Black Sea Journal of Artificial Intelligence

Volume 1 | Issue 2

ISSN: 3108-4281

KAR-YAY

BLACK SEA JOURNAL OF ARTIFICIAL INTELLIGENCE

(BSJ AI)

KAR-YAY

Black Sea Journal of Artificial Intelligence (BSJ AI) is an international a peer-reviewed journal published electronically twice a year in June and December. BSJ AI publishes original research, innovative methodologies, and interdisciplinary studies in the fields of artificial intelligence. The journal provides a global platform for researchers, scientists, engineers, and practitioners to address contemporary and emerging artificial intelligence challenges and to develop sustainable solutions for artificial intelligence. BSJ AI is published electronically and all articles are provided to the readers free of charge.

e-ISSN: 3108-4281

Phone: +90 362 408 25 15

Fax: +90 362 408 25 15

Email: bsjai@blackseapublishers.com

Web site: https://blackseapublishers.online/index.php/ai

Sort of publication: Periodically 2 times (June and December) in a year

Publication date and place: December 15, 2025- Samsun, TÜRKİYE

Publishing kind: Electronically

OWNER

KARYAY Karadeniz Yayımcılık ve Organizasyon Ticaret Limited Company

DIRECTOR IN CHARGE

Prof. Dr. Hasan ÖNDER

REVIEWERS OF THE ISSUE*

Prof. Dr. Çiğdem TAKMA, Ege University, Department of Animal Science, Artificial Intelligence Algorithms, TÜRKİYE

Prof. Dr. Fatih ÖNER, Ordu University, Department of Field Crops, Artificial Intelligence Algorithms, TÜRKİYE

Prof. Dr. Mustafa ŞAHİN, Kahramanmaraş Sütçü İmam University, Department of Agricultural Biotechnology, Artificial Intelligence Algorithms, TÜRKİYE

Assist. Prof. Dr. Emre ENGİN, Ondokuz Mayıs University, Department of Computer Technologies, Artificial Intelligence, TÜRKİYE

Assist. Prof. Dr. Hatice Kübra SEVİNÇ, Karabük University, Department of Software Application Development and Analysis, Machine Learning Algorithms, TÜRKİYE

Asist. Prof. Dr. Nurettin ŞENYER, Samsun University, Department of Software Engineering, Artificial Intelligence, TÜRKİYE

Dr. Hasan Alp ŞAHİN, Ondokuz Mayıs University, Hemp Research Institute, Artificial Intelligence, TÜRKİYE

Dr. Murat VAROL, Zonguldak Bülent Ecevit University, Department of Computer Engineering, Artificial Intelligence, TÜRKİYE

Dr. Tolga TOLUN, Kahramanmaraş Sütçü İmam University, Department of Agricultural Biotechnology, Artificial Intelligence Algorithms, TÜRKİYE

---

* The ranking is arranged alphabetically within the academic title

Volume 1, Issue 2 (2025)

Table of Contents

Research Articles

Black Sea Journal of Artificial Intelligence
doi: 10.5281/zenodo.18213771

Open Access Journal
e-ISSN: 3108-4281

Research Article
Volume 1 - Issue 2: 35-39 / December 2025

ANALYSIS OF OPTIMIZED DATASETS FOR BASIC IMAGE PROCESSING ALGORITHMS

Duygu Sedef ÇALIŞKAN[1]*, Mehmet Serhat ODABAŞ[1], Recai OKTAŞ[2]

[1]Ondokuz Mayıs University, Institute of Graduate, Department of Smart Systems Engineering, Samsun, Türkiye
[2]Ondokuz Mayıs University, Faculty of Engineering, Department of Computer Engineering, Samsun, Türkiye

