

ENGINEERING JOURNAL IJOER

VOLUME-11, ISSUE-11 NOVEMBER 2025



DOWNLOAD NOW



Preface

We would like to present, with great pleasure, the inaugural volume-11, Issue-11, November 2025, of a scholarly journal, *International Journal of Engineering Research & Science*. This journal is part of the AD Publications series *in the field of Engineering, Mathematics, Physics, Chemistry and science Research Development*, and is devoted to the gamut of Engineering and Science issues, from theoretical aspects to application-dependent studies and the validation of emerging technologies.

This journal was envisioned and founded to represent the growing needs of Engineering and Science as an emerging and increasingly vital field, now widely recognized as an integral part of scientific and technical investigations. Its mission is to become a voice of the Engineering and Science community, addressing researchers and practitioners in below areas:

Chemical Engineering	
Biomolecular Engineering	Materials Engineering
Molecular Engineering	Process Engineering
Corrosion Engineering	
Civil Engineering	
Environmental Engineering	Geotechnical Engineering
Structural Engineering	Mining Engineering
Transport Engineering	Water resources Engineering
Electrical Engineering	
Power System Engineering	Optical Engineering
Mechanical Engineering	
Acoustical Engineering	Manufacturing Engineering
Optomechanical Engineering	Thermal Engineering
Power plant Engineering	Energy Engineering
Sports Engineering	Vehicle Engineering
Software Engineering	
Computer-aided Engineering	Cryptographic Engineering
Teletraffic Engineering	Web Engineering
System Engineering	
Mathematics	
Arithmetic	Algebra
Number theory	Field theory and polynomials
Analysis	Combinatorics
Geometry and topology	Topology
Probability and Statistics	Computational Science
Physical Science	Operational Research
Physics	
Nuclear and particle physics	Atomic, molecular, and optical physics
Condensed matter physics	Astrophysics
Applied Physics	Modern physics
Philosophy	Core theories

Chemistry	
Analytical chemistry	Biochemistry
Inorganic chemistry	Materials chemistry
Neurochemistry	Nuclear chemistry
Organic chemistry	Physical chemistry
Other Engineering Areas	
Aerospace Engineering	Agricultural Engineering
Applied Engineering	Biomedical Engineering
Biological Engineering	Building services Engineering
Energy Engineering	Railway Engineering
Industrial Engineering	Mechatronics Engineering
Management Engineering	Military Engineering
Petroleum Engineering	Nuclear Engineering
Textile Engineering	Nano Engineering
Algorithm and Computational Complexity	Artificial Intelligence
Electronics & Communication Engineering	Image Processing
Information Retrieval	Low Power VLSI Design
Neural Networks	Plastic Engineering

Each article in this issue provides an example of a concrete industrial application or a case study of the presented methodology to amplify the impact of the contribution. We are very thankful to everybody within that community who supported the idea of creating a new Research with IJOER. We are certain that this issue will be followed by many others, reporting new developments in the Engineering and Science field. This issue would not have been possible without the great support of the Reviewer, Editorial Board members and also with our Advisory Board Members, and we would like to express our sincere thanks to all of them. We would also like to express our gratitude to the editorial staff of AD Publications, who supported us at every stage of the project. It is our hope that this fine collection of articles will be a valuable resource for *IJOER* readers and will stimulate further research into the vibrant area of Engineering and Science Research.



Mukesh Arora
(Chief Editor)

Board Members

Mr. Mukesh Arora (Editor-in-Chief)

BE (Electronics & Communication), M.Tech (Digital Communication), currently serving as Assistant Professor in the Department of ECE.

Prof. Dr. Fabricio Moraes de Almeida

Professor of Doctoral and Master of Regional Development and Environment - Federal University of Rondonia.

Dr. Parveen Sharma

Dr Parveen Sharma is working as an Assistant Professor in the School of Mechanical Engineering at Lovely Professional University, Phagwara, Punjab.

Prof. S. Balamurugan

Department of Information Technology, Kalaignar Karunanidhi Institute of Technology, Coimbatore, Tamilnadu, India.

Dr. Omar Abed Elkareem Abu Arqub

Department of Mathematics, Faculty of Science, Al Balqa Applied University, Salt Campus, Salt, Jordan, He received PhD and Msc. in Applied Mathematics, The University of Jordan, Jordan.

Dr. AKPOJARO Jackson

Associate Professor/HOD, Department of Mathematical and Physical Sciences, Samuel Adegboyega University, Ogwa, Edo State.

Dr. Ajoy Chakraborty

Ph.D.(IIT Kharagpur) working as Professor in the department of Electronics & Electrical Communication Engineering in IIT Kharagpur since 1977.

Dr. Ukar W. Soelistijo

Ph D, Mineral and Energy Resource Economics, West Virginia State University, USA, 1984, retired from the post of Senior Researcher, Mineral and Coal Technology R&D Center, Agency for Energy and Mineral Research, Ministry of Energy and Mineral Resources, Indonesia.

Dr. Samy Khalaf Allah Ibrahim

PhD of Irrigation &Hydraulics Engineering, 01/2012 under the title of: "Groundwater Management under Different Development Plans in Farafra Oasis, Western Desert, Egypt".

Dr. Ahmet ÇİFCİ

Ph.D. in Electrical Engineering, Currently Serving as Head of Department, Burdur Mehmet Akif Ersoy University, Faculty of Engineering and Architecture, Department of Electrical Engineering.

Dr. M. Varatha Vijayan

Annauniversity Rank Holder, Commissioned Officer Indian Navy, Ncc Navy Officer (Ex-Serviceman Navy), Best Researcher Awardee, Best Publication Awardee, Tamilnadu Best Innovation & Social Service Awardee From Lions Club.

Dr. Mohamed Abdel Fatah Ashabrawy Moustafa

PhD. in Computer Science - Faculty of Science - Suez Canal University University, 2010, Egypt.

Assistant Professor Computer Science, Prince Sattam bin AbdulAziz University ALkharj, KSA.

Prof.S.Balamurugan

Dr S. Balamurugan is the Head of Research and Development, Quants IS & CS, India. He has authored/co-authored 35 books, 200+ publications in various international journals and conferences and 6 patents to his credit. He was awarded with Three Post-Doctoral Degrees - Doctor of Science (D.Sc.) degree and Two Doctor of Letters (D.Litt) degrees for his significant contribution to research and development in Engineering.

Dr. Mahdi Hosseini

Dr. Mahdi did his Pre-University (12th) in Mathematical Science. Later he received his Bachelor of Engineering with Distinction in Civil Engineering and later he Received both M.Tech. and Ph.D. Degree in Structural Engineering with Grade "A" First Class with Distinction.

Dr. Anil Lamba

Practice Head – Cyber Security, EXL Services Inc., New Jersey USA.

Dr. Anil Lamba is a researcher, an innovator, and an influencer with proven success in spearheading Strategic Information Security Initiatives and Large-scale IT Infrastructure projects across industry verticals. He has helped bring about a profound shift in cybersecurity defense. Throughout his career, he has parlayed his extensive background in security and a deep knowledge to help organizations build and implement strategic cybersecurity solutions. His published researches and conference papers has led to many thought provoking examples for augmenting better security.

Dr. Ali İhsan KAYA

Currently working as Associate Professor in Mehmet Akif Ersoy University, Turkey.

Research Area: Civil Engineering - Building Material - Insulation Materials Applications, Chemistry - Physical Chemistry – Composites.

Dr. Parsa Heydarpour

Ph.D. in Structural Engineering from George Washington University (Jan 2018), GPA=4.00.

Dr. Heba Mahmoud Mohamed Afify

Ph.D degree of philosophy in Biomedical Engineering, Cairo University, Egypt worked as Assistant Professor at MTI University.

Dr. Kalpesh Sunil Kamble (Ph.D., P.Eng., M.Tech, B.E. (Mechanical))

A distinguished academic with a Ph.D. in Mechanical Engineering and 13 Years of extensive teaching and research experience. He is currently a Assistant professor at the SSPM's COE, Kankavli and contributes to several undergraduate and masters programs across Maharashtra, India.

Dr. Aurora Angela Pisano

Ph.D. in Civil Engineering, Currently Serving as Associate Professor of Solid and Structural Mechanics (scientific discipline area nationally denoted as ICAR/08—"Scienza delle Costruzioni"), University Mediterranea of Reggio Calabria, Italy.

Dr. Faizullah Mahar

Associate Professor in Department of Electrical Engineering, Balochistan University Engineering & Technology Khuzdar. He is PhD (Electronic Engineering) from IQRA University, Defense View, Karachi, Pakistan.

Prof. Viviane Barrozo da Silva

Graduated in Physics from the Federal University of Paraná (1997), graduated in Electrical Engineering from the Federal University of Rio Grande do Sul - UFRGS (2008), and master's degree in Physics from the Federal University of Rio Grande do Sul (2001).

Dr. S. Kannadhasan

Ph.D (Smart Antennas), M.E (Communication Systems), M.B.A (Human Resources).

Dr. Christo Ananth

Ph.D. Co-operative Networks, M.E. Applied Electronics, B.E Electronics & Communication Engineering Working as Associate Professor, Lecturer and Faculty Advisor/ Department of Electronics & Communication Engineering in Francis Xavier Engineering College, Tirunelveli.

Dr. S.R.Boselin Prabhu

Ph.D, Wireless Sensor Networks, M.E. Network Engineering, Excellent Professional Achievement Award Winner from Society of Professional Engineers Biography Included in Marquis Who's Who in the World (Academic Year 2015 and 2016). Currently Serving as Assistant Professor in the department of ECE in SVS College of Engineering, Coimbatore.

Dr. Balasubramanyam, N

Dr.Balasubramanyam, N working as Faculty in the Department of Mechanical Engineering at S.V.University College of Engineering Tirupati, Andhra Pradesh.

Dr. PAUL P MATHAI

Dr. Paul P Mathai received his Bachelor's degree in Computer Science and Engineering from University of Madras, India. Then he obtained his Master's degree in Computer and Information Technology from Manonmanium Sundaranar University, India. In 2018, he received his Doctor of Philosophy in Computer Science and Engineering from Noorul Islam Centre for Higher Education, Kanyakumari, India.

Dr. M. Ramesh Kumar

Ph.D (Computer Science and Engineering), M.E (Computer Science and Engineering).

Currently working as Associate Professor in VSB College of Engineering Technical Campus, Coimbatore.

Dr. Maheshwar Shrestha

Postdoctoral Research Fellow in DEPT. OF ELE ENGG & COMP SCI, SDSU, Brookings, SD Ph.D, M.Sc. in Electrical Engineering from SOUTH DAKOTA STATE UNIVERSITY, Brookings, SD.

Dr. D. Amaranatha Reddy

Ph.D. (Postdoctoral Fellow, Pusan National University, South Korea), M.Sc., B.Sc. : Physics.

Dr. Dibya Prakash Rai

Post Doctoral Fellow (PDF), M.Sc., B.Sc., Working as Assistant Professor in Department of Physics in Pachhungga University College, Mizoram, India.

Dr. Pankaj Kumar Pal

Ph.D R/S, ECE Deptt., IIT-Roorkee.

Dr. P. Thangam

PhD in Information & Communication Engineering, ME (CSE), BE (Computer Hardware & Software), currently serving as Associate Professor in the Department of Computer Science and Engineering of Coimbatore Institute of Engineering and Technology.

Dr. Pradeep K. Sharma

PhD., M.Phil, M.Sc, B.Sc, in Physics, MBA in System Management, Presently working as Provost and Associate Professor & Head of Department for Physics in University of Engineering & Management, Jaipur.

Dr. R. Devi Priya

Ph.D (CSE), Anna University Chennai in 2013, M.E, B.E (CSE) from Kongu Engineering College, currently working in the Department of Computer Science and Engineering in Kongu Engineering College, Tamil Nadu, India.

Dr. Sandeep

Post-doctoral fellow, Principal Investigator, Young Scientist Scheme Project (DST-SERB), Department of Physics, Mizoram University, Aizawl Mizoram, India- 796001.

Dr. Roberto Volpe

Faculty of Engineering and Architecture, Università degli Studi di Enna "Kore", Cittadella Universitaria, 94100 – Enna (IT).

Dr. S. Kannadhasan

Ph.D (Smart Antennas), M.E (Communication Systems), M.B.A (Human Resources).

Research Area: Engineering Physics, Electromagnetic Field Theory, Electronic Material and Processes, Wireless Communications.

Mr. Bhavinbhai G. Lakhani

An expert in Environmental Technology and Sustainability, with an M.S. from NYIT. Their specialization includes Construction Project Management and Green Building. Currently a Project Controls Specialist Lead at DACK Consulting Solutions, they manage project schedules, resolve delays, and handle claim negotiations. Prior roles as Senior Project Manager at FCS Group and Senior Project Engineer at KUNJ Construction Corp highlight their extensive experience in project estimation, resource management, and on-site supervision.

Mr. Omar Muhammed Neda

Department of Electrical Power Engineering, Sunni Diwan Endowment, Iraq.

Mr. Amit Kumar

Amit Kumar is associated as a Researcher with the Department of Computer Science, College of Information Science and Technology, Nanjing Forestry University, Nanjing, China since 2009. He is working as a State Representative (HP), Spoken Tutorial Project, IIT Bombay promoting and integrating ICT in Literacy through Free and Open Source Software under National Mission on Education through ICT (NMEICT) of MHRD, Govt. of India; in the state of Himachal Pradesh, India.

Mr. Tanvir Singh

Tanvir Singh is acting as Outreach Officer (Punjab and J&K) for MHRD Govt. of India Project: Spoken Tutorial - IIT Bombay fostering IT Literacy through Open Source Technology under National Mission on Education through ICT (NMEICT). He is also acting as Research Associate since 2010 with Nanjing Forestry University, Nanjing, Jiangsu, China in the field of Social and Environmental Sustainability.

