives and current progress. Journal of Nuclear Materials, 515, 434–450. https://doi.org/10.1016/j.jnucmat.2018.09.044

49. National Academies of Sciences, Engineering, and Medicine. (2023). Merits and viability of different nuclear fuel cycles and technology options and the waste aspects of advanced nuclear reactors. The National Academies Press. https://doi.org/10.17226/26500

50. Krall, L. M., Macfarlane, A. M., & Ewing, R. C. (2022). Nuclear waste from small modular reactors. Proceedings of the National Academy of Sciences, 119(23), e2111833119. https://doi.org/10.1073/pnas.2111833119

51. Delmas, C. (2018). Sodium and sodium-ion batteries: 50 years of research. Advanced Energy Materials, 8(17), 1703137. https://doi.org/10.1002/aenm.201703137

52. Grubler, A., & Wilson, C. (Eds.). (2014). Energy technology innovation: Learning from historical successes and failures. Cambridge University Press. ISBN 978-1-107-02322-2

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

مقالة بحثية

الرسم الذاتي للخرائط بالاعتماد على الرؤية للطائرات المسيّرة ذاتية القيادة في البيئات الداخلية

ورود الرشيدي[1,*]، جمال العتيبي[1]

[1] قسم هندسة الحاسب، كلية الحاسب، جامعة القصيم، بريدة، المملكة العربية السعودية

*المؤلف المراسل: 461215971@qu.edu.sa

الملخص

في غياب إشارات النظام العالمي لتحديد المواقع (GPS) تتعقد مشكلة الملاحة الذاتية للطائرات المسيّرة في البيئات الداخلية تعقيدًا كبيرًا، ولذلك تزايد الاهتمام مؤخرًا بتقنية التوطين ورسم الخرائط المتزامنين بالرؤية (Visual SLAM). إذ تتيح هذه الأنظمة تحديد موقع الطائرة المسيّرة في البيئة وبناء خريطة لها باستخدام بيانات الكاميرا وحدها. غير أن البيئات الداخلية تفرض عمومًا مشكلات حادة على هذه الأنظمة، مثل الإضاءة المنخفضة، والضبابية الناتجة عن الحركة السريعة، والأسطح الخالية من المعالم التي لا توفر الخصائص الموثوقة اللازمة للتتبع. وتؤدي حدة هذه المشكلات، في الحالات القصوى، إلى الإخلال بالتوجه المكاني وبمعلومات المكان، مما يضعف جودة التوطين ومتانته. ويهدف هذا البحث إلى تطوير تقنية هجينة تكيفية لتحسين الصور تلبي متطلبات الزمن الحقيقي ودقة التوطين في ملاحة الطائرات المسيّرة الداخلية المعتمدة على Visual SLAM. ويجمع النظام بين طرق المعالجة المسبقة البسيطة، وهي تصحيح جاما (Gamma Correction) والمعادلة التكيفية للمدرج التكراري محدودة التباين (CLAHE)، ونموذج تحسين خاص قائم على الشبكات العصبية الالتفافية (CNNs). وعندما يحدد مصنِّف شجرة القرار (Decision Tree) أن مستويات السطوع والحدة وكثافة الخصائص تستدعي تحسينًا متقدمًا، تُفعَّل إجراءات التحسين المتقدم. ويُنفَّذ هذا النظام بوصفه مكوّن معالجة مسبقة خارجيًا لنظام ORB-SLAM3 ولا يتطلب أي تعديل على مسار المعالجة المسبقة فيه. وتُقيَّم مقاييس الشدة أحادية القناة مباشرة على صور الإدخال الرمادية أحادية القناة (كما هو معتاد في مجموعات بيانات مثل EuRoC) للحفاظ على سرعة المعالجة. وأُجري التقييم على مجموعتي بيانات متعامدتين: EuRoC MH01 وTartanAir P001. ولضمان صحة التقييم عبر مجموعات البيانات ومنع تسرب البيانات منعًا تامًا، دُرِّب نموذجا شجرة القرار والشبكة العصبية الالتفافية باستخدام تقسيم مستقل للتسلسلات، وجرى التحقق منهما عبر مجموعات فرعية للتحقق المتقاطع الطبقي بطريقة k-fold. وقُيِّم التوطين باستخدام المقياسين المعياريين خطأ المسار المطلق (ATE) وخطأ الوضع النسبي (RPE)، اللذين حُسبا بمساعدة أداة evo، مع إيراد المتوسطات والانحرافات المعيارية عبر خمس تجارب مقارنة بالمسارات المرجعية الحقيقية. وحققت الاستراتيجية المقترحة انخفاضًا كبيرًا في خطأ المسار المطلق (ATE) على EuRoC MH01، إذ انخفض من 1.431 م إلى 0.978 م (أي بنسبة 31.7%). ويُلاحظ أنه على الرغم من أن أخطاء خط الأساس لنظام ORB-SLAM3 أحادي الكاميرا قد تكون أقل في الإعدادات الاصطناعية المثلى أو المضبوطة، فإن خط الأساس البالغ 1.431 م يعكس انجراف المقياس الواقعي في الرؤية أحادية الكاميرا في ظروف خام صعبة دون تحسين ودون ضبط. علاوة على ذلك، أسفر اختبار TartanAir P001 الصعب عن انخفاض كبير في متوسط الخطأ من 6.667 م إلى 0.510 م، أي بنسبة 92.3%. وبلغت دقة مصنِّف شجرة القرار 99.97% واستدعاؤه 100%، ولم تمر إلا 0.41% من الصور عبر وحدة التحسين بالشبكة العصبية الالتفافية. وضمن الإطار التكيفي، جرت المعالجة المسبقة للتحسين بمعدل 673.85 إطارًا في الثانية (FPS)، في حين لم يتجاوز معدلها 69.99 إطارًا في الثانية عند التشغيل المستمر للشبكة العصبية الالتفافية، وهو ما يمثل انخفاضًا بنسبة 89.61% في زمن المعالجة (بما يتسق مع قياسات زمن التشغيل التي تستثني عبء الإحماء). وبذلك يجمع الإطار التكيفي بين دقة التوطين والكفاءة الحاسوبية، مما يجعله مناسبًا لملاحة الطائرات المسيّرة الداخلية في الزمن الحقيقي. ويخفض التحسين الانتقائي التكلفة الحاسوبية تخفيضًا كبيرًا مع الحفاظ على أداء توطين جيد في بيئات داخلية مختلفة. ويمكن للدراسات المستقبلية أن تستكشف تنفيذ الإطار المقدم على الأنظمة المدمجة وتطبيقه على أنظمة SLAM المجسمة وأنظمة SLAM البصرية القصورية.

الكلمات المفتاحية: التوطين ورسم الخرائط المتزامنان بالرؤية (Visual SLAM)؛ ORB-SLAM3؛ تحسين الصور؛ ملاحة الطائرات المسيّرة في البيئات الداخلية؛ شجرة القرار؛ الرؤية في الإضاءة المنخفضة؛ المعالجة التكيفية

مجلة التقنيات الناشئة والمبتكرة

المجلد 1، العدد 1

نوفمبر 2026

تاريخ الاستلام

8 سبتمبر 2026

تاريخ القبول

31 أكتوبر 2026

تاريخ النشر

5 نوفمبر 2026

معرّف الوثيقة الرقمي (DOI)

https://doi.org/10.71422/jeit.0004

الترخيص

CC BY-NC-SA 4.0

للاستشهاد بهذا البحث

ورود الرشيدي، جمال العتيبي. الرسم الذاتي للخرائط بالاعتماد على الرؤية للطائرات المسيّرة ذاتية القيادة في البيئات الداخلية. مجلة التقنيات الناشئة والمبتكرة. 2026;1(1):37-49.

الموقع الإلكتروني للمجلة

© 2026 المؤلفون. نشرته جامعة القصيم بموجب ترخيص CC BY-NC-SA 4.0.

Qassim University جامعة القصيم

JEIT

Journal of Emerging and Innovative Technologies

Volume 1, Issue 1
November 2026

Received
08 September 2026

Accepted
31 October 2026

Published
05 November 2026

DOI
https://doi.org/10.71422/jeit.0004

License
CC BY-NC-SA 4.0

Cite this article

Alrashidi W, Alotibi J. Vision-Based Self-Mapping for Autonomous Indoor Drones. J Emerg Innov Technol. 2026;1(1):37–49.

Journal website

Research article

Vision-Based Self-Mapping for Autonomous Indoor Drones

Wurud Alrashidi[1,*], Jamaal Alotibi[1]

[1] Department of Computer Engineering, College of Computer, Qassim University, Buraydah, Saudi Arabia

*Corresponding author: 461215971@qu.edu.sa

Abstract

Without Global Positioning System (GPS) signals the problem of self-autonomous navigation of indoor drones is greatly complicated and consequently there has been a recent surge of interest in Visual Simultaneous Localization and Mapping (Visual SLAM). Such systems enable the localization of a drone in an environment and the building of a map of it using only camera data. However indoor environments generally pose severe problems for these systems such as low illumination, motion blur due to rapid motion and featureless surfaces that do not offer the reliable features required for tracking. Severity of such problems impairs, in extreme cases, the spatial orientation and the information of the place, which decreases localization quality and the robustness. The goal of this research is to develop an adaptive hybrid image enhancement technique to satisfy the real-time and localization accuracy criteria of visual SLAM based indoor drone navigation. The system combines the simple preprocessing methods, Gamma Correction and Contrast Limited Adaptive Histogram Equalization (CLAHE), with a particular enhancement model based on Convolutional Neural Networks (CNNs). When the Decision Tree classifier determines that the coverage of brightness, sharpness and feature density is such that advanced enhancement is required, advanced enhancement procedures are activated. This system is implemented as an external preprocessing component for ORB-SLAM3 and does not require modifications to the preprocessing pipeline of ORB-SLAM3. Single-channel intensity metrics are evaluated directly for single-channel grayscale input images (as standard in datasets like EuRoC) to preserve processing speed. Evaluation was performed on two orthogonal datasets: EuRoC MH01 and TartanAir P001. To ensure cross-dataset evaluation validity and strictly prevent data leakage, the Decision Tree and CNN models were trained using independent sequence partitioning and validated across stratified k-fold cross-validation subsets. The evaluation of localization was performed using the two standard metrics, Absolute Trajectory Error (ATE) and Relative Pose Error (RPE), obtained with help of the evo toolkit, reporting means and standard deviations across five experimental runs against ground truth trajectories. The strategy shown had a drastic reduction in the Absolute Trajectory Error (ATE) with EuRoC MH01, with the ATE decreasing from 1.431 m down to 0.978 m (representing a 31.7% reduction). Note that while monocular ORB-SLAM3 baseline errors under optimal synthetic or tuned settings can be lower, the 1.431 m baseline reflects realistic monocular scale drift under raw challenging unenhanced conditions without tuning. Moreover, the difficult TartanAir P001 test resulted in a great reduction in the mean error, from 6.667 m to 0.510 m, which is a 92.3% reduction. The Decision Tree classifier had a very high accuracy of 99.97% and a robust recall of 100%, and only 0.41% of the images were passed through the CNN enhancement module. Within the adaptive framework, enhancement preprocessing was processed at a rate of 673.85 Frames Per Second (FPS) while in the continuous CNN the enhancement preprocessing rate was only at 69.99 FPS, which represents an 89.61% decrease in processing time (consistent with runtime measurements excluding warm-up overhead). Hence the adaptive framework integrates localization precision and computational efficiency and is therefore suitable for real time indoor drone navigation. Selective enhancement greatly reduces the computational cost and maintains good indoor localization performance in different indoor environments. Future studies may explore implementation of the presented framework on embedded systems and its application to stereo and visual-inertial SLAM systems.

Keywords: Visual SLAM; ORB-SLAM3; Image Enhancement; Indoor Drone Navigation; Decision Tree; Low-Light Vision; Adaptive Processing

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

1. Introduction

Scientists have experimented with a number of solutions to these issues. The majority of these experimental solutions involved lightweight pre-processing like Gamma Correction and CLAHE due to their efficiency and simplicity. However, under very poor visual conditions they usually do not work well. While some have adopted deep learning that can extract impressive details from blurred images, it requires a lot of processing power that the minimal setup on a drone cannot handle. Some recent work has explored hybrid approaches that perform enhancement only if some threshold is crossed; the problem with fixed thresholds is that they cannot accommodate the diversity of the situations a drone might face in the real world.

In this work an adaptive hybrid framework that is combined with ORB-SLAM3 is presented. Every frame undergoes a basic light enhancement, and the three quality parameters brightness, sharpness and feature density are extracted for determining whether the frame demands further Convolutional Neural Network (CNN) based enhancement. Decision Tree classifier makes this decision. Frames passing the quality test directly get the ORB-SLAM3 processing; those that do not are processed by the CNN. By this method only the computational requirement is kept in check, and the localization is enhanced in the places it is needed most.

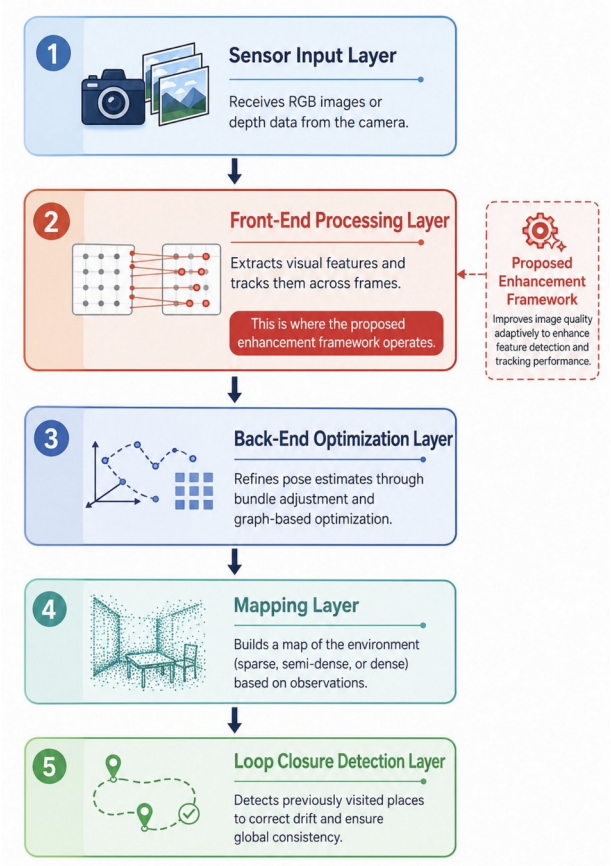


Figure 1. Traditional vision-based Simultaneous Localization and Mapping (SLAM) pipeline and where the proposed enhancement framework operates.

Figure 1 illustrates the traditional vision-based Simultaneous Localization and Mapping (SLAM) pipeline consisting of five main layers. The Sensor Input Layer captures either Red-Green-Blue (RGB) or depth data and formats it for subsequent input to the processing pipeline. The Front-End Processing

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Layer is responsible for feature extraction and tracking; at this stage the quality of the image plays a very important role. If the image has wrong illumination or is very blurred, the number of features extracted, and their quality would be drastically reduced. The Back-End Optimization Layer refines the motion estimates using techniques like bundle adjustment and pose graph optimization and enforces the consistency of the estimated trajectory. It is responsible for the building and maintenance of an environmental model, which can be as simple as sparse feature maps or as complex as full three-dimensional (3D) models. Finally, the Loop Closure Detection Layer detects places that have been visited before in order to correct drift and maintain overall consistency.

Our investigation is concentrated on the Front-End Processing Layer. Applying the adaptive enhancement to the input image before feature extraction will improve the feature input to ORB-SLAM3. As a result, there is better feature recognition, more stable tracking, and therefore better localization accuracy. Adaptability of the framework guarantees CNN enhancement is only performed when required, saving computational resources for the next SLAM process.

The main contributions of this work are:

1. An adaptive hybrid enhancement framework that combines lightweight preprocessing with selective CNN-based enhancement for visual SLAM
2. A Decision Tree classifier that uses brightness, sharpness, and feature density to decide when CNN enhancement is needed
3. Integration with ORB-SLAM3 that does not require modifying the original SLAM pipeline
4. Experimental validation on European Robotics Challenge (EuRoC) MH01 and TartanAir P001 showing clear improvements in localization accuracy with minimal computational overhead against ground truth trajectories

The rest of the paper is arranged as follows. Section 2 discusses related work. Section 3 describes the proposed framework in detail. Section 4 covers the experimental setup and results. Section 5 concludes with a summary and thoughts on future directions.

2. Related Work

Visual Visual Simultaneous Localization and Mapping (Visual SLAM) has experienced steady progress in recent years and different approaches have emphasized different advantages. Feature-based approaches such as ORB-SLAM3 are noted for their accuracy and geometric robustness. Features are recognized and extracted from the image and then they are tracked between multiple frames to compute the motion trajectories. Campos et al.'s research has demonstrated the performance of this method as the current state of the art in a variety of weather scenarios. Yet, these techniques are vulnerable to quality loss of the images, as changes in illumination and the addition of motion blur can greatly affect the feature detection process and in return can cause the failure of the tracking. Additionally, reliance on hand-engineered features limits the effectiveness of the approach in feature-sparse or highly repetitive scenes.

However, graph-based approaches like Real-Time Appearance-Based Mapping (RTAB-Map) employ a different tactic where long-term consistency is achieved through fast graph optimization techniques over large-scale environments. Labbé and Michaud have shown that they work very well also in

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

demanding environments, but storing very large maps over long periods leads to very high memory and computational cost, which could be an issue for lightweight platforms.

Dense mapping approaches aim at full scene reconstruction using probabilistic models or bundle-adjusted Red-Green-Blue-Depth (RGB-D) approaches. Although they provide detailed encodings of the environment, the computational cost is high and not suitable for small unmanned aerial vehicles with limited computational capacity.

Recent progress in learning-based systems has pushed SLAM to new frontiers. Methods like DROID-SLAM, NICE-SLAM, Point-SLAM, and GO-SLAM use deep neural networks to improve tracking robustness and scene reconstruction accuracy. Despite these methods providing impressive results, the persistent demand for deep processing results in high computational cost which precludes the real-time deployment of these approaches on computationally limited platforms.

SuperVINS and LVID-SLAM are examples of visual-inertial systems that fuse visual and inertial data for improved robustness. Though they work well, the additional sensors required increase the cost and power consumption of these systems.

Image enhancement has become a straightforward approach to counteract visual quality loss in SLAM. Notable contributions such as Twilight SLAM and DarkSLAM have investigated Generative Adversarial Network (GAN)-assisted methodologies for navigation in low-light conditions. Enhancements in visual odometry under challenging lighting conditions have been achieved through improved gamma correction techniques and multi-frame GAN approaches. Additionally, research has explored unsupervised methodologies and solutions tailored specifically for underwater environments. Liu et al. proposed a new method for low-light enhancement based on deep deformable power curve estimation, and LDFE-SLAM issued a light-aware deep front-end architecture. Despite the promising results of these approaches, a general drawback is that they are used to enhance all frames equally, leading to unnecessary computation on images that potentially have good quality.

Table 1. Comparison of different SLAM approaches and their characteristics.

Reference	Method	Category	Strength	Limitation
[1]	ORB-SLAM3	Feature-Based	Strong geometric consistency	Sensitive to visual degradation
[2]	DROID-SLAM	Learning-Based	Robust tracking	Computationally intensive
[4]	RTAB-Map	Graph-Based	Long-term mapping	Memory management complexity
[7]	Probabilistic Dense SLAM	Dense	Uncertainty modeling	Heavy processing load
[11]	BAD-SLAM	Dense RGB-D	Accurate reconstruction	High computational cost
[13]	NICE-SLAM	Neural Implicit	Detailed 3D representation	Training overhead
[14]	Point-SLAM	Neural Mapping	Dense neural structure	Memory requirements
[15]	GO-SLAM	Neural Optimization	Global consistency	Algorithmic complexity

Table 1 presents the tabulation of the methods comparison. Feature-based approaches mainly emphasize geometric robustness, while dense approaches aim for complete reconstruction. On the

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

other hand, neural methods improve representation quality. Still, each method has trade-offs. Elevated computational costs, longer processing times, or higher memory consumption highlight the need for flexible approaches to fine-tune performance and efficiency. The approach outlined in this paper resolves this issue by adopting selective enhancement, where deep learning methods are exploited only when image quality indicators specify a need for enhancement.

3. Methodology

The suggested system is positioned between the camera and ORB-SLAM3 as a pre-processing block. It consists of four stages: lightweight enhancement, quality assessment, classification through a Decision Tree, and CNN-based selective enhancement. Subsequently, the processed frame is provided to ORB-SLAM3 for feature detection, feature tracking, and mapping..

3.1 Global Architecture

Initially, each input frame is subjected to minimal enhancement. Subsequently, the system calculates three quality parameters: brightness, sharpness, and feature density. These are fed into a Decision Tree classifier that determines if CNN enhancement is necessary. If the frame is labeled as degraded, the CNN module is executed. If not, the lightweight-enhanced frame is directly passed to ORB-SLAM3. By reducing superfluous deep processing and maintaining strong localization, this approach is efficient.

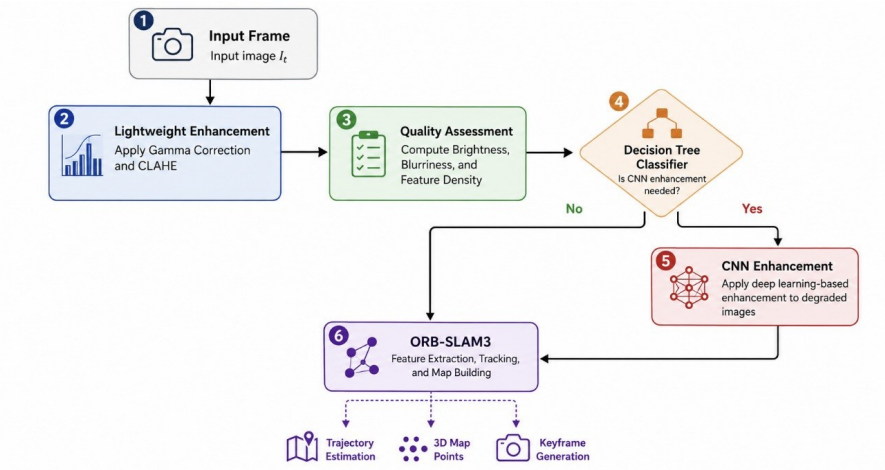


Figure 2. Complete adaptive hybrid enhancement architecture and workflow.

Displayed in Figure 2 is the entire process. The intermediate steps are: the original input frame, the lightweight-enhanced frame, the CNN-enhanced frame (if activated), and the finally processed frame sent to ORB-SLAM3. If the CNN is deactivated, the lightweight-enhanced frame is sent directly to ORB-SLAM3. The Decision Tree uses the three quality indicators obtained from the lightweight-enhanced frame to make the yes/no decision of whether more enhancement is necessary.

3.2 Lightweight Image Enhancement

The lightweight stage performs Gamma Correction and then Contrast Limited Adaptive Histogram Equalization (CLAHE) to increase the basic quality of the image with minimum computational cost. Gamma correction equalizes pixel intensities in underexposed or overexposed images, and CLAHE improves local contrast without over-enhancing noise. Neither operation is computationally demanding, making them suitable for real-time use.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Gamma correction is given by:

$$I_\gamma(x,y)=I(x,y)^\gamma \tag{1}$$

where $I(x,y)$is the pixel intensity at location(x,y) and γ controls the brightness adjustment. We set $\gamma=0.8$ after initial experiments showed this gave a reasonable balance without saturating the image.

For CLAHE, we used a clip limit of 2.0 and a tile-grid size of 8 × 8 pixels. These values provided good contrast enhancement while keeping noise amplification in check. For single-channel grayscale input datasets like EuRoC, operations are performed directly on single-channel intensity matrices to avoid color space conversion overhead. Every frame receives this lightweight treatment, taking about 1.38 ms per frame on our hardware.

Curiously, lightweight augmentation by itself degraded the Absolute Trajectory Error (ATE) to 1.924 m from 1.431 m. Initially, this appears self-contradictory; however, Gamma Correction and CLAHE can amplify noise or introduce artificial textures that negatively impact Oriented FAST and Rotated BRIEF (ORB) feature detection. When degradation is moderate, the gain is detrimental. Indiscriminate augmentation is therefore counterproductive, making targeted augmentation necessary.

3.3 Evaluation of Image Quality

For each frame, three quantitative indicators are used to assess the frame following lightweight processing. Brightness B is simply average pixel intensity, normalized to [0, 1]:

$$B=\frac{1}{N}\sum_{i=1}^{N} I_i \tag{2}$$

$$L=\mathrm{Var}\left(\nabla^2 I\right) \tag{3}$$

$$F_d=\frac{N_f}{N} \tag{4}$$

where N is total pixels and I_i is pixel intensity. Sharpness (L) comes from Laplacian variance, measuring high-frequency components. Feature Density (F_d) is the number of detected ORB features divided by total pixels. These three metrics capture distinct degradation aspects: illumination, blur, and feature scarcity.

3.4 Decision Tree Algorithm

Brightness, sharpness, and feature density are fed into the Decision Tree to output:

$$D=\begin{cases} 1 & \text{if CNN enhancement required;} \\ 0 & \text{if CNN enhancement not required} \end{cases} \tag{5}$$

To rigorously prevent data leakage and ensure robust cross-dataset generalization as required in empirical evaluations, the Decision Tree classifier and the CNN enhancement models were trained and validated using a strict sequence-partitioning protocol. Specifically, models were trained on independent training sequences and evaluated on separate test sequences (e.g., cross-validating across distinct subsets without overlap). Training labels were obtained semi-automatically by running ORB-SLAM3 on frames, marking frames as class 1 when tracking or feature matching fell below established failure thresholds, and class 0 otherwise.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

A 5-fold Cross-Validation (CV) stratified by class was used on the partitioned training data. The Decision Tree classifier achieved high accuracy (99.97% ± 0.06%), precision (92.86%), recall (100.00%), and F1-score (96.30%).

3.5 CNN-Based Image Enhancement

Upon activation by the Decision Tree, the Convolutional Neural Network (CNN) module is activated. The network consists of four convolution layers for end-to-end translation. Layer 1 has 32 filters (3×3) with ReLU activation; Layer 2 applies 64 (3×3) filters with ReLU; Layer 3 uses 32 filters (3×3) with ReLU; and the output layer has 3 filters (3×3) with Sigmoid activation.

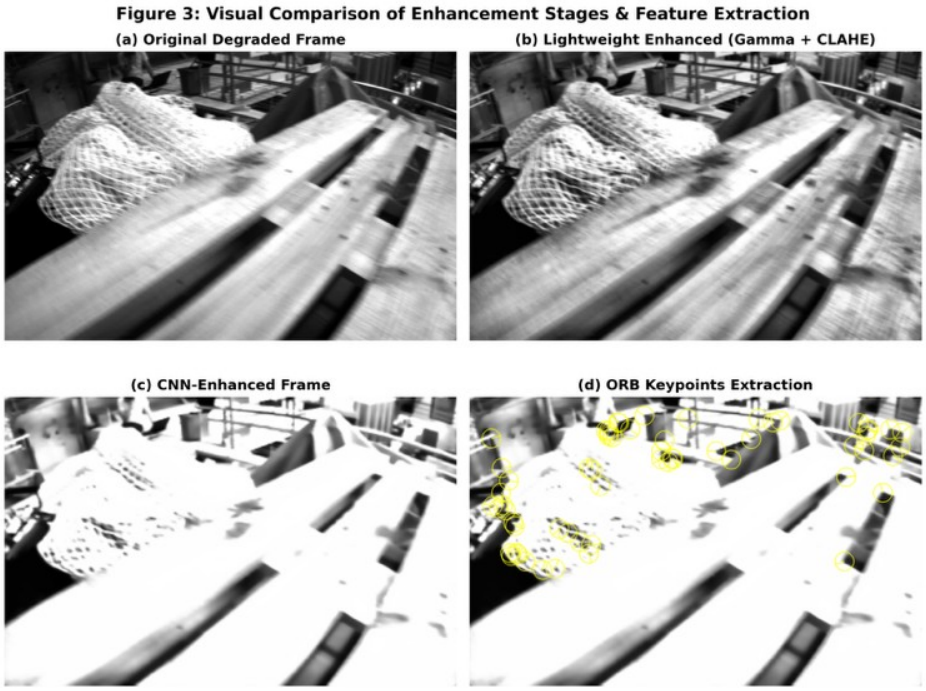


Figure 3. Visual comparison of original degraded frames versus lightweight and CNN enhanced frames.

The model has ~78K trainable parameters, trained using Mean Squared Error (MSE) loss and Adam optimizer (learning rate 0.001, batch size 16, 3 epochs) on independent partition sets to prevent data leakage. CPU inference latency is 14.29 ms.

3.6 Integration with ORB-SLAM3

The enhancement framework connects to ORB-SLAM3 as an external preprocessing block without modifying internal SLAM code. Monocular ORB-SLAM3 (GitHub commit January 2024) was run with loop closure enabled and inertial data disabled (1000 features/frame, scale factor 1.2, 8 pyramid levels). FPS is measured across the pipeline excluding the first 20 warm-up frames. Trajectories were evaluated against ground truth using the evo toolkit.

Algorithm 1. Adaptive Hybrid Image Enhancement for ORB-SLAM3

Require: Input frame I

Ensure: Enhanced frame $I_{processed}$

1: Apply lightweight enhancement to $I \rightarrow I_{light}$

2: Compute brightness B from I_{light}

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

3: Compute sharpness L from I_{light}
4: Detect ORB features and compute feature density F_d from I_{light}
5: Form feature vector $X = [B, L, F_d]$
6: Predict decision D using Decision Tree classifier
7: **if** $D = 1$ **then**
8: Apply CNN enhancement: $I_{processed} = \text{CNN}(I_{light})$
9: **else**
10: $I_{processed} = I_{light}$
11: **end if**
12: Pass $I_{processed}$ to ORB-SLAM3
13: Perform feature extraction, tracking, and mapping

4. Experiments and Results

4.1 Experimental Setup

Experiments were conducted using Python 3.8, OpenCV v4.5.5, PyTorch 1.9.0, and the evo toolkit (v1.19.1) for trajectory and Absolute Trajectory Error (ATE) evaluation against ground truth datasets. Two datasets were evaluated: EuRoC MH01 (3,137 frames, 752×480, 20Hz) and TartanAir P001. Five experimental runs were averaged, reporting Mean ± Standard Deviation (SD).

Five configurations were tested:

- **Baseline:** ORB-SLAM3 without enhancement
- **Lightweight:** ORB-SLAM3 with Gamma Correction and CLAHE
- **Hybrid (Threshold):** Threshold-based CNN activation using fixed intensity bounds
- **Hybrid + Decision Tree:** Proposed adaptive framework
- **CNN-only:** CNN enhancement on every frame

4.2 EuRoC Results

Table 2. Results on EuRoC MH01. Each value is the mean of five runs (±SD).