Abstract: In recent years, rapid advances in artificial intelligence and computer vision have significantly enhanced object detection systems, which are now widely used in fields such as autonomous driving, surveillance, and sports analytics. This study focuses on evaluating and comparing four state-of-the-art object detection architectures (YOLOv11, YOLOv12, Roboflow 3.0, and RF-DETR) to determine their effectiveness in real-time detection of basketball players. A publicly available dataset containing 170 annotated images from basketball game scenarios was obtained from the Roboflow platform. Each model was trained using identical hyperparameter configurations to ensure a fair comparison, and its performance was evaluated using mAP@50, Precision, and Recall metrics. The results demonstrate that RF-DETR achieved the highest overall accuracy (mAP@50 = 91.5%), while YOLOv11 showed the best balance between recall (84.3%) and precision (90.2%), making it ideal for real-time applications. These findings underscore the increasing capability of modern AI models to perform reliable object detection in complex and dynamic environments. As deep learning technologies continue to evolve, such comparative studies provide essential insights for selecting the most efficient architectures for real-world implementations.

Keywords: Image processing, Dataset, YOLO

***Corresponding author:** Ondokuz Mayıs University, Institute of Graduate, Department of Smart Systems Engineering, Samsun, Türkiye
E mail: sedefcalkn@gmail.com (D. S. ÇALIŞKAN)
Duygu Sedef ÇALIŞKAN https://orcid.org/0009-0005-7588-3498
Mehmet Serhat ODABAŞ https://orcid.org/0000-0002-1863-7566
Recai OKTAŞ https://orcid.org/0000-0003-3282-3549
Received: September 10, 2025
Accepted: December 05, 2025
Published: December 15, 2025
Cite as: Çalışkan, D. S., Odabaş, M. S., & Oktaş, R. (2025). Analysis of optimized datasets for basic image processing algorithms. *Black Sea Journal of Artificial Intelligence*, *1*(2), 35–39.

1. Introduction

Image processing, a fundamental branch of artificial intelligence, enables computers to interpret and analyze visual information from images and videos (Zhao et al., 2019; Sun et al., 2024). With the integration of deep learning, it has evolved into data-driven models that perform complex tasks such as detection, segmentation, and object tracking across diverse fields like healthcare, automation, and surveillance (Tao et al., 2022). Object detection is a crucial component of computer vision systems, enabling the automated identification and localization of objects within images or video frames (Jegham et al., 2024). It plays a vital role in various computer vision applications, including autonomous driving, surveillance, and sports analytics. This advanced technology serves as a foundation for a wide range of applications across multiple industries, including robotics, security, and healthcare. Therefore, object tracking is a crucial component of video analysis, as it enhances object detection by ensuring temporal consistency across consecutive frames (Trigka and Dritsas, 2025).

YOLO (You Only Look Once) is one of the leading and most popular object detection approaches for performing object detection. Unlike traditional two-stage methods, YOLO processes the entire image at once and simultaneously predicts class probabilities and bounding box coordinates, which significantly improves speed. Various techniques have been designed for crowd detection in videos and images (Bochkovskiy et al., 2020). Providing a balance between high accuracy and low latency (Esteva et al., 2017). In recent years, deep learning–based object detection models, such as YOLO and transformer-based architectures, like Detection Transformers (DETR), have significantly advanced detection accuracy and inference speed (Bochkovskiy et al., 2020; Carion et al., 2020; Simic and Gavrovska, 2025). However, detecting small or overlapping objects under varying lighting and background conditions remains a persistent challenge (Khalili and Smyth 2024; Xiaozheng et al., 2025).

To address these limitations, this study compares four state-of-the-art architectures (YOLOv11, YOLOv12, Roboflow 3.0, and RF-DETR) using a unified dataset obtained from the Roboflow platform. The selected dataset consists of basketball-player images designed for real-time detection tasks. Each model was trained under identical conditions, and their performance was evaluated based on precision, recall, and mAP@50 metrics to determine the most effective architecture for accurate and real-time sports-oriented object detection.

2. Materials and Methods

2.1. Dataset

Historically, object detection datasets are created by collecting a large collection of images and sourcing annotators to label objects within a fixed set of classes (Ciaglia et al., 2022). The Roboflow platform facilitates this process by allowing users to upload, annotate, and share datasets with high accuracy. As noted in (Sakib et al., 2023), with Roboflow's assistance, all objects in the dataset were annotated with great detail and accuracy, ensuring that each object was accurately classified and its location was determined using bounding boxes.