Mr. Abilash

MTech in VLSI, BTech in Electronics & Telecommunication engineering through A.M.I.E.T.E from Central Electronics Engineering Research Institute (C.E.E.R.I) Pilani, Industrial Electronics from ATI-EPI Hyderabad, IEEE course in Mechatronics, CSHAM from Birla Institute Of Professional Studies.

Mr. Varun Shukla

M.Tech in ECE from RGPV (Awarded with silver Medal By President of India), Assistant Professor, Dept. of ECE, PSIT, Kanpur.

Mr. Shrikant Harle

Presently working as a Assistant Professor in Civil Engineering field of Prof. Ram Meghe College of Engineering and Management, Amravati. He was Senior Design Engineer (Larsen & Toubro Limited, India).

Mr. Zairi Ismael Rizman

Senior Lecturer, Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM) (Terengganu) Malaysia Master (Science) in Microelectronics (2005), Universiti Kebangsaan Malaysia (UKM), Malaysia. Bachelor (Hons.) and Diploma in Electrical Engineering (Communication) (2002), UiTM Shah Alam, Malaysia.

Mr. Ronak

Qualification: M.Tech. in Mechanical Engineering (CAD/CAM), B.E.

Presently working as a Assistant Professor in Mechanical Engineering in ITM Vocational University, Vadodara. Mr. Ronak also worked as Design Engineer at Finstern Engineering Private Limited, Makarpura, Vadodara.

Table of Contents

Volume-11, Issue-11, November 2025

S. No	Title	Page No.
1	A Comprehensive Review of Gauze and Mop Counting Automation Systems for Orthopedic Surgical Safety Authors: Mr. Harsh Kotak; Dr. Nitesh Patel; Mr. Ronak Gandhi  DOI: https://dx.doi.org/10.5281/zenodo.17761497  Digital Identification Number: IJOER-NOV-2025-3	01-11
2	Fusion Strategies for Multi-Class Stock Movement Prediction: Balancing Temporal, Spatial, and Tabular Models Authors: Yiwei Chang; Jinguo Lian  DOI: https://dx.doi.org/10.5281/zenodo.17761530  Digital Identification Number: IJOER-NOV-2025-4	12-27

A Comprehensive Review of Gauze and Mop Counting Automation Systems for Orthopedic Surgical Safety

Mr. Harsh Kotak^{1*}; Dr. Nitesh Patel²; Mr. Ronak Gandhi³

¹M.Tech Research Scholar, CAD/CAM, Mechanical Engineering Department, Faculty of Engineering and Technology, Parul University Vadodara, Gujarat – 391760

²Assistant Professor, Mechanical Engineering Department, Faculty of Engineering and Technology, Parul University Vadodara, Gujarat – 391760

³Head, Entrepreneurship Development Cell and Assistant Professor, Mechanical Engineering Department, Faculty of Engineering and Technology, ITM Vocational University, Vadodara, Gujarat – 391760

*Corresponding Author

Received: 01 November 2025/ Revised: 08 November 2025/ Accepted: 13 November 2025/ Published: 30-11-2025

Copyright @ 2025 International Journal of Engineering Research and Science

This is an Open-Access article distributed under the terms of the Creative Commons Attribution

Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted

Non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract— Retained surgical items (RSIs), particularly gauze pieces and mops, remain a preventable yet persistent threat to patient safety in orthopedic operations. Manual counting protocols, though standardized worldwide, are vulnerable to human fatigue, distraction, and workflow complexity conditions frequently intensified in lengthy orthopedic procedures that involve multiple instruments, draping layers, and substantial blood loss. To address these challenges, diverse automation systems have been developed using bar-coding, radio-frequency identification (RFID), radiofrequency detection (RFD) wands, computer vision (CV), and sensor-fusion architectures. This paper presents a comprehensive review of gauze and mop counting automation systems with a specific focus on orthopedic surgical safety. A systematic literature search was conducted across Scopus, PubMed, IEEE Xplore, and ScienceDirect databases for the period January 2010 to May 2025 using PRISMA based screening criteria. Each study was analyzed for detection accuracy, workflow integration, sterility, human factors, and compliance with international standards (ISO 13485, 14971, IEC 60601).

The review identifies that while RFID and RFD technologies achieve high detection sensitivity, they face interference and sterilization constraints; computer vision approaches offer real-time potential but remain limited by dataset variability and occlusion. Few studies report on orthopedic specific validation or multimodal fusion strategies. The synthesis highlights critical research gaps in interoperability, calibration, regulatory validation, and human-automation collaboration.

Building upon these findings, the paper proposes a robotics oriented framework integrating RFID and CV within a closed-loop counting ecosystem capable of edge-level decision making and standardized audit trails. Such an approach could substantially enhance count accuracy, reduce intra-operative delays, and strengthen traceability. The review thus provides a consolidated evidence base and future roadmap toward intelligent, standards-compliant counting automation for safer orthopedic surgery. This review identifies critical pathways for future research in robotics-assisted surgical safety systems.

Keywords— Surgical Safety; RFID; Computer Vision; Retained Surgical Items; Orthopedic Surgery; Automation Framework.

I. INTRODUCTION

Retained surgical items (RSIs) primarily gauze pieces and mops inadvertently left inside the patient's body remain a serious but preventable source of postoperative complications. The issue persists across surgical specialties despite stringent manual counting and verification protocols. The frequency of RSIs is estimated at one per 1,000 - 1,500 intra-abdominal surgeries, with underreporting masking the true burden (Steelman et al., 2018). RSIs lead to foreign body reactions, infections, reoperations, and legal actions, posing direct threats to both patient safety and hospital credibility (Rupp et al., 2012).

In orthopedic surgery, the risk is significantly magnified. These procedures are often long, involve deep cavities, heavy draping, use of metal implants, and multiple team handovers all of which compound the cognitive load on the surgical team (Zejnnullahu et al., 2017). Gauze and mop counts are conventionally conducted manually by scrub and circulating nurses, but studies show that up to 80% of RSI cases occur even when manual counts are reportedly correct (Cima et al., 2011). Fatigue, interruptions, multitasking, and environmental distractions remain unavoidable in real-world operation theatres, limiting human reliability (Steelman & Petersen, 2012).

Recent advances in automation and robotics have introduced new paradigms to enhance surgical safety. Technologies such as RFID tags, radiofrequency detection (RFD) mats, barcode tracking systems, and AI-driven computer vision algorithms have demonstrated potential to detect, track, and verify surgical consumables in real time (Cima et al., 2011; Steelman et al., 2011; Abo-Zahhad et al., 2024). From a robotics and automation viewpoint, these solutions exemplify sensor integration, real-time control, and human-machine interaction, domains central to modern intelligent healthcare systems.

However, despite multiple innovations, no single system yet achieves universal adoption. RFID tags may be hindered by metallic interference from orthopedic implants (Inaba et al., 2016); RFD devices may produce false alarms due to proximity artifacts; and vision-based systems often underperform in cluttered, blood-contaminated fields (Lázaro-Andreu et al., 2022). Furthermore, most commercial and academic efforts have generalized surgical use cases, overlooking orthopedic-specific environmental and ergonomic challenges.

Therefore, a systematic, robotics-oriented review is required to consolidate existing evidence, evaluate technology readiness, identify performance and regulatory gaps, and propose a coherent framework for next-generation automation in orthopedic operation theatres. This paper aims to fill that gap by analyzing current gauze and mop counting technologies and highlighting pathways toward standardized, intelligent, and safe surgical ecosystems.

II. LITERATURE REVIEW

2.1 Manual Counting Systems:

Despite nearly two decades of standardized counting protocols, retained surgical items (RSIs) persist as a major source of morbidity and litigation. Manual count accuracy is compromised by fatigue, interruptions, multitasking, and environmental distractions (Steelman et al., 2018; Cochran, 2022). Orthopedic operating rooms are particularly vulnerable due to longer procedure durations, heavy draping, and the use of multiple mops and gauze packs (Zejnnullahu et al., 2017; Puvanesarajah et al., 2019). Process-aided methods such as sponge holders, whiteboards, and count bags minimize cognitive load but rely entirely on human compliance (Fencel, 2016). De Vries et al. (2010) and Urbach et al. (2014) observed that even with WHO surgical-safety checklists, residual errors occur when workflow stress is high. These limitations motivated development of automation-assisted systems to enhance verification integrity.

2.2 Barcode and Data-Matrix Technologies:

Early automation focused on optical identification. Cima et al. (2011) demonstrated that bar-coded or data-matrix-tagged sponges improve traceability and reconciliation accuracy by >98 %. Subsequent multi-center studies confirmed sustained elimination of unreconciled counts over 18 months of continuous use (Cima et al., 2011). However, the requirement for line-of-sight scanning and additional handling introduces workflow interruptions, especially in high-throughput orthopedic theatres (Grant & Lin, 2020). Comparative analyses by Sirihorachai et al. (2021) and Peng et al. (2023) classified optical codes as cost-effective but limited to “surface verification” tasks, unsuitable for deep-field detection once sponges are saturated or obscured.

2.3 RFID-Based Systems:

Radio-frequency identification (RFID) has emerged as the most intensively studied technology for RSI prevention since 2010. Clinical and bench investigations reported detection sensitivities of 99–100 % for passive tags positioned within the surgical cavity (Steelman, Cullen, & Anderson, 2011; Steelman, Alasagheir, & Petersen, 2012; Rupp et al., 2012; Kranzfelder et al., 2012). Wiederkehr et al. (2014) validated performance in porcine models, while Inaba et al. (2016) confirmed effectiveness

during emergency surgeries. RFID advantages include non-line-of-sight detection, rapid scanning, and electronic audit trails (Schnock et al., 2017; Hendricks et al., 2022). Nevertheless, tag cost, sterilization tolerance, antenna placement, and metal interference remain barriers to orthopedic deployment where implants and instruments dominate the field (Parlak, Marsic, & Burd, 2012). Cost benefit models (Cima et al., 2011) indicate positive returns only at high surgical volumes.

Recent RF engineering efforts address these challenges through miniaturized antennas, multi-band readers, and anti-collision protocols (Hendricks et al., 2022). Yet, most validations remain small-scale; large orthopedic-specific trials are absent, underscoring a translational gap.

2.4 RFD Systems:

RFD systems employ passive tags activated by external readers at closure. Steelman et al. (2011) and Rupp et al. (2012) showed that RFD scanning detects sponges through tissue up to 9 cm deep. Inaba et al. (2016) demonstrated reliable performance in live surgical settings, significantly reducing search times and unreconciled counts. However, false alarms can occur due to orientation, saline pooling, or proximity to metal (Steelman et al., 2012). Comparative reviews categorize RFD as an adjunctive safeguard effective at final closure but unable to track items throughout the operation (Peng et al., 2023). Integration of RFD mats with OR tables offer hands-free detection yet raises electromagnetic-compatibility and sterilization concerns (Rupp et al., 2012).

2.5 Computer Vision and AI Approaches:

Parallel advances in medical imaging and machine learning have enabled vision-based counting. Early image processing systems tracked gauze on instrument tables (de la Fuente López, Borrás, & Vidal, 2020), later extended to deep learning architectures capable of object segmentation under occlusion (Chávez, Naranjo, & Albiol, 2020; Lázaro-Andreu et al., 2022). Recent work leverages YOLOv7/v8 detectors and transformer networks for real-time instrument and gauze recognition with >95 % mAP (Ran et al., 2023; Xu, Li, & Hu, 2025). Multi-camera and depth-sensor systems enhance robustness against illumination variation (Weidert, Gruetzner, & Mutschler, 2023; Haider et al., 2025).

AI-driven methods show promise for closed loop verification but face hurdles in dataset representativeness, annotation effort, and generalization across specialties (Deol et al., 2024; Abo-Zahhad, Elhoseny, & Abd-Elkader, 2024). Orthopedic workflows introduce additional complexity like blood contamination, metallic reflections, and layered draping impair visual cues (Puvanesarajah et al., 2019). Consequently, CV systems require multimodal redundancy or domain-adaptation strategies to reach clinical reliability (Chávez et al., 2020).

Beyond intra-operative detection, AI also aiding diagnostic identification of retained items in postoperative imaging. Yamaguchi et al. (2022) reported computer-aided diagnosis for sponge visualization on CT, demonstrating the complementary role of AI in both prevention and detection phases.

2.6 Sensor Fusion and IOT Integration:

Emerging research proposes combining RF and vision modalities to exploit complementary strengths. Multi-sensor prototypes integrate RFID for non-line-of-sight detection and CV for visual confirmation, reducing false positives (Hendricks et al., 2022; Deol et al., 2024). IoT architectures with edge computing enable low-latency analytics and cloud-based audit trails (Haider et al., 2025). These frameworks align with the trend toward smart operating rooms, where robotic subsystems and hospital information systems communicate through interoperable standards such as HL7 and FHIR (Parlak et al., 2012; Kooijmans et al., 2024).

Nevertheless, most sensor-fusion studies remain at simulation or prototype level, with limited orthopedic-specific evidence. Reported accuracy ranges from 93 % to 99 %, but testing conditions seldom replicate blood contamination, metallic implants, or prolonged operative time typical of orthopedic procedures (Sirihorachai et al., 2021). Furthermore, heterogeneous performance metrics and absence of standardized evaluation protocols impede cross-study comparison (Chávez et al., 2020; Lázaro-Andreu et al., 2022).

2.7 Human Factors and Regulatory Considerations:

All automation systems must coexist with human workflow. AORN and ISO/IEC standards emphasize usability engineering and alarm management (AORN Guidelines Advisory Board, 2022; Steelman et al., 2018). Human-machine interaction studies highlight the need for intuitive interfaces, graded alarm priorities, and minimal additional steps to avoid non-adherence (Cochran, 2022; Kooijmans et al., 2024). Cost analyses reveal that RFID/RFD systems become cost effective when institutional RSI rates exceed 1 per 5,000 cases or when litigation costs are included (Regenbogen et al., 2009; Grant & Lin, 2020).