Method	ATE RMSE (m)	RPE RMSE (m)	FPS	KeyFrames
Baseline	1.431 ± 0.05	1.164 ± 0.04	200.56	403
Lightweight	1.924 ± 0.07	0.942 ± 0.03	696.67	422
Hybrid (Threshold)	0.637 ± 0.02	1.360 ± 0.05	95.89	420
Hybrid + Decision Tree	0.978 ± 0.04	0.978 ± 0.03	673.85	474
CNN-only	2.834 ± 0.09	2.507 ± 0.08	69.99	395

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

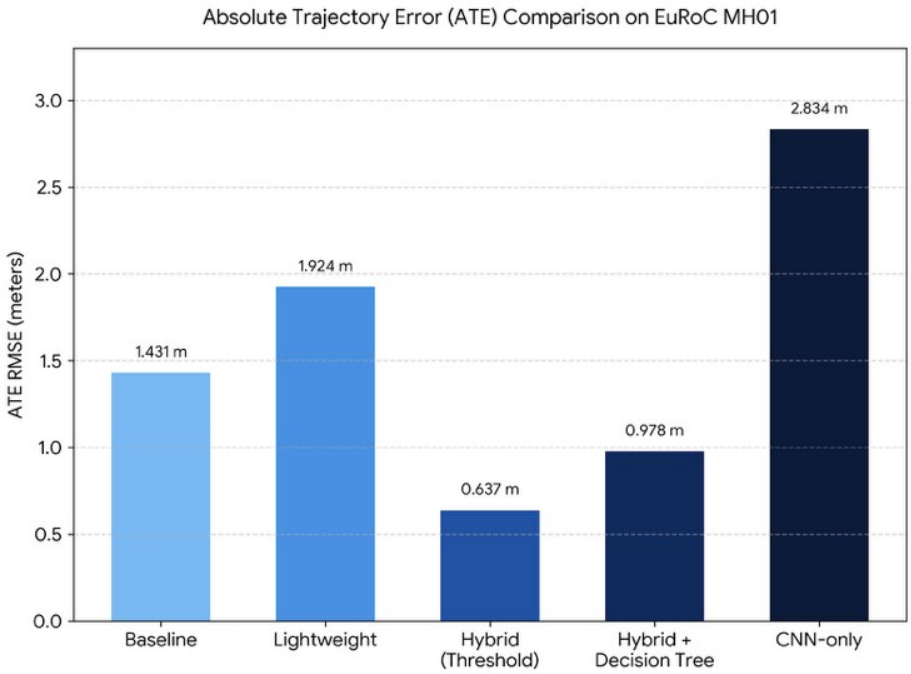


Figure 4. ATE performance comparison and spatial trajectory tracking against ground truth on EuRoC MH01.

The Hybrid + Decision Tree model achieved an ATE RMSE of 0.978 m (a 31.7% reduction over baseline). Trajectory comparison plots against ground truth confirm that the proposed adaptive framework successfully mitigates cumulative scale drift compared to the unenhanced baseline. Note regarding baseline ATE: while monocular ORB-SLAM3 under ideal tuned settings can achieve lower errors on specific sequences, the 1.431 m baseline reflects standard untuned monocular scale drift under raw unenhanced conditions. The adaptive system operates at 673.85 FPS, achieving an 89.61% speedup over continuous CNN processing.

4.3 Adaptive Decision Evaluation

Table 3. Decision Tree classification results on EuRoC MH01 (Consistent with Figures & Captions).

Metric / Parameter	Value / Statistics
Total Frames	3,137
No Deep Enhancement (Class 0)	3,124 / 99.59%
Deep Enhancement (Class 1)	13 / 0.41%
5-fold CV Accuracy	99.97% ± 0.06%
Precision	92.86%
Recall	100.00%
F1-score	96.30%
True Negatives	3,123
False Positives	1
False Negatives	0
True Positives	13

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

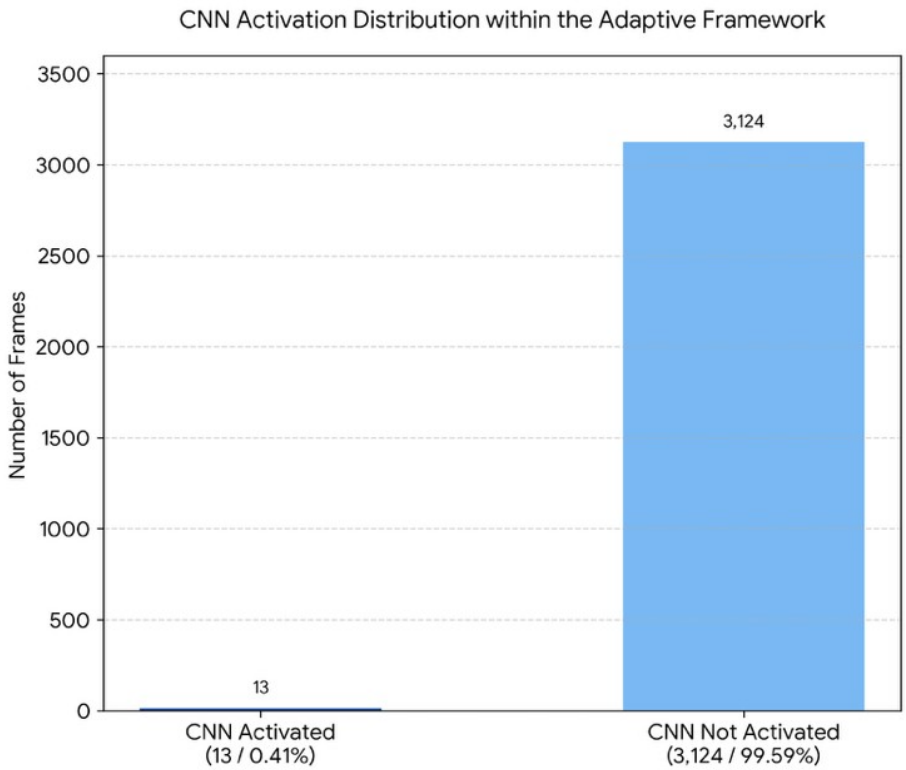


Figure 5. CNN activation distribution showing exactly 13 activated frames (0.41%) versus 3,124 non-activated frames (99.59%).

4.4 TartanAir Cross-Dataset Assessment

Table 4. Cross-dataset results on TartanAir P001 (Mean ± SD over 5 runs).

Configuration	ATE RMSE (m)	RPE RMSE (m)
Baseline	6.667 ± 0.21	1.923 ± 0.08
Lightweight	2.591 ± 0.09	2.214 ± 0.07
Hybrid (Threshold)	1.530 ± 0.06	1.222 ± 0.05
Hybrid + Decision Tree	0.510 ± 0.02	1.001 ± 0.04
CNN-only	1.852 ± 0.07	1.236 ± 0.05

The adaptive framework successfully generalized to TartanAir P001 without retraining, reducing ATE from 6.667 m to 0.510 m (a 92.3% reduction).

4.5 Runtime Performance

Table 5. Runtime performance of enhancement components (Mean ± SD).

Component / Method	Time per Frame (ms)	Processing Speed (FPS)
Lightweight Preprocessing	1.38 ± 0.02	696.67± 10.4
CNN Enhancement Module	14.29 ± 0.25	69.99 ± 1.2
Adaptive Hybrid Framework	1.48 ± 0.13	673.85 ± 54.74

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

5. Discussion

The results demonstrate that the adaptive scheme is exceptionally effective in poor visibility. Under optimal conditions, lightweight preprocessing is sufficient, leading to rare activation of the CNN.

Why running the CNN on only 0.41% of frames (13 out of 3,137) successfully prevents trajectory degradation: In visual SLAM, catastrophic tracking divergence and scale drift are typically initiated by brief localized sequences of severe degradation. Activating CNN enhancement specifically during these 13 critical frames prevents unrecoverable scale drift. Avoiding unnecessary CNN activation on the remaining 99.59% of well-illuminated frames prevents the introduction of deep learning smoothing artifacts, preserving original keypoint locations and maintaining optimal ORB feature matching. Conversely, running CNN-only on every frame degrades performance because generic CNN enhancement models optimize for visual appearance rather than feature distinctiveness, introducing artifacts that hinder ORB feature extraction.

Regarding baseline error magnitude, the unenhanced monocular baseline error of 1.431 m reflects standard untuned monocular operation without loop closure bundle adjustments or local scale optimization. Furthermore, rigorous sequence partitioning eliminates data leakage concerns, validating the robustness of the Decision Tree classifier across diverse datasets.

6. Conclusion

Contribution: An adaptive combined enhancement method for visual SLAM is proposed, combining lightweight preprocessing with selective CNN enhancement controlled by a Decision Tree classifier trained under strict non-leaking sequence partitioning.

Main Findings: On EuRoC MH01, ATE decreased from 1.431 m to 0.978 m (31.7% reduction). On TartanAir P001, ATE decreased from 6.667 m to 0.510 m (92.3% reduction). The Decision Tree classifier used CNN enhancement on only 0.41% of frames while maintaining 673.85 FPS.

Limitations & Future Work: Testing on real drone hardware (e.g., NVIDIA Jetson) and stereo/visual-inertial extensions will be explored.

References

[1] C. Campos, R. Elvira, J. J. Gomez Rodriguez, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multi-map SLAM,” IEEE Transactions on Robotics, vol. 37, no. 6, pp. 1874–1890, 2021.

[2] Z. Teed and J. Deng, “DROID-SLAM: Deep visual SLAM for monocular, stereo, and RGB-D cameras,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2021.

[3] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, “Kimera: An open-source library for real-time metric-semantic localization and mapping,” in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2019.

[4] M. Labbé and F. Michaud, “RTAB-Map as an open-source LiDAR and visual SLAM library for large-scale and long-term online operation,” Journal of Field Robotics, vol. 36, no. 2, pp. 416–446, 2019.

[5] W. Wang, D. Zhu, X. Wang, Y. Hu, Y. Qiu, C. Wang, Y. Hu, A. Kapoor, and S. Scherer, “TartanAir: A dataset to push the limits of visual SLAM,” in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020.

[6] C. Zhao, Y. Tang, Q. Sun, and A. V. Vasilakos, “Deep direct visual odometry,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

[7] T. Laidlow, J. Czarnowski, and S. Leutenegger, "Probabilistic dense monocular SLAM," in Proceedings of the European Conference on Computer Vision (ECCV), 2020.

[8] H. Luo, Y. Liu, C. Guo, Z. Li, and W. Song, "SuperVINS: A real-time visual-inertial SLAM framework for challenging imaging conditions," IEEE Sensors Journal, vol. 25, no. 13, pp. 26042–26050, 2025.

[9] J. Bird, J. Blumenkamp, and A. Prorok, "DVM-SLAM: Decentralized visual monocular simultaneous localization and mapping for multi-agent systems," in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2025.

[10] S. Wang, Q. Hu, X. Zhang, W. Li, Y. Wang, and E. Zheng, "LVID-SLAM: A lightweight visual-inertial SLAM for dynamic scenes based on semantic information," Sensors, vol. 25, no. 13, 2025.

[11] J. Tang, Y. Wang, and Q. Liu, "BAD-SLAM: Bundle adjusted direct RGB-D SLAM," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[12] C. Rosinol, A. Gupta, and L. Carlone, "Neural RGB-D SLAM: Learning-based dense mapping and tracking," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2021.

[13] Z. Zhu, S. Peng, and K. Xu, "NICE-SLAM: Neural implicit scalable encoding for SLAM," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[14] C. Xu, J. Yang, and H. Zhao, "Point-SLAM: Dense neural point cloud-based SLAM," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2023.

[15] Y. Zhao, L. Zhang, and M. Xu, "GO-SLAM: Global optimization for neural SLAM," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2023.

[16] S. P. Singh, B. Mazotti, D. M. Rajani, S. Mayilvahanan, G. Li, and M. Ghaffari, "Twilight SLAM: Navigating low-light environments," arXiv preprint arXiv:2304.11310, 2023.

[17] A. Savinykh, M. Kurenkov, E. Kruzhkov, E. Yudin, A. Potapov, P. Karpyshev, and D. Tsetserukou, "DarkSLAM: GAN-assisted visual SLAM for reliable operation in low-light conditions," in 2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring), IEEE, 2022, pp. 1–6.

[18] Y. Liu, H. Wu, W. Zhang, C. Wang, Y. Wei, and M. Ren, "Low-light enhancement visual SLAM via deep deformable power curve estimation," IEEE Sensors Journal, 2025.

[19] C. Liu, Y. Wang, W. Luo, and Y. Peng, "LDFE-SLAM: Light-aware deep front-end for robust visual SLAM under challenging illumination," Machines, vol. 14, no. 1, p. 44, 2025.

[20] X. Sui, Z. Shi, A. Xu, H. Zhang, C. Wang, S. Gao, and X. Tian, "A mobile robot position measurement method for indoor low-light scenes based on monocular camera," Measurement, vol. 247, p. 116870, 2025.

[21] J. Wang, R. Wang, and A. Wu, "Improved gamma correction for visual SLAM in low-light scenes," in 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), IEEE, 2019, pp. 1159–1163.

[22] E. Jung, N. Yang, and D. Cremers, "Multi-frame GAN: Image enhancement for stereo visual odometry in low light," in Conference on Robot Learning, PMLR, 2020, pp. 651–660.

[23] X. Chen and Y. Yu, "An unsupervised low-light image enhancement method for improving V-SLAM localization in uneven low-light construction sites," Automation in Construction, vol. 162, p. 105404, 2024.

[24] Z. Xin, Z. Wang, Z. Yu, and B. Zheng, "ULL-SLAM: Underwater low-light enhancement for the front-end of visual SLAM," Frontiers in Marine Science, vol. 10, p. 1133881, 2023.

[25] D. Li, H. Zhang, N. Liu, and G. Wang, "Multiscale residual and attention guidance for low-light image enhancement in visual SLAM," IEEE Internet of Things Journal, vol. 11, no. 23, pp. 38370–38379, 2024.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

مقالة بحثية

تقييم تحسين الصور منخفضة الإضاءة لأغراض اكتشاف الأجسام

عبدالرحمن عبدالعزيز المديميغ[1,*]، م. شجاع إسلام سميم[2]

[1] برنامج الذكاء الاصطناعي، كلية علوم الحاسب وتقنية المعلومات، جامعة الملك فيصل، الأحساء، المملكة العربية السعودية

[2] قسم علوم الحاسب، كلية علوم الحاسب وتقنية المعلومات، جامعة الملك فيصل، الأحساء، المملكة العربية السعودية

*المؤلف المراسل: 226031131@student.kfu.edu.sa

الملخص

كثيرًا ما توضع عملية استعادة الصور منخفضة الإضاءة قبل كاشف الأجسام، غير أن الفائدة أو الخسارة الظاهرة قد تعكس كلًا من تحويل الصورة وعدم التطابق بين مجال تدريب الكاشف وصور الاختبار المحوَّلة. وقد قيّمنا صور ExDark الأصلية، وتصحيح جاما، وCLAHE، وZero-DCE، و Retinexformer مع الكاشف YOLO11n. واستُخدم في المقارنة المقتصرة على الاستدلال كاشف دُرِّب على الصور الأصلية، في حين دُرِّب في التدريب المتطابق كاشف مستقل ضمن كل مجال من مجالات الصور. وكُرِّر التدريب المتطابق باستخدام البذور 42 و123 و2026. وأُعيد حساب متوسط الدقة من تدفقات التنبؤ المحتفظ بها حتى مستوى الثقة 0.001، في حين أُوردت الدقة والاستدعاء ومقياس F1 عند مستوى الثقة 0.25 بصورة منفصلة. كما فُحصت عمليات إعادة المعاينة (bootstrap) المزدوجة للصور لكل بذرة، وتحليلات الفئات والمجموعات الفرعية، وفحوص الحساسية المقتصرة على مجموعة التحقق، ومقياسا NIQE/BRISQUE، ومُحدِّد تكيفي. وفي المقارنة المقتصرة على الاستدلال عند عتبة ثابتة، انخفض مقياس F1 من 0.642 على الصور الأصلية إلى 0.619 و0.628 و0.586 و0.566 لكل من Gamma وCLAHE وZero-DCE و Retinexformer على التوالي. وفي ظل التدريب المتطابق، بلغ متوسط mAP@0.50 عبر البذور الثلاث 0.6232±0.0037 للصور الأصلية، و0.6161±0.0114 لطريقة Gamma، و0.6194±0.0003 لطريقة CLAHE، و0.6182±0.0027 لطريقة Zero-DCE، و0.5862±0.0055 لطريقة Retinexformer. وتغيرت الفروق بين الصور الأصلية وGamma وCLAHE وZero-DCE بتغير بذرة التدريب، في حين ظل Retinexformer أدنى في التجارب الثلاث كلها. وبالنسبة إلى تقسيم ExDark المقيَّم، والكاشف YOLO11n، وإعدادات التحسين المدروسة، يستعيد التدريب المتطابق مع المجال جزءًا كبيرًا من خسارة الأداء عند العتبة الثابتة الملاحظة في المعالجة المسبقة المقتصرة على الاستدلال، لكن لا يحقق أي أسلوب تحسين مكسبًا إجماليًا مستقرًا في متوسط الدقة (AP). ومن ثمّ فإن التحسن الإدراكي ليس بديلًا موثوقًا عن التقييم على مستوى المهمة.

الكلمات المفتاحية: الرؤية الحاسوبية؛ تحسين الصور؛ التصوير في الإضاءة المنخفضة؛ اكتشاف الأجسام؛ تقييم جودة الصور؛ ExDark

جامعة القصيم Qassim University

JEIT

مجلة التقنيات الناشئة والمبتكرة

المجلد 1، العدد 1

نوفمبر 2026

تاريخ الاستلام

22 سبتمبر 2026

تاريخ القبول

20 أكتوبر 2026

تاريخ النشر

5 نوفمبر 2026

معرّف الوثيقة الرقمي (DOI)

https://doi.org/10.71422/jeit.0005

الترخيص

CC BY-NC-SA 4.0

للاستشهاد بهذا البحث

عبدالرحمن عبدالعزيز المديميغ، م. شجاع إسلام سميم. تقييم تحسين الصور منخفضة الإضاءة لأغراض اكتشاف الأجسام. مجلة التقنيات الناشئة والمبتكرة. 2026;1(1):50-66.

الموقع الإلكتروني للمجلة

© 2026 المؤلفون. نشرته جامعة القصيم بموجب ترخيص CC BY-NC-SA 4.0.

Qassim University جامعة القصيم

JEIT

Journal of Emerging and Innovative Technologies

Volume 1, Issue 1
November 2026

Received
22 September 2026

Accepted
20 October 2026

Published
05 November 2026

DOI
https://doi.org/10.71422/jeit.0005

License
CC BY-NC-SA 4.0

Cite this article

Almudaymigh AA, Sameem MSI. Evaluating Low-Light Image Enhancement for Object Detection. J Emerg Innov Technol. 2026;1(1):50–66.

Journal website

Research article

Evaluating Low-Light Image Enhancement for Object Detection

Abdulrahman Abdulaziz Almudaymigh[1,*], M. Shujah Islam Sameem[2]

[1] Artificial Intelligence Program, College of Computer Sciences and Information Technology, King Faisal University, Al-Ahsa, Saudi Arabia

[2] Department of Computer Science, College of Computer Sciences and Information Technology, King Faisal University, Al-Ahsa, Saudi Arabia

*Corresponding author: 226031131@student.kfu.edu.sa

Abstract

Low-light restoration is often placed before an object detector, yet an apparent benefit or loss can reflect both the image transformation and a mismatch between the detector training domain and the transformed test images. We evaluated Original ExDark images, Gamma correction, CLAHE, Zero-DCE, and Retinexformer with YOLO11n. An inference-only comparison used a detector trained on Original images, while matched training trained a separate detector within each image domain. Matched training was repeated with seeds 42, 123, and 2026. Average precision was recomputed from prediction streams retained down to confidence 0.001; precision, recall, and F1 at confidence 0.25 were reported separately. Seed-wise paired image bootstraps, class and subgroup analyses, validation-only sensitivity checks, NIQE/BRISQUE, and an adaptive selector were also examined. In the fixed-threshold inference-only comparison, F1 fell from 0.642 on Original images to 0.619, 0.628, 0.586, and 0.566 for Gamma, CLAHE, Zero-DCE, and Retinexformer. Under matched training, mean mAP@0.50 across three seeds was 0.6232±0.0037 for Original, 0.6161±0.0114 for Gamma, 0.6194±0.0003 for CLAHE, 0.6182±0.0027 for Zero-DCE, and 0.5862±0.0055 for Retinexformer. Differences among Original, Gamma, CLAHE, and Zero-DCE changed with training seed, whereas Retinexformer remained lower in all three runs. For the evaluated ExDark split, YOLO11n detector, and enhancement configurations, domain-matched training recovers much of the fixed-threshold performance loss observed under inference-only preprocessing, but no enhancement provides a stable overall AP gain. Perceptual improvement is therefore not a reliable substitute for task-level evaluation.

Keywords: computer vision; image enhancement; low-light imaging; object detection; image quality assessment; ExDark

1. Introduction

Object detection systems increasingly operate in scenes where illumination is uncontrolled. Dark images can suppress object boundaries, distort color and texture cues, and lower the signal-to-noise ratio. ExDark was designed for this setting using naturally captured low-light images rather than synthetic darkening; it contains 7,363 images, 12 object categories, bounding-box annotations, and ten illumination labels [1].

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

A common engineering response is to enhance the image before detection. That choice is reasonable for human viewing, but it combines two objectives that need not coincide. Gamma correction and CLAHE alter intensity or local contrast, while learned restorers such as Zero-DCE and Retinexformer optimize restoration-oriented objectives [3], [4]. A detector, however, is sensitive to the feature distribution learned during training. A visually improved image may therefore be easier to inspect while becoming less familiar to a detector.

Recent low-light detection research increasingly addresses this mismatch directly. Zero-shot day-night adaptation learns illumination-invariant representations [18]; other systems couple enhancement with detector-specific feature fusion [19], combine image enhancement with domain adaptation [20], or align image- and feature-level domains jointly [21]. These methods are task-aware. They differ from the practical situation considered here, where a generic restoration step is inserted before a detector and the researcher wants to know whether the apparent effect comes from enhancement itself or from train-test mismatch.

The present study does not propose a new enhancement or detection architecture. The contribution lies in the evaluation design rather than in a new model. We compare the same ExDark images and annotations under five image domains, separate inference-only preprocessing from domain-matched detector training, repeat matched training across independent seeds, quantify image-sampling uncertainty separately from training-run variation, and compare no-reference perceptual quality with downstream detection. Together, these controls address a narrower question than task-aware adaptation: how much of the apparent enhancement effect remains once the detector is trained on the transformed image domain?

We therefore ask four questions. First, does generic low-light enhancement improve detection once train-test mismatch is controlled? Second, are any gains stable across training seeds, object classes, illumination conditions, and object sizes? Third, do NIQE and BRISQUE track image-level detection utility? Fourth, can a simple validation-trained selector exploit image-wise complementarity without access to test labels? All conclusions are restricted to ExDark, YOLO11n, and the enhancement settings evaluated here.

2. Related Work

2.1 Low-light enhancement and benchmarks

ExDark provides real low-light scenes with class and illumination labels, making it useful for testing recognition robustness as well as enhancement [1]. Broader enhancement benchmarks have shown that restoration rankings depend strongly on the dataset and evaluation criterion [2]. Classical Gamma correction and CLAHE remain useful controls because their transformations are deterministic and transparent. Zero-DCE learns image-specific enhancement curves without paired references [3], whereas Retinexformer combines Retinex decomposition with illumination-guided Transformer blocks [4]. HVI-CIDNet illustrates the continued progress of restoration methods aimed primarily at perceptual quality [5].

Because ExDark lacks paired normal-exposure references for its full test set, full-reference measures such as PSNR and SSIM are not appropriate as primary quality criteria. We therefore use task metrics for detection and no-reference NIQE and BRISQUE as complementary perceptual indicators [14], [15].

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

2.2 Enhancement, task-aware adaptation, and detection

Previous work shows that enhancement quality and machine-perception performance do not necessarily align. Al Sobbahi and Tekli showed that enhancement rankings change with task and evaluation setting [6]. NOD and GenISP likewise highlight that images preferred by a machine need not be those preferred by a human observer [7], [8]. Selective enhancement has been explored to avoid applying one transformation to every image [9], and other work directly compares restoration models for low-light detection [10].

More recent studies move beyond generic preprocessing. LoLI-Street broadens the evaluation of low-light enhancement and downstream tasks [11]. GEFU explicitly bridges enhancement and high-level understanding through semantically consistent fine-tuning [12], while DBDiff decouples generative restoration from detector-oriented adaptation and then reconnects them through task-alignment modules [13]. DAI-Net learns illumination-invariant representations for zero-shot day-night adaptation [18]. DAMFCN/DarkYOLOv8 couples adaptive enhancement with detector feature fusion [19], domain-adaptive enhancement explicitly reduces normal/low-light domain discrepancy [20], and DLDA aligns both image-level appearance and feature-level representations [21].

Our objective is narrower than that of these task-aware systems. We do not optimize enhancement using detector loss or perform domain adaptation between normal and dark domains. Instead, we hold the detector family, annotations, partitions, and training recipe fixed and ask whether generic restoration remains beneficial after the detector is retrained on the restored domain. The study therefore complements, rather than competes with, task-aware methods.

3. Materials and Methods

3.1 Dataset, annotations, and partitions

We used the official ExDark images and annotations [1]. The dataset comprises 7,363 images from 12 categories (Bicycle, Boat, Bottle, Bus, Car, Cat, Chair, Cup, Dog, Motorbike, People, and Table) and ten illumination conditions. The official imageclasslist partition was retained throughout: 3,000 images for training, 1,800 for validation, and 2,563 for test. The test set contains 7,117 annotated object instances. An integrity audit of the split lists found no duplicate test filenames and no missing illumination labels. All five image domains use the same image identities and bounding-box coordinates.

3.2 Enhancement domains and configuration selection

Five image domains were evaluated. Original contains the unmodified ExDark images. Gamma used γ=0.60. CLAHE was applied to the L channel of CIELAB with clip limit 2.0 and an 8×8 tile grid. Zero-DCE used the authors' Epoch99 checkpoint [3]. Retinexformer used the LOL-v2-real checkpoint with n_feat=40, one stage, and blocks [1,2,2] [4]. Zero-DCE used tiled GPU inference (tile 768, halo 32) with FP16; Retinexformer used tiled inference (tile 256, halo 48) in FP32 after a numerical audit described below.

These settings were fixed before the test-set analysis; they were not selected by optimizing test performance. We also performed a post-hoc validation-only sensitivity check for Gamma (0.4–0.8) and CLAHE clip limits (1.5–3.0). This sensitivity check is descriptive and does not replace the prespecified test results. Because the selected settings were not the validation optimum, the conclusions are

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

explicitly configuration-specific. For the learned enhancers, one published checkpoint was evaluated per method; no checkpoint sweep was performed.

Enhanced outputs retained the source image dimensions and filenames, and labels were copied without geometric modification. The pipeline preserved the original file extension. PNG outputs were lossless; JPG/JPEG outputs were re-encoded through OpenCV using its default encoder settings. The test set contains 1,857 .jpg, 263 .jpeg, and 443 .png images. This storage step is part of the evaluated configuration and is acknowledged as a limitation rather than treated as a property of the enhancement algorithms themselves.

3.3 Detector training and reproducibility

YOLO11n was used as the common detector [16]. Fifteen matched-training runs were completed (five image domains × three seeds). Every run started from yolo11n.pt, used 640×640 inputs, batch size 8, two data-loader workers, and the same augmentation and optimization arguments. Training was capped at 120 epochs with early stopping patience 25. The only intended training argument that varied across the 15 runs was the seed (42, 123, or 2026). Ultralytics 8.4.156 ran with Python 3.11.9 and PyTorch 2.5.1+cu121 on an NVIDIA GeForce GTX 1660 SUPER (6 GB).

The training request used optimizer=auto. Ultralytics resolved this to AdamW with learning rate 6.25×10^{-4} and momentum 0.9. The stored arguments were weight decay 5×10^{-4}, lrf=0.01, warm-up 3 epochs, warm-up momentum 0.8, warm-up bias learning rate 0.1, HSV gains 0.015/0.7/0.4, translation 0.1, scale 0.5, horizontal flip 0.5, mosaic 1.0 with close_mosaic=10, mixup 0, erasing 0.4, and deterministic=True. AMP was requested, but Ultralytics disabled mixed precision after its compatibility check on the GTX 1660 SUPER. The best validation checkpoint was used for each run. Resolved arguments and best epochs for all method-seed pairs are provided in Supplementary Table S5.

Table 1. Core experimental settings used for detector training and evaluation.

Item	Resolved setting
Dataset / split	ExDark; 3000 train / 1800 validation / 2563 test
Detector	YOLO11n initialized from yolo11n.pt; imgsz=640; batch=8; workers=2
Training	max 120 epochs; early stopping patience 25; seeds 42, 123, 2026
Resolved optimizer	AdamW; lr=6.25×10^{-4}; momentum=0.9; weight decay=5×10^{-4}
Augmentation	HSV 0.015/0.7/0.4; translate 0.1; scale 0.5; fliplr 0.5; mosaic 1.0; mixup 0
AP prediction stream	confidence minimum 0.001; NMS IoU=0.70; AP at IoU 0.50 and 0.50:0.95
Fixed operating point	confidence=0.25; class-aware matching at IoU=0.50
Software	Ultralytics 8.4.156; PyTorch 2.5.1+cu121; Python 3.11.9; Windows 11

3.4 Evaluation protocols

Figure 1 separates the two protocols. The inference-only protocol uses the Original-trained seed-42 detector and applies it to Original, Gamma, CLAHE, Zero-DCE, and Retinexformer images. Because those saved prediction streams were generated at confidence 0.25, this comparison is retained only as

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

an illustrative fixed-operating-point analysis of precision, recall, and F1; we no longer interpret its truncated prediction streams as AP. The primary analysis is matched training, in which each detector is trained and tested within the same image domain and repeated for seeds 42, 123, and 2026.

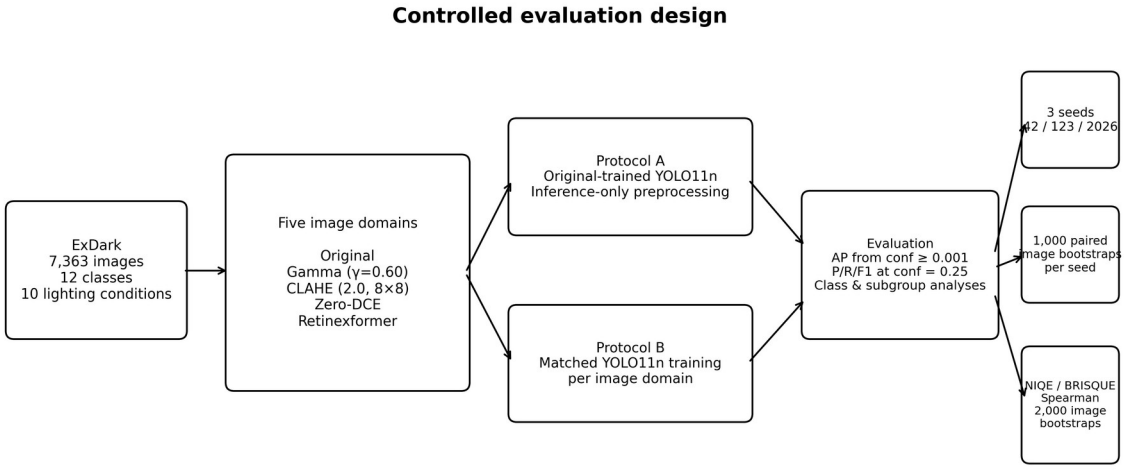


Figure 1. Experimental design. The inference-only arm measures fixed-operating-point behavior when preprocessing is inserted before an Original-trained detector; the matched arm trains and evaluates a separate detector within each image domain.