In this study, an open-source dataset for detecting basketball players was obtained from the Roboflow platform. Each image contained multiple players in different poses and lighting conditions, and their positions were labeled with bounding boxes. Using Roboflow's bounding box tool and the generated annotation files, the ground truth bounding boxes in each image were automatically constructed, ensuring consistent and highly accurate annotations (Yigit, 2025). The selected dataset was deemed more suitable for object detection models because it represents a dynamic and challenging environment characterized by multiple overlapping targets, rapid movement, and complex backgrounds, similar to a basketball game. This makes it suitable for evaluating YOLO and transformer-based architectures in terms of accuracy, robustness, and real-time performance.

The YOLO family approaches object detection as a single-stage regression problem, directly estimating the class probabilities p(ci) and location coordinates (x, y, w, h) of objects in an image (Ultralytics, 2024).

Overall detection confidence is defined as given in Equation 1:

$$Confidence = P(Object)\ x\ IoU_{pred}^{truth} \quad (1)$$

YOLOv11 and YOLOv12 are the latest versions of the Ultralytics series, which balance high accuracy with real-time performance using a CSPDarknet-based backbone and PANet/ConvNeXt-based feature fusion layers (Audu and Ndirmbula, 2024).

Roboflow 3.0 enhances the model's generalization performance by combining the optimization processes of the YOLO algorithm with advanced data preprocessing and data augmentation techniques.

RF-DETR (Roboflow-DETR) employs a transformer-based encoder-decoder architecture, capturing the contextual relationships between objects in the image through attention mechanisms (Redmon et al., 2016).

2.2. Dataset Properties

The dataset consists of a total of 170 annotated images, each depicting basketball game scenarios captured from YouTube videos at various angles and lighting conditions. It includes nine object classes, with "Player" being the primary class. The object categories are: Ball, Hoop, Period, Player, Referee, Shot Clock, Team Name, Team Points, and Time Remaining. The dataset segmentation was given in Figure 1.

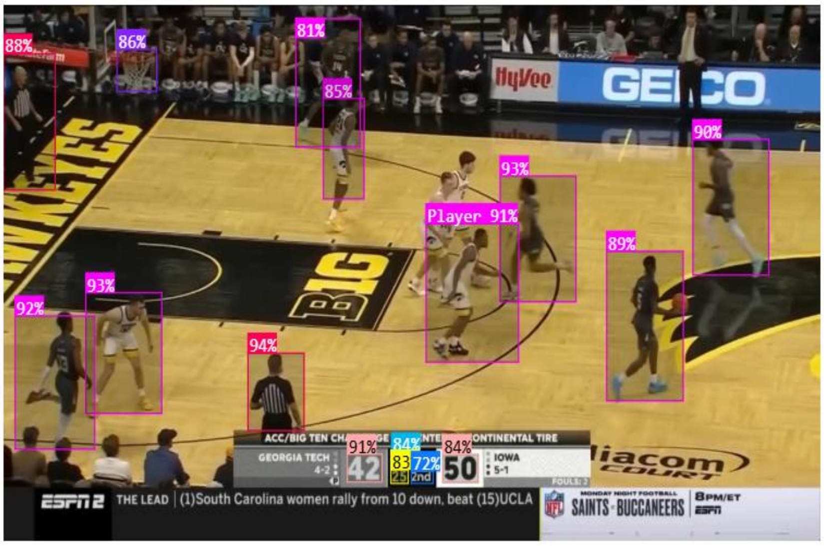


Figure 1. Dataset segmentation.

Each image was annotated with bounding boxes that precisely delineate object boundaries. The dataset was divided into 67% training, 19% validation, and 14% testing subsets to ensure balanced model evaluation. All images were standardized to a resolution of 640 × 640 pixels during preprocessing to maintain consistency across all trained models.

2.3. Data Augmentation

To improve the generalization capability of the models, various data augmentation techniques were applied through the Roboflow platform. These include horizontal and vertical flipping, brightness and contrast adjustment, rotation, and Gaussian blur. Such augmentations simulate real-world variations, helping the models to become more robust against scale, pose, and lighting changes commonly encountered in sports environments.