From an automation standpoint, the literature converges on several design imperatives: (i) multimodal sensing for redundancy, (ii) edge-AI inference to maintain low latency, (iii) adherence to ISO 14971 risk-management and IEC 60601-1 safety standards, (iv) traceability through interoperable OR-IoT infrastructure, and (v) validation through multicenter, orthopedic-specific trials.

Between 2010 and 2025, gauze- and mop-counting automation evolved from optical identifiers to intelligent, connected systems integrating sensors, algorithms, and hospital data streams. Yet no single technology fully satisfies the orthopedic domain's demands for robustness, sterility, usability, and regulatory compliance.

Collectively, the reviewed studies reveal that while detection accuracy has improved significantly, orthopedic-specific validation and regulatory readiness remain underdeveloped.

Research Gap:

Based on the literature synthesis, the following research gaps were identified, guiding the aims and objectives of this review.

- 1) Lack of integrated multimodal sensing frameworks that combine RFID and CV under orthopedic conditions.
- 2) Limited validation protocols addressing sterility, occlusion, and metallic interference.
- 3) Insufficient regulatory and interoperability alignment in most published prototypes.
- 4) Underexplored human factors, engineering and workflow integration.
- 5) Minimal cross-validation of AI models on orthopedic-specific datasets.
- 6) Absence of an open, reproducible framework for Ph.D. level extension and validation.

Aims and Objectives:

The primary aim of this review is to critically evaluate, classify, and synthesize existing gauze and mop counting automation systems, with a focused lens on orthopedic surgical safety, and to propose a robotics-oriented, multimodal framework that enhances accuracy, usability, and regulatory compliance through intelligent sensing and automation.

- 1) Conduct a systematic review (2010–2025) of RFID, barcode, and CV-based RSI prevention systems.
- 2) Develop a unified taxonomy of sensing, decision-making, and usability parameters.
- 3) Identify and classify existing technological and regulatory limitations.
- 4) Align proposed systems with international standards (ISO 13485, ISO 14971, IEC 60601, IEC 62366, HL7/FHIR).
- 5) Propose a conceptual RFID–CV hybrid framework with edge-AI integration.
- 6) Outline a research roadmap for future doctoral validation.

III. PROPOSED METHODOLOGY

3.1 Concept Overview:

In response to the identified research gaps, this study conceptually proposes a robotics oriented, multimodal counting framework that integrates Radio Frequency Identification (RFID) and Computer Vision (CV) technologies to enable intelligent, real time gauze and mop tracking during orthopedic surgeries.

The proposed model is presented as a conceptual framework for future development serving as a blueprint for research scholars and engineers aiming to implement, validate, and clinically translate such systems.

Its novelty lies in merging sensing, automation, and safety regulation principles into a unified workflow that aligns with orthopedic surgical realities as shown in Figure 1.

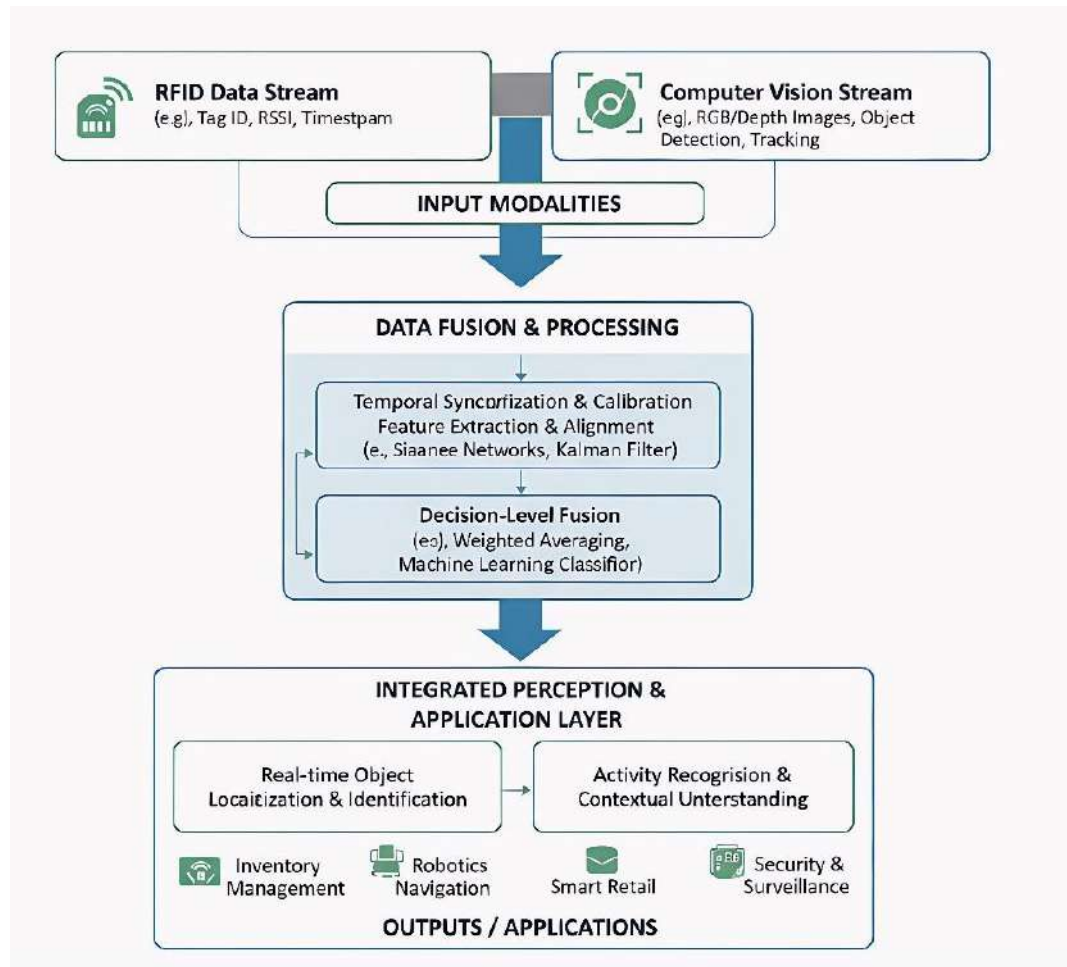


FIGURE 1: Conceptual architecture of the proposed RFID-CV multimodal framework

3.2 Conceptual System Architecture:

The framework is designed with four conceptual layers, representing the operational flow of an automated counting ecosystem as tabulated in Table 1:

3.2.1 Sensing Layer (Perception):

- Passive sterilizable RFID tags embedded in gauze/mops provide nonvisual identification.
- Overhead or side-mounted cameras use deep learning based CV algorithms for object detection and count tracking.
- Together, these systems provide redundant sensing under diverse operating conditions, improving reliability.

3.2.2 Data Fusion and Edge Processing Layer:

- Both RFID and CV data streams are processed in parallel.
- Fusion logic integrates results using a weighted-confidence algorithm or Bayesian inference model to confirm gauze/mop presence.
- Edge-AI hardware (e.g., Jetson or equivalent) is conceptually proposed for near real time processing, ensuring minimal latency.

3.2.3 Decision Support and Alerting Layer:

- The system performs automatic reconciliation between initial and final counts.

- Any mismatch triggers a graded alert sequence (visual/auditory), prompting staff verification before wound closure.
- Decision outcomes and timestamps are logged digitally for audit purposes.

3.2.4 Integration and Compliance Layer:

- Conceptually integrates with hospital information systems via standard communication protocols such as HL7 and FHIR.
- Aligns with ISO 14971 (risk management), IEC 60601 (electrical safety), and IEC 62366 (usability) to ensure regulatory readiness.

TABLE 1

FUNCTIONAL MAPPING OF PROPOSED FRAMEWORK LAYERS, OBJECTIVES, AND COMPLIANCE REFERENCES

Framework Layer	Primary Objective	Key Functional Components	Regulatory/Compliance References
1. Sensing Layer (Perception)	Provide redundant, high reliability tracking of gauze and mops.	Passive sterilizable RFID tags and readers; Deep-learning based Computer Vision (CV) algorithms (object detection/tracking) via overhead/side-mounted cameras.	ISO 13485 (Quality Management System for Medical Devices) related to sensor reliability.
2. Data Fusion and Edge Processing Layer	Confirm object presence and location in near-real-time with minimal latency.	Parallel processing of RFID and CV streams; Weighted-confidence algorithm or Bayesian Inference for data fusion; Edge-AI hardware (e.g., NVIDIA Jetson or equivalent).	IEC 60601-1 (Electrical Safety) and related standards for hardware integration in the surgical environment.
3. Decision Support and Alerting Layer	Ensure automatic and accurate reconciliation of surgical counts and prompt staff action upon discrepancy.	Automatic count reconciliation logic (Initial vs. Final count); Graded alert sequence (visual/auditory); Digital logging of decision outcomes and timestamps for audit.	ISO 14971 (Risk Management) for defining and mitigating risks associated with count errors; IEC 62366 (Usability) for alert design.
4. Integration and Compliance Layer	Facilitate seamless operational workflow and ensure regulatory readiness for clinical translation.	Integration with Hospital Information Systems (HIS); Standard protocols like HL7 and FHIR for data exchange; Alignment with safety and usability standards.	HL7/FHIR (Interoperability); ISO 14971 (Risk Management); IEC 60601 (Electrical Safety); IEC 62366 (Usability/Human Factors).

3.3 Conceptual Workflow in Orthopedic Surgery:

The proposed model as evidenced in Figure 2 can be embedded at three critical surgical checkpoints:

- **Preoperative (Initial Count):** Each gauze/mop is RFID-scanned and visually registered in the system inventory.
- **Intraoperative (Active Tracking):** Continuous sensing monitors item transitions between the sterile field, wound cavity, and waste zones. The system conceptually maintains an active count database through multimodal confirmation.
- **Postoperative (Final Verification):** Before closure, the fusion algorithm reconciles total detected items with the initial registry. Any unverified count triggers a system-generated critical alert requiring manual resolution. This workflow ensures closed-loop traceability without disrupting existing surgical roles or responsibilities.

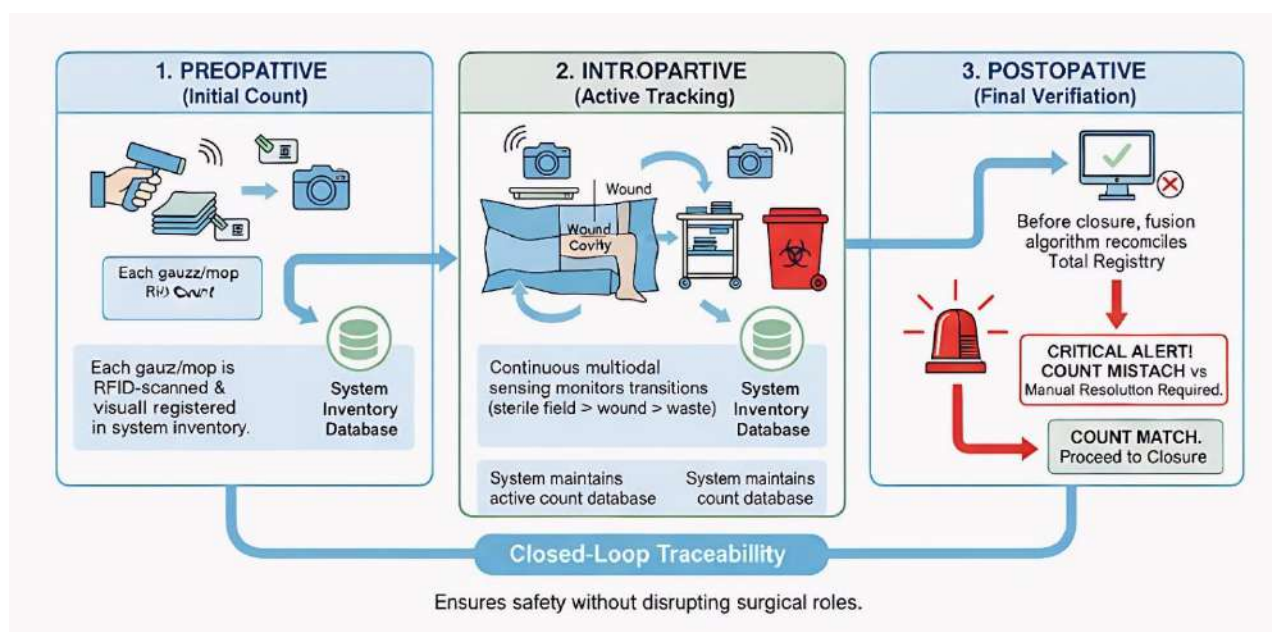


FIGURE 2: Proposed integration of the multimodal counting framework within orthopedic surgical workflow

3.4 Research Implementation Roadmap:

While this review does not perform experimental validation, it provides a conceptual implementation roadmap for future doctoral or research projects as shown in Figure 3:

1) Phase 1 – Algorithm Development:

Development of modular AI models for visual recognition and RFID event handling, with simulated data fusion for proof of concept.

2) Phase 2 – Prototype Design:

Construction of a benchtop prototype integrating RFID readers and camera modules for controlled testing in a laboratory environment.

3) Phase 3 – Validation and Optimization:

Future researchers can extend this framework for clinical evaluation under ethical approval, focusing on orthopaedic-specific constraints (metallic interference, draping, blood contamination).

This roadmap defines how the framework may be realized, without implying direct execution in the current study.