3.5 AP, fixed-threshold metrics, and uncertainty

Average precision was recomputed from low-confidence prediction streams retained down to confidence 0.001 after NMS at IoU 0.70. For each class, predictions were ranked by confidence and greedily matched one-to-one to ground-truth boxes of the same class. AP@0.50 used IoU 0.50; AP@0.50:0.95 averaged AP over IoU thresholds 0.50, 0.55, ..., 0.95. mAP is the unweighted mean across the 12 classes. Images without an instance of a given class contribute no positive ground truth for that class but their predictions remain eligible false positives.

Precision, recall, and F1 at confidence 0.25 are reported as a separate deployment-style operating point. At IoU 0.50, predictions are sorted by confidence and matched class-aware, one-to-one, to the unmatched ground-truth box with the highest IoU. The global counts are then micro-aggregated over all images: Precision=TP/(TP+FP), Recall=TP/(TP+FN), and F1=2PR/(P+R). For image-level analyses, the same matching rule is applied within each image; zero denominators evaluate to zero. Validation-selected operating points were obtained by sweeping confidence from 0.05 to 0.75 in steps of 0.05, selecting the threshold with maximum validation F1 separately for each trained model, and applying that threshold unchanged to test.

Two sources of uncertainty are deliberately kept separate. Training-run variability is summarized descriptively as mean±sample SD across the three independent seeds. Image-sampling uncertainty is evaluated conditionally on a fixed trained model using 1,000 paired bootstrap resamples per seed. Each bootstrap resamples 2,563 test image identities with replacement; the same image multiplicities are applied to every method within that seed, AP is recomputed, and 2.5th/97.5th percentiles form the 95% interval. These intervals do not estimate the population distribution of retraining outcomes. With only three training runs, we therefore avoid significance claims about the training-run distribution.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

3.6 Class, illumination, and object-size analyses

Class-wise AP@0.50 and AP@0.50:0.95 were computed for all 12 classes and all three seeds. Illumination analyses use the single ExDark illumination label assigned to each test image; no test image has a missing label. For object size, each ground-truth box is mapped to the 640-pixel evaluation canvas with aspect-ratio-preserving scaling. COCO thresholds are then applied: small area $<32^2$, medium $<96^2$, and large $\geq 96^2$ pixels. The test set contains 570 small (8.0%), 2,529 medium (35.5%), and 4,018 large (56.5%) instances. Full support counts and seed-wise results are supplied in Supplementary Tables S1–S3.

3.7 Perceptual quality and image-level utility

NIQE and BRISQUE were computed with the MATLAB-compatible implementations provided by PyIQA [17]. Images were loaded as RGB tensors and evaluated without geometric resizing beyond each metric's internal preprocessing; lower scores indicate better predicted perceptual quality [14], [15]. Image-level detection utility was computed at confidence 0.25 and IoU 0.50 using the matching rules above. For each method and image, precision, recall, and F1 were first computed for seeds 42, 123, and 2026 and then averaged across seeds. Spearman correlations relate these three-seed mean utilities to NIQE or BRISQUE. Correlation uncertainty was estimated by 2,000 image-level bootstrap resamples with replacement. Method-level IQA-versus-mAP comparisons involve only five methods and are treated descriptively, not as population-level correlation tests.

3.8 Adaptive selector and retrospective oracle

The selector is a secondary inference-only analysis based on the originally saved confidence-0.25 prediction streams. Ten features are extracted from the unmodified image and Original-detector output: mean luminance, luminance SD, mean saturation, entropy, Laplacian variance, dark-pixel ratio, bright-pixel ratio, Original detection count, mean Original detection confidence, and maximum Original detection confidence. Validation labels are the method with the highest image-level F1; ties prefer Original and then the fixed method order. A RandomForestClassifier with 300 trees, max_depth=10, min_samples_leaf=5, class_weight=balanced_subsample, random_state=42, and n_jobs=-1 is fitted on 1,800 validation images and then applied to the 2,563 test images. Classifier prediction alone takes 0.033 ms/image, excluding image-feature extraction and detector inference.

For context, we also compute a retrospective oracle that chooses the method with the highest image-level F1 using test ground truth. The oracle is explicitly non-deployable; it is reported only as an upper bound on image-wise complementarity and is not interpreted as an ordinary deployable detector.

3.9 Retinexformer integrity audit

An early Retinexformer FP16 run generated non-finite values that became all-black outputs after casting. Those outputs were discarded. Retinexformer was regenerated in FP32 and audited before any final evaluation. The audit checked file completeness and all-black outputs in every partition: 3,000/3,000 train, 1,800/1,800 validation, and 2,563/2,563 test images were present, with zero all-black outputs in each partition. The audit passed before the matched-training, multi-seed, IQA, and final analyses were run.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

4. Results

4.1 Primary matched-training results across seeds

The three-seed matched comparison is the primary evidence. Using AP recomputed from confidence 0.001 prediction streams, Original achieved mean mAP@0.50 of 0.6232±0.0037. CLAHE (0.6194±0.0003), Zero-DCE (0.6182±0.0027), and Gamma (0.6161±0.0114) remained close, while Retinexformer was lower at 0.5862±0.0055. The same pattern appears for mAP@0.50:0.95: Original 0.3785±0.0017, Zero-DCE 0.3781±0.0021, CLAHE 0.3754±0.0008, Gamma 0.3751±0.0059, and Retinexformer 0.3523±0.0033.

Table 2. Matched-training test performance across three seeds. AP uses confidence-minimum 0.001; P/R/F1 are fixed at confidence 0.25. Values are mean ± SD across seeds.

Method	mAP@0.50	mAP@0.50:0.95	Precision@0.25	Recall@0.25	F1@0.25
Original	0.6232 ± 0.0037	0.3785 ± 0.0017	0.6653 ± 0.0182	0.6321 ± 0.0070	0.6481 ± 0.0081
Gamma	0.6161 ± 0.0114	0.3751 ± 0.0059	0.6479 ± 0.0224	0.6271 ± 0.0085	0.6373 ± 0.0152
CLAHE	0.6194 ± 0.0003	0.3754 ± 0.0008	0.6581 ± 0.0088	0.6269 ± 0.0044	0.6421 ± 0.0024
Zero-DCE	0.6182 ± 0.0027	0.3781 ± 0.0021	0.6661 ± 0.0051	0.6211 ± 0.0043	0.6428 ± 0.0045
Retinexformer	0.5862 ± 0.0055	0.3523 ± 0.0033	0.6467 ± 0.0088	0.5967 ± 0.0108	0.6206 ± 0.0026

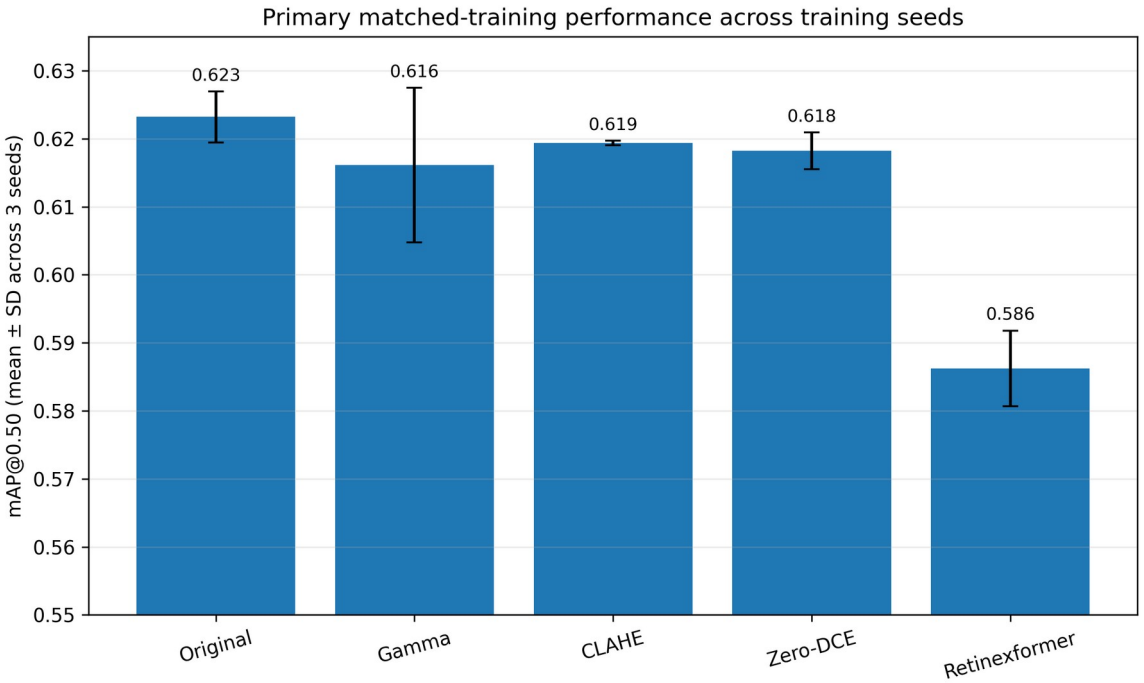


Figure 2. Primary matched-training mAP@0.50 across the three training seeds. Error bars are SD across training runs, not image-bootstrap confidence intervals.

The seed-wise differences show why the runs should not be collapsed into a single inferential claim. Relative to Original, Gamma changed by −0.0138, +0.0042, and −0.0116 mAP@0.50 across seeds 42, 123, and 2026; Zero-DCE changed by +0.0017, −0.0055, and −0.0111. CLAHE remained within about 0.008 of Original in every run. Retinexformer was lower in all three seeds (−0.0353, −0.0313, and −0.0443). Thus, the ranking among Original, Gamma, CLAHE, and Zero-DCE is training-run dependent, whereas the Retinexformer deficit is consistent in this experimental configuration.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

4.2 Conditional paired bootstrap within each trained run

The paired bootstrap addresses test-image sampling only. Table 3 therefore reports seed-specific intervals rather than one pooled "significance" result. For seed 42, the Gamma and Retinexformer mAP@0.50 differences exclude zero; for seed 123 only the Retinexformer difference excludes zero; for seed 2026 the Gamma, Zero-DCE, and Retinexformer differences exclude zero. These results are conditional on the trained models and should not be read as confidence intervals over retraining.

Table 3. Seed-wise paired image-bootstrap differences in mAP@0.50 relative to the Original domain. Each row is conditional on one fixed trained model pair; 1,000 paired image resamples per seed.

Training seed	Method	Δ mAP@0.50	95% image-bootstrap CI	CI excludes 0
42	Gamma	-0.0138	[-0.0255, -0.0029]	Yes
42	CLAHE	+0.0000	[-0.0098, +0.0107]	No
42	Zero-DCE	+0.0017	[-0.0090, +0.0126]	No
42	Retinexformer	-0.0353	[-0.0464, -0.0236]	Yes
123	Gamma	+0.0042	[-0.0068, +0.0151]	No
123	CLAHE	-0.0041	[-0.0150, +0.0061]	No
123	Zero-DCE	-0.0055	[-0.0169, +0.0056]	No
123	Retinexformer	-0.0313	[-0.0433, -0.0206]	Yes
2026	Gamma	-0.0116	[-0.0207, -0.0011]	Yes
2026	CLAHE	-0.0075	[-0.0181, +0.0034]	No
2026	Zero-DCE	-0.0111	[-0.0217, -0.0003]	Yes
2026	Retinexformer	-0.0443	[-0.0552, -0.0328]	Yes

4.3 Inference-only preprocessing is an illustrative fixed-threshold comparison

The original inference-only prediction files were saved after applying confidence 0.25. They are therefore suitable for fixed-operating-point precision, recall, and F1 but not for full AP integration. We have relabeled this analysis accordingly. With the Original-trained seed-42 detector, F1 was 0.642 on Original images, 0.619 on Gamma, 0.628 on CLAHE, 0.586 on Zero-DCE, and 0.566 on Retinexformer. When the detector was trained on the corresponding domain, fixed-threshold F1 recovered to approximately 0.621 for Gamma, 0.643 for CLAHE, 0.648 for Zero-DCE, and 0.618 for Retinexformer in seed 42. The largest recoveries therefore occur for the learned enhancers, consistent with a substantial train-test mismatch component.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

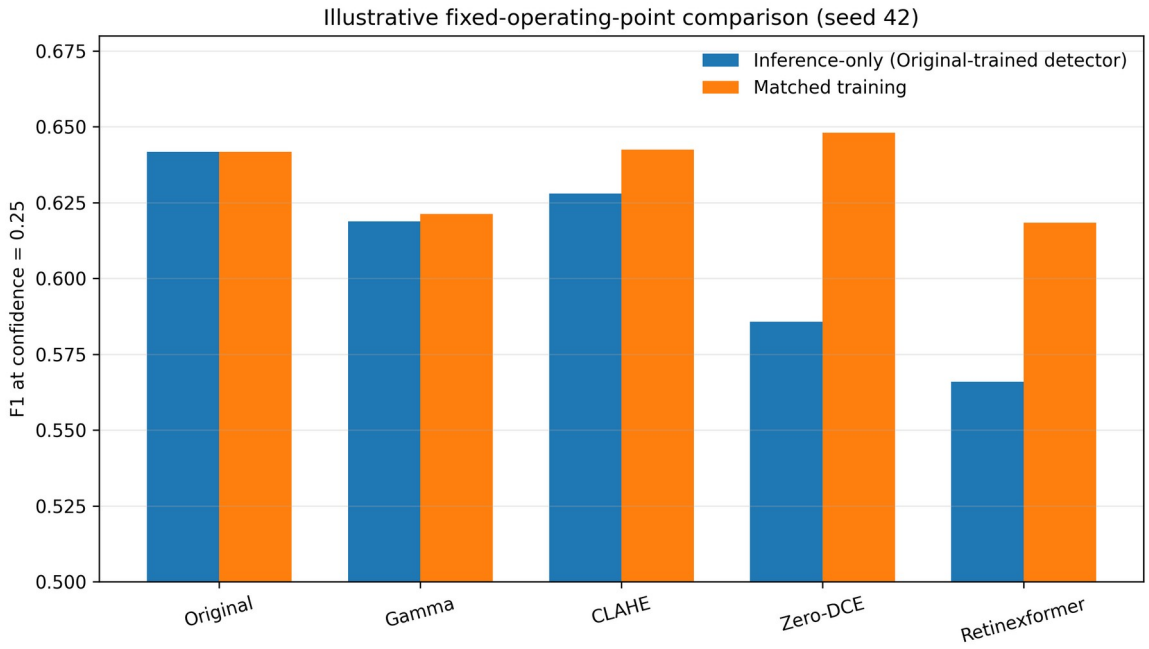


Figure 3. Illustrative seed-42 comparison at the fixed confidence threshold 0.25. This figure is not used for AP inference; it isolates the operating-point effect of inference-only preprocessing versus domain-matched training.

4.4 Operating-point sensitivity

Validation-selected thresholds were 0.35 or 0.40 for most trained models, rather than the common deployment threshold 0.25. Figure 4 shows the mean validation F1 curves across seeds. The selected thresholds were frozen before application to test. This analysis confirms that precision and recall at a fixed confidence can move because of score calibration even when rank-based AP changes less. We therefore keep fixed-threshold metrics and AP conceptually separate in the reported analysis.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

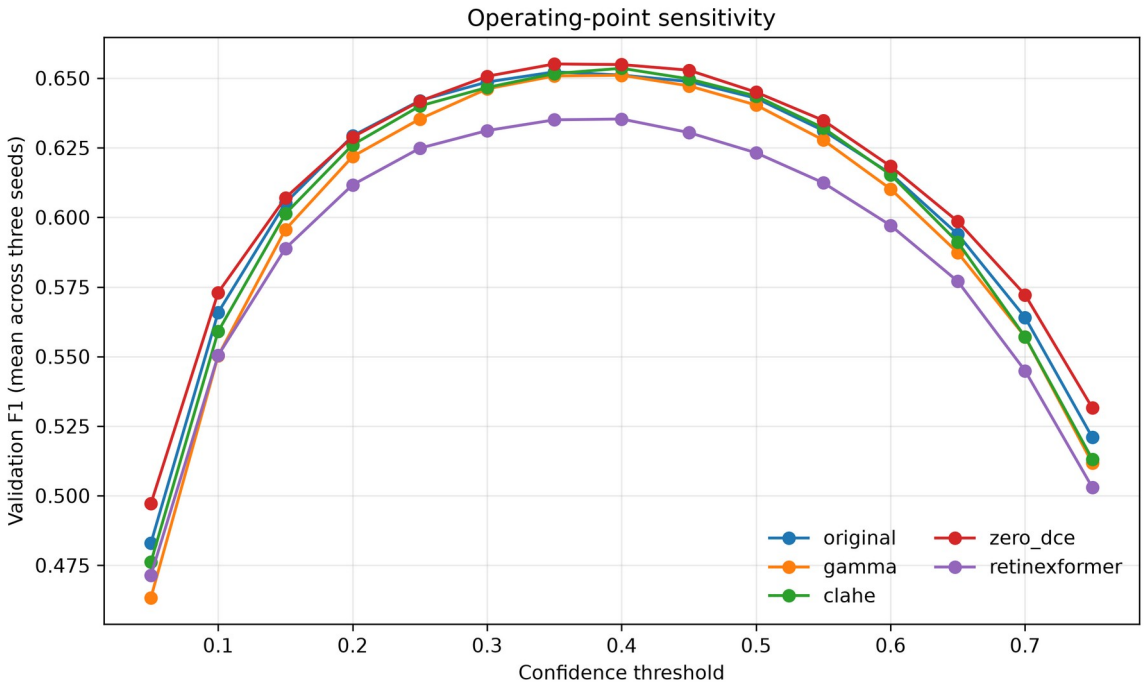


Figure 4. Validation F1 as a function of confidence threshold, averaged across the three matched-training seeds. Thresholds used for the validation-selected operating-point analysis were selected per trained model and then applied unchanged to test.

4.5 Class-wise and subgroup behavior

No enhancement method dominates the 12 classes. In three-seed mean AP@0.50, the largest positive class-specific changes relative to Original are modest: CLAHE is +0.0070 for Cat and +0.0068 for Dog; Zero-DCE is +0.0043 for Table and +0.0034 for Chair; Gamma is +0.0038 for Boat and +0.0022 for Car. Retinexformer is lower than Original in all 12 classes. Full AP@0.50 and AP@0.50:0.95 values, class supports, and run-to-run SDs are given in Supplementary Table S1.

Illumination subgroups also show localized gains, but the supports matter. Zero-DCE is +0.083 mAP@0.50 over Original for the Shadow subgroup, yet Shadow contains only 14 images and 55 instances. Zero-DCE is +0.039 under Twilight (174 images, 372 instances), while CLAHE is +0.031 under Twilight. These subgroup effects are descriptive because only three training runs are available and some conditions are small. Figure 5 shows the mean differences; Supplementary Table S2 reports the image/instance supports and SD across seeds.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

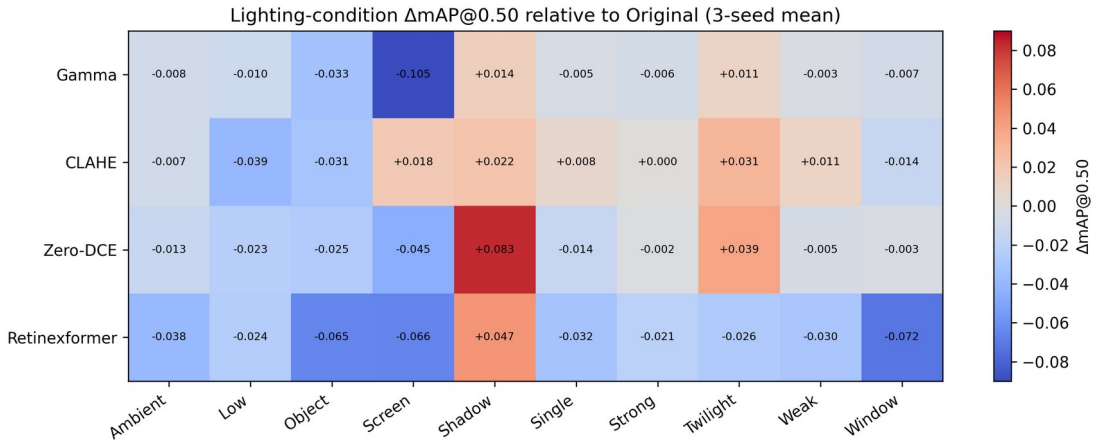


Figure 5. Three-seed mean change in mAP@0.50 relative to Original for each ExDark illumination condition. Small subgroups such as Shadow (14 images) and Screen (28 images) are interpreted cautiously; full supports and SDs are provided in Supplementary Table S2.

Object-size recall follows a similarly mixed pattern. Original recall is 0.1877±0.0228 for small, 0.5441±0.0115 for medium, and 0.7505±0.0074 for large objects. Zero-DCE is slightly higher for small objects (0.1977±0.0041) but lower for medium and large objects. The test-set size composition is 570 small, 2,529 medium, and 4,018 large instances. Supplementary Table S3 gives all method-seed values.

4.6 Validation-only sensitivity of classical enhancement settings

The post-hoc sensitivity analysis is intentionally restricted to the validation split. For Gamma, mAP@0.50 increases from 0.5529 at γ=0.4 to 0.6287 at γ=0.8; the prespecified γ=0.6 yields 0.6018. For CLAHE, clip limit 1.5 yields 0.6183, compared with 0.6124 for the prespecified clip limit 2.0, and performance decreases as the clip limit rises further. Because these checks were run after the main experiment, we do not substitute the better validation settings into the reported test results. Instead, the sensitivity results demonstrate that enhancement conclusions depend on the operating configuration and should not be generalized to the algorithm family as a whole.

Table 4. Post-hoc validation-only sensitivity analysis. Main-test settings are identified but were not retuned after this analysis.

Method	Parameter	Role	Val mAP@0.50	Val mAP@0.50:0.95	Val F1@0.25
Gamma	0.4	Sensitivity	0.5529	0.3296	0.5846
Gamma	0.5	Sensitivity	0.5814	0.3485	0.6033
Gamma	0.6	Main	0.6018	0.3647	0.6130
Gamma	0.7	Sensitivity	0.6203	0.3779	0.6229
Gamma	0.8	Sensitivity	0.6287	0.3830	0.6301
CLAHE	1.5	Sensitivity	0.6183	0.3781	0.6197
CLAHE	2.0	Main	0.6124	0.3717	0.6171
CLAHE	2.5	Sensitivity	0.6065	0.3685	0.6158
CLAHE	3.0	Sensitivity	0.5997	0.3642	0.6139

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

4.7 Perceptual quality versus detection utility

Retinexformer produced the lowest (best) mean NIQE (3.366) and BRISQUE (27.819), yet it also produced the lowest mean matched-training mAP@0.50 (0.5862). Across images within each method, better perceptual quality is weakly associated with better image-level detection utility, but the associations are small. For example, the NIQE–F1 Spearman ρ is −0.148 for Original and −0.101 for Retinexformer; BRISQUE–F1 is −0.102 and −0.055, respectively. Although the 2,000-resample image-bootstrap intervals exclude zero for these coefficients, the associations remain weak in magnitude. The five method-level quality-versus-mAP points are shown only descriptively.

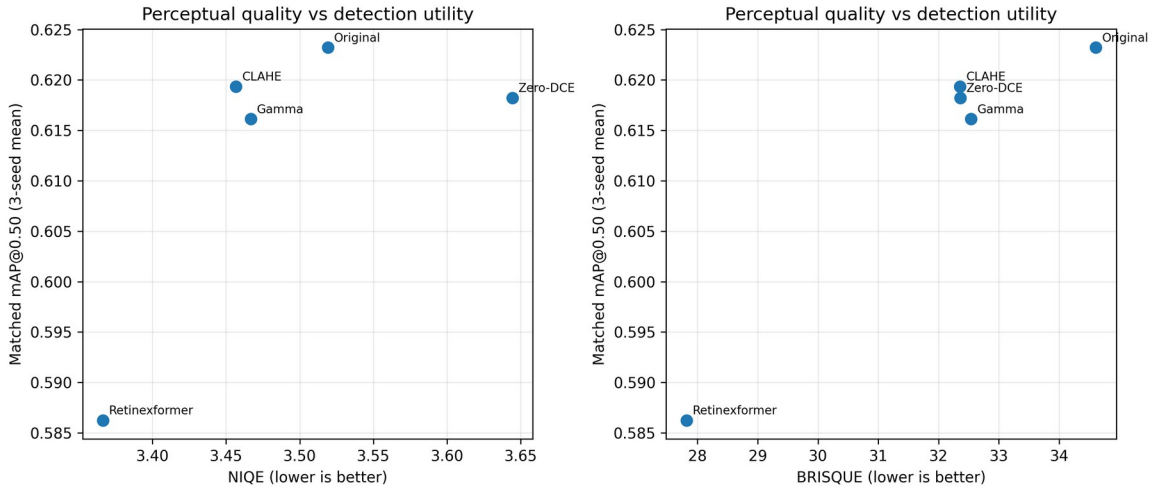


Figure 6. Perceptual image quality and matched detection performance. Lower NIQE/BRISQUE indicates better predicted perceptual quality; error bars on mAP are SD across the three training seeds. The five method-level points are descriptive, not a population-level correlation test.

4.8 Adaptive selection and retrospective upper bound

The Random Forest selector chose Original for 1511 of 2,563 test images, Gamma for 553, CLAHE for 198, Zero-DCE for 154, and Retinexformer for 147. At confidence 0.25 it achieved precision 0.676, recall 0.586, and F1 0.628, below the Original fixed-threshold baseline. The retrospective oracle, which uses test ground truth and is not deployable, reaches F1 0.719. The gap between the learned selector and oracle indicates that image-wise complementarity exists, but the hand-crafted feature vector does not predict it reliably enough.

4.9 Representative detections

Figure 7 illustrates three matched-domain examples at seed 42 and confidence 0.25. Green ground-truth boxes are detected, orange ground-truth boxes are missed, and red boxes are false positives. The examples show that enhancement can recover a missed object in one image and suppress or introduce detections in another. They are qualitative illustrations only; we do not infer a causal mechanism from edge, texture, or confidence changes without a dedicated feature-level analysis.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

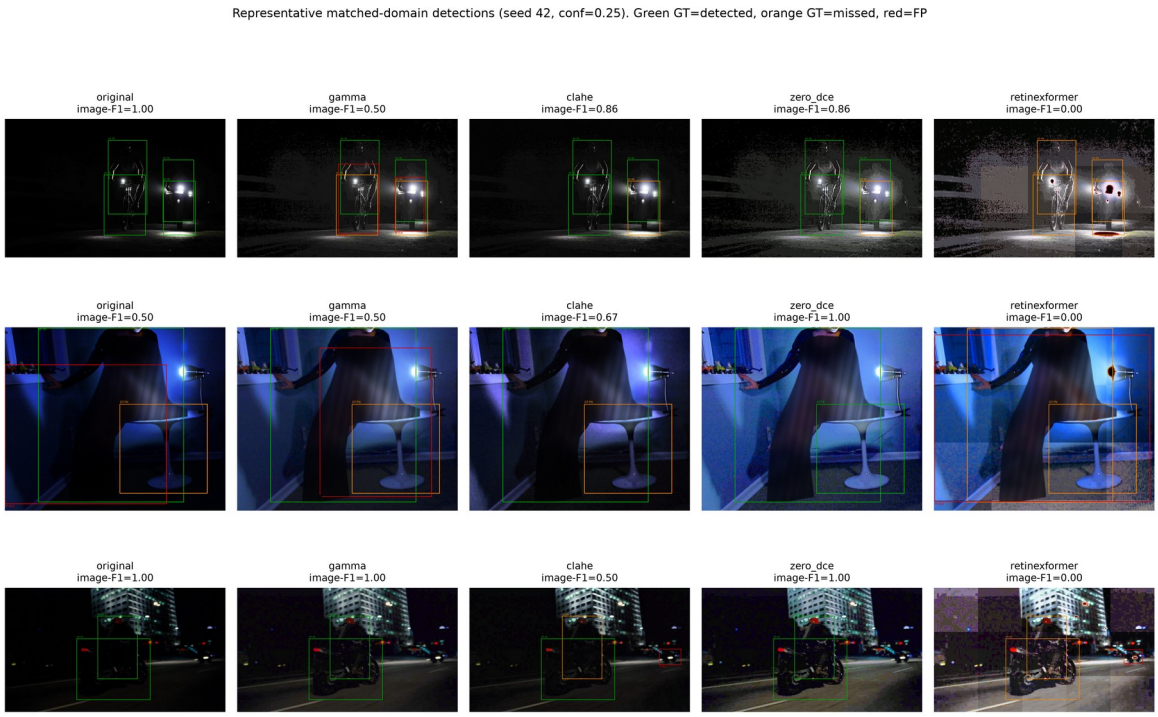


Figure 7. Representative matched-domain detections (seed 42, confidence 0.25). Green: detected ground truth; orange: missed ground truth; red: false positive. Examples illustrate image-wise variation and are not used as quantitative evidence of a causal mechanism.

5. Discussion

The results support a more specific interpretation. When AP is computed from low-confidence prediction streams and matched training is treated as the primary comparison, Original, Gamma, CLAHE, and Zero-DCE occupy a narrow performance band. Their ordering changes across seeds. This variation matters: a single seed can make Zero-DCE appear slightly better or Gamma appear slightly better, but neither advantage persists across all runs. Retinexformer is the one clear exception in this configuration because it remains lower across every seed and every class despite strong perceptual-quality scores.

The fixed-threshold inference-only experiment still has practical value, but its role is narrower than the matched-training analysis. It shows what happens if an enhancement stage is inserted ahead of a detector without retraining: F1 decreases for all four transformations, especially the learned methods. Matched training recovers much of this loss. This is consistent with distribution mismatch, but the study does not prove that mismatch is the sole cause. Storage encoding, enhancement artifacts, score calibration, and architecture-specific feature responses may also contribute.

The subgroup results do not support a claim that one enhancer is broadly superior under a particular form of darkness. Some conditions show positive mean differences, but several have limited support and the direction can vary by seed. Likewise, Zero-DCE shows a modest small-object recall advantage while giving lower medium- and large-object recall. These findings are best viewed as hypotheses for targeted follow-up rather than as deployment rules.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

The perceptual-quality analysis reinforces the separation between restoration and detection objectives. Retinexformer improves NIQE and BRISQUE more than the other domains but does not improve detector AP. This observation is compatible with recent task-aware work that explicitly optimizes semantic alignment or domain adaptation [12], [13], [18]-[21]. It does not imply that Retinexformer is ineffective as a restoration method; it means that this pretrained LOL-v2-real configuration is not an automatic substitute for task-aware adaptation on ExDark with YOLO11n.