2.4. Model Architecture and Training Settings

In this study, four different object detection architectures (YOLOv11, YOLOv12, Roboflow 3.0, and RF-DETR) were trained and evaluated under identical experimental conditions. Each model was selected to represent a different approach to object detection: the YOLO family for real-time convolutional architectures, and RF-DETR for transformer-based detection (Chandana and Ramachandra, 2022; Zhang et al., 2023). Models were trained using the Roboflow platform, which provides an integrated training pipeline with GPU acceleration. Each training session lasted approximately 40–30 minutes, depending on the architecture and model complexity. The same dataset version, preprocessing pipeline, and augmentation techniques were applied across all models to ensure a fair comparison.

2.5. Hyperparameters

All models (YOLOv11, YOLOv12, Roboflow 3.0, and RF-DETR) were trained using the same hyperparameter configuration to ensure a fair and unbiased comparison (Table 1). This uniform setup ensures that any observed performance differences are due solely to the architectural design of each model, not to differences in training conditions.

Table 1. Training hyperparameters and configuration

Image size	640 × 640 pixels
Batch size	16
Epochs	50
Optimizer	Adam optimizer with default learning-rate scheduling
Checkpoint	MS COCO pretrained weights were used to initialize all models

3. Results and Discussion

3.1. Evaluation Metrics

All four models were evaluated on the same test subset of the dataset to ensure a fair comparison. Model performance was assessed using three key metrics: mAP@50, Precision, and Recall. The mAP@50 metric measures the overall detection accuracy by calculating how precisely each model identifies and localizes objects. Precision indicates how many of the detected objects were correct, while Recall measures how many of the actual objects in the images were successfully detected. Using the same dataset and training configuration ensured that performance differences resulted only from variations in the model architectures.

The results indicate that RF-DETR achieved the highest overall detection accuracy, with a mAP@50 score of 91.5%, followed by YOLOv11 (87.3%), YOLOv12 (87.0%), and Roboflow 3.0 (86.2%). In terms of precision, YOLOv11 recorded the best performance (90.2%), closely followed by Roboflow 3.0 (90.1%) and RF-DETR (90.9%), demonstrating that all models maintained a low rate of false detections. Regarding recall, YOLOv11 again outperformed the others (84.3%), suggesting that it detected a greater proportion of true objects compared to the other architectures. Figure 2 shows the output of the proposed object detection model on a test frame from the basketball dataset.

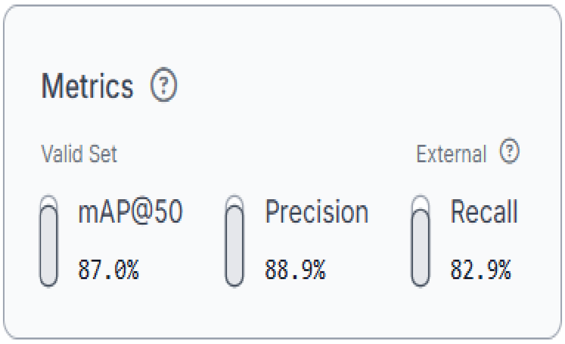
The bounding boxes demonstrate the model's ability to correctly detect multiple object classes, such as Player, Ball, and Hoop, with high confidence levels (e.g., 90%, 83%, and 98%). This qualitative example visually confirms the quantitative findings presented in Table 2, showing that the model can accurately identify and localize objects in complex, real-world scenes.

4. Conclusion

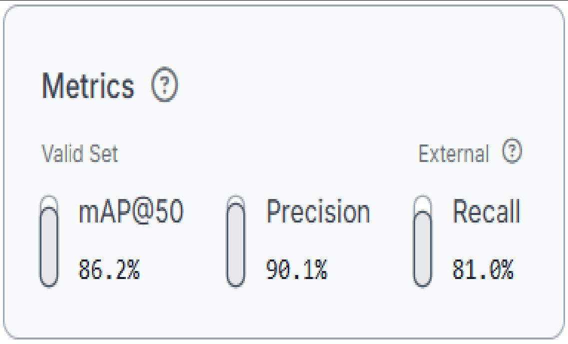
This study presented a comparative evaluation of four modern object detection architectures (YOLOv11, YOLOv12, Roboflow 3.0, and RF-DETR) trained in the same experimental environment to assess their effectiveness in real-time basketball player detection. The results showed that RF-DETR achieved the highest detection accuracy thanks to its attention-based transformative structure, while YOLOv11 offered the most balanced performance in terms of precision, recall, and inference speed, making it a strong candidate for real-time sports-focused applications. These findings highlight the growing maturity of contemporary AI models in handling visually complex scenarios involving rapid movements, occlusions, and varying lighting conditions.