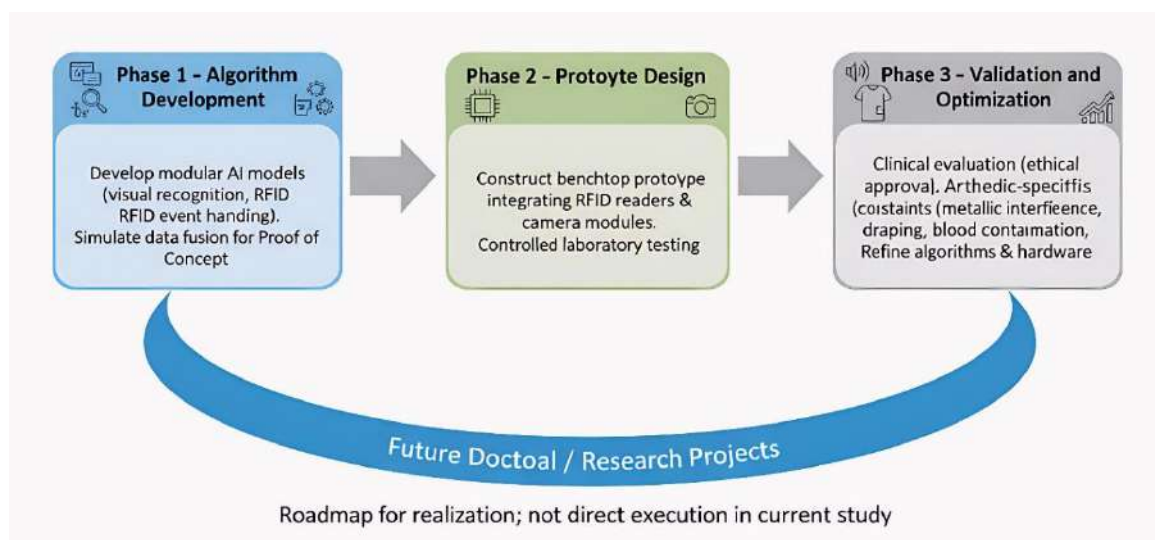


FIGURE 3: Conceptual three phase roadmap for future research and validation of the proposed system

3.5 Anticipated Benefits and Research Contributions:

The conceptual framework aims to provide the following academic and practical contributions as summarized in Table 2:

- **Multimodal redundancy:** Combining RFID and CV ensure high detection confidence, even under occlusion or electromagnetic noise.
- **Human–machine synergy:** Designed for minimal cognitive load and alignment with surgical counting checkpoints.
- **Regulatory readiness:** Framework integrates standard compliance considerations early in design, enabling smoother translation.
- **Scalability and adaptability:** Modular architecture allows easy expansion to instrument tracking and robotic surgical platforms.
- **Research utility:** Offers a structured roadmap for postgraduate and doctoral scholars to implement, refine, and validate multimodal OR automation systems.

TABLE 2
PROJECTED BENEFITS AND RESEARCH CONTRIBUTIONS OF THE PROPOSED FRAMEWORK

Contribution/Benefit	Description/Impact
Multimodal Redundancy	Combines RFID and CV to ensure high detection confidence, even under common surgical challenges like occlusion or electromagnetic noise.
Human–Machine Synergy	Designed for minimal cognitive load on surgical staff and aligns directly with existing surgical counting checkpoints.
Regulatory Readiness	Integrates standard compliance considerations (e.g., ISO, IEC) early in the design process, enabling smoother clinical translation and adoption.
Scalability and Adaptability	Features a modular architecture that allows for easy expansion to other critical tracking needs, such as surgical instrument tracking and integration with robotic surgical platforms.
Research Utility	Offers a structured roadmap and conceptual blueprint for postgraduate and doctoral scholars to implement, refine, and validate multimodal Operating Room (OR) automation systems.

IV. CONCLUSION

Automation in surgical safety has progressed substantially, yet orthopedic RSIs remain a challenge due to human error and environmental complexity. Reviewing RFID, RFD, and AI-based approaches reveals strong potential but limited clinical translation. This paper proposes a conceptual multimodal framework combining RFID and computer vision, grounded in robotics and automation principles. The system architecture emphasizes multimodal redundancy, regulatory compliance, and human–machine synergy. While implementation is beyond current scope, this work provides a research blueprint for doctoral scholars aiming to achieve verifiable, deployable, and standard-compliant automation systems for surgical safety. Ultimately, this review bridges engineering and medicine laying a path toward data-driven, intelligent operating rooms capable of achieving zero retained surgical items. The proposed review contributes a conceptual roadmap that robotics researchers and healthcare engineers can extend into experimental validation, thereby strengthening surgical safety automation frameworks.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] de Vries, E. N., Dijkgraaf, M. G. W., Smorenburg, S. M., Velema, J. P., Hobler, A., Pieper, H. A., van Kampen, R. J. (2010). The effect of a comprehensive surgical safety system on patient outcomes. *New England Journal of Medicine*, 363(20), 1928–1937. <https://doi.org/10.1056/NEJMsa0911535>
- [2] Urbach, D. R., Govindarajan, A., Saskin, R., Wilton, A. S., & Baxter, N. N. (2014). Introduction of surgical safety checklists in Ontario, Canada. *New England Journal of Medicine*, 370(11), 1029–1038. <https://doi.org/10.1056/NEJMsa1308261>

- [3] Fencil, J. L. (2016). Guideline implementation: Prevention of retained surgical items. *AORN Journal*, 104(1), 37–48. <https://doi.org/10.1016/j.aorn.2016.05.005>
- [4] Cochran, K. (2022). Guidelines in practice: Prevention of unintentionally retained surgical items. *AORN Journal*, 116(5), 427–440. <https://doi.org/10.1002/aorn.13804>
- [5] Steelman, V. M., Shaw, C., Shine, L., & Hardy-Fairbanks, A. (2018). Retained surgical sponges: A descriptive study of 319 occurrences and contributing factors (2012–2017). *Patient Safety in Surgery*, 12, Article 20. <https://doi.org/10.1186/s13037-018-0166-0>
- [6] Zejnullahu, V. A., Bicaj, B., Hamza, A. R., & Zejnullahu, V. A. (2017). Retained surgical foreign bodies after surgery. *Open Access Macedonian Journal of Medical Sciences*, 5(1), 97–100. <https://doi.org/10.3889/oamjms.2017.005>
- [7] Rabie, M. E., Hosni, A., Al Safty, A. M., & Altonbary, A. (2016). Gossypiboma revisited: A never ending issue. *International Journal of Surgery Case Reports*, 19, 87–91. <https://doi.org/10.1016/j.ijscr.2015.12.002>
- [8] Kooijmans, A. M., Vermeulen, H., & van Oostveen, C. J. (2024). Surgical instrument counting: Current practice and staff perspectives. *Perioperative Care and Operating Room Management*, 33, Article 100358. <https://doi.org/10.1016/j.pcorm.2024.100358>
- [9] Sirihorachai, R., Raksamani, K., & Sriphiromya, P. (2021). Interventions for the prevention of retained surgical items: A systematic review. *Patient Safety in Surgery*, 15, Article 24. <https://doi.org/10.1186/s13037-021-00290-2>
- [10] Grant, E. K., & Lin, I. (2020). Reducing the risk of unintended retained surgical sponges. *Perioperative Care and Operating Room Management*, 20, Article 100098. <https://doi.org/10.1016/j.pcorm.2020.100098>
- [11] Cima, R. R., Kollengode, A., Clark, J., Poola, G., Pearson, L., & Kor, D. J. (2011). Using a data-matrix-coded sponge counting system across a surgical practice: Impact after 18 months. *Joint Commission Journal on Quality and Patient Safety*, 37(2), 51–58. [https://doi.org/10.1016/S1553-7250\(11\)37007-9](https://doi.org/10.1016/S1553-7250(11)37007-9)
- [12] Steelman, V. M., Cullen, J. J., & Anderson, C. A. (2011). Sensitivity of detection of radiofrequency surgical sponges: A prospective, cross over study. *American Journal of Surgery*, 201(2), 233–237. <https://doi.org/10.1016/j.amjsurg.2010.05.001>
- [13] Steelman, V. M., Alasagheirin, M. H., & Petersen, D. R. (2012). Assessment of radiofrequency device sensitivity for detection of retained surgical sponges through the torso. *AORN Journal*, 95(5), 556–571. <https://doi.org/10.1016/j.aorn.2011.12.016>
- [14] Rupp, C. C., Kagarise, M. J., Nelson, S. M., Zegers, S. J., & Zins, J. E. (2012). Effectiveness of a radiofrequency detection system as an adjunct to manual counting protocols. *Journal of the American College of Surgeons*, 215(4), 543–549. <https://doi.org/10.1016/j.jamcollsurg.2012.06.012>
- [15] Kranzfelder, M., Schneider, A., Gillen, S., Feussner, H., & Wilhelm, D. (2012). Real-time monitoring for detection of retained surgical sponges using RFID: A pilot study. *Journal of Surgical Research*, 176(1), 19–24. <https://doi.org/10.1016/j.jss.2011.06.062>
- [16] Wiederkehr, J. C., Nobre, R. F., Garrido, T. C., Junior, A. T., da Silva, A. M. L. F., & Wiederkehr, J. C. (2014). Radio-frequency identification of surgical sponges in the abdominal cavity of swine. *Annals of Medicine and Surgery*, 3(2), 46–52. <https://doi.org/10.1016/j.amsu.2014.01.001>
- [17] Inaba, K., Okoye, O., Aksoy, H., Reddy, S., Barmparas, G., & Demetriades, D. (2016). The role of radio frequency detection system embedded surgical sponges in preventing retained surgical sponges. *Annals of Surgery*, 264(4), 599–604. <https://doi.org/10.1097/SLA.0000000000001860>
- [18] Schnock, K. O., Biggs, J., Bates, D. W., & Raut, R. (2017). Evaluating the impact of RFID technology on the incidence of retained surgical sponges. *Patient Safety in Surgery*, 11, Article 20. <https://doi.org/10.1186/s13037-017-0134-3>
- [19] Parlak, S., Marsic, I., & Burd, R. S. (2012). Introducing RFID in dynamic, time-critical clinical settings. *International Journal of Medical Informatics*, 81(10), 713–723. <https://doi.org/10.1016/j.ijmedinf.2012.06.003>
- [20] Hendricks, W., Fermil, S., Eguia, A., Jajoria, P., Jethani, A., Lam, D., & Kothari, M. (2022). Evaluation of a novel RFID intraoperative tracking system for instruments. *Sensors*, 22(15), Article 5728. <https://doi.org/10.3390/s22155728>
- [21] Peng, J., Gao, M., Ran, X., Zhang, H., Cui, Y., & Wei, R. (2023). Effectiveness of radiofrequency scanning technology in preventing RSIs: An integrative review. *Journal of Clinical Nursing*, 32(21–22), 6862–6876. <https://doi.org/10.1111/jocn.16447>
- [22] de la Fuente López, E., Borràs, A., & Vidal, R. (2020). Automatic gauze tracking in laparoscopic surgery using image processing. *Computer Methods and Programs in Biomedicine*, 190, Article 105367. <https://doi.org/10.1016/j.cmpb.2019.105367>
- [23] Chávez, G., Naranjo, V., & Albiol, A. (2020). Computer vision methods for counting surgical instruments: An analysis. *Surgical Innovation*, 27(6), 650–659. <https://doi.org/10.1177/1553350620956425>
- [24] Lázaro-Andreu, A., Sánchez-González, P., Calvo, R., Sifre, G., & Monzón-Nebot, J. (2022). Gauze detection and segmentation in minimally invasive surgery videos. *Healthcare*, 10(13), Article 2472. <https://doi.org/10.3390/healthcare10132472>
- [25] Ran, B., Pang, M., Liu, Y., Lv, Y., Huang, H., & Liu, Q. (2023). Surgical instrument detection algorithm based on improved YOLOv7x. *Sensors*, 23(11), Article 5037. <https://doi.org/10.3390/s23115037>
- [26] Weidert, S., Grützner, P. A., & Mutschler, C. (2023). A curated 3D surgical instrument collection for vision/XR research. *Scientific Data*, 10, Article 71. <https://doi.org/10.1038/s41597-023-02684-0>
- [27] Deol, E. S., Dhanireddy, A., Choi, E., Shah, S., Dykstra, J., & Shah, R. (2024). AI model for automated surgical instrument detection and counting. *Patient Safety in Surgery*, 18, Article 18. <https://doi.org/10.1186/s13037-024-00406-y>
- [28] Abo-Zahhad, M., Elhoseny, M., & Abd-Elkader, A. (2024). Minimization of retained surgical items using AI/ML: A review. *BioData Mining*, 17, Article 18. <https://doi.org/10.1186/s13040-024-00367-z>