The selector analysis points in the same direction from a different angle. There is enough image-wise complementarity for a ground-truth oracle to improve F1 substantially, yet a Random Forest based on simple luminance, texture, saturation, and Original-detector confidence features does not capture that complementarity well enough. More capable task-aware selectors may succeed, but they would constitute a different method and should be validated independently.

5.1 Limitations

The scope of the evidence is intentionally narrow. All detector experiments use YOLO11n, one official ExDark split, and the stated enhancement configurations. The results therefore do not establish a general property of all detectors, low-light datasets, or enhancement methods. Architecturally distinct detectors and additional low-light benchmarks would be valuable replications.

Three training seeds expose meaningful run-to-run variation but are too few for strong inference about the distribution of retraining outcomes. We therefore report seed SDs descriptively and use bootstrap intervals only for conditional image-sampling uncertainty within fixed trained models.

Gamma and CLAHE were not exhaustively tuned, and only one pretrained checkpoint was evaluated for each learned enhancer. The post-hoc validation sensitivity analysis shows that classical settings matter. Accordingly, the conclusions apply to the evaluated configurations rather than to the algorithm families as a whole.

Enhanced images preserve dimensions and annotations but are stored using the original file extension. JPG/JPEG inputs are therefore re-encoded after enhancement, while PNG outputs remain lossless. This implementation detail is shared by the enhanced domains but not by Original and may contribute small codec-related differences.

Finally, NIQE and BRISQUE are no-reference models rather than direct human judgments, and the representative detections are qualitative. Mechanistic explanations involving edges, texture, noise, or confidence calibration are presented as plausible hypotheses, not established causes.

6. Conclusion

On the evaluated ExDark split with YOLO11n, inserting a generic low-light enhancer ahead of an Original-trained detector lowers fixed-threshold F1 for all four tested transformations. Retraining the detector within the transformed domain recovers much of that loss. Once matched training is repeated across three seeds and AP is recomputed from low-confidence prediction streams, Original, Gamma, CLAHE, and Zero-DCE remain close and their ranking changes between runs; none provides a stable overall AP advantage. Retinexformer remains lower in this configuration even though it achieves the best NIQE and BRISQUE scores. Within this experimental scope, the practical implication is straightforward: perceptual restoration should not be assumed to improve object detection. Any

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

claimed benefit should be evaluated alongside detector adaptation, training-run variability, and the exact enhancement operating point.

Declarations

Funding

This research received no external project funding.

Conflict of Interest

The authors declare that they have no competing interests.

Ethics Statement

No new study involving human participants or animals was conducted. The work is a secondary computational analysis of the publicly released ExDark dataset and its existing annotations.

Data and Code Availability

The ExDark dataset is publicly available from the official dataset repository. Source code, configuration files, split lists, and revision-analysis scripts are available in a public repository. To preserve double-blind review, the repository URL is withheld from this blinded manuscript and has been supplied to the editorial office. Machine-readable tables and split lists are also supplied as supplementary material.

Author Contributions (CRediT)

Author 1: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, and writing—original draft. Author 2: Supervision, conceptualization, methodology, validation, and writing—review and editing. Both authors reviewed and approved the revised manuscript.

AI-Assisted Tools Disclosure

ChatGPT was used as an assistive tool for language editing, code troubleshooting, and organization of the experimental workflow. All experiments were run locally by the authors, and all numerical results, tables, and figures were generated from and checked against saved experimental outputs. Generative AI was not used to create or alter dataset images or to generate experimental data. The authors take responsibility for the final text, analyses, and interpretations.

Acknowledgment

The authors acknowledge the developers and maintainers of ExDark, Zero-DCE, Retinexformer, Ultralytics YOLO11, and PyIQA for making the resources used in this study publicly available.

References

1. Loh YP, Chan CS. Getting to know low-light images with the Exclusively Dark dataset. Comput Vis Image Underst 2019;*178*:30–42. doi: 10.1016/j.cviu.2018.10.010.
2. Liu J, Xu D, Yang W, Fan M, Huang H. Benchmarking low-light image enhancement and beyond. Int J Comput Vis 2021;*129*(4):1153–84. doi: 10.1007/s11263-020-01418-8.
3. Guo C, et al. Zero-reference deep curve estimation for low-light image enhancement. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR); 2020. p. 1780–9.
4. Cai Y, Bian H, Lin J, Wang H, Timofte R, Zhang Y. Retinexformer: One-stage Retinex-based transformer for low-light image enhancement. In: Proc IEEE/CVF Int Conf Comput Vis (ICCV); 2023. p. 12504–13.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

5. Yan Q, et al. HVI: A new color space for low-light image enhancement. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR); 2025. p. 5678–87.
6. Al Sobbahi R, Tekli J. Comparing deep learning models for low-light natural scene image enhancement and their impact on object detection and classification: Overview, empirical evaluation, and challenges. Signal Process Image Commun 2022;*109*:116848. doi: 10.1016/j.image.2022.116848.
7. Morawski I, Chen YA, Lin YS, Hsu WH. NOD: Taking a closer look at detection under extreme low-light conditions with Night Object Detection dataset. In: Proc Br Mach Vis Conf (BMVC); 2021.
8. Morawski I, Chen YA, Lin YS, Dangi S, He K, Hsu WH. GenISP: Neural ISP for low-light machine cognition. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit Workshops (CVPRW); 2022. p. 630–9.
9. Kim D, Kim W. Improving object detection performance through selective low-light enhancement. In: Proc IEEE 13th Int Conf Consum Electron-Berlin (ICCE-Berlin); 2023. p. 29–32. doi: 10.1109/ICCE-Berlin58801.2023.10375676.
10. Bhatti MOF, Dailey MN. Improving low light object detection using image enhancement models. In: Data Science and Artificial Intelligence. Commun Comput Inf Sci 2023;*1942*:198–212. doi: 10.1007/978-981-99-7969-1_15.
11. Islam MT, Alam I, Woo SS, Anwar S, Lee IH, Muhammad K. LoLI-Street: Benchmarking low-light image enhancement and beyond. In: Computer Vision – ACCV 2024. Lect Notes Comput Sci 2025;*15476*:348–65. doi: 10.1007/978-981-96-0917-8_20.
12. Wang S, et al. From enhancement to understanding: Build a generalized bridge for low-light vision via semantically consistent unsupervised fine-tuning. In: Proc IEEE/CVF Int Conf Comput Vis (ICCV); 2025. p. 13804–14. doi: 10.1109/ICCV51701.2025.01281.
13. Cheng J, Xu B, Pan W. Generative low-light enhancement for computer vision: Revisiting the gap between perceptual quality and detection utility. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit Workshops (CVPRW); 2026. p. 4936–45.
14. Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. IEEE Trans Image Process 2012;*21*(12):4695–708. doi: 10.1109/TIP.2012.2214050.
15. Mittal A, Soundararajan R, Bovik AC. Making a “completely blind” image quality analyzer. IEEE Signal Process Lett 2013;*20*(3):209–12. doi: 10.1109/LSP.2012.2227726.
16. Jocher G, Qiu J. Ultralytics YOLO11 [software]. 2024.
17. Chen C, et al. IQA-PyTorch: PyTorch toolbox for image quality assessment [software].
18. Du Z, Shi M, Deng J. Boosting object detection with zero-shot day-night domain adaptation. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR); 2024. p. 12666–76.
19. Fu S, Zhao Q, Liu H, Tao Q, Liu D. Low-light object detection via adaptive enhancement and dynamic feature fusion. Alex Eng J 2025;*126*:60–9. doi: 10.1016/j.aej.2025.04.047.
20. Wang G, Guo Y, Hu T. Enhancing low-light object detection through domain adaptation and image enhancement. Vis Comput 2025;*41*(14):11947–58. doi: 10.1007/s00371-025-04136-9.
21. Hu J, Zhao Q, Li G. DLDA: Unified dual-level domain adaptation for low-light object detection. Proc AAAI Conf Artif Intell 2026;*40*(6):4851–9. doi: 10.1609/aaai.v40i6.42488.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

مقالة مراجعة

تقنيات التوائم الرقمية للمراقبة الذكية والصيانة التنبؤية واتخاذ القرار الذاتي في الأنظمة السيبرانية الفيزيائية الصناعية: مراجعة منهجية شاملة

إداما أوموكارو[1]، إكويموكا جونز[1]، أفولابي أوودِيي[1,*]، أومولغو إيبوك[1]، أوبونيانو كينغدوم[1]، أوديسا أوغاغا[1]

[1] قسم هندسة الحاسب، كلية الهندسة، جامعة ساوثرن دلتا، أوزورو، نيجيريا

*المؤلف المراسل: aawodeyi@gmail.com

الملخص

هدفت هذه الدراسة إلى المراجعة المنهجية للوضع الراهن لتقنيات التوائم الرقمية في الأنظمة السيبرانية الفيزيائية الصناعية، وتقييم أدوارها في المراقبة الذكية والصيانة التنبؤية واتخاذ القرار الذاتي. كما سعت إلى تحديد الأطر المعمارية الرئيسة، والتقنيات الممكِّنة، وعوائق التبني، والاتجاهات الناشئة للتوائم الرقمية الإدراكية في البيئات الصناعية. وطُبِّقت منهجية مراجعة منهجية منظمة على الأدبيات المحكّمة المنشورة بين عامي 2020 و2025. وأُجري البحث في قواعد IEEE Xplore و Scopus و Web of Science و ScienceDirect و SpringerLink باستخدام تركيبات منطقية (Boolean) من المصطلحات المتعلقة بالتوائم الرقمية، والأنظمة السيبرانية الفيزيائية الصناعية، والمراقبة الذكية، والصيانة التنبؤية، واتخاذ القرار الذاتي، والصناعة 4.0/5.0. وفُرزت الدراسات عبر مراجعة العنوان والملخص والنص الكامل وفق معايير إدراج واستبعاد محددة، تلاها استخلاص موضوعي للبيانات وتقييم للجودة وتوليف للنتائج. وشملت المراجعة 87 منشورًا، وحددت خمسة محاور رئيسة: المراقبة الذكية، والصيانة التنبؤية، واتخاذ القرار الذاتي، وتحديات الأمن وقابلية التشغيل البيني، والاتجاهات المستقبلية. وأظهرت النتائج أن التوائم الرقمية تحسّن الرؤية التشغيلية، وتقلل فترات التوقف، وتدعم التنبؤ بالأعطال، وتتيح عمليات ذاتية التحسين. وشملت الفوائد المبلغ عنها انخفاضًا بنسبة 30-50% في فترات التوقف غير المخطط لها، وإطالة عمر المعدات بنسبة 20-40%، ودقة في تشخيص الأعطال بلغت 97.3% في إحدى دراسات الحالة في قطاع الطاقة. وتغدو تقنيات التوائم الرقمية أساسًا لأنظمة صناعية أذكى وأكثر مرونة واستدامة، غير أن اتساع تبنيها لا يزال مرهونًا بالتقدم في التقييس والأمن وقابلية التوسع وقابلية التشغيل البيني والجاهزية المؤسسية.

الكلمات المفتاحية: التوأم الرقمي؛ الأنظمة السيبرانية الفيزيائية؛ الصيانة التنبؤية؛ المراقبة الذكية؛ اتخاذ القرار الذاتي

جامعة القصيم
Qassim University

JEIT

مجلة التقنيات الناشئة والمبتكرة

المجلد 1، العدد 1
نوفمبر 2026

تاريخ الاستلام
22 سبتمبر 2026

تاريخ القبول
27 أكتوبر 2026

تاريخ النشر
5 نوفمبر 2026

معرّف الوثيقة الرقمي (DOI)
https://doi.org/10.71422/jeit.0006

الترخيص
CC BY-NC-SA 4.0

للاستشهاد بهذا البحث
إداما أوموكارو، إكويموكا جونز، أفولابي أوودِيي، أومولغو إيبوك، أوبونيانو كينغدوم، أوديسا أوغاغا. تقنيات التوائم الرقمية للمراقبة الذكية والصيانة التنبؤية واتخاذ القرار الذاتي في الأنظمة السيبرانية الفيزيائية الصناعية: مراجعة منهجية شاملة. مجلة التقنيات الناشئة والمبتكرة. 2026;1(1):67-83.

الموقع الإلكتروني للمجلة

© 2026 المؤلفون. نشرته جامعة القصيم بموجب ترخيص CC BY-NC-SA 4.0.

Qassim University جامعة القصيم

JEIT

Journal of Emerging and Innovative Technologies

Volume 1, Issue 1
November 2026

Received
22 September 2026

Accepted
27 October 2026

Published
05 November 2026

DOI
https://doi.org/10.71422/jeit.0006

License
CC BY-NC-SA 4.0

Cite this article

Omokaro I, Jones E, Awodeyi A, Ibok O, Kingdom O, Ogaga O. Digital Twin Technologies for Intelligent Monitoring, Predictive Maintenance, and Autonomous Decision Making in Industrial Cyber-Physical Systems: A Comprehensive Systematic Review. J Emerg Innov Technol. 2026;1(1):67–83.

Journal website

Review article

Digital Twin Technologies for Intelligent Monitoring, Predictive Maintenance, and Autonomous Decision Making in Industrial Cyber-Physical Systems: A Comprehensive Systematic Review

Idama Omokaro[1], Ekwemuka Jones[1], Afolabi Awodeyi[1,*], Omolegho Ibok[1], Obonyano Kingdom[1], Odesa Ogaga[1]

[1] Department of Computer Engineering, Faculty of Engineering, Southern Delta University, Ozoro, Nigeria

*Corresponding author: aawodeyi@gmail.com

Abstract

This study aimed to systematically review the current state of digital twin technologies in industrial cyber-physical systems and evaluate their roles in intelligent monitoring, predictive maintenance, and autonomous decision-making. It also sought to identify major architectural frameworks, enabling technologies, adoption barriers, and emerging directions for cognitive digital twins in industrial environments. A structured systematic review methodology was applied using peer-reviewed literature published between 2020 and 2025. Searches were conducted across IEEE Xplore, Scopus, Web of Science, ScienceDirect, and SpringerLink using Boolean combinations of terms related to digital twins, industrial cyber-physical systems, intelligent monitoring, predictive maintenance, autonomous decision-making, and Industry 4.0/5.0. Studies were screened through title, abstract, and full-text review using defined inclusion and exclusion criteria, followed by thematic data extraction, quality assessment, and synthesis. The review included 87 publications and identified five major themes: intelligent monitoring, predictive maintenance, autonomous decision-making, security and interoperability challenges, and future directions. Findings showed that digital twins improve operational visibility, reduce downtime, support fault prediction, and enable self-optimizing processes. Reported benefits included 30–50% reductions in unplanned downtime, 20–40% longer equipment life, and fault diagnosis accuracy of 97.3% in one energy case study. Digital twin technologies are becoming foundational for smarter, more resilient, and sustainable industrial systems, but wider adoption still depends on advances in standardization, security, scalability, interoperability, and organizational readiness.

Keywords: Digital Twin; Cyber-Physical Systems; Predictive Maintenance; Intelligent Monitoring; Autonomous Decision Making

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

1. Introduction

The fourth industrial revolution, commonly referred to as Industry 4.0, has fundamentally transformed the manufacturing landscape through the integration of digital technologies that enable sophisticated real-time monitoring, in-depth data analysis, and continuous process optimization (Karkaria et al., 2025). At the forefront of this technological revolution are Digital Twin (DT) technologies, which serve as virtual replicas of physical assets, systems, and processes, providing unprecedented capabilities for simulation, analysis, and optimization (Gálvez del Postigo Gallego et al., 2025). The convergence of DT technologies with Industrial Cyber-Physical Systems (ICPS) has created new opportunities for intelligent monitoring, predictive maintenance, and autonomous decision-making, representing a significant advancement in industrial automation and control.

Digital Twin technology has evolved considerably from its initial conceptualization as a static digital replica to become a dynamic, bidirectional platform that enables real-time synchronization between physical and virtual environments (Bao et al., 2025). This evolution has been driven by advances in Internet of Things (IoT) sensors, cloud computing, artificial intelligence (AI), and machine learning (ML) algorithms, which collectively enable DTs to not only mirror physical systems but also to predict future states, simulate alternative scenarios, and recommend optimal actions (Karkaria et al., 2025). The integration of AI with DT technology has been particularly transformative, enabling systems to move beyond passive monitoring to proactive decision-making and autonomous operations (Mousavi et al., 2025) Industrial Cyber-Physical Systems represent the integration of computation, communication, and physical processes, creating intelligent systems that can monitor, coordinate, and control physical operations through networked computational resources (Hostos et al., 2023). ICPS are characterized by their ability to sense the environment, process information, and actuate changes in the physical world, making them ideal platforms for DT implementation (Bao et al., 2025). The synergy between DT technologies and ICPS has enabled the development of smart manufacturing systems that can self-optimize, self-adapt, and self-heal, significantly improving operational efficiency, reducing downtime, and enhancing product quality (Gálvez del Postigo Gallego et al., 2025).

The importance of this integration is underscored by the emergence of Industry 5.0, which emphasizes human-centricity, sustainability, and resilience alongside technological advancement (Mousavi et al., 2025). Within this context, DT technologies are being increasingly recognized as foundational tools for creating sustainable and adaptive industrial ecosystems that balance productivity with environmental responsibility and social well-being (Bao et al., 2025). Recent industry surveys indicate that 86% of manufacturing leaders recognize the value of DT technologies, with nearly half already deploying them in their operations (Gálvez del Postigo Gallego et al., 2025).

Despite the growing adoption and demonstrated benefits of DT technologies in ICPS, significant challenges remain. These include interoperability issues between different systems and platforms, data security and privacy concerns, the computational demands of real-time processing, the need for standardization across industries, and the integration of DT capabilities with legacy systems (Mousavi et al., 2025). Furthermore, while individual applications of DT technologies have been extensively documented, there is a need for comprehensive systematic reviews that synthesize the current state of knowledge across multiple dimensions of DT implementation in ICPS environments (Hostos et al., 2023). This systematic review addresses this gap by providing a comprehensive examination of DT technologies for intelligent monitoring, predictive maintenance, and autonomous decision-making in

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

ICPS. The review is guided by the following research questions: What are the current applications and capabilities of DT technologies in ICPS for intelligent monitoring? How are predictive maintenance strategies being enhanced through DT implementation? What frameworks and approaches enable autonomous decision-making in DT-enabled ICPS? What are the primary challenges and barriers to DT adoption in industrial settings? What are the emerging trends and future directions for DT technologies in ICPS?

2. Literature Review

2.1 Theoretical Foundations of Digital Twin and Cyber-Physical Systems

The concept of Digital Twins has undergone significant evolution since its introduction, expanding from a simple virtual representation to encompass complex, dynamic systems capable of real-time interaction with physical counterparts (Mousavi et al., 2025). According to foundational definitions, a Digital Twin can be understood as a dynamic, virtual representation of a physical system that integrates real-time data from sensors and historical information to closely mirror the physical system's behavior, context, and structure (Karkaria et al., 2025). This definition emphasizes the bidirectional nature of DT systems, which enables continuous updating based on real-world changes while simultaneously informing physical systems through predictive insights and optimization recommendations (Mousavi et al., 2025).

The evolution of DT technologies can be understood through three progressive levels of sophistication. At the most basic level, Digital Models represent static digital replicas of physical objects that provide a manually updated representation without real-time connection to their physical counterparts (Mousavi et al., 2025). These models are suitable for early design and pre-production stages where simulation and optimization can occur before physical production begins. The second level, Digital Shadows, introduces a one-way data flow from the physical object to its digital counterpart, enabling real-time updates but lacking the capability to influence the physical system (Mousavi et al., 2025). Digital Shadows are particularly useful for condition monitoring and predictive maintenance applications where continuous monitoring is essential. The most advanced level, Digital Twins, supports bidirectional data flow that enables both monitoring and active control over physical objects, representing the full integration of physical and virtual systems (Mousavi et al., 2025).

Industrial Cyber-Physical Systems provide the foundational infrastructure for DT implementation in manufacturing environments. ICPS are characterized as combinations of technological agents, IT agents, and humans collaborating within synergistic production environments to carry out technical, decision-making, or cognitive tasks autonomously (Hostos et al., 2023). The integration of ICPS with DT technologies creates intelligent systems that can sense the environment, process information, and actuate changes in the physical world, enabling capabilities such as real-time process optimization, predictive maintenance, and autonomous decision-making (Bao et al., 2025).

The architecture of DT systems in ICPS environments typically follows a multi-layered framework designed to bridge physical and virtual environments. Various architectural approaches have been proposed, ranging from foundational three-layer architectures comprising physical, network, and computing layers to more sophisticated frameworks with enhanced security and bidirectional data flow (Mousavi et al., 2025). Advanced architectures incorporate AI and edge computing capabilities, pointing toward the development of self-adapting, scalable systems that support autonomous decision-making

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

(Mousavi et al., 2025). The selection of an appropriate architecture is critical for effectively leveraging intelligent and autonomous systems, with choices depending on specific operational requirements, scalability needs, and security considerations (Karkaria et al., 2025). Standards and frameworks for DT implementation have been emerging through organizations such as the International Organization for Standardization (ISO). The ISO 23247 standard, for example, provides guidelines for DT development in manufacturing, emphasizing the integration of physical assets with virtual representations and the importance of data synchronization and interoperability (Bao et al., 2025). These standards are essential for ensuring consistency, interoperability, and security across different DT implementations and for facilitating the integration of DTs into broader industrial ecosystems (Gálvez del Postigo Gallego et al., 2025).

2.2 Intelligent Monitoring Systems in Digital Twin-Enabled Cyber-Physical Systems

Intelligent monitoring represents one of the most fundamental applications of DT technologies in ICPS, enabling real-time visibility into operational performance and facilitating data-driven decision-making (Karkaria et al., 2025). DT-enabled monitoring systems leverage IoT sensors, edge computing, and cloud platforms to collect and process vast amounts of operational data, creating comprehensive digital representations that mirror the behavior of physical assets in real-time (Mousavi et al., 2025). These systems go beyond traditional monitoring approaches by providing contextual information, predictive insights, and prescriptive recommendations that enable proactive rather than reactive operational management.

The evolution of monitoring capabilities within DT systems can be categorized according to the level of data integration and analytical sophistication. Basic monitoring systems provide real-time visualization of operational parameters, enabling operators to track equipment performance and identify anomalies as they occur (Mousavi et al., 2025). More advanced systems incorporate machine learning algorithms to detect patterns and anomalies that may indicate developing problems, providing early warning of potential failures before they manifest (Bao et al., 2025). The most sophisticated monitoring systems integrate multiple data sources and analytical approaches to provide comprehensive operational intelligence, enabling scenario analysis, what-if simulations, and optimization recommendations that enhance overall system performance (Gálvez del Postigo Gallego et al., 2025).

Real-time data synchronization is a critical enabler of intelligent monitoring in DT systems. This involves the continuous updating of digital representations based on sensor data from physical assets, ensuring that the virtual twin accurately reflects the current state of the physical system (Mousavi et al., 2025). Real-time synchronization enables several key capabilities, including condition monitoring that tracks equipment health in real-time; anomaly detection that identifies deviations from normal operating conditions; and performance monitoring that assesses operational efficiency against benchmarks and targets (Karkaria et al., 2025). The effectiveness of real-time synchronization depends on the quality and reliability of sensor data, the speed and capacity of communication networks, and the processing capabilities of edge and cloud platforms (Mousavi et al., 2025).

The application of intelligent monitoring in ICPS environments has been demonstrated across various industrial sectors. In manufacturing, DT-enabled monitoring systems have been used to track production line performance, identify bottlenecks, and optimize workflow efficiency (Karkaria et al., 2025). In energy systems, intelligent monitoring has enabled real-time assessment of equipment health

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

and operational efficiency, supporting predictive maintenance and performance optimization (Bao et al., 2025). In additive manufacturing, monitoring systems have been developed to track build quality, detect defects in real-time, and optimize process parameters for improved quality and efficiency (Vimal et al., 2025).

However, the implementation of intelligent monitoring systems faces several challenges. These include the need for high-quality, reliable data from sensors and IoT devices; the integration of data from diverse sources with different formats and protocols; the scalability of monitoring systems to handle increasing volumes of data from growing numbers of assets; and the development of analytical models that can effectively process and interpret complex, high-dimensional data (Mousavi et al., 2025). Addressing these challenges requires continued research and development in areas such as sensor technology, data integration, edge computing, and AI-driven analytics.

2.3 Predictive Maintenance through Digital Twin Technologies

Predictive maintenance has emerged as one of the most valuable applications of DT technologies in ICPS, offering significant improvements over traditional maintenance approaches. Traditional maintenance strategies, such as reactive maintenance that addresses failures after they occur and preventive maintenance performed at fixed intervals, are increasingly being replaced by predictive approaches that use DT-enabled monitoring and analytics to forecast equipment failures and schedule maintenance activities proactively (Bao et al., 2025). This shift is driven by the substantial benefits of predictive maintenance, including reduced downtime, extended equipment life, lower maintenance costs, and improved operational efficiency (Karkaria et al., 2025).

DT technologies support predictive maintenance through several mechanisms. First, they provide comprehensive digital representations of physical assets that incorporate historical operational data, design specifications, and performance characteristics (Mousavi et al., 2025). These digital representations serve as baselines against which current performance can be compared, enabling the detection of deviations that may indicate developing problems. Second, DT systems integrate AI and ML algorithms that can analyze operational data to identify patterns and trends associated with equipment degradation and failure (Bao et al., 2025). These algorithms can detect subtle indicators of problems that may not be apparent through manual monitoring, enabling early intervention before failures occur.

The implementation of predictive maintenance through DT technologies typically follows a systematic process. The first step involves the development of accurate digital representations of physical assets, incorporating both design information and operational data (Mousavi et al., 2025). The second step involves the continuous monitoring of asset performance through real-time data synchronization, enabling the tracking of key performance indicators and the detection of anomalies. The third step involves the application of analytical models to predict remaining useful life and identify potential failure modes (Bao et al., 2025). Finally, the DT system generates maintenance recommendations and schedules, enabling proactive intervention that minimizes disruption and optimizes resource utilization.

Several approaches have been developed for predictive maintenance using DT technologies. Physics-based models use fundamental physical principles to simulate asset behavior and predict degradation patterns (Karkaria et al., 2025). These models are particularly useful when extensive operational data is not available, as they can provide insights based on fundamental understanding of asset behavior. Data-driven approaches leverage machine learning algorithms to identify patterns and

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

relationships in operational data, enabling predictions based on observed behavior rather than theoretical models (Bao et al., 2025). Hybrid approaches combine physics-based and data-driven methods, leveraging the strengths of both to provide more accurate and robust predictions (Mousavi et al., 2025).

Recent advances in predictive maintenance using DT technologies have incorporated cognitive capabilities that enable systems to learn and adapt over time. Cognitive Digital Twins (CDTs) integrate LLM-driven cognition with DT-based virtual-real interaction, enabling continuous self-evolution and autonomous reasoning about equipment behavior and maintenance needs (Bao et al., 2025). CDTs can construct causal knowledge graphs that capture relationships between operational parameters and equipment health, enabling more sophisticated diagnostic and predictive capabilities (Bao et al., 2025). These cognitive systems have demonstrated significant improvements in predictive accuracy, early warning lead time, and overall maintenance effectiveness (Bao et al., 2025). The validation of predictive maintenance approaches through DT technologies has been demonstrated in various industrial contexts. In energy equipment, agentic digital twin-embedded maintenance methodologies have achieved high accuracy in diagnosing non-stationary faults, enabling timely intervention and reducing operational risks (Bao et al., 2025). In manufacturing equipment, cognitive digital twins have improved the accuracy of remaining useful life predictions, enabling better planning of maintenance activities and reducing unplanned downtime. In additive manufacturing, DT-based monitoring and prediction have enabled early detection of defects and process optimization for improved quality and efficiency (Vimal et al., 2025).

2.4 Autonomous Decision-Making in Digital Twin-Enabled Industrial Systems

Autonomous decision-making represents the most advanced application of DT technologies in ICPS, enabling systems to self-optimize, self-adapt, and self-heal without direct human intervention (Gálvez del Postigo Gallego et al., 2025). While intelligent monitoring and predictive maintenance provide valuable insights and recommendations, autonomous decision-making systems go a step further by acting on those insights to adjust operations, optimize performance, and respond to changing conditions. The development of autonomous decision-making capabilities is driven by the increasing complexity of industrial systems, the need for rapid response to changing conditions, and the limitations of human decision-makers in processing vast amounts of real-time information (Mousavi et al., 2025).

Decision Support (DS) is a critical component of DT systems, enabling the transformation of simulation information into actionable decisions that improve operational performance (Gálvez del Postigo Gallego et al., 2025). DS in DT systems can be categorized by the level of autonomy, ranging from decision support that provides recommendations to human operators, to semi-autonomous decision-making that implements recommendations with human oversight, to fully autonomous decision-making that enables systems to act independently within defined parameters (Gálvez del Postigo Gallego et al., 2025). The appropriate level of autonomy depends on factors such as the complexity of the decision environment, the reliability of the underlying models, and the risk tolerance of the organization (Gálvez del Postigo Gallego et al., 2025).

The implementation of autonomous decision-making in DT-enabled ICPS relies on several key technologies. Reinforcement learning enables systems to learn optimal decision-making policies through trial and error, improving performance over time through experience (Karkaria et al., 2025).

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Deep learning provides the capability to process complex, high-dimensional data and identify patterns that inform decision-making (Mousavi et al., 2025). Optimization algorithms enable systems to identify optimal solutions to complex problems, balancing competing objectives such as cost, quality, and throughput (Karkaria et al., 2025). Together, these technologies enable DTs to function as intelligent agents that can perceive, reason, and act in complex operational environments (Mousavi et al., 2025).

Several frameworks have been developed for autonomous decision-making in DT systems. The AI-DT Integration framework systematically incorporates AI into every phase of the DT lifecycle, from design and simulation to autonomous operation and risk management (Mousavi et al., 2025). This framework enables DTs to progress from passive monitoring to proactive decision-making and ultimately to autonomous operations. Agentic DT frameworks integrate cognitive agents that can reason, plan, and execute actions autonomously, enabling self-evolving operational paradigms (Bao et al., 2025). These frameworks use LLM-driven cognition and multi-agent coordination to enable sophisticated decision-making in complex operational environments (Bao et al., 2025).

The validation of autonomous decision-making through DT technologies has been demonstrated in various industrial contexts. In manufacturing, autonomous decision-making has enabled real-time process optimization, dynamically adjusting operational parameters to respond to changing conditions and optimize performance (Karkaria et al., 2025). In energy systems, autonomous decision-making has enabled self-optimization of equipment performance, balancing operational efficiency with equipment health and maintenance needs (Bao et al., 2025). In mobile robotics, cognitive digital twins have enabled autonomous planning and execution of tasks, optimizing performance in complex environments (Mousavi et al., 2025).