Figure 2. Qualitative detection results on basketball frames.

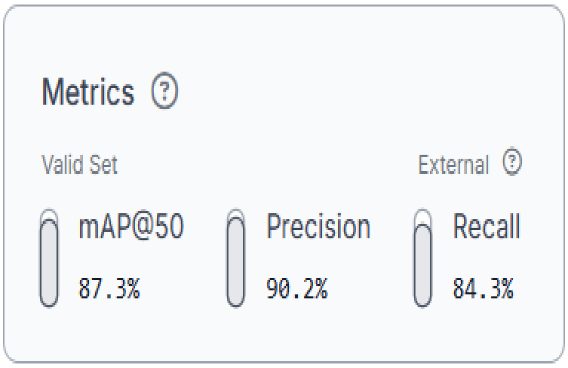
Table 2. Evaluation metrics



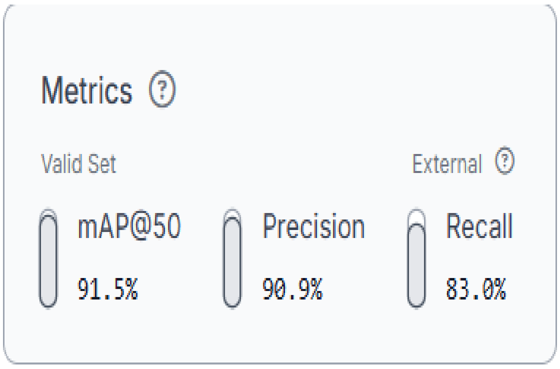
YOLOv12 Object Detection (Fast)



Roboflow 3.0 Object Detection (Fast)



YOLOv11 Object Detection (Fast)



RF-DETR (Base)

Despite these advancements, challenges remain, particularly in detecting small or overlapping targets and ensuring stability in dynamic environments. Future work could expand the dataset, incorporate multiple camera perspectives, evaluate lightweight architectures on edge devices, or explore hybrid CNN-transformer models to further improve accuracy and computational efficiency. Moreover, integrating advanced tracking algorithms could enable complete detection and tracking pipelines supporting next-generation sports analytics and automated broadcast systems. Furthermore, embedding temporal attention mechanisms could improve frame-to-frame consistency in fast-paced sequences, while domain adaptation strategies could help maintain robustness across different arenas and recording conditions. Ultimately, extending these approaches to multimodal frameworks that combine video, sensor, or contextual data could lead to more holistic and intelligent decision-making systems for real-world applications.

Author Contributions

The percentages of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	D.S.Ç.	M.S.O.	R.O.
C	40	30	30
D	40	30	30
S	40	30	30
DCP	40	30	30
DAI	40	30	30
L	40	30	30
W	40	30	30
CR	40	30	30
SR	40	30	30
PM	40	30	30
FA	40	30	30

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

Audu, A. I., & Ndirmbula, S. M. (2024). Machine learning based real-time detection of ripe and unripe sorghum. *Nigerian Journal of Engineering Science and Technology Research*, *10*(2), 124–137.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). *YOLOv4: Optimal speed and accuracy of object detection*. arXiv. https://doi.org/10.48550/arXiv.2004.10934

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In A. Vedaldi, H. Bischof, T. Brox, & J. M. Frahm (Eds.), *Computer Vision – ECCV 2020* (pp. 213–229). Springer. https://doi.org/10.1007/978-3-030-58452-8_13

Chandana, R. K., & Ramachandra, A. C. (2022). *Real time object detection system with YOLO and CNN models: A review*. arXiv. https://doi.org/10.48550/arXiv.2208.00773