- [29] Xu, Z., Li, Y., & Hu, C. (2025). Surgical tool detection in open surgery with RET-YOLOv8. *Biomedical Signal Processing and Control*, 93, Article 106117. <https://doi.org/10.1016/j.bspc.2025.106117>
- [30] Haider, S. A., Amer, H., Alarifi, I., Aljohani, N., Alzahrani, M., & Alghunaim, S. (2025). Multimodal AI for instrument recognition. *Bioengineering*, 12(1), Article 72. <https://doi.org/10.3390/bioengineering12010072>
- [31] Yamaguchi, S., Kamei, Y., Yamashita, K., Ogawa, K., Uramoto, H., & Takenaka, H. (2022). Computer-aided diagnosis for retained surgical sponge on imaging. *Journal of Surgical Research*, 276, 171–178. <https://doi.org/10.1016/j.jss.2021.12.053>
- [32] AORN Guidelines Advisory Board. (2022). Retained surgical items: Guideline quick view. *AORN Journal*, 115(1), 8–11. <https://doi.org/10.1002/aorn.13632>
- [33] Puvanesarajah, V., Rao, S. S., Sood, A., Atta, S. B., Levin, A. S., & Fritz, J. (2019). Extremity gossypiboma mimicking sarcoma: Case report and review. *Skeletal Radiology*, 48(4), 629–635. <https://doi.org/10.1007/s00256-018-3059-5>
- [34] Yash, W., & Gupta, A. (2020). Gossypiboma retained for ten years in the thigh: A case report. *Journal of Orthopaedic Case Reports*, 10(5), 9–12. <https://doi.org/10.13107/jocr.2020.v10.i05.1816>
- [35] Alsuhaime, M. A., Alkhalifah, M. A., & Amin, M. (2023). Retained surgical item (gossypiboma): A case report and review. *Annals of Medicine and Surgery*, 85, Article 104784. <https://doi.org/10.1016/j.amsu.2023.104784>
- [36] Biswas, R. S., Ganguly, S., Saha, M. L., Ghosh, B., & Patra, R. K. (2012). Gossypiboma and surgeon: Current medicolegal aspects. *Indian Journal of Surgery*, 74(5), 318–322. <https://doi.org/10.1007/s12262-012-0419-4>
- [37] Grant, R. E., Berjohn, C., & Lamba, R. (2014). Radiographic identification of retained surgical sponges. *Clinical Imaging*, 38(5), 701–705. <https://doi.org/10.1016/j.clinimag.2014.03.006>
- [38] Steelman, V. M., & Petersen, D. R. (2012). Retained surgical items: Causes and prevention strategies. *AORN Journal*, 95(3), 306–318. <https://doi.org/10.1016/j.aorn.2011.12.015>
- [39] Trunfio, T. A., & Fay, K. (2011). Enhancing surgical counts with technology: Review. *AORN Journal*, 93(3), 300–307. <https://doi.org/10.1016/j.aorn.2010.11.039>
- [40] Cima, R. R., Kollengode, A., Storseth, M., & Clark, J. (2011). Using RF technology to reduce search time and unreconciled counts. *AORN Journal*, 93(3), 294–303. <https://doi.org/10.1016/j.aorn.2010.11.037>
- [41] Fabri, P. J., & Zayas-Castro, J. L. (2014). Human error in surgery and patient safety. *Surgery*, 156(6), 1069–1071. <https://doi.org/10.1016/j.surg.2014.08.043>
- [42] Feldman, L. S., Fuchshuber, P. R., Jones, D. B., & Schwaitzberg, S. (2016). Surgical technology and operative safety. *Surgical Endoscopy*, 30(10), 4469–4476. <https://doi.org/10.1007/s00464-016-4901-1>
- [43] Henaux, P. L., & Hübner, M. (2013). RFID in hospitals: A practical review. *International Journal of Healthcare Technology and Management*, 14(1–2), 3–20. <https://doi.org/10.1504/IJHTM.2013.055735>
- [44] Rosen, M. A., Dietz, A. S., & Pronovost, P. J. (2018). Teamwork in healthcare and patient safety. *BMJ Quality & Safety*, 27(7), 547–550. <https://doi.org/10.1136/bmjqs-2018-008461>
- [45] Morikawa, S., Inubushi, T., & Namiki, H. (2012). RFID tag performance after autoclave cycles. *Sensors and Actuators A: Physical*, 173(1), 29–34. <https://doi.org/10.1016/j.sna.2011.11.024>
- [46] Wang, L., Zhang, Y., & Chen, X. (2015). Anti-collision protocols for passive UHF RFID: A survey. *IEEE Communications Surveys & Tutorials*, 17(2), 749–773. <https://doi.org/10.1109/COMST.2014.2376516>
- [47] Raval, M. V., & Bhananker, S. M. (2015). Patient safety in the operating room. *Anesthesiology Clinics*, 33(4), 679–692. <https://doi.org/10.1016/j.anclin.2015.07.000>
- [48] Chan, B., & Bergsland, J. (2014). OR workflow and human factors. *Surgical Innovation*, 21(1), 65–72. <https://doi.org/10.1177/1553350613497097>
- [49] Sun, R., Hu, C., & Li, Y. (2021). Edge computing for medical image analysis: A review. *Computer Methods and Programs in Biomedicine*, 200, Article 105936. <https://doi.org/10.1016/j.cmpb.2020.105936>
- [50] Sadozai, F., Dagiukas, T., & Manzanara, A. (2022). Cybersecurity in healthcare IoT: A review. *Computer Networks*, 212, Article 109032. <https://doi.org/10.1016/j.comnet.2022.109032>
- [51] Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- [52] van der Schaar, M., Alaa, A. M., Someh, M., & Mordan, H. (2021). How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Machine Learning*, 110(1), 1–14. <https://doi.org/10.1007/s10994-020-05928-w>
- [53] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243. <https://doi.org/10.1136/svn-2017-000101>
- [54] Azad, T. D., Han, J., & Veeravagu, A. (2019). Machine learning in surgery: From knowledge to action. *Surgery*, 165(6), 1111–1115. <https://doi.org/10.1016/j.surg.2019.02.009>
- [55] Monkhouse, S., & Pugh, C. (2020). Simulation in surgical education. *BJS Open*, 4(6), 1181–1187. <https://doi.org/10.1002/bjs5.50373>
- [56] Chung, A. Y., & Chung, H. J. (2015). Human factors in surgical innovation. *Annals of Surgery*, 262(1), e1–e2. <https://doi.org/10.1097/SLA.0000000000001206>
- [57] Carayon, P., Wooldridge, A., Hegg, P. S., & Alarcon, A. (2014). Human factors systems approach to healthcare quality and patient safety. *Applied Ergonomics*, 45(1), 14–25. <https://doi.org/10.1016/j.apergo.2013.04.022>
- [58] Russ, S., & Sevdalis, N. (2013). Safety checklist implementation: Lessons for practice. *BMJ Quality & Safety*, 22(12), 1025–1031. <https://doi.org/10.1136/bmjqs-2013-001859>

- [59] Schiff, G. D., Singh, H., & Graber, M. L. (2015). Diagnostic error in medicine: Analysis and proposals. *BMJ Quality & Safety*, 24(7), 425–433. <https://doi.org/10.1136/bmjqs-2015-004314>
- [60] Arora, S., Sevdalis, N., Nestel, D., Dajani, D., Kneebone, R., Darzi, A., & Green, J. S. (2010). Interruptions and multitasking in surgery: An observational study. *Annals of Surgery*, 252(1), 171–178. <https://doi.org/10.1097/SLA.0b013e3181dc3659>
- [61] O'Connor, P., Reddin, C., O'Sullivan, M., Shanahan, D., & O'Brien, S. (2013). Surgical teams, human factors and outcomes. *Surgeon*, 11(3), 114–119. <https://doi.org/10.1016/j.surge.2012.10.002>
- [62] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. *OSDI*, 265–283.
- [63] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. <https://doi.org/10.48550/arXiv.1804.02767>
- [64] Bozek, J., Chlipala, J., & Li, R. (2021). Medical device cybersecurity: Standards and practice. *IEEE Access*, 9, 97735–97752. <https://doi.org/10.1109/ACCESS.2021.3094353>
- [65] Roediger, H., Karpicke, J., & Luria, M. (2013). Memory, attention, and interruptions in clinical environments. *Journal of Applied Research in Memory and Cognition*, 2(4), 223–229. <https://doi.org/10.1016/j.jarmc.2013.08.005>
- [66] Gupta, A., & Kumar, N. (2021). Edge intelligence in healthcare: A survey. *IEEE Access*, 9, 128809–128830. <https://doi.org/10.1109/ACCESS.2021.3112266>
- [67] Rieke, N., Hancox, J., Li, W., Zhou, B., Ding, Y., Liu, Y., ... Chen, J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3, Article 119. <https://doi.org/10.1038/s41746-020-00323-1>
- [68] International Organization for Standardization. (2019). *Medical devices — Application of risk management to medical devices* (ISO Standard No. 14971:2019). <https://doi.org/10.3403/30331006>
- [69] International Electrotechnical Commission. (2012). *Medical electrical equipment — Part 1: General requirements for basic safety and essential performance* (IEC Standard No. 60601-1:2012). <https://doi.org/10.3403/30127453>
- [70] International Electrotechnical Commission. (2014). *Electromagnetic compatibility — Requirements and tests* (IEC Standard No. 60601-1-2:2014). <https://doi.org/10.3403/30127454>
- [71] International Electrotechnical Commission. (2015). *Application of usability engineering to medical devices* (IEC Standard No. 62366-1:2015). <https://doi.org/10.3403/30189148>
- [72] International Organization for Standardization. (2016). *Medical devices — Quality management systems — Requirements for regulatory purposes* (ISO Standard No. 13485:2016). <https://doi.org/10.3403/30224212>
- [73] HL7 International. (2021). *FHIR release 4.0.1: Specification for healthcare interoperability*. <https://hl7.org/fhir/>.

Fusion Strategies for Multi-Class Stock Movement Prediction: Balancing Temporal, Spatial, and Tabular Models

Yiwei Chang^{1*}; Jinguo Lian²

Department of Mathematics & Statistics, University of Massachusetts Amherst, Amherst, MA, USA

*Corresponding Author

Received: 04 November 2025/ Revised: 13 November 2025/ Accepted: 19 November 2025/ Published: 30-11-2025

Copyright @ 2025 International Journal of Engineering Research and Science

This is an Open-Access article distributed under the terms of the Creative Commons Attribution

Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted

Non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract— Accurate short-horizon stock-movement forecasting remains a central problem in computational finance, where even small directional errors can accumulate into significant trading risk. The most challenging regime is the neutral state—intervals with minor price changes that are easily masked by noise. To address this challenge, we compare three complementary learning paradigms and their combinations across multiple lookback horizons for three representative equities (AAPL, GOOG, TSLA). We evaluate Long Short-Term Memory (LSTM) networks for temporal dynamics, Convolutional Neural Networks (CNNs) on polar-transformed price images for spatial pattern extraction, and XGBoost on tabular technical indicators for structured feature learning.

Empirical results (Appendices A–C) reveal distinct horizon-dependent behaviors: CNNs excel at ultra-short windows ($W = 1-3$) with perfect accuracy and neutral-F1 ≈ 1.00 but deteriorate rapidly as horizons lengthen; LSTMs gain overall accuracy with longer windows ($W = 30-60$) but lose sensitivity to neutral segments; and XGBoost remains the most stable single model, maintaining accuracy $\approx 0.89-0.93$, low loss $\approx 0.4-0.6$, and neutral-F1 $\approx 0.89-0.96$ across assets.

Building on these complementary patterns, we propose fusion frameworks that integrate CNN and XGBoost outputs through weighted voting, cascaded thresholds, and probability-smoothed blending. The best configuration—probability-smoothed fusion—achieves roughly a 3–4 percentage-point improvement in neutral-F1 over the strongest standalone model while preserving comparable accuracy and calibration loss. The LSTM is retained solely as a benchmark to illustrate sequence-model trade-offs and is not included in the fusion.

Together, the results demonstrate that combining spatial and tabular perspectives yields more balanced recognition of neutral states without sacrificing directional accuracy. Accuracy measures overall correctness, loss captures probabilistic calibration, and F1 quantifies class-wise precision–recall balance. Viewed jointly, these metrics show that CNN–XGBoost fusion produces smoother and more interpretable predictions across assets and horizons. Such stability can reduce overtrading during ambiguous market phases, improving risk-adjusted decision-making in algorithmic trading strategies.

Keywords— Short-horizon stock prediction, CNN–XGBoost fusion, polar-coordinate transformation, financial time-series classification, probabilistic calibration, neutral-class F1.

I. INTRODUCTION

Predicting short-horizon stock direction remains a central challenge in financial modeling and algorithmic trading. Near-term price movements are affected by volatility, microstructure effects, and external shocks, producing highly non-stationary data that obscure clear predictive patterns [7]. Among the three outcome classes—up, down, and neutral—the neutral state is the most difficult to detect because its price changes are small and easily masked by random market noise. Misclassifying neutral periods as directional often results in unnecessary trades, increasing turnover and reducing risk-adjusted returns.

Recent machine-learning methods have advanced the ability to model financial time series from multiple perspectives. Long Short-Term Memory (LSTM) networks capture sequential dependencies and are widely used for temporal forecasting [1, 8]. Convolutional Neural Networks (CNNs), when applied to time-series segments rendered as polar-coordinate images, recognize local spatial patterns resembling technical chart structures [2]. XGBoost, a gradient-boosted ensemble of decision trees, remains a strong baseline for structured financial indicators and is known for its stability and resilience to class imbalance [3].

Each modeling family offers a distinct view of the market but also exhibits horizon-specific limitations, as demonstrated by the empirical findings in Appendices A–C. The LSTM’s accuracy improves with longer lookback windows, yet its ability to identify neutral movements deteriorates. The CNN captures short-term spatial cues effectively—particularly at the one-day horizon—but its performance degrades sharply as the temporal window expands. XGBoost, by contrast, maintains steady accuracy, low loss, and high neutral-F1 scores across horizons, making it the most reliable standalone option in this study.

These complementary strengths motivate a targeted fusion design. In this work, we combine CNN and XGBoost outputs to integrate short-horizon pattern recognition with stable probabilistic calibration, while the LSTM is retained solely as a non-fused baseline. This separation clarifies the specific contribution of sequential modeling and helps benchmark fusion performance against traditional temporal approaches [4].

Our contributions are fourfold:

- 1) We conduct a systematic evaluation of LSTM, CNN, and XGBoost across multiple lookback windows (1–60 days) for three representative equities—Apple (AAPL), Alphabet (GOOG), and Tesla (TSLA).
- 2) We develop fusion strategies—weighted voting, cascaded thresholds, and probability smoothing—to merge CNN and XGBoost predictions.
- 3) We show that the best fusion configurations improve neutral-class F1 by up to approximately four percentage points compared to the strongest single model, while maintaining comparable accuracy and loss.
- 4) We interpret accuracy, loss, and F1 jointly: accuracy measures global correctness, loss reflects probability calibration, and F1 quantifies precision–recall balance, particularly for neutral detection.

The remainder of this paper is organized as follows: Section 2 describes the data sources, preprocessing pipeline, and model structures; Section 3 presents results for single models and fusion strategies across assets and time horizons; and Section 4 concludes with practical implications for trading and risk management, limitations, and avenues for future work

II. MATERIALS AND METHODS

2.1 Data Source:

We employ daily OHLCV (open, high, low, close, volume) data for three highly liquid technology equities—Apple Inc. (AAPL), Alphabet Inc. (GOOG), and Tesla, Inc. (TSLA)—spanning June 10, 2020 to June 10, 2025. Data were retrieved from Yahoo Finance, a widely used source for academic research in quantitative finance. These stocks were selected to represent actively traded, high-volatility securities within the same sector, providing sufficiently diverse dynamics for cross-model evaluation while avoiding issues of thin trading or missing records.