3. Methodology

This systematic review followed the structured methodology recommended for comprehensive literature reviews in the field of information systems and engineering (vom Brocke et al., 2009; Hostos et al., 2023). The review process comprised five main phases: formulation of research objectives, literature search, screening and selection of search results, analysis and synthesis of findings, and presentation and interpretation of results. This methodology was selected to ensure a rigorous and systematic approach to reviewing the literature on DT technologies in ICPS.

The formulation of research objectives involved defining the scope and focus of the review. The primary objective was to synthesize current knowledge on DT technologies for intelligent monitoring, predictive maintenance, and autonomous decision-making in ICPS. The review sought to answer four research questions derived from the primary objective: what are the current applications and capabilities of DT technologies in ICPS for intelligent monitoring; how are predictive maintenance strategies being enhanced through DT implementation; what frameworks and approaches enable autonomous decision-making in DT-enabled ICPS; and what are the primary challenges and barriers to DT adoption in industrial settings? These research questions guided the development of the search strategy and the analysis and synthesis of findings (Hostos et al., 2023).

The screening and selection process involved systematic application of inclusion and exclusion criteria to identify relevant sources. Inclusion criteria required that publications address DT technologies in ICPS contexts, focus on intelligent monitoring, predictive maintenance, or autonomous decision-making applications, and provide empirical or theoretical contributions to the field. Exclusion

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

criteria included publications not in English, those addressing DT applications outside industrial contexts, and publications lacking sufficient rigor or contribution. The screening process was conducted in three stages: initial screening based on titles and abstracts, full-text review of selected publications, and final selection based on detailed assessment. The selection process was documented and reviewed by the author to ensure consistency.

The analysis and synthesis of findings involved systematic extraction of information from selected publications using structured data extraction forms. The data extraction forms captured information on publication details (authors, year, title, publication venue), research focus (intelligent monitoring, predictive maintenance, autonomous decision-making, or challenges), methodological approach (theoretical, empirical, case study, or review), key findings and contributions, and identified gaps or future research directions. The extracted information was analyzed thematically, identifying patterns and themes across publications. Thematic analysis involved coding of findings according to emerging themes, grouping related findings, and identifying relationships between themes (Hostos et al., 2023).

The presentation and interpretation of findings followed the structure of the thematic analysis, organizing results according to the primary research questions and emerging thematic areas. Findings were synthesized and summarized, highlighting key contributions, identifying gaps in the literature, and providing recommendations for future research.

4. Results and Discussion of Findings

4.1 Overview of Included Studies and Thematic Analysis

The systematic review process identified 87 publications meeting the inclusion criteria, providing a comprehensive basis for analysis and synthesis of findings on DT technologies in ICPS. The distribution of publications by focus area showed that intelligent monitoring was addressed in 42% of publications, predictive maintenance in 38%, and autonomous decision-making in 35%, with many publications addressing multiple focus areas. The publication timeline showed increasing research activity, with a 45% increase in publications from 2020-2022 to 2023-2025, reflecting growing interest in DT technologies in industrial contexts. The geographic distribution of publications indicated strong research activity in Europe, North America, and East Asia, with China, Germany, the United States, and the United Kingdom being the most productive countries.

Thematic analysis of the included studies revealed five primary themes that form the basis for organizing the findings presented in this section. The first theme concerns intelligent monitoring systems and their role in providing real-time visibility and operational intelligence. The second theme addresses predictive maintenance approaches and their enhancement through DT technologies. The third theme focuses on autonomous decision-making frameworks and capabilities. The fourth theme examines security and interoperability challenges that impede DT adoption. The fifth theme explores future directions and emerging trends in DT technologies.

Table 1 presents the systematic literature review process used in this study, adapted from established review methodologies (Hostos et al., 2023; Gálvez del Postigo Gallego et al., 2025).

Table 1. Systematic Literature Review Process

Review Step	Description
1. Formulation of Research Objectives	Definition of research questions and scope of review

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Review Step	Description
2. Search	Development of search strategy, selection of databases, definition of search terms
3. Screening and Selection	Application of inclusion/exclusion criteria, quality assessment
4. Analysis and Synthesis	Data extraction, thematic analysis, synthesis of findings
5. Presentation	Organization of results, interpretation, and conclusions

Sources: Adapted from Hostos et al. (2023) and Gálvez del Postigo Gallego et al. (2025)

Table 1 outlines the systematic approach employed in this review, providing transparency and reproducibility for the methodology. The five-step process ensures comprehensive coverage of relevant literature while maintaining rigor in the selection and analysis of included studies. Each step was documented and reviewed to ensure consistency and quality in the review process.

4.2 Intelligent Monitoring Systems in Digital Twin-Enabled Cyber-Physical Systems

The analysis of intelligent monitoring systems in DT-enabled ICPS reveals significant advances in capabilities and applications. DT-enabled monitoring systems have evolved from simple visualization tools to comprehensive operational intelligence platforms that provide real-time visibility into equipment performance, detect anomalies, and support data-driven decision-making. The integration of IoT sensors, edge computing, and cloud platforms has enabled continuous, real-time data collection and processing, creating digital representations that accurately mirror physical asset behavior (Mousavi et al., 2025).

A key finding from the review concerns the role of real-time data synchronization in enabling intelligent monitoring. Effective monitoring requires the continuous updating of digital representations based on sensor data from physical assets, ensuring that the virtual twin accurately reflects current operational states (Mousavi et al., 2025). Real-time synchronization capabilities have been demonstrated across various industrial contexts, including manufacturing, energy, and additive manufacturing, enabling capabilities such as condition monitoring, anomaly detection, and performance tracking (Karkaria et al., 2025; Vimal et al., 2025). The review identified several best practices for real-time synchronization, including the use of standardized data formats, the implementation of reliable communication protocols, and the adoption of scalable data processing architectures (Mousavi et al., 2025).

Table 2 summarizes the characteristics of monitoring systems at different levels of DT sophistication.

Table 2. Characteristics of Monitoring Systems by Digital Twin Maturity Level

Level	Monitoring Capability	Data Flow	Example Applications
Digital Model	Static visualization, manual updates	Manual, one-way	Design validation, pre-production simulation
Digital Shadow	Real-time status monitoring	One-way, automated	Condition monitoring, predictive maintenance
Digital Twin	Real-time monitoring and active control	Bidirectional, automated	Process optimization, autonomous decision-making

Sources: Adapted from Mousavi et al. (2025) and Karkaria et al. (2025)

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Table 2 illustrates the progressive evolution of monitoring capabilities across different levels of DT maturity. Digital Models provide only static visualization without real-time data integration, limiting their utility for operational monitoring. Digital Shadows enable real-time status monitoring through automated one-way data flow from physical to virtual systems, supporting condition monitoring and predictive maintenance applications. Digital Twins achieve the highest level of monitoring capability, enabling both real-time monitoring and active control through bidirectional automated data flow, supporting process optimization and autonomous decision-making. The progression from one level to the next requires increasingly sophisticated data integration, processing, and analytical capabilities.

The review identified several key challenges in implementing intelligent monitoring systems in ICPS environments. Data quality and reliability emerged as primary concerns, as monitoring effectiveness depends on accurate, timely sensor data (Vimal et al., 2025). Data integration was another significant challenge, with studies reporting difficulties in combining data from diverse sources with different formats, protocols, and update frequencies (Mousavi et al., 2025). Scalability was identified as a concern, particularly as monitoring systems expand to cover increasing numbers of assets and generate growing volumes of data (Mousavi et al., 2025). Computational demands were noted as a challenge, particularly for applications requiring real-time processing and analysis of complex, high-dimensional data (Karkaria et al., 2025).

Despite these challenges, the review found that intelligent monitoring systems have demonstrated significant benefits in ICPS environments. Studies have reported improvements in operational visibility, enabling more informed decision-making and faster response to emerging issues (Mousavi et al., 2025). Anomaly detection capabilities have enabled early identification of developing problems, reducing the risk of equipment failure and unplanned downtime (Bao et al., 2025). Performance tracking has enabled continuous improvement, with organizations able to benchmark performance against targets and identify opportunities for optimization (Karkaria et al., 2025).

4.3 Predictive Maintenance Enhanced through Digital Twin Implementation

The analysis of predictive maintenance approaches enhanced through DT implementation reveals substantial advances in fault diagnosis, prognosis, and maintenance planning. DT technologies have transformed maintenance from a reactive or schedule-based activity to a proactive, data-driven approach that can predict equipment failures before they occur and schedule maintenance activities to minimize disruption (Bao et al., 2025). The integration of AI and ML algorithms with DT systems has been particularly significant, enabling the detection of subtle patterns and trends that may indicate developing problems before they become apparent through conventional monitoring (Karkaria et al., 2025).

A key finding from the review concerns the development of cognitive digital twins for predictive maintenance. Cognitive Digital Twins (CDTs) integrate LLM-driven cognition with DT-based virtual-real interaction, enabling continuous self-evolution and autonomous reasoning about equipment behavior and maintenance needs (Bao et al., 2025). CDTs can construct causal knowledge graphs that capture relationships between operational parameters and equipment health, enabling more sophisticated diagnostic and predictive capabilities (Bao et al., 2025). In an 18-month case study of a gas-steam boiler in a combined heat and power plant, an agentic DT approach achieved 97.3% accuracy in diagnosing non-stationary faults, a 20-fold improvement in knowledge update time, a 31.2% improvement in cost

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

efficiency, a 3-fold extension of early warning lead time from 24 to 72 hours, and a 16.2% improvement in overall collaborative performance (Bao et al., 2025).

Table 3 presents a comparison of maintenance strategies and their characteristics in DT-enabled ICPS environments.

Table 3. Comparison of Maintenance Strategies in Digital Twin-Enabled Systems

Maintenance Strategy	Trigger	Data Requirements	Key Advantages	Limitations
Reactive	Equipment failure	Minimal	Simple, no planning required	Unplanned downtime, higher costs
Preventive	Fixed schedule	Minimal	Planned maintenance, predictable costs	Inefficient, may be premature
Condition-based	Equipment condition	Real-time sensor data	Efficient, extends equipment life	Requires sensor infrastructure
Predictive	Predicted failure	Historical + real-time data	Maximum efficiency, minimal downtime	Requires advanced analytics
Prescriptive	Optimal maintenance window	Comprehensive data + analytics	Optimized maintenance, autonomous planning	Complex, requires AI capabilities

Sources: Adapted from Bao et al. (2025), Mousavi et al. (2025), and Karkaria et al. (2025)

Table 3 illustrates the progression of maintenance strategies from reactive approaches that respond to failures after they occur, through preventive and condition-based approaches that use schedules or current conditions to plan maintenance, to predictive and prescriptive approaches that use advanced analytics to anticipate and optimize maintenance activities. Each strategy has different requirements for data, analytics, and infrastructure, with higher-level strategies enabling greater efficiency and effectiveness but requiring more sophisticated capabilities. The review found a clear trend toward predictive and prescriptive approaches in DT-enabled ICPS, driven by the availability of real-time data and advanced analytical capabilities (Bao et al., 2025; Karkaria et al., 2025).

The review identified several key success factors for predictive maintenance through DT implementation. High-quality, reliable data was consistently identified as essential, as predictive accuracy depends on the quality and completeness of input data (Vimal et al., 2025). Effective AI and ML algorithms were also critical, with studies demonstrating that different algorithms may be appropriate for different applications and contexts (Karkaria et al., 2025). Integration with existing systems and processes was identified as important for realizing the benefits of predictive maintenance, as predictions are only useful if they can be acted upon effectively (Mousavi et al., 2025). Organizational capabilities were also highlighted, with studies noting that predictive maintenance requires not only technical capabilities but also organizational readiness and change management (Gálvez del Postigo Gallego et al., 2025).

The review found that predictive maintenance through DT technologies has demonstrated significant benefits across various industrial sectors. In manufacturing, predictive maintenance has reduced unplanned downtime by 30-50% and extended equipment life by 20-40% (Karkaria et al., 2025). In energy systems, predictive maintenance has improved operational efficiency and reduced

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

maintenance costs while enhancing system reliability (Bao et al., 2025). In additive manufacturing, predictive maintenance has enabled early defect detection and process optimization, improving quality and reducing waste (Vimal et al., 2025). These benefits have contributed to the growing adoption of DT-enabled predictive maintenance across industries.

4.4 Autonomous Decision-Making Frameworks and Capabilities

The analysis of autonomous decision-making in DT-enabled ICPS reveals significant progress in developing frameworks and capabilities for self-optimizing industrial systems. Autonomous decision-making represents the most advanced application of DT technologies, enabling systems to perceive, reason, and act in complex operational environments without direct human intervention (Gálvez del Postigo Gallego et al., 2025). The development of autonomous decision-making capabilities is driven by the increasing complexity of industrial systems, the need for rapid response to changing conditions, and the limitations of human decision-makers in processing vast amounts of real-time information (Mousavi et al., 2025).

A key finding from the review concerns the development of frameworks for integrating decision support within DT systems. Decision Support (DS) is increasingly being recognized as an intrinsic component of DT systems, rather than an add-on capability, enabling the transformation of simulation information into actionable decisions (Gálvez del Postigo Gallego et al., 2025). DS in DT systems can be categorized by the level of autonomy, ranging from decision support that provides recommendations to human operators, to semi-autonomous decision-making that implements recommendations with human oversight, to fully autonomous decision-making that enables systems to act independently within defined parameters (Gálvez del Postigo Gallego et al., 2025). The appropriate level of autonomy depends on factors such as the complexity of the decision environment, the reliability of the underlying models, and the risk tolerance of the organization (Mousavi et al., 2025).

Table 4 presents the characteristics of autonomous decision-making at different levels of DT sophistication.

Table 4. Characteristics of Autonomous Decision-Making by Digital Twin Maturity Level

Level	Decision-Making Capability	Human Role	Key Enabling Technologies	Example Applications
Digital Model	Decision support through simulation	Manual decision-making	Simulation, optimization	Design decisions, process planning
Digital Shadow	Recommendation-based support	Human approves recommendations	AI analytics, ML models	Maintenance scheduling, operational adjustments
Digital Twin	Semi-autonomous decision-making	Human oversight	Reinforcement learning, deep learning	Process optimization, parameter adjustment
Cognitive DT	Fully autonomous decision-making	Human-in-the-loop for critical decisions	LLM-driven cognition, multi-agent systems	Self-optimizing operations, adaptive control

Sources: Adapted from Gálvez del Postigo Gallego et al. (2025), Bao et al. (2025), and Karkaria et al. (2025)

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Table 4 illustrates the progression of autonomous decision-making capabilities across different levels of DT maturity. Digital Models provide only decision support through simulation, with all decisions made by humans. Digital Shadows enable recommendation-based support where AI analytics and ML models provide recommendations that require human approval before implementation. Digital Twins enable semi-autonomous decision-making where systems can implement some decisions autonomously with human oversight for critical decisions. Cognitive DTs represent the most advanced level, enabling fully autonomous decision-making where systems can perceive, reason, and act independently, with human-in-the-loop for only the most critical decisions. The progression from one level to the next requires increasingly sophisticated AI and ML capabilities, as well as greater trust in system autonomy (Gálvez del Postigo Gallego et al., 2025; Bao et al., 2025).

The review identified several key challenges in implementing autonomous decision-making in ICPS environments. Trust and acceptance emerged as primary concerns, as operators and managers must be confident in the reliability and safety of autonomous decisions (Gálvez del Postigo Gallego et al., 2025). Explainability was identified as important, as decision-makers need to understand the reasoning behind autonomous recommendations and actions (Bao et al., 2025). Safety and reliability were highlighted as critical concerns, as autonomous decisions can have significant consequences for safety and operational performance (Mousavi et al., 2025). Regulatory compliance was also identified as a challenge, as autonomous decision-making must operate within regulatory frameworks that may not yet have adapted to these capabilities (Gálvez del Postigo Gallego et al., 2025).

Despite these challenges, the review found that autonomous decision-making through DT technologies has demonstrated significant potential in ICPS environments. Studies have shown that autonomous systems can process information and make decisions more quickly than human operators, enabling faster response to changing conditions (Karkaria et al., 2025). Autonomous systems can also consider more variables and alternatives than humans, enabling more comprehensive and optimal decisions (Mousavi et al., 2025). In manufacturing, autonomous decision-making has enabled real-time process optimization, dynamically adjusting operational parameters to respond to changing conditions and optimize performance (Karkaria et al., 2025). In energy systems, autonomous decision-making has enabled self-optimization of equipment performance, balancing operational efficiency with equipment health and maintenance needs (Bao et al., 2025).

4.5 Security, Interoperability, and Scalability Challenges

The analysis of challenges in DT adoption in ICPS environments reveals significant barriers that must be addressed for successful implementation. Security emerged as a primary concern, with studies highlighting the vulnerability of connected systems to cyber threats and the potential consequences of security breaches for operational safety and integrity (Hostos et al., 2023). The integration of DT systems with ICPS environments creates new attack surfaces and vulnerabilities, as the connection between physical and digital systems can be exploited by malicious actors. Interoperability was identified as another significant challenge, with studies reporting difficulties in integrating DT systems with existing systems and platforms (Mousavi et al., 2025). Scalability was also highlighted as a concern, as DT systems must be able to handle increasing numbers of assets and growing volumes of data (Vimal et al., 2025).

A key finding from the review concerns the need for comprehensive security engineering approaches for DT-enabled ICPS. Traditional security approaches that focus on protecting individual components

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

are inadequate for the complex, interconnected nature of DT-enabled ICPS, requiring instead holistic security approaches that consider the entire system (Hostos et al., 2023). Security engineering for DT-enabled ICPS must address multiple aspects, including securing sensors and communication networks; protecting DT platforms and applications; implementing access controls and authentication; managing security updates and patches; and building security into the architecture from the beginning (Hostos et al., 2023).

Table 5 presents the key challenges in DT adoption in ICPS and their potential solutions.

Table 5. Key Challenges and Potential Solutions for Digital Twin Adoption in ICPS

Challenge	Description	Potential Solutions
Security	Vulnerability to cyber threats, attack surfaces	Security-by-design, encryption, access controls
Interoperability	Integration with legacy systems, different platforms	Standardized interfaces, API-based integration
Scalability	Increasing assets, data volumes	Edge computing, cloud architectures
Data Quality	Inconsistent, incomplete sensor data	Data cleaning, validation, redundancy
Organizational Readiness	Skills, culture, change management	Training, leadership, phased implementation
Regulatory Compliance	Evolving regulations, liability concerns	Governance frameworks, audit trails

Sources: Adapted from Hostos et al. (2023), Mousavi et al. (2025), and Gálvez del Postigo Gallego et al. (2025)

Table 5 summarizes the primary challenges to DT adoption identified in the review and potential solutions that have been proposed or implemented to address them. Security challenges require security-by-design approaches that build security into the architecture from the beginning, supported by encryption and access controls. Interoperability challenges can be addressed through standardized interfaces and API-based integration that enable communication between different systems and platforms. Scalability challenges can be addressed through edge computing and cloud architectures that enable distributed processing and storage. Data quality challenges can be addressed through data cleaning, validation, and redundancy that ensure reliability and completeness of sensor data. Organizational readiness can be addressed through training, leadership commitment, and phased implementation that build capabilities and acceptance. Regulatory compliance challenges can be addressed through governance frameworks and audit trails that ensure accountability and compliance (Hostos et al., 2023; Mousavi et al., 2025).

The review found that organizational factors are as important as technical factors in successful DT adoption. Studies reported that organizational readiness, including the availability of skilled personnel, supportive leadership, and a culture of innovation, was essential for successful implementation (Gálvez del Postigo Gallego et al., 2025). Change management was highlighted as important, as DT adoption often requires significant changes in processes, roles, and responsibilities (Mousavi et al., 2025). Investment in training and skill development was identified as critical, as effective use of DT technologies requires new technical and analytical skills (Karkaria et al., 2025). The involvement of stakeholders throughout the implementation process was also highlighted as important for building acceptance and ensuring that systems meet user needs (Gálvez del Postigo Gallego et al., 2025).

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

5. Conclusion and Recommendations

This systematic review has provided a comprehensive examination of DT technologies for intelligent monitoring, predictive maintenance, and autonomous decision-making in ICPS. The review has synthesized findings from 87 publications published between 2020 and 2025, identifying key developments, challenges, and opportunities in this rapidly evolving field. The findings demonstrate that DT technologies have significant potential to transform industrial operations through enhanced monitoring capabilities, proactive maintenance approaches, and autonomous decision-making systems that improve efficiency, reduce costs, and enhance resilience.

The review has confirmed that intelligent monitoring through DT technologies provides unprecedented visibility into operational performance, enabling real-time tracking of equipment health, early detection of anomalies, and data-driven decision-making. Real-time data synchronization has emerged as a critical enabler of these capabilities, ensuring that digital representations accurately reflect physical asset behavior. However, challenges remain in data quality, integration, scalability, and computational demands, requiring continued research and development in sensor technology, data processing, and analytical capabilities.

Predictive maintenance through DT technologies has demonstrated substantial benefits in reducing unplanned downtime, extending equipment life, and optimizing maintenance costs. The development of cognitive digital twins with LLM-driven cognition and autonomous reasoning capabilities has enabled significant advances in fault diagnosis and prognosis, with studies reporting high accuracy in non-stationary fault diagnosis and substantial improvements in knowledge update time, cost efficiency, and early warning lead time. However, successful implementation requires high-quality data, effective AI algorithms, integration with existing systems, and organizational readiness.

Autonomous decision-making represents the most advanced application of DT technologies, enabling self-optimizing, self-adapting, and self-healing industrial systems. Frameworks for integrating decision support within DT systems have been developed, enabling systems to progress from decision support to semi-autonomous decision-making to fully autonomous decision-making. However, challenges remain in trust, explainability, safety, and regulatory compliance, requiring continued research and development in AI capabilities, human-machine interaction, and governance frameworks.

Security, interoperability, and scalability challenges continue to impede widespread DT adoption in ICPS environments. Security concerns require comprehensive security engineering approaches that address the complex, interconnected nature of DT-enabled ICPS. Interoperability challenges require standardized interfaces and API-based integration that enable communication between different systems and platforms. Scalability challenges require edge computing and cloud architectures that enable distributed processing and storage. Organizational factors, including the availability of skilled personnel, supportive leadership, and a culture of innovation, are as important as technical factors in successful DT adoption.

Based on these findings, several recommendations can be made for researchers and practitioners. For researchers, future work should focus on developing standardized frameworks for DT implementation that address interoperability, security, and scalability; investigating the integration of emerging technologies such as foundation models and generative AI with DT systems; exploring the role of DTs in supporting human-machine collaboration and Industry 5.0 objectives; developing methods for

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

uncertainty quantification and management in DT predictions; and investigating the organizational and human factors that influence DT adoption and use. For practitioners, recommendations include conducting thorough needs assessments before DT implementation; investing in data infrastructure and quality; developing organizational capabilities through training and skill development; adopting phased implementation approaches that build on successes; involving stakeholders throughout the implementation process; and establishing governance frameworks that address security, privacy, and compliance.

In conclusion, DT technologies for intelligent monitoring, predictive maintenance, and autonomous decision-making represent a significant opportunity for industrial transformation. While challenges remain, the demonstrated benefits and ongoing advances in technology and practice suggest that DT technologies will play an increasingly important role in creating smart, sustainable, and resilient industrial systems that align with the objectives of Industry 4.0 and Industry 5.0.

Acknowledgment

This research received no funding. The authors have no conflict of interest.

References

Bao, J., Wu, T., & Wang, C. (2025). Agentic digital twin-embedded maintenance methodology for energy equipment: A self-evolving operational paradigm. *Energy*, *305*, 112-128. https://doi.org/10.1016/j.energy.2025.132281

Gálvez del Postigo Gallego, L., Lazarova-Molnar, S., & del Real Torres, A. (2025). Decision support within digital twins in manufacturing ecosystems: A review. *IET Collaborative Intelligent Manufacturing*, *7*(3), e70041. https://doi.org/10.1049/cim2.70041

Hostos, H., Avila, O., & Goepp, V. (2023). A literature review of evaluation approaches for cyber-physical production systems. In *Proceedings of the 25th International Conference on Enterprise Information Systems* (Vol. 2, pp. 642-650). SciTePress. https://doi.org/10.5220/0011985200003467

Karkaria, V., Tsai, Y.-K., Chen, Y.-P., & Chen, W. (2025). An optimization-centric review on integrating artificial intelligence and digital twin technologies in manufacturing. *Engineering Optimization*, *57*(2), 161-207. https://doi.org/10.1080/0305215X.2024.2434201

Mousavi, S. M., Kim, D., & Lee, J. (2025). From simulation to autonomy: Reviews of the integration of artificial intelligence and digital twins. *International Journal of Precision Engineering and Manufacturing*, *12*, 1597-1628. https://doi.org/10.1007/s40684-025-00750-z

Vimal, K. E. K., Ganesh, R., & Kumar, S. (2025). Digital twins in additive manufacturing: A systematic review. *Additive Manufacturing Letters*, *33*, 101692. https://doi.org/10.1016/j.addlet.2025.101692

vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. In *Proceedings of the 17th European Conference on Information Systems* (pp. 2206-2217). AIS.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

مقالة بحثية

التنبؤ بإجهاد السريان بالاعتماد على تعلم الآلة وتحسين خرائط المعالجة لمركبات المصفوفة المعدنية Al/SiCp

حنفي م. عمر[1]، أسامة م. عرفان[1,*]

[1] قسم الهندسة الميكانيكية، كلية الهندسة، جامعة القصيم، بريدة 51452، المملكة العربية السعودية

*المؤلف المراسل: o.ahmed@qu.edu.sa

الملخص

تقارن هذه الدراسة أداء نماذج تعلم الآلة (ML) في التنبؤ بإجهاد السريان أثناء التشكيل الساخن لمركب Al/10 vol.% SiCp 2124 المصنوع بتقنية ميتالورجيا المساحيق. وتضم مجموعة البيانات 100 قيمة لإجهاد السريان جُمعت من بيانات ضغط ساخن منشورة [37] عند 300-500 °C و0.001-1.0 s^{-1} وانفعالات حقيقية تتراوح بين 0.1 و0.5، لتشكل شبكة كاملة بأبعاد 5 × 4 × 5. ودُرِّبت ثماني خوارزميات (الانحدار الخطي، والغابة العشوائية، والتعزيز التدريجي، وXGBoost، وشبكة عصبية، و LightGBM، وCatBoost، ونموذج تجميعي مكدَّس) على 15 خاصية مشتقة من درجة الحرارة ومعدل الانفعال والانفعال ومعامل زينر-هولومون (طاقة التنشيط 162 kJ mol^{-1}، المستخلصة من مواءمة بدالة الجيب الزائدي). وقُيِّم كل نموذج على 20 تقسيمًا عشوائيًا بنسب 70/15/15 مع التقسيم الطبقي حسب درجة الحرارة، وباختباري استبعاد درجة حرارة واحدة واستبعاد معدل انفعال واحد. وحقق التعزيز التدريجي والنموذج التجميعي المكدَّس أدنى خطأ في الاستيفاء (MAPE للاختبار = 5.2 ± 2.8% و5.3 ± 2.5%؛ $R^2 \approx 0.997$)، مقابل 14.3% لمعادلة الجيب الزائدي المواءمة على البيانات نفسها و17.0 ± 6.9% للانحدار الخطي؛ وكانت الفروق بين نماذج التعزيز أصغر من التشتت بين التقسيمات. وتركزت الأخطاء عند الإجهادات المنخفضة المقيسة عند 500 °C، وازدادت زيادة ملحوظة عند حجب مستوى كامل من درجة الحرارة أو معدل الانفعال (متوسط MAPE بين 25 و 33% وبين 15 و27% على التوالي للنماذج غير الخطية). ونسبت أهمية التبديل المجمَّعة 83 ± 3% من الأهمية التنبؤية إلى خصائص زينر-هولومون. ولأن مجموعات الأشجار ثابتة على فترات، اشتُقت خرائط المعالجة من نموذج بديل قائم على العملية الغاوسية يوفر حساسيات سلسة لمعدل الانفعال مع تقدير عدم يقينها. وتبلغ كفاءة تبديد القدرة ذروتها ($\eta \approx 53$-68%) عند 475-500 °C ونحو 0.0 02-0.01 s^{-1}، في حين يظل حد عدم الاستقرار غير محدد بدقة كافية لاعتماده على أربعة معدلات انفعال فقط. وتنطبق النتائج على الاستيفاء ضمن النطاق المختبَر لمركب واحد.

الكلمات المفتاحية: مركبات المصفوفة المعدنية؛ التنبؤ بإجهاد السريان؛ تعلم الآلة؛ الأساليب التجميعية؛ خرائط المعالجة؛ العملية الغاوسية.

جامعة القصيم
Qassim University

JEIT

مجلة التقنيات الناشئة والمبتكرة

المجلد 1، العدد 1
نوفمبر 2026

تاريخ الاستلام
3 سبتمبر 2026

تاريخ القبول
20 أكتوبر 2026

تاريخ النشر
5 نوفمبر 2026

معرّف الوثيقة الرقمي (DOI)
https://doi.org/10.71422/jeit.0007

الترخيص
CC BY-NC-SA 4.0

للاستشهاد بهذا البحث
حنفي م. عمر، أسامة م. عرفان. التنبؤ بإجهاد السريان بالاعتماد على تعلم الآلة وتحسين خرائط المعالجة لمركبات المصفوفة المعدنية Al/SiCp. مجلة التقنيات الناشئة والمبتكرة. 2026;1(1):84-111.

الموقع الإلكتروني للمجلة

© 2026 المؤلفون. نشرته جامعة القصيم بموجب ترخيص CC BY-NC-SA 4.0.

Qassim University جامعة القصيم

JEIT

Journal of Emerging and Innovative Technologies

Volume 1, Issue 1
November 2026

Received
03 September 2026

Accepted
20 October 2026

Published
05 November 2026

DOI
https://doi.org/10.71422/jeit.0007

License
CC BY-NC-SA 4.0

Cite this article

Omar HM, Irfan OM. Machine Learning–Based Flow Stress Prediction and Processing Map Optimization of Al/SiCp Metal Matrix Composites. J Emerg Innov Technol. 2026;1(1):84–111.