Ciaglia, F., Zuppichini, F. S., Guerrie, P., McQuade, M., & Solawetz, J. (2022). *Roboflow 100: A rich, multi-domain object detection benchmark*. arXiv. https://doi.org/10.48550/arXiv.2211.13523

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). *YOLO evolution: A comprehensive benchmark and architectural review of YOLOv12, YOLO11, and their previous versions*. arXiv. https://doi.org/10.48550/arXiv.2411.00201

Khalili, B., & Smyth, A. W. (2024). *SOD-YOLOv8—Enhancing YOLOv8 for small object detection in traffic scenes*. arXiv. https://doi.org/10.48550/arXiv.2408.04786

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. https://doi.org/10.1109/CVPR.2016.91

Sakib, K. S., Abir, T. I., Kareena, S. S., & Alam, R. (2023). *Autonomous object detection dataset: A study on Bangladeshi roadways* [Undergraduate thesis, Brac University]. Brac University Repository.

Simic, N., & Gavrovska, A. (2025). Comparative analysis of YOLOv11 and YOLOv12 for AI-powered aerial people detection. *12th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN)*, 1–4.

Sun, Y., Sun, Z., & Chen, W. (2024). The evolution of object detection methods. *Engineering Applications of Artificial Intelligence*, *133*, 108458. https://doi.org/10.1016/j.engappai.2024.108458

Tao, X., Gong, X., Zhang, X., Yan, S., & Adak, C. (2022). Deep learning for unsupervised anomaly localization in industrial images: A survey. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–21. https://doi.org/10.1109/TIM.2022.3196436

Trigka, M., & Dritsas, E. (2025). A comprehensive survey of machine learning techniques and models for object detection. *Sensors*, *25*(1), 214. https://doi.org/10.3390/s25010214

Ultralytics. (2024). *YOLOv12 official release documentation*. https://docs.ultralytics.com (Erişim tarihi: 3 Eylül 2025).

Xiaozheng, Z., Zhongjun, Y., & Huaici, Z. (2025). DCS-YOLOv8: A lightweight context-aware network for small object detection in UAV remote sensing imagery. *Remote Sensing*, *17*(17), 2989. https://doi.org/10.3390/rs17172989

Yigit, G. (2025). Crowd detection: Leveraging YOLO for human recognition. *Turkish Journal of Engineering*, *9*(3), 571–577. https://doi.org/10.31127/tuje.1627839

Zhang, J., Zhang, H., Liu, B., Qu, G., Wang, F., Zhang, H., & Shi, X. (2023). Small object intelligent detection method based on adaptive recursive feature pyramid. *Heliyon*, *9*(7), e17730. https://doi.org/10.1016/j.heliyon.2023.e17730

Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865

Black Sea Journal of Artificial Intelligence
doi: 10.5281/zenodo.18213819

Open Access Journal
e-ISSN: 3108-4281

Research Article
Volume 1 - Issue 2: 40-52 / December 2025

PROMPT ENGINEERING: A BIBLIOMETRIC ANALYSIS

Alperen KOÇYİĞİT[1]*, Nurettin ŞENYER[1]

[1]*Samsun University, Faculty of Engineering, Department of Software Engineering, 55060, Samsun, Türkiye*