2.2 Data Segmentation and Label Definition:

To study short-term predictability at different temporal scales, each price series is segmented into overlapping lookback windows of length $W \in \{1, 3, 7, 15, 30, 60\}$ trading days. These windows capture multiple horizons:

- **1–3 days:** ultra-short-term reactions to microstructure events and news shocks;
- **7–15 days:** short-term swings reflecting sentiment cycles and mean reversion;
- **30–60 days:** broader mini-trends incorporating regime and seasonal effects.

For each window ending at time t , the next-day return is:

$$r_{t+1} = \frac{P_{t+1} - P_t}{P_t} \quad (1)$$

Classes are assigned as

$$\text{up if } r_{t+1} > +0.03, \quad \text{neutral if } |r_{t+1}| \leq 0.03, \quad \text{down if } r_{t+1} < -0.03$$

This $\pm 3\%$ threshold reflects short-horizon volatility typical of large-cap U.S. equities.

2.3 Polar Coordinate Transformation for CNN Input:

Following Sezer and Ozbayoglu [2], each price window $P = \{p_1, p_2, \dots, p_n\}$ is normalized and mapped into polar coordinates to transform sequential variations into spatial structures.

$$p_i^{norm} = \frac{p_i - \min(P)}{\max(P) - \min(P)}, \quad \theta_i = \frac{2\pi i}{n}, \quad r_i = p_i^{norm} \quad (2)$$

This representation converts the one-dimensional sequence into a two-dimensional polar image where radius encodes relative price and angle encodes time. The polar layout preserves order and scale while exposing local curvatures that resemble chart formations commonly used in technical analysis. Empirically (Appendix A–C), such images produce clear distinctions in very short windows ($W=1-3$) but lose sharpness as W grows, explaining the CNN’s degradation at long horizons.

2.4 Feature Construction for XGBoost:

For the gradient-boosted tree model, we derive structured tabular indicators from the same price data: daily returns, moving averages, momentum, relative strength index (RSI), Bollinger Bands, and rolling volatility. All features are standardized to zero mean and unit variance to ensure balanced influence during tree growth. This tabular design captures aggregated financial signals that complement the CNN’s local pattern recognition [6].

2.5 LSTM Baseline:

The Long Short-Term Memory (LSTM) network serves as an **independent baseline** rather than a fused component. It models sequential dependencies directly from the raw normalized price series. Appendix A–C show that the LSTM achieves higher accuracy at longer windows but declining Neutral-F1, highlighting its horizon-dependent bias toward directional trends. Including it as a reference provides a benchmark for assessing how fusion (CNN + XGBoost) differs from traditional sequence modeling [9].

2.6 Handling Class Imbalance:

Neutral days dominate most stock series, creating an imbalanced classification task. To address this:

- **CNN:** applies class-weighted focal loss [5] emphasizing underrepresented classes (especially neutral) and uses early stopping to prevent overfitting;
- **XGBoost:** incorporates class weights directly into the objective function to maintain sensitivity to minority classes.

Both strategies aim to prevent the models from being biased toward majority up/down classes.

2.7 Preprocessing Workflow:

Figure 1 outlines the end-to-end workflow:

- 1) Collect OHLCV data for AAPL, GOOG, TSLA (2020–2025);
- 2) Segment each series into overlapping windows for six horizons;
- 3) Convert each window into two representations—polar images (CNN) and engineered features (XGBoost);
- 4) Apply normalization, class weighting, and focal loss
- 5) Split 80/20 for training/testing with validation monitoring.

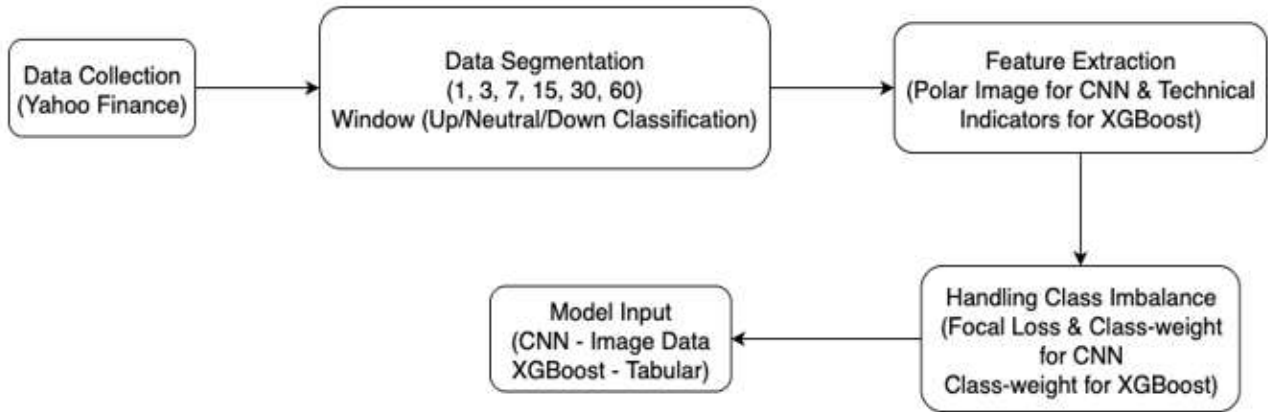


FIGURE 1: Data preprocessing pipeline showing window segmentation, polar image creation for CNN, and tabular feature generation for XGBoost

2.8 Model Structures:

CNN with Polar Input. Polar images ($H \times W_c \times 3$ tensors) are standardized to the range $[0, 1]$ and lightly augmented using rotation, scaling, and brightness jitter. We adopt a compact convolutional backbone with two variants:

- **Tri-class CNN:** outputs probabilities for *down*, *neutral*, and *up*.
- **Binary UD-CNN:** estimates the probability of *up* for fusion cascades.

Training uses the Adam optimizer with early stopping based on validation loss.

- **XGBoost.** We train a multiclass gradient-boosted tree model using the multi:softprob objective to output calibrated class probabilities. Typical hyperparameters include: learning rate = 0.1, max_depth = 6, n_estimators = 100, subsample = 0.8, and column sample = 0.8. Evaluation metrics include multi-class log-loss and overall accuracy.
- **Fusion Scope.** Consistent with the results in Appendices A–C, the fusion framework integrates only CNN and XGBoost. The LSTM remains separate as a sequential benchmark. This choice is empirically motivated: CNN provides strong sensitivity to short-horizon spatial signals, XGBoost offers stable probabilistic calibration, and combining them yields complementary improvements in neutral-class performance.

2.9 Evaluation Metrics:

To evaluate classification performance, we adopt three complementary metrics: **Accuracy**, **Cross-Entropy Loss**, and **F1 Score**. Accuracy reflects overall correctness; loss measures probabilistic calibration; and F1 captures the precision–recall balance, especially critical under class imbalance where the neutral class dominates. Formally:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, Loss = -\frac{1}{N} \sum_{i,k} y_{ik} \log(\hat{p}_{ik}), F1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall}$$

A model with high accuracy, low loss, and balanced F1 across classes is considered both well-calibrated and suitable for decision-making.

2.10 Fusion Strategies:

2.10.1 Soft Voting:

Let $p_{cnn}, p_{xgb} \in \mathbb{R}^3$ denote class-probability vectors for (down, neutral, up). Weighted averaging forms the fused prediction:

$$\hat{p} = \text{normalize}(Wp_{cnn} + (I - W)p_{xgb}) \quad (3)$$

where $W = \text{diag}(w_{\text{down}}, w_{\text{neutral}}, w_{\text{up}})$. Weights $w_{\text{down}}, w_{\text{up}} > 0.5$ emphasize CNN directional detection, while $w_{\text{neutral}} < 0.5$ assigns greater confidence to XGBoost for neutral identification.

2.10.2 Hybrid Fusion Model:

Cascaded Thresholds: A binary CNN provides p_{up} , and XGBoost outputs p_{neutral} . The confidence score $\text{conf}_{ud} = \max(p_{\text{up}}, 1 - p_{\text{up}})$ controls the cascade:

- If $\text{conf}_{ud} \geq \tau_{ud}$, choose CNN's up/down decision.
- Else if $p_{\text{neutral}} \geq \tau_{\text{neu}}$, predict neutral.
- Otherwise, fall back to the higher-confidence CNN output.

Thresholds $(\tau_{ud}, \tau_{\text{neu}})$ are grid-searched per asset.

2.10.3 Probability-Smoothed Fusion:

To further stabilize predictions, cascaded outputs are passed through a probability-smoothing calibration stage. By applying softmax temperature scaling or a confidence threshold τ_{neu} , the resulting probability distribution becomes smoother and better calibrated across all three classes. This reduces overconfidence and aligns CNN and XGBoost probability scales, improving multi-model consistency.

To avoid hard thresholds, we generate a continuously smoothed tri-class distribution:

$$p'_{\text{down}} = 1 - p_{\text{up}} \quad (4)$$

$$p_{\text{not-neutral}} = 1 - p_{\text{neutral}} \quad (5)$$

$$\tilde{p}_{\text{down}} = p_{\text{not-neutral}} \cdot \frac{p'_{\text{down}}}{p_{\text{up}} + p'_{\text{down}}}, \tilde{p}_{\text{up}} = p_{\text{not-neutral}} \cdot \frac{p_{\text{up}}}{p_{\text{up}} + p'_{\text{down}}} \quad (6)$$

$$\hat{P} = \text{normalize}\{\tilde{p}_{\text{down}}, p_{\text{neutral}}, \tilde{p}_{\text{up}}\} \quad (7)$$

Together, these steps form the **Cascaded and Probability-Smoothed Fusion** model, where the cascaded decision structure ensures directional selectivity, and the smoothing process enforces consistent probabilistic calibration across all three classes.

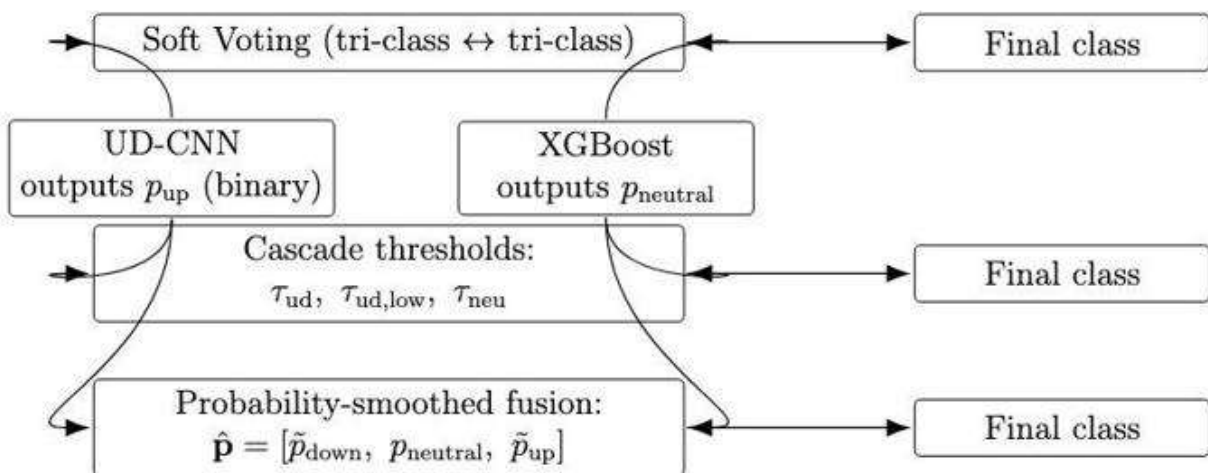


FIGURE 2: Fusion strategies tested: weighted voting, cascaded thresholds, and probability smoothing. All combine CNN and XGBoost outputs; the LSTM is evaluated separately as a reference baseline

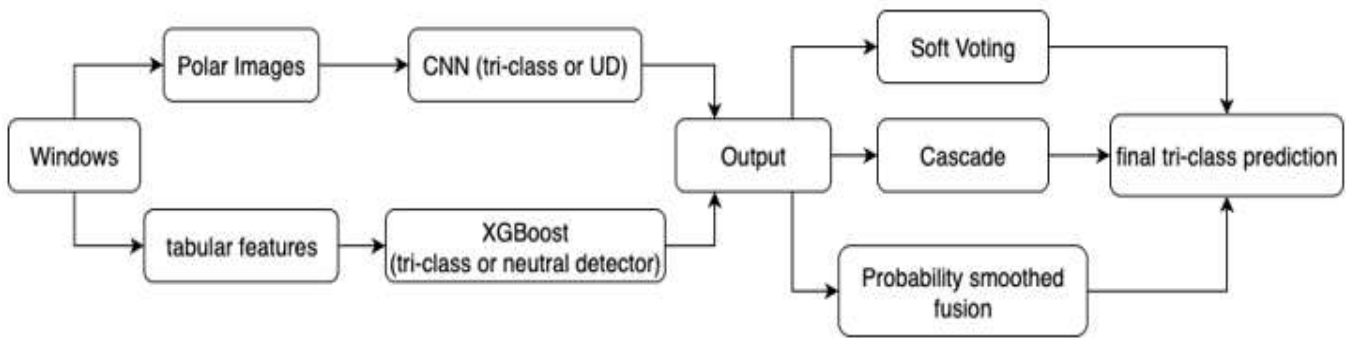


FIGURE 3: Overall architecture: polar images feed the CNN, tabular indicators feed XGBoost, and their probability outputs are fused. The LSTM baseline is run independently for comparison

III. RESULTS AND DISCUSSION

3.1 Single-Model Baselines Across Horizons:

3.1.1 CNN with Polar-Image Input:

When short windows are used, the CNN trained on polar-coordinate images achieves exceptionally strong performance. At $W = 1$ day, it attains perfect accuracy and a Neutral-F1 score of **1.00** across all assets (Appendix A and E). This confirms that the polar transformation enables the network to capture immediate geometric cues in price movement—such as sharp reversals or local plateaus—that are highly discriminative at ultra-short horizons.