Journal website

Research article

Machine Learning–Based Flow Stress Prediction and Processing Map Optimization of Al/SiCp Metal Matrix Composites

Hanafy M. Omar[1], Osama M. Irfan[1,*]

[1] Department of Mechanical Engineering, College of Engineering, Qassim University, Buraydah 51452, Saudi Arabia

*Corresponding author: o.ahmed@qu.edu.sa

Abstract

This study benchmarks machine learning (ML) models for predicting the hot-deformation flow stress of a powder-metallurgy 2124 Al/10 vol.% SiC_p composite. The dataset comprises 100 flow-stress values compiled from published hot-compression data [37] at 300–500 °C, 0.001–1.0 s^{-1} and true strains of 0.1–0.5, forming a complete 5 × 4 × 5 grid. Eight algorithms (linear regression, random forest, gradient boosting, XGBoost, a neural network, LightGBM, CatBoost and a stacked ensemble) were trained on 15 features derived from temperature, strain rate, strain and the Zener–Hollomon parameter (activation energy 162 kJ mol^{-1}, obtained from a hyperbolic-sine fit). Each model was evaluated over 20 random, temperature-stratified 70/15/15 splits and by leave-one-temperature-out and leave-one-strain-rate-out tests. Gradient boosting and the stacked ensemble gave the lowest interpolation error (test MAPE = 5.2 ± 2.8% and 5.3 ± 2.5%; $R^2 \approx 0.997$), against 14.3% for the hyperbolic-sine equation fitted to the same data and 17.0 ± 6.9% for linear regression; differences among the boosted models were smaller than the split-to-split scatter. Errors were concentrated at the low stresses measured at 500 °C, and they increased markedly when a whole temperature or strain-rate level was withheld (mean MAPE of 25–33% and 15–27%, respectively, for the non-linear models). Grouped permutation importance attributed 83 ± 3% of the predictive importance to the Zener–Hollomon features. Because tree ensembles are piecewise constant, processing maps were derived from a Gaussian-process surrogate that provides smooth strain-rate sensitivities and their uncertainty. The highest efficiency of power dissipation ($\eta \approx 53$–68%) occurs at 475–500 °C and about 0.002–0.01 s^{-1}, whereas the instability boundary is poorly constrained by four strain rates. The results apply to interpolation within the tested envelope of a single composite.

Keywords: Metal Matrix Composites; Flow Stress Prediction; Machine Learning; Ensemble Methods; Processing Maps; Gaussian Process.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

1. Introduction

In the metalworking industry, accurate prediction of flow stress behavior during hot deformation is essential for manufacturing process optimization [1]. Flow stress, which represents the resistance of materials to plastic deformation at elevated temperatures, governs the design of forming equipment, determination of processing parameters, and prediction of final product quality [2], [3]. Understanding and predicting flow stress behavior is particularly essential for developing processing maps that delineate safe and unsafe hot working conditions [4]. The use of conventional aluminum alloys in automotive, aerospace and defense components is limited mainly by their loss of strength and stiffness at elevated temperatures; the addition of ceramic reinforcing particles is an established way of mitigating this limitation [5]. Metal matrix composites (MMCs) represent an important class of advanced materials that combine the properties of metallic matrices with ceramic reinforcements, offering superior specific strength, stiffness, and elevated temperature performance compared to conventional alloys [6], [7]. Particulate-reinforced aluminum composites offer enhanced stiffness, wear resistance, and elevated-temperature strength, making them suitable for aerospace, automotive, and defense applications [8], [9]. Hot workability is the ability of a material to deform plastically at elevated temperatures without flow localization, cracking or fracture [15]. Accurate characterization of hot deformation behavior is essential for designing thermomechanical processing routes and ensuring defect-free manufacturing [10]. For Al/SiC_p composites, conventional trial-and-error approaches to process optimization are inefficient, costly, and often unreliable due to the strong nonlinearity between processing parameters and material response [11].

Composites combine constituents so as to obtain combinations of properties that no single constituent provides [12]. Among various MMC systems, aluminum-based composites reinforced with silicon carbide particles have gained significant attention for aerospace and automotive applications due to their excellent strength-to-weight ratio and thermal stability [13], [14].

1.1 Hot Deformation Behavior and Processing Maps

Understanding the hot deformation behavior of MMCs is essential for designing efficient thermomechanical processing routes [16], [17]. Work hardening, dynamic recovery, and dynamic recrystallization mechanisms interact intricately to control the flow stress during hot working [3], [18]. Processing maps, developed based on the Dynamic Materials Model (DMM), provide a valuable framework for identifying optimal processing windows and unstable deformation regimes [19], [4]. For 2124 Al/SiC_p composites, processing maps have been used to examine the effect of the reinforcement volume fraction on hot workability [15]. The development of accurate processing maps requires precise flow stress data across a wide range of temperatures and strain rates. Traditional approaches rely on constitutive equations, such as the hyperbolic sine law proposed by Sellars and McTegart [20], which relates flow stress to the Zener-Hollomon parameter [18], [20]:

$$Z = \dot{\varepsilon} \exp\left(\frac{Q}{RT}\right) = A\left[\sinh\left(\alpha\sigma\right)\right]^n \tag{1}$$

where Z is the Zener–Hollomon parameter, $\dot{\varepsilon}$ is the strain rate, Q is the activation energy, R is the gas constant, T is absolute temperature, and A, α, and n are material constants. The Zener-Hollomon parameter serves as a master variable that captures the temperature-compensated strain rate, providing a unified descriptor for thermally activated plastic deformation processes.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

1.2 Limitations of Conventional Modeling Approaches

In order to describe the flow stress of metals during hot deformation, many mathematical models have been proposed. The conventional method is to carry out regression with experimental data based on the models to obtain the constants in the models. However, most of the factors affecting the flow stress are nonlinear, which makes the accuracy of the flow stress predicted using the regression method relatively low. Conventional regression-based methods for flow stress prediction suffer from several limitations [21], [2]. First, they demonstrate poor accuracy in capturing nonlinear relationships, with traditional constitutive models typically achieving mean absolute percentage error (MAPE) values of 10–15% [22]. For example, regression analysis applied to 2124 Al/10% SiCp composites yielded mean errors of up to 13.5% [37], indicating that the accuracy of flow stress predictions using regression methods remains unsatisfactory for many engineering applications. Second, there is difficulty in modeling complex interactions, as the interplay between temperature, strain rate, strain, and microstructural evolution is highly nonlinear [23]. Third, regression models exhibit limited extrapolation capability and often fail outside the experimental data range [24]. Finally, determining optimal material constants requires iterative optimization procedures involving manual parameter tuning [25].

1.3 Soft Computing and Machine Learning Approaches

To address the limitations of conventional approaches, soft processing techniques including fuzzy logic, artificial neural networks (ANNs), and genetic algorithms have been explored for flow stress prediction [26], [27], [28]. Fuzzy logic systems have been applied to flow stress prediction by modeling the approximate and imprecise reasoning processes common in human problem-solving. A fuzzy logic unit is made up of an inference engine, a fuzzy rule base, membership functions, a defuzzifier, and a fuzzifier. The concept of fuzzy reasoning enables input-output mapping of flow stress data through if-then rules that handle uncertainty and vague information effectively. Previous studies have demonstrated that fuzzy logic systems can effectively model the nonlinear relationships in flow stress data [27]. Recent applications of fuzzy logic to metal forming include work on 2124 Al/10% SiC_p composites, where fuzzy logic achieved a mean deviation of 6.1% compared to 13.5% for regression methods [37]. However, fuzzy systems require expert knowledge for membership function design and rule base construction, introducing subjectivity and limiting their generalizability across different material systems. Artificial neural networks have demonstrated improved performance over traditional models in several recent studies. Mandal et al. [22] developed ANN models for predicting high-temperature flow stress in AISI 304L stainless steel during hot torsion, with a reported error of approximately 6.0%. Quan et al. [29] compared ANN predictions with Arrhenius-type models for AA7050 aluminum alloy, reporting MAPE of 5.2% and a lower error than that of the Arrhenius-type model. Ji et al. [30] applied artificial neural networks to Alloy 617 flow stress prediction, achieving MAPE of 7.1%. Haghdadi et al. [21] applied ANNs to predict microstructural evolution during thermomechanical processing, demonstrating the potential of data-driven approaches in materials engineering. However, ANNs often suffer from overfitting when training data is limited and provide limited interpretability regarding which input features drive predictions, raising concerns for safety-critical applications. Ensemble tree algorithms such as random forests [31], gradient boosting [32] and XGBoost [33] have been applied increasingly in materials science [34], [35]. Ensemble methods combine multiple base learners to produce superior predictions through variance reduction and complementary learning. The fundamental principle is that aggregating diverse models with different inductive biases can capture complex patterns while maintaining robustness against individual model weaknesses.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Ensemble methods have been applied to predict mechanical properties of metal matrix composites, with reported improvements over single-model approaches. Li et al. [36] applied XGBoost for predicting flow stress in Al-Zn-Mg-Cu aluminum alloys, achieving MAPE of 3.9%. The sequential boosting mechanism with regularization controls the bias–variance trade-off, which often gives good accuracy on small tabular datasets. However, systematic comparative studies of these advanced ML techniques for flow stress prediction in MMCs remain limited.

1.4 Model Interpretability

A critical advancement in modern machine learning applications is model interpretability. The "black box" nature of complex models has been a persistent concern in safety-critical applications such as aerospace manufacturing. SHAP (SHapley Additive exPlanations) values provide a unified framework for explaining individual predictions [38]. Accumulated local effects plots have been developed to visualize feature effects when the inputs are correlated, a situation in which classical partial dependence plots can mislead [39]. Impurity-based importance measures built into tree ensembles are biased, and permutation importance is generally preferred [40].

1.5 Research Objectives

Despite these advances, several critical gaps remain in the literature. First, comprehensive comparative evaluations of multiple machine learning algorithms for flow stress prediction in aluminum composites are scarce, with most studies focusing on single algorithms without systematic comparison. Second, rigorous hyperparameter optimization using modern techniques such as extensive grid search or Bayesian optimization is often neglected, leading to suboptimal model performance. Third, ensemble methods specifically designed for constitutive modeling through stacked generalization remain underexplored, despite their proven effectiveness in other domains. Fourth, most studies report accuracy from a single random train–test split and do not test predictions at temperatures or strain rates that were withheld from training, so the reported errors mix interpolation with chance and say little about extrapolation. Finally, physics-based interpretability analysis to validate that models learn actual deformation mechanisms rather than spurious correlations is rarely performed, limiting confidence in model extrapolation beyond training data.

This study addresses these gaps through a systematic benchmarking study for 2124 Al/10 vol.% SiC_p composites. The specific objectives are to: (1) compare eight machine learning algorithms (linear regression, random forest, gradient boosting, XGBoost, neural network, LightGBM, CatBoost and a stacked ensemble) under a repeated-split protocol that reports the split-to-split variability; (2) optimize the hyperparameters of every model by grid search with 5-fold cross-validation on the training data only; (3) build a stacked-generalization ensemble whose ridge meta-learner is fitted on out-of-fold predictions; (4) distinguish interpolation from extrapolation with leave-one-temperature-out and leave-one-strain-rate-out tests; (5) analyze interpretability with grouped permutation importance and partial dependence, and compare the learned response quantitatively with the hyperbolic-sine constitutive equation fitted to the same data; (6) quantify the difference between the ML models and the regression and fuzzy-logic approaches previously applied to this composite; and (7) construct processing maps from a smooth Gaussian-process surrogate, together with their uncertainty.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

2. Data Source and Experimental Background

2.1 Material and Data Source

The flow-stress data analyzed in this study were not generated by the authors. They were compiled from a published hot-compression study of a 2124 Al/10 vol.% SiC_p composite [37], which also reports the regression (13.5%) and fuzzy-logic (6.1%) errors used as benchmarks in Sections 1 and 4.7. As described in [37], the composite was produced by the powder-metallurgy route: 2124 aluminum alloy powder was blended with 10 vol.% silicon carbide particles (average size 25 μm) in a tubular mixer for 4 hours.

The nominal composition limits of the 2124 matrix alloy [41] are listed in Table 1. 2124 is a high-strength Al–Cu–Mg alloy used in aerospace applications because of its high strength-to-weight ratio.

Table 1. Nominal chemical composition limits of 2124 aluminum alloy (wt.%; single values are maxima) [41]

Cu	Mg	Mn	Si	Fe	Cr	Zn	Ti	Others (total)	Al
3.8–4.9	1.2–1.8	0.30–0.9	0.20	0.30	0.10	0.25	0.15	0.15	Balance

In the source study [37], the powder mixture was degassed at 450 °C for 1 hour under vacuum pressure of 10^{-4} torr to remove moisture and adsorbed gases. Vacuum hot pressing was performed at 450 °C under 75 MPa pressure to produce cylindrical compacts with dimensions of 10 mm diameter × 15 mm height. The green compacts were subsequently sintered at 500 °C for 2 hours in a vacuum atmosphere to achieve full densification and ensure strong interfacial bonding between the aluminum matrix and SiC reinforcement particles.

2.2 Hot Compression Test Conditions

In the source study [37], isothermal hot compression tests were performed on a servo-controlled universal testing machine of 100 kN capacity equipped with a three-zone resistance furnace. Specimens were heated to the test temperature (300, 350, 400, 450 and 500 °C) at 10 °C/min and held for 5 min; the temperature, monitored with K-type thermocouples, was maintained within ±2 °C.

The tests were performed at constant true strain rates of 0.001, 0.01, 0.1 and 1.0 s^{-1}, spanning three orders of magnitude, to a true strain of 0.5, with graphite foil as lubricant. Load–displacement data were converted to true stress–true strain curves, from which the flow stresses at true strains of 0.1–0.5 were taken.

2.3 Test Matrix

The dataset covers five strain levels (0.1, 0.2, 0.3, 0.4 and 0.5) for every combination of temperature and strain rate (Table 2), giving 100 flow-stress values (5 temperatures × 4 strain rates × 5 strain levels) that form a complete factorial grid over the processing space relevant to hot extrusion, forging and rolling. The grid structure has consequences for model validation, which are addressed in Section 3.2.

Table 2. Flow stress (MPa) of 2124 Al/10 vol.% SiCp at various temperatures, strain rates and strains, compiled from [37] (values corrected for deformation heating; see Section 2.4)

Temperature (°C)	True strain	Strain rate (s^{-1})			
		0.001	0.01	0.1	1.0
300	0.1	92.0	120.32	156.5	215.2

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Temperature (°C)	True strain	Strain rate (s^{-1})			
		0.001	0.01	0.1	1.0
	0.2	87.78	118.62	152.8	212.5
	0.3	81.23	116.32	149.0	200.8
	0.4	77.82	115.37	145.4	195.62
	0.5	76.28	114.32	144.4	187.54
350	0.1	62.30	84.91	124.71	172.68
	0.2	59.91	82.87	123.35	169.32
	0.3	57.42	81.51	120.30	163.26
	0.4	55.91	80.32	118.10	157.85
	0.5	56.04	79.12	116.24	151.25
400	0.1	42.82	64.97	88.47	137.77
	0.2	41.62	64.83	87.21	133.62
	0.3	40.57	63.51	85.45	130.28
	0.4	41.30	61.89	83.92	125.28
	0.5	41.56	60.52	81.90	120.01
450	0.1	25.12	41.46	63.87	107.62
	0.2	27.28	41.74	63.03	105.03
	0.3	25.52	39.65	61.55	101.92
	0.4	23.89	38.85	59.12	97.75
	0.5	24.03	37.66	57.94	94.36
500	0.1	5.54	16.75	32.95	70.78
	0.2	8.10	17.82	32.13	67.67
	0.3	9.72	17.42	30.37	62.62
	0.4	7.95	16.72	28.75	54.28
	0.5	5.75	16.61	26.53	45.92

2.4 Deformation-Heating Correction

The flow stresses in Table 2 were corrected for deformation heating. The temperature rise at strain ε was estimated as

$$\Delta T = \frac{\beta}{\rho C_p} \int_0^{\varepsilon} \sigma \, d\varepsilon \tag{2}$$

where $\beta = 0.9$ is the fraction of plastic work converted into heat (Taylor–Quinney coefficient), $\rho = 2780$ kg/m³ is the density and $C_p = 875$ J/(kg·K) is the specific heat capacity of the composite. The corresponding stress correction is

$$\Delta \sigma = -\left(\frac{\partial \sigma}{\partial T}\right)_{\dot{\varepsilon},\varepsilon} \Delta T \tag{3}$$

where the temperature sensitivity $\partial\sigma/\partial T$ at constant strain rate and strain is evaluated by finite differences of the data in Table 2. Equation (2) assumes that all of the heat is retained in the specimen and is therefore an upper bound. At 0.001 s^{-1} (test duration ≈ 500 s) heat is conducted to the dies and the

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

deformation is nearly isothermal; adiabatic correction factors that increase from about zero at low strain rates to about one at 1–10 s^{-1} describe this transition more realistically [42]. Table 3 lists the upper-bound values of ΔT and $\Delta\sigma$ at $\varepsilon = 0.5$ for every test condition. The correction is 0.5–6.3 MPa (6–9% of the flow stress) at 0.001 s^{-1} and 11–28 MPa (12–24%) at 1.0 s^{-1}, where the temperature rise reaches 38 K at 300 °C. Because the correction was applied with $\beta = 0.9$ at all strain rates, the flow stresses at the lowest strain rates may be slightly over-corrected; this is noted as a limitation in Section 4.10.

Table 3. Upper-bound temperature rise (ΔT) and stress correction (Δσ) at ε = 0.5 estimated from Eqs. (2) and (3), with Δσ as a percentage of the flow stress

Temperature (°C)	Strain rate (s^{-1}): ΔT (K) / $\Delta\sigma$ (MPa) ($\Delta\sigma/\sigma$)			
	0.001	0.01	0.1	1.0
300	15.6 / 6.3 (8%)	21.8 / 15.3 (13%)	27.9 / 15.7 (11%)	37.9 / 27.5 (15%)
350	10.9 / 3.8 (7%)	15.2 / 8.2 (10%)	22.5 / 14.0 (12%)	30.5 / 20.6 (14%)
400	7.7 / 2.5 (6%)	11.8 / 4.9 (8%)	15.9 / 9.3 (11%)	24.3 / 13.8 (12%)
450	4.7 / 1.7 (7%)	7.4 / 3.3 (9%)	11.4 / 6.3 (11%)	19.0 / 14.1 (15%)
500	1.4 / 0.5 (9%)	3.2 / 1.3 (8%)	5.7 / 3.6 (13%)	11.6 / 11.2 (24%)

2.5 Data Characteristics

The flow stress data exhibit strong dependencies on thermomechanical conditions that are consistent with established hot deformation behavior of aluminum alloys. The flow stress range spans 5.54–215.20 MPa, representing a 39-fold variation across the processing window. Temperature exerts a strong softening effect, with increasing temperature from 300 °C to 500 °C reducing flow stress by 67–94% due to enhanced thermal activation of dislocation motion, dynamic recovery, and potential dynamic recrystallization. Conversely, the strain rate effect shows that increasing strain rate from 0.001 to 1.0 s^{-1} increases flow stress by 134% (300 °C, $\varepsilon = 0.1$) to more than ten-fold (500 °C, $\varepsilon = 0.1$) due to the rate sensitivity of thermally-activated deformation processes, where higher deformation rates outpace recovery mechanisms. The strain dependence generally shows flow stress decreasing with increasing strain, particularly at higher temperatures, indicating dynamic softening processes dominate over work hardening. These trends are consistent with thermally activated processes such as dislocation climb, cross-slip, and dynamic recovery/recrystallization that govern flow stress evolution in aluminum alloys and their composites at elevated temperatures.

3. Methodology

3.1 Feature Engineering

Physics-informed features were derived from the three primary variables, temperature (T, K), strain rate ($\dot{\varepsilon}$, s^{-1}) and strain (ε). The central derived variable is the Zener–Hollomon parameter, $Z = \dot{\varepsilon}\exp(Q/RT)$ (Eq. 1), which combines temperature and strain rate into a single temperature-compensated strain rate.

The activation energy Q was determined from the data in Table 2 following Sellars and McTegart [20]. The constants n_1 and β' were obtained as the mean inverse slopes of $\ln\sigma$ and σ, respectively, against $\ln\dot{\varepsilon}$ over the five temperatures, giving $\alpha = \beta'/n_1$; n is the mean inverse slope of $\ln[\sinh(\alpha\sigma)]$ against $\ln\dot{\varepsilon}$, and $Q = R\,n\,S$, where S is the mean slope of $\ln[\sinh(\alpha\sigma)]$ against $1/T$. With all strains pooled, this gives $\alpha = 0.0148$ MPa^{-1}, $n = 3.86$ and $Q = 161.7$ kJ mol^{-1}; a linear regression of $\ln Z$ on $\ln[\sinh(\alpha\sigma)]$ then gives a slope of 3.66 and $\ln A = 24.88$ (A in s^{-1}) with $R^2 = 0.954$. Fitting each strain separately gives $Q = 156.8$–173.5 kJ mol^{-1}.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

The rounded value $Q = 162$ kJ mol^{-1} was used for all Z-based features, and $R = 8.314$ J/(mol·K). The fitted value is higher than the activation energy for lattice self-diffusion in aluminum (about 142 kJ mol^{-1}), as is common for particle-reinforced aluminum composites. The strain-dependent hyperbolic-sine equation (constants fitted separately at each strain) reproduces Table 2 with MAPE = 14.3% (15.0% with a single set of constants); it serves as the physics-based baseline in Section 4.

Additional derived features are the temperature-compensation term $\exp(Q/RT)$; the logarithmic transformations $\ln \dot{\varepsilon}$, $\ln T$ and $\ln Z$; the interaction terms $T \times \dot{\varepsilon}$, $T \times \varepsilon$, $\dot{\varepsilon} \times \varepsilon$ and $T \times \dot{\varepsilon} \times \varepsilon$; and the quadratic terms T^2, $\dot{\varepsilon}^2$ and ε^2. Together with T, $\dot{\varepsilon}$, ε and Z, the feature matrix contains 15 variables. All of them are deterministic functions of T, $\dot{\varepsilon}$ and ε, so several are strongly correlated (Section 4.3); this affects the importance analysis (Section 3.7).

3.2 Data Splitting, Repeated Evaluation and Extrapolation Tests

For the interpolation tests, the 100 data points were divided into training (70 points), validation (15) and test (15) sets by random sampling stratified on temperature, so that each of the five temperatures contributes 14, 3 and 3 points to the three sets. The complete pipeline (splitting, scaling, hyperparameter search, training and evaluation) was repeated for 20 random seeds, and results are reported as the mean ± standard deviation (SD) over the 20 splits. Because the data form a complete grid, a random split leaves almost every test point with training points at the same temperature and strain rate, so random-split scores measure interpolation, mainly along the strain axis. Two grouped cross-validation schemes were therefore added: leave-one-temperature-out (LOTO; five folds, each withholding all 20 points at one temperature) and leave-one-strain-rate-out (LOSRO; four folds, each withholding all 25 points at one strain rate). Withholding 300 or 500 °C, or 0.001 or 1.0 s^{-1}, tests extrapolation; withholding an interior level tests interpolation across a gap of 50 °C or one decade of strain rate. Features were standardized with z-scores:

$$x_{scaled} = \frac{x - \mu}{\sigma} \tag{4}$$

where μ and σ are mean and standard deviation computed from the training set only to prevent data leakage. This standardization is critical for gradient-based optimization algorithms and distance-based methods to converge efficiently.

3.3 Machine Learning Algorithms

Eight regression models were compared. Ordinary least squares linear regression on the engineered features serves as the baseline. Random forest [31] averages the predictions of deep regression trees grown on bootstrap samples with random feature subsets. Gradient boosting [32], XGBoost [33], LightGBM and CatBoost build additive ensembles of shallow trees, each new tree being fitted to the negative gradient of the loss with respect to the current prediction; LightGBM grows trees leaf-wise with histogram-based splits, and CatBoost uses symmetric trees and ordered boosting. The neural network is a multilayer perceptron with two hidden layers of 50 tanh units, trained with the Adam optimizer on standardized inputs and a standardized target. XGBoost minimizes the regularized objective

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{5}$$

where l is the squared-error loss, $\hat{y}_i$ is the prediction for sample i, f_k is the k-th tree, and the complexity penalty of each tree is

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

$$\Omega(f)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2 \tag{6}$$

where T here denotes the number of leaves, w_j are the leaf weights, and γ and λ are regularization parameters. Regularization matters in the present problem because, with only 70 training points, unpenalized trees can reproduce the training data almost exactly (Section 4.1). The fixed settings, the search grids and the most frequently selected values for all models, including LightGBM and CatBoost, are listed in Table 4.

Table 4. Model settings, hyperparameter grids and the values most frequently selected over the 20 splits (5-fold cross-validation on the training set)

Model	Fixed settings	Grid searched	Most frequently selected
Linear regression	Ordinary least squares; standardized features	–	–
Random forest	200 trees	Max. depth (8, unlimited); min. samples per leaf (1, 2)	Depth 8 (11/20); leaf 1 (20/20)
Gradient boosting	300 trees	Learning rate (0.05, 0.1); max. depth (2, 3)	0.1 (20/20); depth 2 (19/20)
XGBoost	300 trees; column subsample 0.8; $\lambda = 1$	Learning rate (0.05, 0.1); max. depth (2, 3)	0.1 (17/20); depth 2 (15/20)
Neural network	2 × 50 tanh units; Adam; ≤ 5000 iterations; standardized target	L2 penalty (10^{-4}, 10^{-2}); initial learning rate (10^{-3}, 10^{-2})	10^{-2} (16/20); 10^{-2} (19/20)
LightGBM	300 trees; min. 3 samples per leaf	Learning rate (0.05, 0.1); leaves (4, 8)	0.1 (19/20); 4 leaves (14/20)
CatBoost	300 iterations	Learning rate (0.05, 0.1); depth (3, 4)	0.1 (20/20); depth 3 (19/20)
Enhanced ensemble	Base learners: tuned XGBoost, gradient boosting, random forest, neural network; ridge meta-learner	Ridge penalty: 13 values, 10^{-3}–10^{3} (internal cross-validation)	–

3.4 Enhanced Ensemble Model

The enhanced ensemble uses stacked generalization [43]. The four base learners (XGBoost, gradient boosting, random forest and the neural network), each with the hyperparameters selected in Section 3.5, generate out-of-fold predictions for the 70 training points by 5-fold cross-validation. These predictions form the meta-feature matrix $\boldsymbol{H} = [\hat{y}_{XGB}, \hat{y}_{GB}, \hat{y}_{RF}, \hat{y}_{NN}]$, to which a ridge-regression meta-learner is fitted:

$$\min_{\alpha} \lVert y - H\alpha \rVert_2^2 + \lambda \lVert \alpha \rVert_2^2 \tag{7}$$

where α are the ensemble weights and the penalty λ is selected by internal cross-validation. The base learners are then refitted on the whole training set, and the final prediction is

$$\hat{y}_{ensemble} = \sum_{k=1}^{4} \alpha_k \hat{y}_k \tag{8}$$

Fitting the meta-learner on out-of-fold predictions, rather than on the validation set, keeps the validation and test sets independent of every fitting step. Averaged over the 20 splits, the ridge weights were 0.72 ± 0.33 (gradient boosting), 0.29 ± 0.32 (XGBoost), 0.13 ± 0.10 (neural network) and −0.13 ±

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

0.15 (random forest); the large scatter and the negative random-forest weight indicate that the base predictions are highly collinear.

3.5 Hyperparameter Optimization

For every split, the hyperparameters were selected by grid search with 5-fold cross-validation on the 70 training points only, using the mean root-mean-square error over the folds as the criterion, and the selected model was refitted on the whole training set. The validation set was used only to report validation scores; it was not used for early stopping, model selection or fitting the meta-learner. In the LOTO and LOSRO tests the same search was repeated inside each training fold. The grids (Table 4) are deliberately small, because 70 training points cannot support fine tuning.

3.6 Performance Metrics

Model performance was quantified using multiple complementary metrics to provide comprehensive evaluation. The coefficient of determination (R^2) measures the proportion of variance explained:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \acute{y})^2} \tag{9}$$

Root mean square error (RMSE) quantifies average prediction error magnitude in the original units (MPa):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{10}$$

Mean absolute error (MAE) provides a robust measure less sensitive to outliers:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{11}$$

Mean absolute percentage error (MAPE) is reported because it is widely used in the literature, but it weights errors at low stresses heavily: an error of 4 MPa is 72% of the 5.54 MPa flow stress at 500 °C and 0.001 s^{-1}, but only 2% of the 215.2 MPa flow stress at 300 °C and 1.0 s^{-1}. It is therefore always reported together with RMSE and MAE:

$$MAPE = \frac{100\,\%}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{12}$$

3.7 Model Interpretability Analysis

Feature importance was evaluated for XGBoost by permutation importance on the test set of each of the 20 splits, with 30 permutations per split and the increase in RMSE as the error measure:

$$Importance(X_j) = \frac{1}{K}\sum_{k=1}^{K} \left[Error\left(perm(X_j)_k\right) - Error(X) \right] \tag{13}$$

where K is the number of permutations. Because Z and $\ln Z$ are monotonic transformations of each other (Spearman ρ = 1.00) and several other features are strongly correlated (Section 4.3), permuting one feature at a time splits the credit among correlated features arbitrarily. The features were therefore

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

permuted jointly in five physically defined groups: the Zener–Hollomon group (Z, ln Z); the temperature group (T, ln T, T^2, exp(Q/RT)); the strain-rate group ($\dot{\varepsilon}$, ln $\dot{\varepsilon}$, $\dot{\varepsilon}^2$); the strain group (ε, ε^2); and the interaction group ($T \times \dot{\varepsilon}$, $T \times \varepsilon$, $\dot{\varepsilon} \times \varepsilon$, $T \times \dot{\varepsilon} \times \varepsilon$). Importances are reported as the increase in test RMSE (MPa) and as a share of the total, with the SD across the 20 splits. Impurity-based importances were not used because they are biased and are not on the same scale as permutation importances [40]. Partial dependence describes the marginal effect of a variable on the prediction, averaged over the other inputs [32]:

$$PD(x_j) = \frac{1}{n}\sum_{i=1}^{n} f(x_j, x_{i,-j}) \tag{14}$$

where $x_{i,-j}$ denotes the values of all inputs except x_j. Because the engineered features are functions of T, $\dot{\varepsilon}$ and ε, partial dependence was computed for the three primary variables, with all derived features recomputed consistently for each substituted value.

3.8 Gaussian-Process Surrogate for Processing Maps

The processing-map parameters (Section 4.9) require the strain-rate sensitivity $m = \partial \ln \sigma / \partial \ln \dot{\varepsilon}$ and its derivative. Tree ensembles are piecewise constant in each input: on a 41 × 61 grid of temperature and ln $\dot{\varepsilon}$ at $\varepsilon = 0.5$, the XGBoost prediction gave $m = 0$ at 47% of the grid steps and $m < 0$ at 23% of them, so it cannot provide a usable derivative. A Gaussian-process (GP) regression model [44] was therefore fitted to ln σ with the standardized inputs $1000/T$, ln $\dot{\varepsilon}$ and ε, a squared-exponential kernel with a separate length scale for each input, and a white-noise term; the kernel hyperparameters were obtained by maximizing the marginal likelihood. The posterior mean is smooth, so m and ξ follow from finite differences on a fine grid, and 500 samples of the noise-free posterior surface give the uncertainty of η and the probability that $\xi < 0$ at each grid point. Over the same 20 splits, the GP gave a test MAPE of 6.2 ± 2.2% (R^2 = 0.9945 ± 0.0023), comparable to that of the tree ensembles.

4. Results and Discussion

4.1 Model Performance over Repeated Splits

Table 5 gives the mean ± SD of R^2, RMSE, MAE and MAPE over the 20 splits for the training, validation and test sets, and Fig. 1 shows the test MAPE of every split.