However, this advantage diminishes rapidly as the lookback window increases. For AAPL at $W = 30$, accuracy falls to **0.723** and Neutral-F1 to **0.52**, and at $W = 60$ the Neutral class collapses to **0**. Loss values (Appendix B) simultaneously rise toward **0.9**, indicating poor probability calibration at longer horizons. This degradation arises because longer polar sequences lose spatial coherence—images become noisy, overlapping, and visually ambiguous, making them harder for convolutional filters to separate effectively. Therefore, CNNs perform best for ultra-short forecasts but are unreliable for medium- or long-term prediction windows.

3.1.2 LSTM (Sequential Baseline):

The LSTM provides a meaningful temporal benchmark. Accuracy improves steadily with longer horizons—from approximately **0.80** at $W = 3$ to **0.86** at $W = 60$ (Appendix A)—indicating that additional temporal context strengthens directional prediction. However, Neutral-F1 decreases sharply over the same range (AAPL: **0.95** → **0.56**; GOOG: **0.95** → **0.43**), demonstrating an increasing tendency to classify ambiguous days as directional movement.

Loss values decrease with increasing window length (Appendix B), suggesting improved probabilistic calibration, but this does not correct the declining sensitivity to neutral periods. As a result, the LSTM is retained only as a comparative baseline to illustrate the horizon-dependent trade-off between accuracy and class balance.

3.1.3 XGBoost (Tabular Features):

Across all horizons, XGBoost exhibits the most stable and balanced performance. For AAPL at $W = 30$, it achieves an accuracy of **0.89**, loss of **0.50**, and Neutral-F1 of **0.94**—levels that remain similarly strong for GOOG and TSLA. Appendix A reports accuracy levels between **0.89–0.93** for AAPL and GOOG, with moderately lower values (**0.53–0.62**) for TSLA due to greater volatility.

Loss remains consistently low (**0.4–0.6**, Appendix B), confirming strong probability calibration. Appendix E shows Neutral-F1 values remaining between **0.89–0.96** across horizons, making XGBoost the most dependable single model for neutral detection.

These results align with prior findings on ensemble-method advantages reported by Chen and Guestrin [3] and other studies on hybrid financial forecasting frameworks [4].

TABLE 1
REPRESENTATIVE SINGLE-MODEL RESULTS FOR AAPL (30-DAY WINDOW).

Model	Window	Accuracy	Loss	Macro-F1	Up-F1	Neutral-F1	Down-F1
LSTM	30	0.776	0.446	0.68	0.83	0.51	0.73
CNN	30	0.723	0.576	0.71	0.81	0.52	0.8
XGBoost	30	0.89	0.497	0.31	0	0.94	0

3.2 Fusion of CNN and XGBoost:

All fusion experiments combine the outputs of the CNN and XGBoost models, while the LSTM is excluded from fusion and used only as a reference baseline. Three fusion mechanisms were evaluated: soft voting, cascaded thresholds, and probability smoothing. The goal was to assess whether combining spatial representations from CNNs with tabular feature learning from XGBoost could improve prediction stability, calibration, and neutral-state detection across forecasting horizons.

3.2.1 Soft Voting:

In the soft-voting approach, weighted probabilities from CNN and XGBoost are combined based on performance characteristics. Higher weights were assigned to the CNN for movement direction (up and down), while XGBoost received higher emphasis for detecting neutral periods. Grid-based hyperparameter tuning identified optimal balance points, yielding Macro-F1 scores of approximately 0.86 and Neutral-F1 scores near 0.96 in the strongest configuration (GOOG, $W = 15$). Despite these improvements, gains remained inconsistent across assets and horizons. This variability indicates that fixed weighting strategies cannot fully adapt to changing market volatility or asset-specific behavior.

3.2.2 Cascaded Thresholds:

The cascaded framework introduces conditional routing of predictions. The CNN first evaluates directional confidence; if confidence exceeds a defined threshold, the prediction is accepted. Otherwise, the decision is deferred to XGBoost, particularly for potential neutral cases. The best-performing thresholds were (0.8, 0.65), producing highly competitive results in isolated configurations. For example, the cascade achieved a Neutral-F1 of 0.99 for AAPL at $W = 60$ and 0.96 for GOOG at $W = 15$. However, performance consistency decreased when evaluating pooled three-class predictions across all assets. Test loss also fluctuated notably, suggesting that cascaded logic is highly sensitive to hyperparameter selection and asset volatility characteristics.

3.2.3 Probability-Smoothed Fusion:

Probability smoothing replaces binary routing logic with a continuous blending mechanism that mixes CNN directional confidence with XGBoost's neutral probability. This eliminates abrupt decision boundaries and produces smoother predicted distributions. The approach achieved the most stable cross-asset performance. For example, GOOG at $W = 30$ achieved accuracy of 0.919, loss of 0.38, and Neutral-F1 of 0.96. Performance remains robust even at longer windows ($W = 60$), particularly for more volatile assets such as TSLA. On average, across all assets and horizons, Neutral-F1 improvements of approximately 3–4 percentage points were observed relative to the strongest individual model, while accuracy and calibration loss remained comparable. These results indicate that integrating spatial and tabular learning perspectives through probability smoothing yields more balanced three-class performance without compromising overall predictive reliability.

TABLE 2
REPRESENTATIVE FUSION RESULTS FOR GOOG (15-DAY WINDOW)

Model	Window	Accuracy	Loss	Macro-F1	Up-F1	Neutral-F1	Down-F1
Soft Voting	15	0.938	0.457	0.86	0.75	0.96	0.86
Hybrid Fusion Model	15	0.857	0.852	0.85	0.85	0.91	0.8

3.3 Comparative Interpretation:

Accuracy reflects the proportion of correct predictions across all three classes. As shown in Appendix A, XGBoost and the fused approaches maintain the most stable accuracy across forecasting windows. In contrast, the CNN performs strongly only at very short horizons and deteriorates rapidly beyond $W = 3$. The LSTM shows increasing accuracy as the temporal window expands, although this does not translate into balanced classification across all classes.

- Loss represents probabilistic calibration, where lower values indicate that predicted probabilities correspond meaningfully to observed class frequencies. Appendix B demonstrates that XGBoost and the probability-smoothed fusion method achieve the lowest and most consistent loss values, typically between 0.3 and 0.6. This indicates that both models not only predict correctly but also express confidence in a manner aligned with empirical outcomes.
- Neutral-F1 evaluates precision and recall specifically for the neutral class, which is operationally important for preventing unnecessary directional trades. Results in Appendix C show that fusion models improve Neutral-F1 by approximately three to four percentage points compared with the strongest individual model, confirming that combining CNN's sensitivity to short-term structure with XGBoost's feature stability produces measurable classification benefits.

Taken together, the metrics reveal a consistent pattern in model behavior. The CNN is most effective at ultra-short temporal horizons but declines as sequences become longer. The LSTM benefits from extended temporal context but loses precision in distinguishing neutral states. XGBoost remains the most stable and well-calibrated single model across assets and horizons. The fusion models, particularly those incorporating probability smoothing, provide incremental yet consistent improvements in Neutral-F1 while maintaining comparable accuracy and improved calibration behavior.

These results demonstrate that integrating spatially derived representations with structured tabular learning can mitigate, though not fully eliminate, the inherent trade-off between directional accuracy and neutral-state detection in short-horizon financial forecasting.

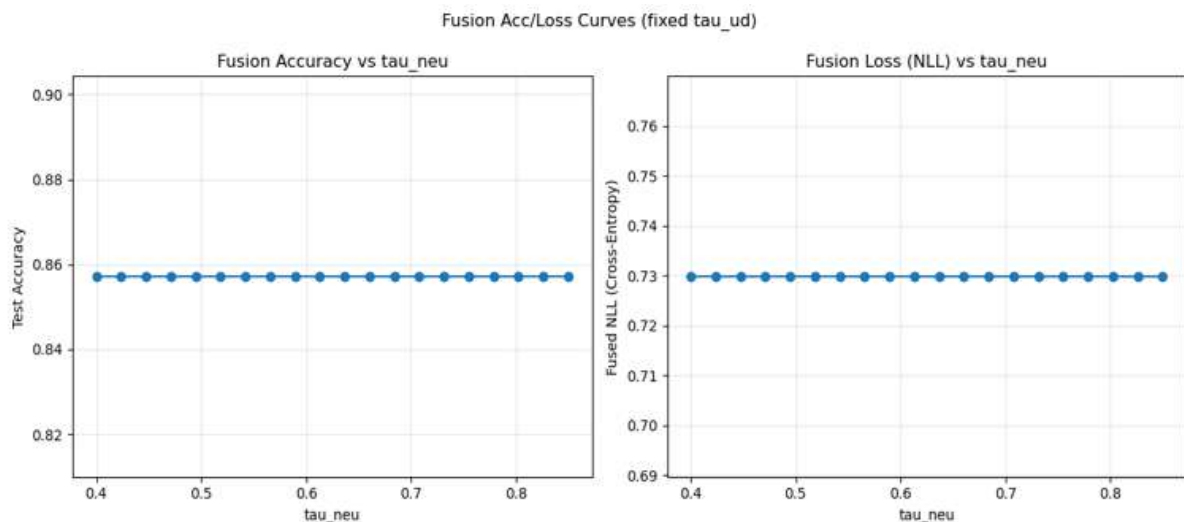


FIGURE 4: Accuracy and loss sensitivity to the neutral threshold τ_{neu} in the Hybrid Fusion

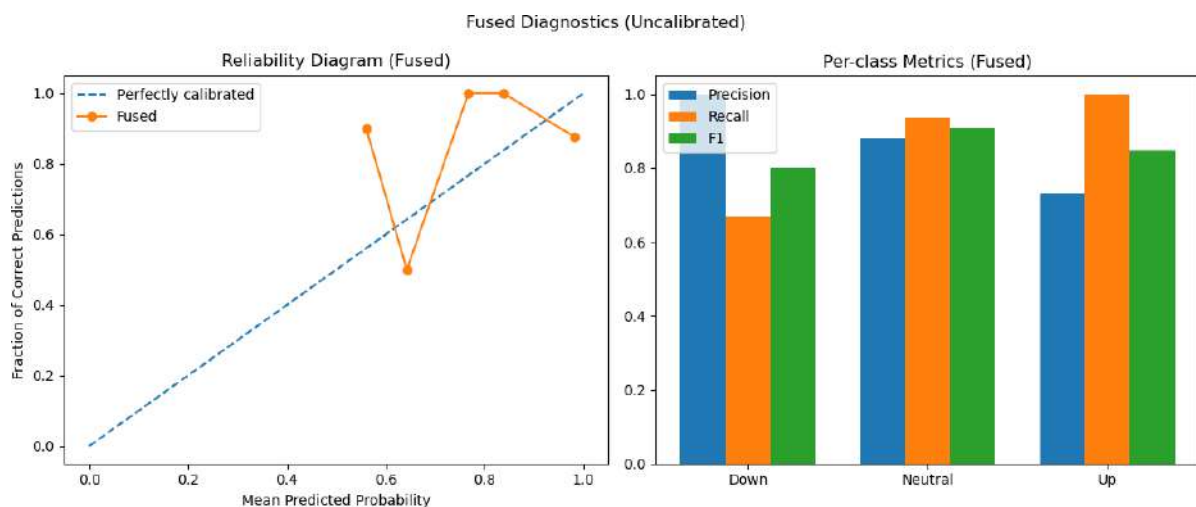


FIGURE 5: Reliability and per-class performance for representative fusion configurations. Probability-smoothed fusion shows the most stable calibration across classes

IV. CONCLUSION AND FUTURE SCOPE

This study investigated short-horizon stock movement prediction with a specific focus on accurately identifying the neutral class across multiple forecasting intervals. Three complementary modeling approaches were analyzed: Convolutional Neural Networks (CNNs) using polar-transformed price images, Long Short-Term Memory (LSTM) networks for sequence learning, and XGBoost trained on structured financial indicators. Results across AAPL, GOOG, and TSLA (Appendices A–F) show that each method exhibits distinct behavior depending on the prediction window, motivating the proposed hybrid design.

The CNN demonstrated exceptional performance at ultra-short windows ($W = 1-3$), achieving perfect accuracy and Neutral-F1 scores near 1.00. However, its effectiveness declined sharply as the temporal window expanded due to increasing spatial complexity in the polar representation. The LSTM exhibited the opposite trend: accuracy improved with longer horizons ($W = 30-60$), yet sensitivity to neutral patterns weakened, leading to directional overbias. XGBoost delivered the most stable and well-balanced performance across all settings, maintaining accuracy between approximately 0.89 and 0.93, loss values near 0.4–0.6, and consistently high Neutral-F1 scores between 0.89 and 0.96. This made XGBoost the most reliable standalone approach.

To leverage complementary strengths, three fusion strategies combining CNN and XGBoost outputs were evaluated: weighted soft voting, cascaded thresholds, and probability-smoothed blending. Among these, probability-smoothed fusion yielded the most robust results, improving Neutral-F1 by approximately 3–4 percentage points relative to the strongest single model while maintaining comparable accuracy and calibration. The LSTM remained part of the study only as a conceptual benchmark to highlight contrast with non-sequential architectures and to help interpret horizon-dependent trade-offs.

Across all experiments, accuracy, loss, and F1 score provided complementary evaluation perspectives. Accuracy measured overall correctness, loss captured probability calibration and confidence alignment, and F1 quantified the precision–recall balance essential for neutral prediction. Together, these metrics show that the CNN–XGBoost fusion approach produces more stable and interpretable forecasts across assets and horizons, particularly for applications where avoiding false directional signals is important.

From an applied standpoint, improved neutral detection can reduce unnecessary trades, thereby lowering transaction costs and improving risk-adjusted returns in algorithmic trading systems. The stability of XGBoost and the incremental yet consistent improvement from fusion methods provide practical insight: combining spatial and tabular representations mitigates overconfidence during volatile regimes, while sequence models remain relevant for broader directional forecasting tasks.

Several limitations warrant further study. The experiments were conducted on three individual equities; extending evaluation to diversified assets, market indices, and higher-frequency intraday data would strengthen generalization. Macroeconomic, sentiment, and options-derived variables were not incorporated and may provide additional predictive value. Finally, fusion weights were static; adaptive, regime-aware ensembles may better capture dynamic market behavior.

Future work may explore:

- 1) Incorporating external signals such as sentiment, macroeconomic indicators, or options-derived features into the polar-CNN architecture;
- 2) Testing alternative spatial encodings to preserve structure at longer windows;
- 3) Developing adaptive or self-tuning fusion systems responsive to volatility shifts; and
- 4) Extending the framework to portfolio-level or multi-asset forecasting.

Advancing these directions may establish the CNN–XGBoost hybrid as a generalizable and interpretable method for robust financial time-series prediction across diverse market environments.

ACKNOWLEDGEMENTS

This research is supported by the UMass Amherst Research Support Fund (RSF). The authors thank the university for providing the computational resources and infrastructure that made this work possible.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.
- [2] O. B. Sezer and A. M. Ozbayoglu, "Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach," *Applied Soft Computing*, vol. 70, pp. 525–538, 2018.
- [3] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
- [4] M. Ballings, D. Van den Poel, N. Hespeels, and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7046–7056, 2015.
- [5] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
- [6] Y. Zhang and L. Wu, "Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8849–8854, 2009.
- [7] K. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1–2, pp. 307–319, 2003.
- [8] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long short-term memory," *PLOS ONE*, vol. 12, no. 7, p. e0180944, 2017.
- [9] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Proceedings of the 24th International Conference on Artificial Intelligence*, 2015, pp. 2327–2333.

APPENDIX

Detailed Accuracy Results:

TABLE 3
TEST ACCURACY OF DIFFERENT MODELS ACROSS TIME WINDOWS FOR AAPL, GOOG, TSLA, AND THE
AGGREGATED THREE-CLASS (3C) SETTING

Model	Window	AAPL Acc	GOOG Acc	TSLA Acc	3C Acc
LSTM	1	0.9	0.896	0.614	0.803
	3	0.809	0.777	0.709	0.754
	7	0.712	0.7	0.732	0.73
	15	0.762	0.734	0.815	0.808
	30	0.776	0.816	0.869	0.835
	60	0.858	0.862	0.895	0.861
CNN	1	1	1	1	1
	3	0.78	0.812	0.571	0.647
	7	0.625	0.653	0.812	0.718
	15	0.796	0.708	0.771	0.912
	30	0.723	0.771	0.792	0.856
	60	0.596	0.809	0.872	0.846
XGBoost	1	0.908	0.885	0.53	0.794
	3	0.892	0.9	0.55	0.811
	7	0.892	0.896	0.552	0.814
	15	0.887	0.907	0.621	0.805
	30	0.89	0.91	0.616	0.784
	60	0.912	0.929	0.561	0.803
CNN+XGBoost (Soft Voting)	1	0.74	0.82	0.46	0.467
	3	0.74	0.833	0.49	0.487
	7	0.812	0.837	0.5	0.487
	15	0.653	0.938	0.479	0.493
	30	0.809	0.646	0.562	0.452
	60	0.979	0.681	0.681	0.448
CNN+XGBoost (Hybrid Fusion Model)	3	0.804	0.784	0.765	0.583
	7	0.638	0.74	0.8	0.647
	15	0.61	0.857	0.86	0.647
	30	0.784	0.919	0.9	0.811
	60	0.875	0.871	0.917	0.891

Detailed Loss Results:

TABLE 4
TEST LOSS OF DIFFERENT MODELS ACROSS TIME WINDOWS FOR AAPL, GOOG, TSLA, AND THE
AGGREGATED THREE-CLASS (3C) SETTING

Model	Window	AAPL Loss	GOOG Loss	TSLA Loss	3C Loss
LSTM	1	0.375	0.402	0.937	0.606
	3	0.465	0.518	0.675	0.601
	7	0.642	0.699	0.598	0.637
	15	0.544	0.557	0.389	0.492
	30	0.446	0.429	0.332	0.384
	60	0.316	0.36	0.303	0.345
CNN	1	0	0	0	0
	3	0.039	0.075	0.056	0.061
	7	0.076	0.05	0.06	0.057
	15	0.043	0.067	0.078	0.027
	30	0.576	0.47	0.524	0.036
	60	0.9	0.53	0.36	0.047
XGBoost	1	0.423	0.493	1.113	0.657
	3	0.432	0.409	1.065	0.598
	7	0.463	0.419	1.017	0.566
	15	0.44	0.389	0.937	0.585
	30	0.497	0.376	0.946	0.634
	60	0.399	0.302	1.091	0.589
CNN+XGBoost (Soft Voting)	1	0.135	0.108	0.248	0.284
	3	0.576	0.509	1.011	1.086
	7	0.565	0.51	0.819	1.086
	15	0.642	0.457	0.773	0.917
	30	0.527	0.586	0.672	0.871
	60	0.569	0.636	0.672	0.909
CNN+XGBoost (Hybrid Fusion Model)	3	0.721	0.467	0.865	0.736
	7	0.462	0.52	0.856	0.497
	15	0.912	0.852	0.768	0.497
	30	0.993	0.38	0.363	0.414
	60	0.747	0.256	0.357	0.304

Detailed Macro-F1 Results:

TABLE 5
TEST MACRO-F1 OF DIFFERENT MODELS ACROSS TIME WINDOWS FOR AAPL, GOOG, TSLA, AND THE
AGGREGATED THREE-CLASS (3C) SETTING

Model	Window	AAPL F1	GOOG F1	TSLA F1	3C F1
LSTM	1	0.32	0.32	0.25	0.3
	3	0.65	0.66	0.72	0.69
	7	0.7	0.69	0.66	0.73
	15	0.77	0.73	0.74	0.79
	30	0.68	0.7	0.61	0.72
	60	0.79	0.74	0.62	0.78
CNN	1	1	1	1	1
	3	0.29	0.59	0.72	0.38
	7	0.65	0.67	0.79	0.75
	15	0.79	0.73	0.71	0.9
	30	0.71	0.58	0.66	0.81
	60	0.24	0.77	0.72	0.73
XGBoost	1	0.32	0.35	0.29	0.3
	3	0.32	0.32	0.34	0.31
	7	0.35	0.38	0.28	0.35
	15	0.36	0.32	0.38	0.34
	30	0.31	0.32	0.35	0.32
	60	0.32	0.32	0.31	0.38
CNN+XGBoost (Soft Voting)	1	0.28	0.3	0.29	0.23
	3	0.28	0.3	0.33	0.25
	7	0.48	0.41	0.48	0.25
	15	0.5	0.86	0.47	0.46
	30	0.7	0.57	0.56	0.43
	60	0.98	0.58	0.68	0.45
CNN+XGBoost (Hybrid Fusion Model)	3	0.42	0.47	0.77	0.58
	7	0.46	0.52	0.78	0.61
	15	0.58	0.85	0.77	0.61
	30	0.79	0.91	0.73	0.7
	60	0.86	0.86	0.79	0.73

Detailed Up-F1 Results:

TABLE 6
TEST UP-F1 OF DIFFERENT MODELS ACROSS TIME WINDOWS FOR AAPL, GOOG, TSLA, AND THE
AGGREGATED THREE-CLASS (3C) SETTING

Model	Window	AAPL F1	GOOG F1	TSLA F1	3C F1
LSTM	1	0	0	0	0
	3	0.58	0.54	0.75	0.61
	7	0.75	0.67	0.84	0.79
	15	0.8	0.81	0.89	0.88
	30	0.83	0.87	0.9	0.88
	60	0.93	0.91	0.94	0.93
CNN	1	-	-	-	-
	3	0	0.57	0.67	0.17
	7	0.62	0.92	0.88	0.76
	15	0.92	0.81	0.91	0.93
	30	0.81	0.92	0.87	0.93
	60	0.71	0.92	0.92	0.93
XGBoost	1	0	0.12	0.06	0
	3	0	0	0.17	0
	7	0.1	0.2	0.06	0.02
	15	0.12	0	0.15	0.06
	30	0	0	0.1	0.02
	60	0	0	0.1	0.09
CNN+XGBoost (Soft Voting)	1	0	0	0.1	0
	3	0	0	0	0.05
	7	0	0	0.36	0.05
	15	0.12	0.75	0.45	0.44
	30	0.5	0.27	0.53	0.4
	60	1	0.38	0.67	0.42
CNN+XGBoost (Hybrid Fusion Model)	3	0.36	0.53	0.77	0.54
	7	0.56	0.67	0.8	0.78
	15	0.6	0.85	0.93	0.78
	30	0.1	0.94	0.95	0.87
	60	0.8	0.9	0.95	0.91

Detailed Neutral-F1 Results:

TABLE 7
TEST NEUTRAL-F1 OF DIFFERENT MODELS ACROSS TIME WINDOWS FOR AAPL, GOOG, TSLA, AND THE
AGGREGATED THREE-CLASS (3C) SETTING

Model	Window	AAPL F1	GOOG F1	TSLA F1	3C F1
LSTM	1	0.95	0.95	0.76	0.89
	3	0.88	0.85	0.64	0.82
	7	0.73	0.72	0.34	0.7
	15	0.65	0.64	0.46	0.66
	30	0.51	0.5	0.07	0.46
	60	0.56	0.43	0	0.49
CNN	1	1	1	1	1
	3	0.86	0.9	0.67	0.77
	7	0.65	0.8	0.67	0.71
	15	0.71	0.63	0.58	0.82
	30	0.52	0.33	0.29	0.69
	60	0	0.5	0.33	0.41
XGBoost	1	0.95	0.94	0.7	0.89
	3	0.95	0.95	0.71	0.9
	7	0.94	0.95	0.71	0.9
	15	0.94	0.95	0.76	0.89
	30	0.94	0.95	0.76	0.89
	60	0.95	0.96	0.72	0.89
CNN+XGBoost (Soft Voting)	1	0.85	0.9	0.63	0.63
	3	0.85	0.91	0.63	0.65
	7	0.91	0.91	0.59	0.65
	15	0.8	0.96	0.52	0.65
	30	0.89	0.77	0.59	0.58
	60	0.99	0.78	0.72	0.48
CNN+XGBoost (Hybrid Fusion Model)	3	0.9	0.88	0.73	0.6
	7	0.82	0.89	0.64	0.33
	15	0.77	0.91	0.5	0.33
	30	0.87	0.96	0.33	0.33
	60	0.78	0.83	0.5	0.32

Detailed Down-F1 Results:

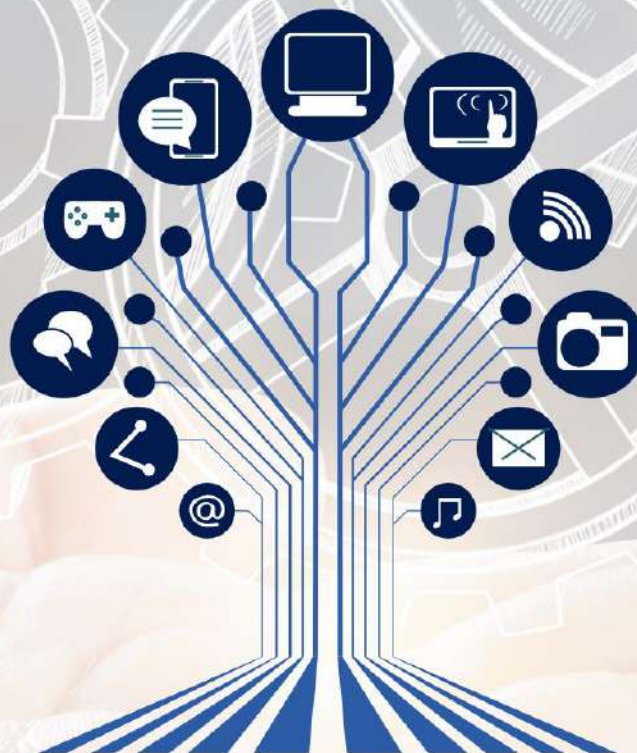
TABLE 8
TEST DOWN-F1 OF DIFFERENT MODELS ACROSS TIME WINDOWS FOR AAPL, GOOG, TSLA, AND THE
AGGREGATED THREE-CLASS (3C) SETTING

Model	Window	AAPL F1	GOOG F1	TSLA F1	3C F1
LSTM	1	0	0	0	0
	3	0.48	0.61	0.76	0.64
	7	0.61	0.68	0.79	0.7
	15	0.84	0.75	0.87	0.83
	30	0.73	0.75	0.87	0.83
	60	0.89	0.87	0.9	0.92
CNN	1	-	-	-	-
	3	0	0.29	0.81	0.2
	7	0.67	0.29	0.81	0.77
	15	0.74	0.75	0.64	0.94
	30	0.8	0.5	0.83	0.82
	60	0	0.89	0.9	0.85
XGBoost	1	0	0	0.12	0.03
	3	0	0	0.13	0.03
	7	0	0	0.07	0.13
	15	0	0	0.22	0.07
	30	0	0	0.18	0.05
	60	0	0	0.1	0.15
CNN+XGBoost (Soft Voting)	1	0	0	0.15	0.05
	3	0	0	0.38	0.05
	7	0.53	0.33	0.5	0.05
	15	0.59	0.86	0.43	0.31
	30	0.71	0.67	0.56	0.32
	60	0.94	0.57	0.64	0.44
CNN+XGBoost (Hybrid Fusion Model)	3	0	0	0.81	0.58
	7	0	0	0.89	0.74
	15	0.38	0.8	0.88	0.74
	30	0.76	0.82	0.92	0.89
	60	1	0.86	0.92	0.96



IJOER
ENGINEERING JOURNAL

International Journal of Engineering Research and Science



Published by
AD Publications

Contact us



+91-7665235235



www.ijoer.com



info@ijoer.com