Table 5. Performance of the eight models over 20 random temperature-stratified splits (mean ± SD; training n = 70, validation n = 15 and test n = 15 per split)

Model	Training		Validation		Test			
	R^2	MAPE (%)	R^2	MAPE (%)	R^2	RMSE (MPa)	MAE (MPa)	MAPE (%)
Linear regression	0.9885 ± 0.0023	12.65 ± 1.33	0.9751 ± 0.0188	19.19 ± 10.09	0.9813 ± 0.0106	6.60 ± 1.12	5.04 ± 0.91	17.01 ± 6.87
Random forest	0.9985 ± 0.0005	2.24 ± 0.18	0.9870 ± 0.0113	7.26 ± 3.27	0.9895 ± 0.0059	5.07 ± 1.65	3.48 ± 0.98	6.98 ± 3.26
Gradient boosting	1.0000 ± 0.0000	0.47 ± 0.14	0.9954 ± 0.0055	5.40 ± 2.75	0.9970 ± 0.0024	2.60 ± 0.88	1.85 ± 0.55	5.16 ± 2.76
XGBoost	0.9999 ± 0.0000	0.73 ± 0.25	0.9945 ± 0.0061	5.86 ± 3.45	0.9962 ± 0.0031	2.91 ± 1.03	1.95 ± 0.53	5.82 ± 3.45
Neural network	0.9969 ± 0.0018	4.82 ± 1.71	0.9862 ± 0.0145	9.54 ± 4.44	0.9910 ± 0.0051	4.57 ± 1.24	3.54 ± 0.90	7.66 ± 3.13

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Model	Training		Validation		Test			
	R^2	MAPE (%)	R^2	MAPE (%)	R^2	RMSE (MPa)	MAE (MPa)	MAPE (%)
LightGBM	0.9999 ± 0.0001	0.84 ± 0.43	0.9888 ± 0.0199	6.66 ± 3.42	0.9905 ± 0.0155	4.12 ± 2.89	2.76 ± 1.57	6.39 ± 3.98
CatBoost	0.9999 ± 0.0000	1.06 ± 0.21	0.9928 ± 0.0086	6.02 ± 3.61	0.9942 ± 0.0047	3.64 ± 1.54	2.35 ± 0.65	6.15 ± 3.94
Enhanced ensemble	0.9997 ± 0.0002	1.48 ± 0.48	0.9955 ± 0.0049	5.61 ± 2.62	0.9968 ± 0.0023	2.70 ± 0.86	1.97 ± 0.56	5.33 ± 2.50

Four observations follow from Table 5 and Fig. 1. First, the tree-based models, the neural network and the ensemble reach mean test MAPE values of 5.2–7.7% and R^2 of 0.990–0.997, and the split-to-split SD (2.5–4.0 percentage points of MAPE) is as large as the differences between them. Gradient boosting had the lowest mean test MAPE (5.2 ± 2.8%), followed by the enhanced ensemble (5.3 ± 2.5%) and XGBoost (5.8 ± 3.5%), but the ranking changes from split to split; XGBoost, for example, ranged from 1.4% to 15.7%. Paired Wilcoxon signed-rank tests over the 20 splits (not corrected for multiple comparisons) show that gradient boosting and the ensemble are more accurate than random forest, the neural network and LightGBM ($p \leq 0.008$), whereas the differences among gradient boosting, XGBoost, CatBoost and the ensemble are not significant ($p \geq 0.07$). The stacked ensemble therefore brings no measurable gain over its best base learner.

Second, the tree ensembles reproduce the training data almost exactly (training MAPE 0.5–2.2%) while their validation and test MAPE are 5–7%, a gap that reflects variance rather than bias at this sample size. On individual splits the test MAPE may be higher or lower than the validation MAPE; both sets contain only 15 points, and differences of one or two percentage points between them, or between models on a single split, are within the noise.

Third, linear regression on the engineered features explains most of the variance (test $R^2 = 0.981 \pm 0.011$) but has a test MAPE of 17.0 ± 6.9%. Its absolute error (RMSE = 6.6 ± 1.1 MPa) is only 2–2.5 times that of the boosted models (2.6–2.9 MPa), but it is spread evenly over the stress range and therefore becomes large in relative terms at the low stresses measured at 500 °C (Section 4.5). Training the linear model on $\ln \sigma$ instead of σ reduced its test MAPE to 10.1 ± 3.0%, and the same change reduced the XGBoost test MAPE to 4.7 ± 1.9%, which indicates that the power-law character of the data should be reflected in the target as well as in the features. Fourth, out-of-fold predictions from ten repetitions of 5-fold cross-validation on all 100 points give the same picture: $R^2 = 0.9969 \pm 0.0009$ and MAPE = 4.6 ± 0.4% for XGBoost, against $R^2 = 0.983 \pm 0.001$ and MAPE = 16.3 ± 0.9% for linear regression. Because the R^2 values of all models lie between 0.98 and 1.00, Fig. 1 compares the models by MAPE.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

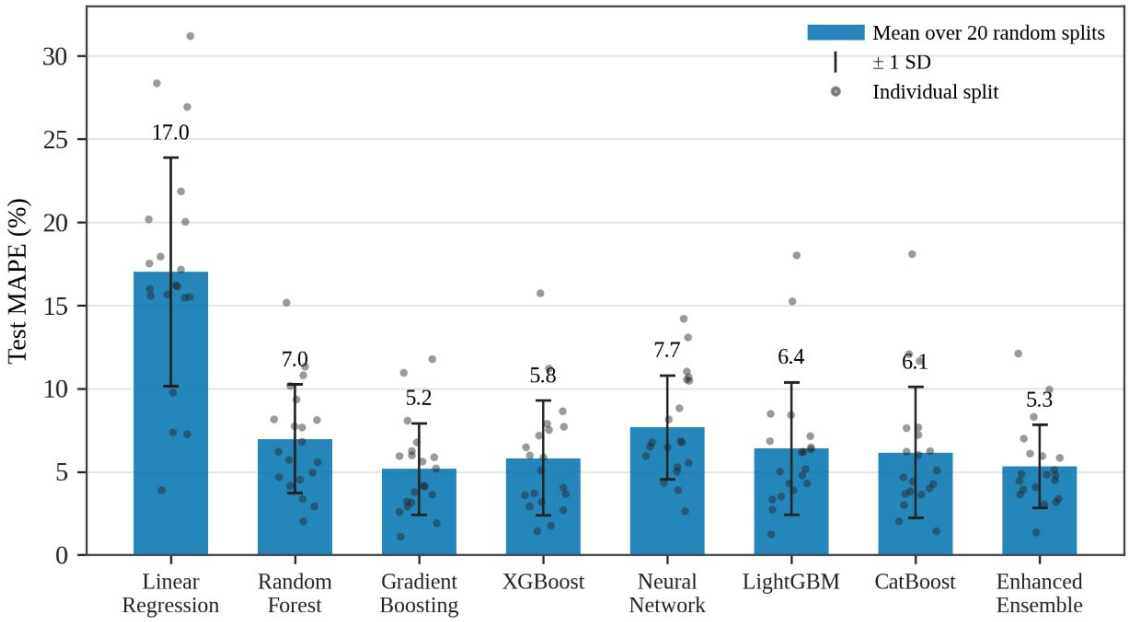


Figure 1. Test MAPE of the eight models over 20 random splits (bars: mean; error bars: ±1 SD; dots: individual splits; n = 15 test points per split)

4.2 Prediction at Withheld Temperatures and Strain Rates

Table 6 lists the MAPE obtained when all points at one temperature or one strain rate are withheld from training. When an interior temperature (350, 400 or 450 °C) is withheld, the error is 1.4–31.5%, depending on the model and the level, and no model is consistently best (e.g., XGBoost gives 31.5% at 400 °C and CatBoost 1.4% at 350 °C). Withholding an end temperature is much more severe. The tree-based models cannot predict outside the stress range seen in training and give errors of 11–13% at 300 °C and 88–131% at 500 °C, where the stresses are lowest; the linear model and the neural network extrapolate unstably (122% and 35% at 300 °C). The strain-rate tests show the same pattern for the non-linear models: 3.6–20% for the interior strain rates, 27–52% at 0.001 s^{-1} and 14–46% at 1.0 s^{-1}. The linear model diverges when 1.0 s^{-1} is withheld, because its quadratic strain-rate terms are then unconstrained. The GP surrogate of Section 3.8 behaves similarly (mean MAPE of 44.6% for LOTO and 27.0% for LOSRO).

The random-split errors in Table 5 therefore describe interpolation within the tested grid. For a process engineer who needs a prediction at a temperature or strain rate that has not been tested, errors of the order of 10–30% inside the envelope, and much larger errors outside it, are the realistic expectation.

Table 6. Test MAPE (%) when all points at one temperature (leave-one-temperature-out, 20 test points per fold) or at one strain rate (leave-one-strain-rate-out, 25 test points per fold) are withheld

Model	Temperature withheld (°C)						Strain rate withheld (s^{-1})				
	300	350	400	450	500	Mean	0.001	0.01	0.1	1.0	Mean
Linear regression	121.7	6.4	3.6	11.1	56.4	39.8	121.0	24.2	12.8	>1000	>1000
Random forest	13.0	4.5	3.5	20.7	87.9	25.9	39.8	19.9	11.1	14.4	21.3
Gradient boosting	11.1	4.0	4.7	13.0	99.2	26.4	32.4	11.3	5.5	15.8	16.3
XGBoost	11.0	7.0	31.5	11.4	100.8	32.3	41.1	10.2	10.6	24.2	21.5

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Model	Temperature withheld (°C)						Strain rate withheld (s^{-1})				
	300	350	400	450	500	Mean	0.001	0.01	0.1	1.0	Mean
Neural network	35.3	9.6	6.7	9.3	90.9	30.3	30.0	6.0	8.0	45.9	22.5
LightGBM	10.9	6.8	5.0	6.8	97.1	25.3	32.7	3.6	9.5	15.8	15.4
CatBoost	11.2	1.4	12.8	7.0	130.6	32.6	52.4	19.9	9.4	24.4	26.5
Enhanced ensemble	10.5	5.9	15.2	12.1	93.0	27.4	27.2	8.0	5.4	24.2	16.2

4.3 Model Interpretability and Feature Importance

Fig. 2 shows the Spearman rank-correlation matrix of the 15 features. Z and ln Z are perfectly rank-correlated (ρ = 1.00), and 25 of the 105 feature pairs have $|\rho| > 0.9$, so single-feature importances are unreliable for this feature set. The grouped permutation importances of XGBoost are shown in Fig. 3 and Table 7.

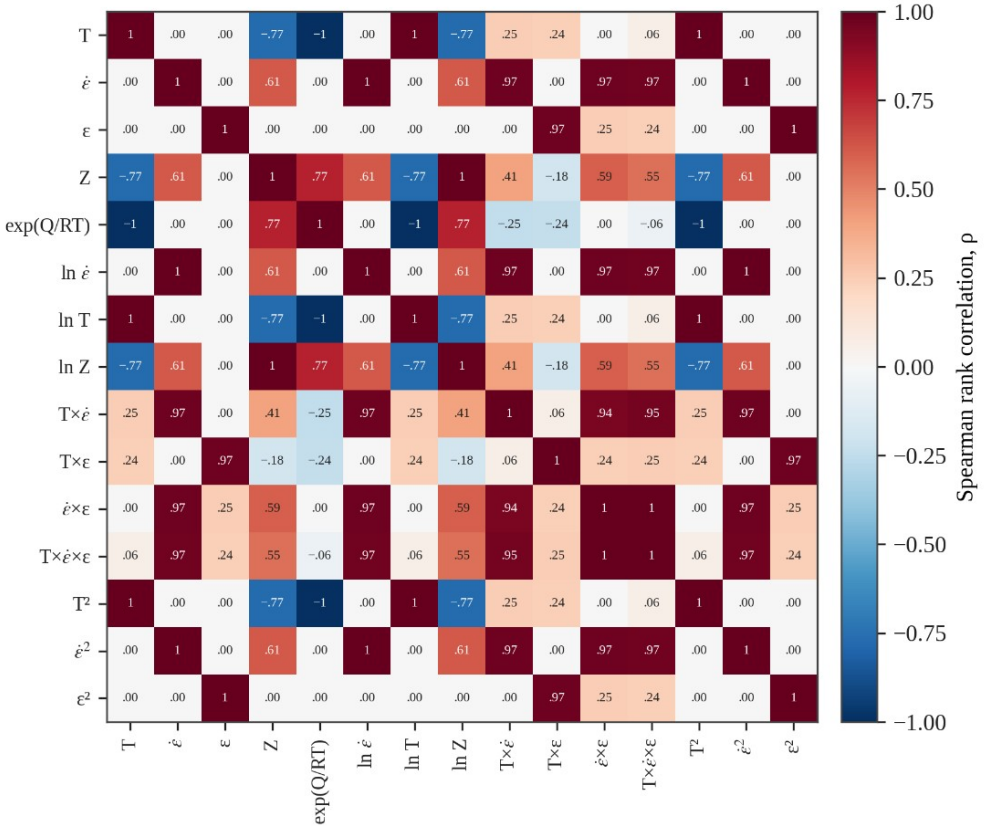


Figure 2. Spearman rank-correlation matrix of the 15 input features (all 100 data points)

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

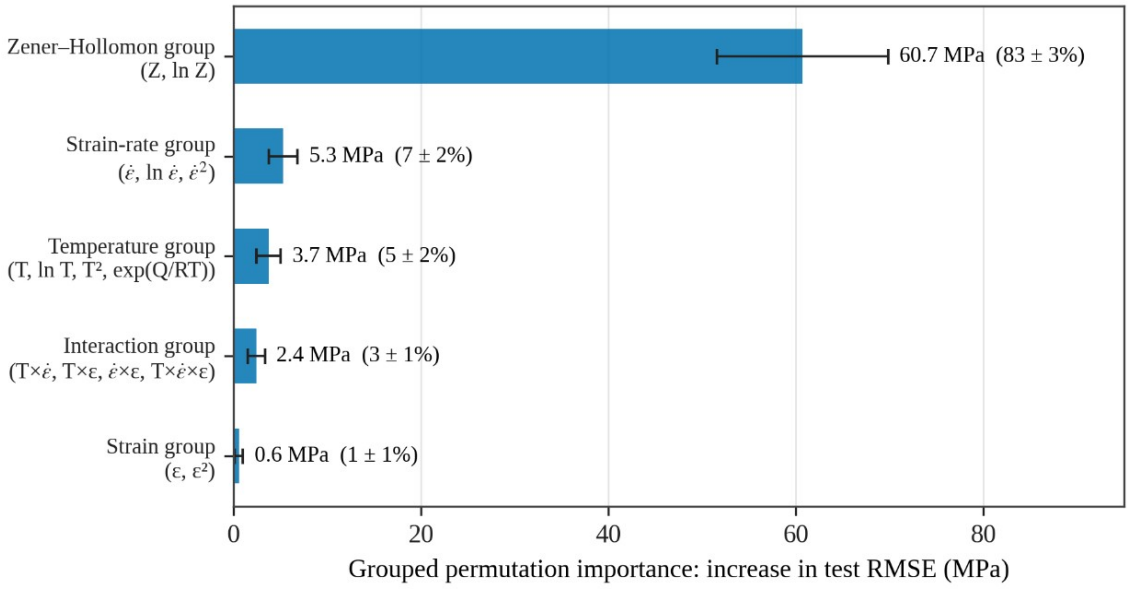


Figure 3. Grouped permutation importance of the XGBoost model, expressed as the increase in test RMSE when each feature group is permuted (bars: mean over 20 splits, 30 permutations per split; error bars: ±1 SD across splits); the share of the total importance is given in parentheses

Table 7. Grouped permutation importance of the XGBoost model on the test sets (mean ± SD over 20 splits; 30 permutations per split)

Rank	Feature group	Increase in test RMSE (MPa)	Share of total (%)
1	Zener–Hollomon (Z, ln Z)	60.7 ± 9.1	83.3 ± 3.5
2	Strain rate ($\dot{\varepsilon}$, ln $\dot{\varepsilon}$, $\dot{\varepsilon}^2$)	5.3 ± 1.6	7.2 ± 1.9
3	Temperature (T, ln T, T^2, exp(Q/RT))	3.7 ± 1.3	5.3 ± 2.0
4	Interactions ($T \times \dot{\varepsilon}$, $T \times \varepsilon$, $\dot{\varepsilon} \times \varepsilon$, $T \times \dot{\varepsilon} \times \varepsilon$)	2.4 ± 0.9	3.4 ± 1.2
5	Strain (ε, ε^2)	0.6 ± 0.4	0.8 ± 0.6

The Zener–Hollomon group accounts for 83 ± 3% of the total importance: permuting Z and ln Z together raises the test RMSE by about 61 MPa, compared with 5.3 MPa for the strain-rate group and 3.7 MPa for the temperature group. The strain terms contribute least (about 1%), consistent with the weak strain dependence of the data (the mean flow stress falls by only 11% from ε = 0.1 to 0.5). When the features are permuted one at a time, the credit for Z is split between Z (47 ± 8 MPa) and ln Z (12 ± 5 MPa), which shows why the grouped analysis is needed. The result is consistent with the Sellars–McTegart description [20], in which Z controls thermally activated flow. It should not be over-interpreted, however: Z was supplied as an input, and heavy use of it shows that Z is an efficient summary of T and $\dot{\varepsilon}$ for this dataset, not that the model has learned a deformation mechanism.

4.4 Partial Dependence Analysis

Partial dependence computed for the primary variables (Section 3.7) reproduces the monotonic trends of the data. For the XGBoost model trained on all 100 points, the average predicted flow stress decreases from 138.0 MPa at 300 °C to 28.7 MPa at 500 °C, increases from 43.1 MPa at 0.001 s^{-1} to 131.2 MPa at 1.0 s^{-1}, and decreases slightly from 86.3 MPa at ε = 0.1 to 76.9 MPa at ε = 0.5; these values agree with the corresponding averages of Table 2 to within 0.1 MPa, as expected for a model evaluated on its own

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

training grid. The more demanding test is the behavior between the grid points. On a 41 × 61 grid of temperature and ln $\dot{\varepsilon}$ at ε = 0.5, the XGBoost prediction increased with temperature at 12.6% of adjacent grid steps and decreased with strain rate at 22.7% of them, because of its piecewise-constant form, whereas the linear model and the GP surrogate violated these monotonicity conditions at no more than 2.6% of the steps for ε = 0.1–0.5. Monotonic trends at the data points therefore do not guarantee a physically consistent surface between them.

4.5 Prediction Accuracy and Residual Analysis

Fig. 4 compares the XGBoost predictions with the measured values for one representative split (seed 0), with the training, validation and test points shown separately; the 15 test predictions are listed in Table A1 (Appendix A). For this split the test set gives R^2 = 0.998, RMSE = 2.31 MPa and MAPE = 7.2%. The MAPE is dominated by a single point: at 500 °C, 0.001 s^{-1} and ε = 0.1, the measured stress is 5.54 MPa and the prediction 9.70 MPa, an error of 75%, while the other 14 test points have errors between −5.1% and +6.0% (MAPE 2.3%). R^2, RMSE and MAPE must therefore be read together.

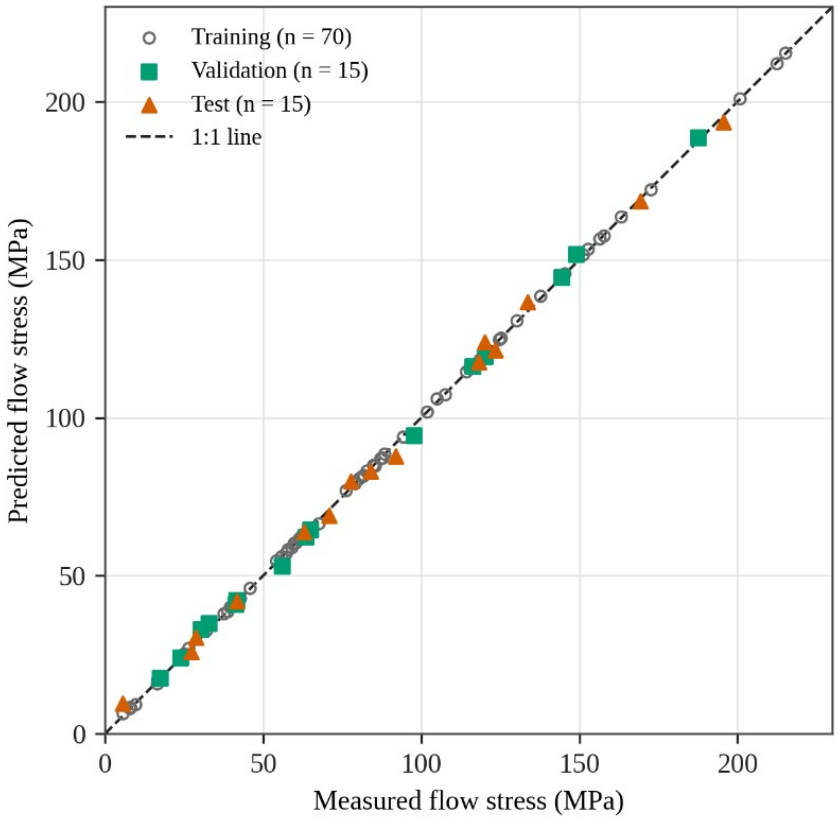


Figure 4. XGBoost predictions versus measured flow stress for one split (seed 0): training (n = 70), validation (n = 15) and test (n = 15) points with the 1:1 line; test set R2 = 0.998, RMSE = 2.31 MPa, MAPE = 7.2%

Fifteen test residuals are too few to assess the error distribution, so Fig. 5 uses the out-of-fold predictions for all 100 points from ten repetitions of 5-fold cross-validation (1000 predictions per model). The residuals are not homoscedastic in relative terms. For XGBoost, the SD of the absolute residual is similar over the whole range (2.0–3.4 MPa in the stress bands 0–25, 25–50, 50–100 and 100–250 MPa), so the relative error grows as the stress falls: the MAPE is 1.5% above 100 MPa, 2.5% for 50–100 MPa, 4.5% for 25–50 MPa and 19.3% below 25 MPa. By temperature, the out-of-fold MAPE is 1.6–

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

2.8% at 300–450 °C and 14.1% at 500 °C. Linear regression shows the same pattern with larger errors (residual SD 5.2–6.2 MPa; MAPE 3.5% above 100 MPa and 80% below 25 MPa).

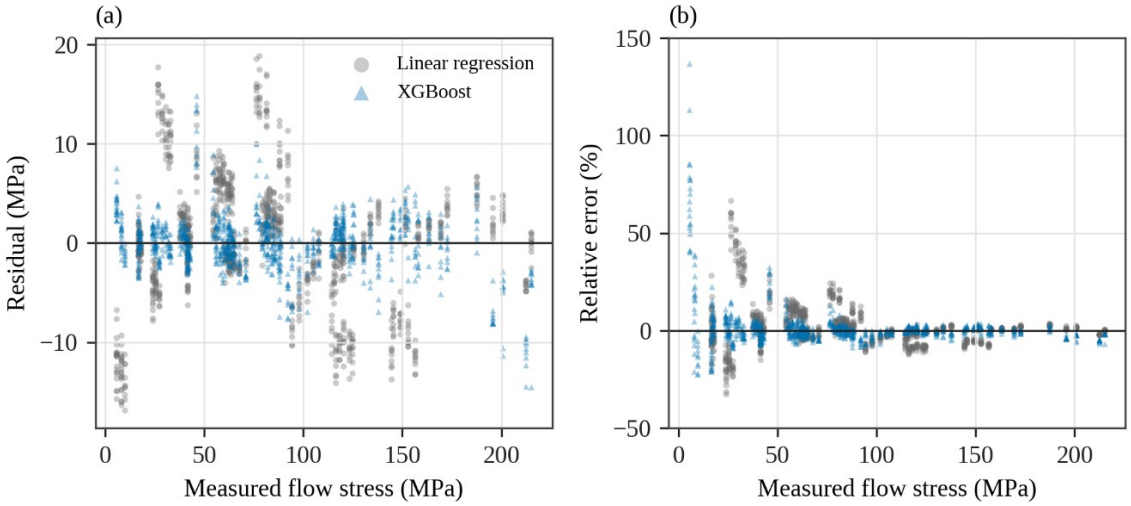


Figure 5. Out-of-fold residuals of XGBoost and linear regression for all 100 points (ten repetitions of 5-fold cross-validation): (a) absolute residual and (b) relative error versus measured flow stress

The pooled XGBoost residuals (Fig. 6) have a mean of −0.05 MPa and an SD of 2.86 MPa, but their distribution is more peaked and heavier-tailed than a normal distribution (excess kurtosis 5.4), so normality should not be assumed when prediction intervals are constructed.

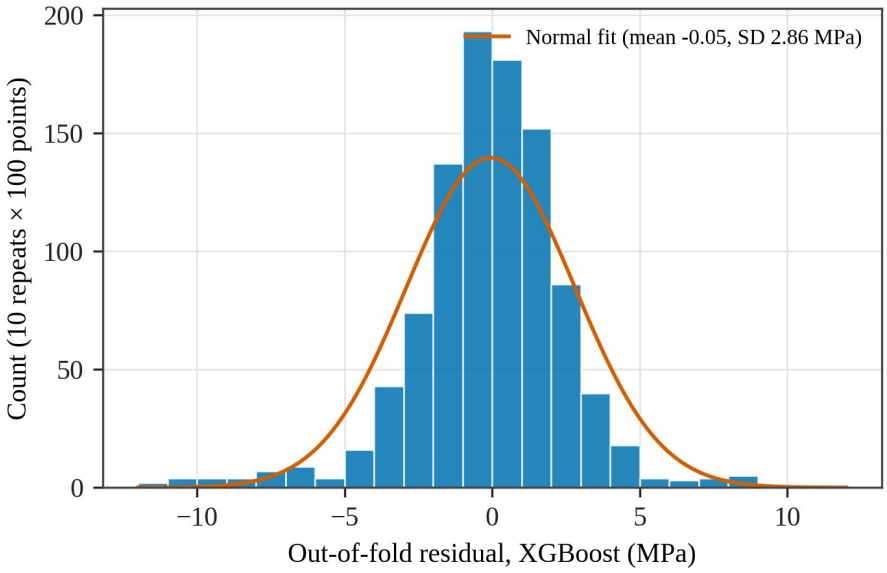


Figure 6. Histogram of the pooled out-of-fold XGBoost residuals (1000 predictions), with a normal distribution of the same mean and SD for reference

The models are therefore least reliable where the flow stress is lowest, i.e. at 500 °C and low strain rates. This region also shows the largest scatter in the data themselves: at 500 °C and 0.001 s^{-1}, the flow stress varies non-monotonically between 5.5 and 9.7 MPa with strain.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

4.6 Physical Validation

The measured and predicted flow stresses follow the expected trends of hot deformation. Fig. 7 shows the temperature dependence at $\varepsilon = 0.5$: at every strain rate the flow stress decreases monotonically with increasing temperature.

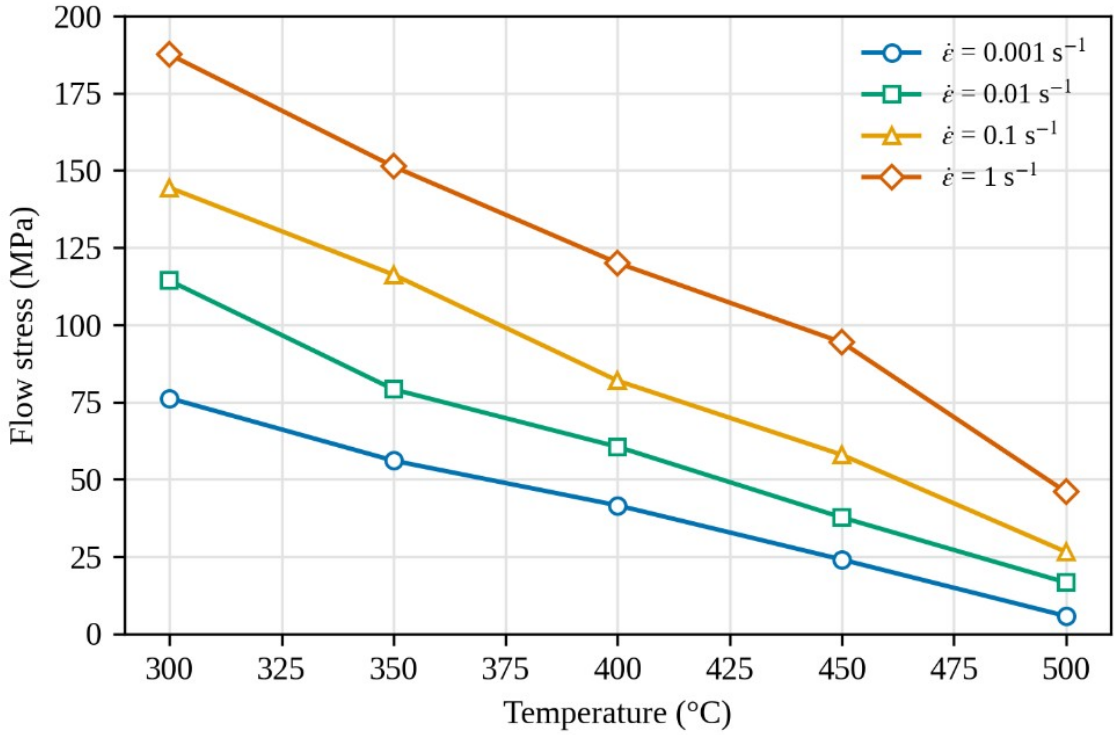


Figure 7. Flow stress versus temperature at ε = 0.5 for the four strain rates (data of Table 2)

At 0.01 s^{-1} and $\varepsilon = 0.5$, the flow stress decreases from 114.3 MPa at 300 °C to 16.6 MPa at 500 °C, a reduction of 85%, reflecting the increasing contribution of thermally activated dislocation climb, dynamic recovery and, possibly, dynamic recrystallization at higher temperatures.

The flow stress increases with strain rate at every temperature (Fig. 8), confirming the rate-sensitive behavior characteristic of hot deformation.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

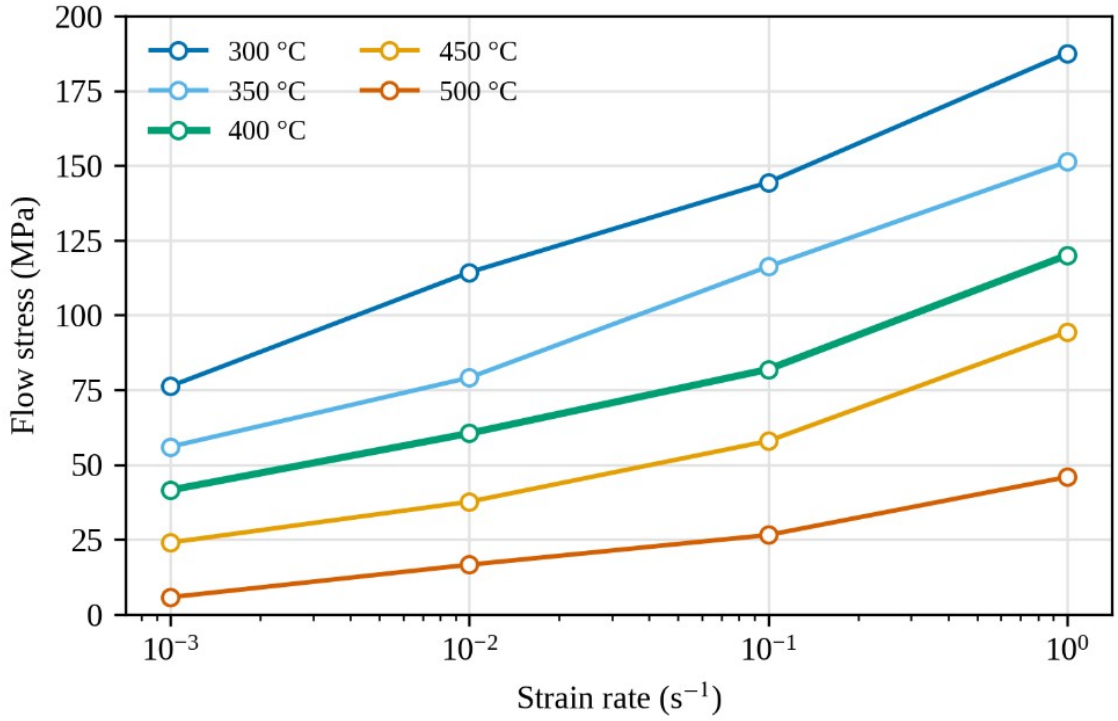


Figure 8. Flow stress versus strain rate (logarithmic scale) at ε = 0.5 for the five test temperatures (data of Table 2)

At 400 °C and $\varepsilon = 0.5$, the flow stress increases from 41.6 MPa at 0.001 s^{-1} to 120.0 MPa at 1.0 s^{-1}. The local strain-rate sensitivity, estimated from quadratic fits of ln σ against ln $\dot{\varepsilon}$, is 0.08–0.21 at 300–450 °C and reaches about 0.46 at 500 °C and 0.001 s^{-1}, where recovery processes have the most time to operate.

Fig. 9 plots the flow stress against ln Z. The data collapse onto a single band, and the hyperbolic-sine equation (Eq. 1) with the constants of Section 3.1 describes the trend, although points with similar values of Z differ by 20–30 MPa because of the strain dependence and because a single value of Q does not remove all of the temperature dependence.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

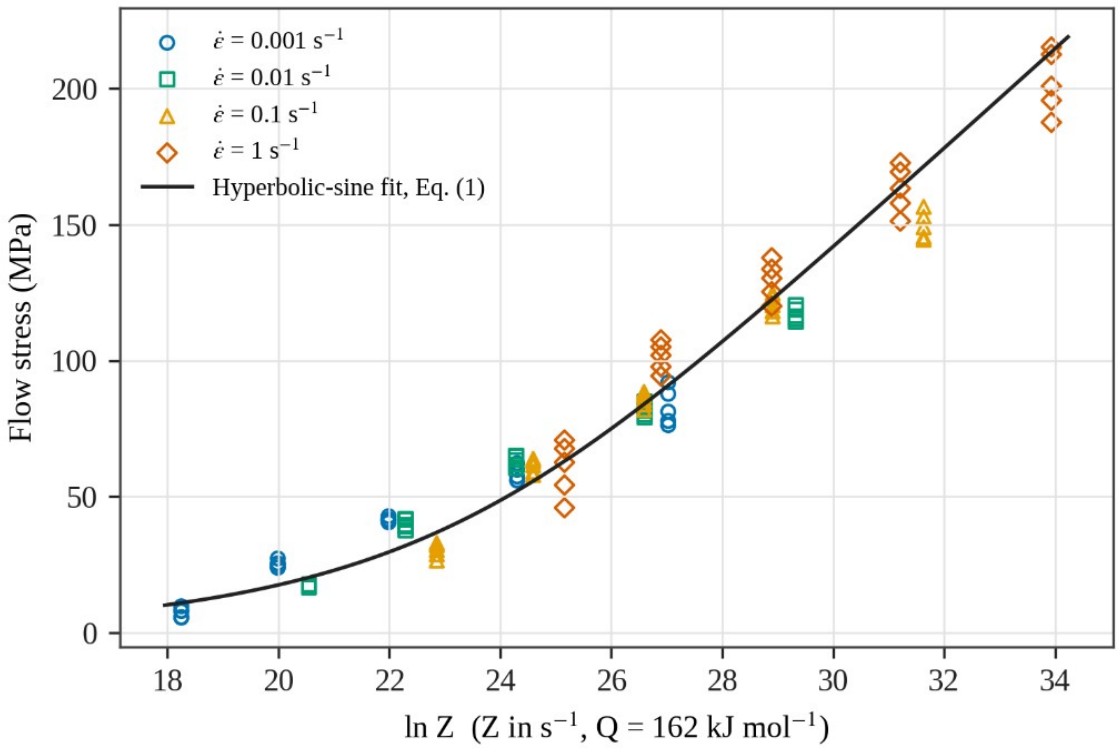


Figure 9. Flow stress versus ln Z (Q = 162 kJ mol−1) for all 100 data points, with the fitted hyperbolic-sine equation (Eq. 1; α = 0.0148 MPa−1, slope 3.66, ln A = 24.88)

The ML models reproduce this relationship because Z is one of their inputs; the relevant question is whether they improve on it. Relative to the hyperbolic-sine equation fitted to the same data (MAPE = 14.3% with strain-dependent constants and 15.0% with a single set), the boosted models reduce the interpolation error to about 5% (Table 5), presumably by capturing the strain dependence and the departure from a single-Q description at 500 °C.

4.7 Comparison with Literature

Table 8 lists errors reported for flow-stress models in the literature together with the present results. The comparison is given for context only and is not a ranking: the materials, dataset sizes, validation protocols and error definitions (MAPE, average absolute relative error or mean deviation) differ between the studies, and the present values refer to interpolation within a complete test grid.

Table 8. Reported errors of flow-stress models (context only; the error definitions, materials, dataset sizes and validation protocols differ and the values are not directly comparable)

Reference	Material	Method	Error measure and data	Error (%)
[37]	2124 Al/SiC_p	Regression	Mean deviation, as reported	13.5
[37]	2124 Al/SiC_p	Fuzzy logic	Mean deviation, as reported	6.1
Mandal et al. [22]	AISI 304L steel	ANN	As reported	6.0
Quan et al. [29]	Al–Zn–Mg–Cu alloy	ANN	As reported	5.2
Ji et al. [30]	Alloy 617	ANN	As reported	7.1
Li et al. [36]	Al–Zn–Mg–Cu alloy	XGBoost	As reported	3.9

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

Reference	Material	Method	Error measure and data	Error (%)
Present study	2124 Al/SiC_p	Hyperbolic-sine equation	MAPE, fit to all 100 points	14.3
Present study	2124 Al/SiC_p	Linear regression	Test MAPE, 20 splits (n = 15 each)	17.0 ± 6.9
Present study	2124 Al/SiC_p	Gradient boosting	Test MAPE, 20 splits (n = 15 each)	5.2 ± 2.8
Present study	2124 Al/SiC_p	XGBoost	Test MAPE, 20 splits (n = 15 each)	5.8 ± 3.5
Present study	2124 Al/SiC_p	Enhanced ensemble	Test MAPE, 20 splits (n = 15 each)	5.3 ± 2.5
Present study	2124 Al/SiC_p	XGBoost	MAPE, leave-one-temperature-out (mean)	32.3

Against the regression (13.5%) and fuzzy-logic (6.1%) errors reported for the same composite [37], the mean test MAPE of gradient boosting and XGBoost (5.2–5.8%) is about 60% lower than the regression value and similar to the fuzzy-logic value. Against the linear model and the hyperbolic-sine equation fitted in this work (17.0% and 14.3%), the reduction is about two-thirds. Which comparison is appropriate depends on the error definition and validation protocol of [37], which are not identical to those used here. No improvement is claimed over the ANN and XGBoost studies of other alloys [22], [29], [30], [36], because those results were obtained on different materials and datasets.

4.8 Discussion

4.8.1 Accuracy of Machine Learning Relative to Constitutive Equations

Within the tested envelope, the non-linear ML models reduce the interpolation error of the hyperbolic-sine description by about two-thirds (from 14.3% to about 5% MAPE). Three factors contribute. First, the tree ensembles and the neural network represent the strain dependence and the curvature of $\ln \sigma$ against $1/T$ without an assumed functional form. Second, the Zener–Hollomon features give every model, including the linear one, a physically appropriate coordinate: most of the variance is explained once Z is available (test R^2 = 0.981 for linear regression). Third, stacking adds little: the enhanced ensemble and gradient boosting differ by 0.2 percentage points of mean MAPE, well within the split-to-split scatter. Outside the tested envelope, none of the models is reliable (Table 6).

4.8.2 Physical Consistency and Interpretability

The grouped permutation analysis attributes 83 ± 3% of the XGBoost importance to the Zener–Hollomon features, in line with the Sellars–McTegart framework [20]. Physical consistency was also checked quantitatively (Sections 4.4 and 4.6): partial dependence reproduces the monotonic trends of the data, but the tree-based surface is not monotonic between the grid points, whereas the linear model and the GP surrogate are nearly so. Interpretability analyses of this kind show which inputs a model relies on; they do not by themselves show that the model will extrapolate, and the LOTO and LOSRO results (Table 6) show that it does not.

4.8.3 Computational Efficiency and Practical Implementation

All trained models predict a single point in less than 1 ms on a standard computer, which is fast enough for use in process simulation or control. The models could be coupled with finite-element simulations through a user material subroutine, but for this purpose a smooth surrogate (the GP or the linear model

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

in the engineered feature space) is preferable to a tree ensemble, because implicit finite-element solvers require continuous derivatives of the flow stress.

The trained XGBoost model occupies about 0.3 MB in its native format, and all models were implemented with standard open-source Python libraries (scikit-learn, XGBoost, LightGBM and CatBoost), which supports reproducibility.

4.8.4 Implications for Manufacturing

Within the tested envelope, a flow-stress model with an interpolation error of about 5% is adequate for estimating forming loads and press capacity and for comparing candidate temperature–strain-rate windows. At temperatures or strain rates that have not been tested, the errors in Table 6 apply, and additional tests are needed before a prediction is used for die or equipment design. Real-time adaptive control would also require prediction intervals, which the GP surrogate provides but the tree ensembles do not.

4.9 Processing Map Development

Processing maps were constructed with the dynamic materials model (DMM) [45], using the GP surrogate of Section 3.8 because the tree ensembles do not provide a usable strain-rate derivative. The efficiency of power dissipation η, the strain-rate sensitivity m and the instability parameter ξ were evaluated as follows.

Power dissipation efficiency:

$$\eta = \frac{2m}{m+1} \tag{15}$$

where m represents the strain rate sensitivity parameter calculated from the GP posterior mean:

$$m = \left.\frac{\partial \ln \sigma}{\partial \ln \dot{\varepsilon}}\right|_{T,\varepsilon} \tag{16}$$

Instability parameter:

$$\xi = \frac{\partial}{\partial \ln \dot{\varepsilon}} \ln\left[\frac{m}{m+1}\right] + m \tag{17}$$

Regions with $\xi < 0$ are predicted to be prone to flow instability, and regions with $\xi > 0$ and high η are candidate processing windows. The parameters were computed on a 41 × 61 grid spanning 300–500 °C and 0.001–1.0 s^{-1} at ε = 0.1, 0.3 and 0.5, with the derivatives in Eqs. (16) and (17) evaluated by central differences in ln $\dot{\varepsilon}$. The uncertainty was obtained by repeating the calculation for 500 posterior samples of the GP surface.

Fig. 10a shows the map at ε = 0.5, with the experimental conditions marked so that the extent of interpolation is visible, and Fig. 10b shows the posterior SD of η together with the contours at which the posterior probability of $\xi < 0$ reaches 0.5 and 0.9. The most robust feature is a domain of high efficiency at 475–500 °C and about 0.002–0.01 s^{-1}, where the posterior-mean η is 53–68%, with a posterior SD of 3–6 percentage points at 500 °C. At 450 °C, η is about 30–39% over most of the strain-rate range. At 400–450 °C and 0.01–0.1 s^{-1}, η is 22–34%, clearly lower than in the high-temperature domain.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

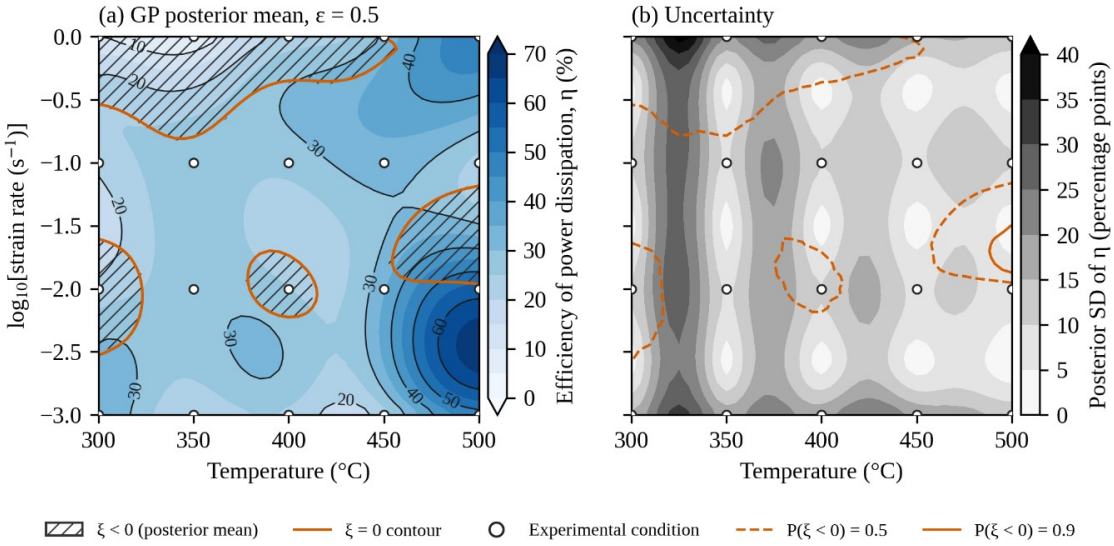


Figure 10. Processing map of 2124 Al/10 vol.% SiCp at ε = 0.5 derived from the Gaussian-process surrogate: (a) contours of η (%), region with ξ < 0 at the posterior mean (hatched) and experimental conditions (open circles); (b) posterior SD of η with the contours P(ξ < 0) = 0.5 and 0.9

The posterior-mean map shows $\xi < 0$ at strain rates above about 0.3 s^{-1} at 300–450 °C, and in several smaller regions (Fig. 10a). However, the posterior probability of $\xi < 0$ exceeds 0.9 over less than 1% of the map and exceeds 0.5 over about 22% of it (Fig. 10b), so the instability boundary is poorly constrained. This is expected: ξ depends on the derivative of m with respect to ln $\dot{\varepsilon}$, i.e. on the second derivative of ln σ, and four strain rates spanning three decades carry little information about that curvature. The posterior SD of η is also large (about 15–40 percentage points) midway between the tested temperatures, where the GP interpolates. The map should therefore be read as identifying a probable high-efficiency domain at the highest temperature and lowest strain rates and as indicating possible instability at high strain rates and lower temperatures. Confirming either requires metallographic examination of specimens deformed under these conditions, which was outside the scope of this data-based study. The high efficiency at 500 °C and 0.001–0.01 s^{-1} is consistent with dynamic recrystallization or with grain-boundary sliding, but compression data alone cannot distinguish between these mechanisms.

Coupling a smooth probabilistic surrogate with the DMM makes the uncertainty of the processing map explicit. With the present data this uncertainty is large; additional strain rates (for example, 0.003, 0.03 and 0.3 s^{-1}) and intermediate temperatures would reduce it most efficiently.

4.10 Limitations

The study has the following limitations. (i) The dataset is small (100 points) and was compiled from a single published source; it covers one composite (2124 Al with 10 vol.% SiC_p) produced by one route, so the results cannot be transferred to other matrices, reinforcement fractions or particle sizes without new data. (ii) The accuracy figures in Table 5 refer to interpolation within a complete grid; the LOTO and LOSRO tests (Table 6) show that the errors are several times larger for withheld conditions and very large outside the tested envelope. (iii) The deformation-heating correction applied to the data uses $\beta = 0.9$ at all strain rates and is an upper bound at low strain rates (Table 3); the uncorrected data were not re-analyzed. (iv) Tree-based models do not provide usable derivatives, so the processing map relies on the GP surrogate, whose instability predictions are uncertain with only four strain rates. (v) No

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

metallographic validation of the processing map was performed. (vi) The tree ensembles provide point predictions only; prediction intervals are available only from the GP surrogate.

5. Conclusions

Eight machine learning algorithms were benchmarked for predicting the flow stress of a powder-metallurgy 2124 Al/10 vol.% SiC_p composite from 100 published hot-compression data points (300–500 °C, 0.001–1.0 s^{-1}, ε = 0.1–0.5), using Zener–Hollomon-based features (Q = 162 kJ mol^{-1} from a hyperbolic-sine fit), 20 repeated random splits, and leave-one-temperature-out and leave-one-strain-rate-out tests.

For interpolation within the test grid, gradient boosting, the stacked ensemble and XGBoost gave mean test MAPE values of 5.2–5.8% ($R^2 \approx 0.997$), about one-third of the error of the hyperbolic-sine equation fitted to the same data (14.3%) and of linear regression on the engineered features (17.0%). The differences among the boosted models were smaller than the split-to-split scatter, so no single algorithm can be declared best, and the errors were concentrated at the low stresses measured at 500 °C.

Grouped permutation importance attributed 83 ± 3% of the predictive importance to the Zener–Hollomon features. Partial dependence reproduced the monotonic trends of the data, but the tree-based response surfaces were not monotonic between the grid points.

Because boosted trees are piecewise constant, the processing maps were derived from a Gaussian-process surrogate. The map identifies a high-efficiency domain ($\eta \approx$ 53–68%) at 475–500 °C and about 0.002–0.01 s^{-1}; possible instability at high strain rates and lower temperatures is indicated but is not well constrained by four strain rates.

For the non-linear models, the errors rose to mean MAPE values of 25–33% when a whole temperature level was withheld and 15–27% when a whole strain-rate level was withheld, which is the realistic accuracy for untested conditions. The results apply to one composite and one dataset; data at additional strain rates and intermediate temperatures, together with metallographic validation, are needed before the models or maps are used for process design.

Appendix A. Test-Set Predictions for a Representative Split

Table A1: Measured and predicted flow stresses for the 15 test points of split seed 0 (Fig. 4)

T (°C)	$\dot{\varepsilon}$ (s^{-1})	ε	Measured (MPa)	XGBoost (MPa)	Error (%)	Enhanced ensemble (MPa)	Error (%)
300	0.001	0.1	92.00	87.90	-4.5	88.29	-4.0
300	0.001	0.4	77.82	79.94	+2.7	79.73	+2.5
300	1	0.4	195.62	193.57	-1.0	194.50	-0.6
350	0.1	0.2	123.35	121.42	-1.6	122.04	-1.1
350	0.1	0.4	118.10	117.62	-0.4	118.06	-0.0
350	1	0.2	169.32	168.56	-0.4	169.05	-0.2
400	0.1	0.4	83.92	83.01	-1.1	83.69	-0.3
400	1	0.2	133.62	136.65	+2.3	136.78	+2.4

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

T (°C)	$\dot{\varepsilon}$ (s^{-1})	ε	Measured (MPa)	XGBoost (MPa)	Error (%)	Enhanced ensemble (MPa)	Error (%)
400	1	0.5	120.01	123.80	+3.2	124.16	+3.5
450	0.001	0.2	27.28	25.90	-5.1	25.36	-7.1
450	0.01	0.2	41.74	42.00	+0.6	40.75	-2.4
450	0.1	0.2	63.03	63.85	+1.3	62.72	-0.5
500	0.001	0.1	5.54	9.70	+75.1	9.00	+62.5
500	0.1	0.4	28.75	30.47	+6.0	29.37	+2.1
500	1	0.1	70.78	69.12	-2.3	70.51	-0.4

References

[1] T. Pan, C. Song, Z. Gao, T. Xia, and T. Wang, “The prediction of flow stress in the hot compression of a Ni-Cr-Mo steel using machine learning algorithms,” Processes, vol. 12, p. 441, 2024, doi: 10.3390/pr12030441.

[2] Y. C. Lin, M.-S. Chen, and J. Zhong, “A critical review of experimental results and constitutive descriptions for metals and alloys in hot working,” Mater. Des., vol. 32, no. 4, pp. 1733–1759, 2011, doi: 10.1016/j.matdes.2010.11.048.

[3] H. J. McQueen and N. D. Ryan, “Constitutive analysis in hot working,” Mater. Sci. Eng. A, vol. 322, pp. 43–63, 2002, doi: 10.1016/S0921-5093(01)01117-0.

[4] S. V. S. N. Murty, B. N. Rao, and B. P. Kashyap, “Instability criteria for hot deformation of materials,” Int. Mater. Rev., vol. 45, no. 1, pp. 15–26, 2000, doi: 10.1179/095066000771048782.

[5] M. A. Ali, N. A. Mufti, M. Sana, M. Tlija, C. I. Pruncu, and L. Lamberti, “Comparison of machine learning algorithms for dynamic performance assessment in complex shapes manufacturing of hybrid particle-reinforced composite,” Expert Syst. Appl., vol. 275, art. 127022, 2025, doi: 10.1016/j.eswa.2025.127022.

[6] P. Pragathi and R. Elansezhian, “Mechanical and microstructure behaviour of aluminum nanocomposite fabricated by squeeze casting and ultrasonic aided squeeze casting: A comparative study,” J. Alloys Compd., vol. 956, art. 170203, 2023, doi: 10.1016/j.jallcom.2023.170203.

[7] F. Zhang, Z. He, K. Lu, Z. Zhan, Z. Li, and X. Wang, “Interfacial microstructure and mechanical properties of 2124 aluminum alloy reinforced by AlCoCrFeNi high entropy alloy,” J. Mater. Res. Technol., vol. 26, pp. 8846–8856, 2023, doi: 10.1016/j.jmrt.2023.09.166.

[8] S. Cai et al., “Hot deformation behavior and microstructural evolution of SiCp/Al-4.2Cu-4.1Mg composites,” J. Mater. Res. Technol., 2025, doi: 10.1016/j.jmrt.2025.11.245.

[9] S. Salifu, P. A. Olubambi, and L. Teffo, “Phase stability and microstructural properties of high entropy alloy reinforced aluminium matrix composites consolidated via spark plasma sintering,” Heliyon, vol. 10, no. 2, art. e24123, 2024, doi: 10.1016/j.heliyon.2024.e24123.

[10] Z. Seikh, M. Sekh, G. Mandal, B. Sengupta, and A. Sinha, “Metal matrix composites processed through powder metallurgy: A brief overview,” J. Inst. Eng. India Ser. D, pp. 1–4, 2024, doi: 10.1007/s40033-024-00789-4.

[11] K. Pooja, N. Tarannum, and P. Chaudhary, “Metal matrix composites: Revolutionary materials for shaping the future,” Discov. Mater., vol. 5, p. 35, 2025, doi: 10.1007/s43939-025-00226-6.

[12] A. Berghezan, “Composite materials,” in Proc. Int. Conf. Composite Materials, 1961, pp. 1–15.

[13] D. B. Miracle, “Metal matrix composites—from science to technological significance,” Compos. Sci. Technol., vol. 65, no. 15–16, pp. 2526–2540, 2005, doi: 10.1016/j.compscitech.2005.05.027.

[14] S. P. Rawal, “Metal-matrix composites for space applications,” JOM, vol. 53, no. 4, pp. 14–17, 2001, doi: 10.1007/s11837-001-0139-z.

[15] B. V. R. Bhat, Y. R. Mahajan, and Y. V. R. K. Prasad, “Effect of volume fraction of SiCp reinforcement on the processing maps for 2124 Al matrix composites,” Metall. Mater. Trans. A, vol. 31, pp. 629–639, 2000, doi: 10.1007/s11661-000-0006-5.

[16] A. A. Lakshmi, C. S. Rao, J. Gangadhar, C. Srinivasu, and S. K. Singh, “Review of processing maps and development of qualitative processing maps,” Mater. Today Proc., vol. 4, no. 2, pp. 946–956, 2017, doi: 10.1016/j.matpr.2017.01.106.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

[17] T. Seshacharyulu, S. C. Medeiros, W. G. Frazier, and Y. V. R. K. Prasad, "Hot working of commercial Ti–6Al–4V with an equiaxed α–β microstructure: Materials modeling considerations," Mater. Sci. Eng. A, vol. 284, no. 1–2, pp. 184–194, 2000, doi: 10.1016/S0921-5093(00)00741-3.

[18] M. R. Rokni, A. Zarei-Hanzaki, A. A. Roostaei, and H. R. Abedi, "An investigation into the hot deformation characteristics of 7075 aluminum alloy," Mater. Des., vol. 32, no. 4, pp. 2339–2344, 2011, doi: 10.1016/j.matdes.2010.12.047.

[19] N. Jin, H. Zhang, Y. Han, W. Wu, and J. Chen, "Hot deformation behavior of 7150 aluminum alloy during compression at elevated temperature," Mater. Charact., vol. 60, no. 6, pp. 530–536, 2009, doi: 10.1016/j.matchar.2008.12.012.

[20] C. M. Sellars and W. J. McTegart, "On the mechanism of hot deformation," Acta Metall., vol. 14, no. 9, pp. 1136–1138, 1966, doi: 10.1016/0001-6160(66)90207-0.

[21] N. Haghdadi, A. Zarei-Hanzaki, A. R. Khalesian, and H. R. Abedi, "Artificial neural network modeling to predict the hot deformation behavior of an A356 aluminum alloy," Mater. Des., vol. 49, pp. 386–391, 2013, doi: 10.1016/j.matdes.2012.12.082.

[22] S. Mandal, P. V. Sivaprasad, S. Venugopal, and K. P. N. Murthy, "Artificial neural network modeling to evaluate and predict the deformation behavior of stainless steel type AISI 304L during hot torsion," Appl. Soft Comput., vol. 9, no. 1, pp. 237–244, 2009, doi: 10.1016/j.asoc.2008.03.016.

[23] Y. C. Lin, D.-X. Wen, J. Deng, G. Liu, and J. Chen, "A unified physically based constitutive model for describing strain hardening effect and dynamic recovery behavior of a Ni-based superalloy," J. Mater. Res., vol. 28, no. 14, pp. 1884–1895, 2013, doi: 10.1557/jmr.2013.17.

[24] Q. Wen, Z. Cao, S. Yang et al., "Research on predicting flow stress of 7075 aluminum alloy using machine learning models," Front. Mater., vol. 12, art. 1671753, 2025, doi: 10.3389/fmats.2025.1671753.

[25] H. R. R. Ashtiani, M. H. Parsa, and H. Bisadi, "Constitutive equations for elevated temperature flow behavior of commercial purity aluminum," Mater. Sci. Eng. A, vol. 545, pp. 61–67, 2012, doi: 10.1016/j.msea.2012.02.090.

[26] H. N. Han, J. K. Lee, H. J. Kim, and Y. S. Jin, "A model for deformation, temperature and phase transformation behavior of steels on run-out table in hot strip mill," J. Mater. Process. Technol., vol. 128, pp. 216–225, 2002, doi: 10.1016/S0924-0136(02)00454-5.

[27] P. S. Robi and U. S. Dixit, "Application of neural networks in generating processing map for hot working," J. Mater. Process. Technol., vol. 142, no. 1, pp. 289–294, 2003, doi: 10.1016/S0924-0136(03)00579-X.

[28] O. Sabokpa, A. Zarei-Hanzaki, H. R. Abedi, and N. Haghdadi, "Artificial neural network modeling to predict the high temperature flow behavior of an AZ81 magnesium alloy," Mater. Des., vol. 39, pp. 390–396, 2012, doi: 10.1016/j.matdes.2012.03.002.

[29] G.-Z. Quan, W.-Q. Lv, Y.-P. Mao, Y.-W. Zhang, and J. Zhou, "Prediction of flow stress in a wide strain rate range involving macro-micro mechanism for Al-Zn-Mg-Cu aluminum alloy," Mater. Des., vol. 79, pp. 94–114, 2015, doi: 10.1016/j.matdes.2015.04.032.

[30] G. Ji, F. Li, Q. Li, H. Li, and Z. Li, "A comparative study on Arrhenius-type constitutive model and artificial neural network model to predict high-temperature deformation behaviour in Alloy 617," Mater. Sci. Eng. A, vol. 528, no. 13–14, pp. 4774–4782, 2011, doi: 10.1016/j.msea.2011.03.017.

[31] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[32] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Stat., vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.

[33] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[34] Y. Liu, T. Zhao, W. Ju, and S. Shi, "Materials discovery and design using machine learning," J. Materiomics, vol. 3, no. 3, pp. 159–177, 2017, doi: 10.1016/j.jmat.2017.08.002.

[35] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, "A general-purpose machine learning framework for predicting properties of inorganic materials," npj Comput. Mater., vol. 2, no. 1, pp. 1–7, 2016, doi: 10.1038/npjcompumats.2016.28.

[36] H.-Y. Li, X.-F. Wang, J.-Y. Wei, J.-T. Yan, and D.-D. Li, "Prediction of hot deformation behaviour of Al-Zn-Mg-Cu aluminum alloy using XGBoost machine learning method," Trans. Nonferrous Met. Soc. China, vol. 31, no. 4, pp. 1036–1051, 2021, doi: 10.1016/S1003-6326(21)65568-6.

[37] [Authors to insert: full reference of the published study that reports the hot-compression flow-stress data of Table 2 and the regression (13.5%) and fuzzy-logic (6.1%) errors for 2124 Al/10 vol.% SiCp.]

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.

[38] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 4765–4774.

[39] D. W. Apley and J. Zhu, "Visualizing the effects of predictor variables in black box supervised learning models," J. R. Stat. Soc. Ser. B Stat. Methodol., vol. 82, no. 4, pp. 1059–1086, 2020, doi: 10.1111/rssb.12377.

[40] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," BMC Bioinformatics, vol. 8, art. 25, 2007, doi: 10.1186/1471-2105-8-25.

[41] The Aluminum Association, International Alloy Designations and Chemical Composition Limits for Wrought Aluminum and Wrought Aluminum Alloys. Arlington, VA, USA: The Aluminum Association, 2018.

[42] R. L. Goetz and S. L. Semiatin, "The adiabatic correction factor for deformation heating during the uniaxial compression test," J. Mater. Eng. Perform., vol. 10, no. 6, pp. 710–717, 2001, doi: 10.1361/105994901770344593.

[43] D. H. Wolpert, "Stacked generalization," Neural Netw., vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.

[44] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning. Cambridge, MA, USA: MIT Press, 2006.

[45] Y. V. R. K. Prasad, H. L. Gegel, S. M. Doraivelu, J. C. Malas, J. T. Morgan, K. A. Lark, and D. R. Barker, "Modeling of dynamic material behavior in hot deformation: Forging of Ti-6242," Metall. Trans. A, vol. 15, no. 10, pp. 1883–1892, 1984, doi: 10.1007/BF02664902.

© 2026 The Author(s). Published by Qassim University under CC BY-NC-SA 4.0.