Abstract: Prompt engineering addresses the design and optimization of natural language instructions for controlling large language model behavior, representing a critical domain within artificial intelligence and human-computer interaction. Despite explosive research growth following widespread adoption of large language models, systematic analyses of the field's intellectual structure, publication patterns, and collaborative networks remain limited. This study conducted a comprehensive bibliometric analysis of prompt engineering research from 2020 to 2025 using Web of Science Core Collection as the data source. A systematic search strategy retrieved 4,890 publications from 1,538 sources authored by 16,052 researchers across computer science and artificial intelligence domains. Analysis employed the Bibliometrix package (version 4.3.5) in R (version 4.5.1) to examine publication trends, author productivity, institutional contributions, geographic distribution, thematic structure, collaboration patterns, and citation impact through performance analysis, keyword co-occurrence networks, and science mapping techniques. The field demonstrated explosive expansion with 125.1% annual growth rate, exhibiting three developmental phases: emergence phase (2020-2021), acceleration phase (2022-2023), and explosion phase (2024-2025) when output reached 2,093 publications annually. The Chinese Academy of Sciences led institutional productivity with 345 publications, while China dominated national output with 1,584 documents representing 32.67% of the corpus. Geographic analysis revealed quality-quantity trade-offs with Singapore achieving the highest average citation impact (37.37 citations per document) despite modest volume. Author analysis identified Zhang Y as most productive (43 publications) while collaboration metrics indicated 4.9 co-authors per document and 26.44% international co-authorship rate. Keyword analysis revealed "large language models" (946 occurrences) and "prompt engineering" (733 occurrences) as dominant themes with three distinct thematic clusters: core prompting methodologies, machine learning foundations, and application domains. Network visualization confirmed integration of few-shot learning, chain-of-thought prompting, and in-context learning techniques into large language model applications. IEEE Access dominated publication venues with 176 articles, while natural language processing conferences (ACL, EMNLP, NeurIPS) emerged as primary dissemination channels. Citation analysis identified foundational contributions in instruction following and chain-of-thought reasoning alongside contemporary methodological innovations. The findings reveal prompt engineering's rapid crystallization as a distinct research domain emphasizing practical techniques over theoretical foundations, while concentration around specific models indicates potential fragmentation risks requiring unified frameworks transcending particular implementations.

Keywords: In-context learning, Natural language processing, Publication trends, Research collaboration, Thematic analysis

***Corresponding author:** Samsun University, Faculty of Engineering, Department of Software Engineering, 55060, Samsun, Türkiye
E mail: muh.alperenkocyigit@gmail.com (A. KOÇYİĞİT)
Alperen KOÇYİĞİT https://orcid.org/0009-0004-2781-3543
Nurettin ŞENYER https://orcid.org/0000-0002-2324-9285

Received: November 02, 2025
Accepted: December 05, 2025
Published: December 15, 2025

Cite as: Koçyiğit, A., & Şenyer, N. (2025). Prompt engineering: A bibliometric analysis. *Black Sea Journal of Artificial Intelligence*, *1*(2), 40–52.

1. Introduction

The rapid evolution of artificial intelligence has transformed how we interact with computational systems, with large language models emerging as powerful tools capable of understanding and generating human-like text. These sophisticated neural networks, trained on vast corpora of textual data, have demonstrated remarkable versatility across diverse natural language processing tasks. At the heart of effectively harnessing these models lies a critical skill: prompt engineering—the art and science of crafting precise instructions to guide model behavior and optimize output quality.

Prompt engineering represents a fundamental paradigm shift in human-computer interaction, where natural language becomes the primary interface for controlling and directing artificial intelligence systems. The field's significance stems from a well-documented phenomenon: large language models exhibit substantial sensitivity to the specific wording, structure, and context provided in prompts, with seemingly minor modifications often yielding dramatically different results (Sasson Lazovsky et al., 2024). Research has demonstrated that effective prompt design can substantially enhance model performance, while poorly constructed prompts may produce irrelevant or nonsensical responses (Sasson Lazovsky et al., 2024). The emergence of prompt engineering as a distinct research domain parallels the rapid advancement of transformer-based language models, particularly following the introduction of GPT-3 in 2020 (Brown et al., 2020) and the subsequent release of ChatGPT in November 2022 (Vaira et al., 2023).

These developments catalyzed widespread public and academic interest in large language models, transforming them from specialized research tools into ubiquitous technologies accessible to millions of users. This democratization of access has simultaneously amplified both the potential impact and the critical need for systematic understanding of effective prompting strategies. Contemporary prompt engineering encompasses several sophisticated techniques that extend beyond simple instruction-giving. Zero-shot prompting provides task descriptions without examples, relying on the model's pre-existing knowledge (Jovanovic and Voss, 2025). Few-shot learning incorporates demonstration examples within prompts to guide model behavior, with research showing that even the ordering of these examples can significantly influence final outputs (Cheong, 2025). Chain-of-thought prompting breaks complex problems into intermediate reasoning steps, enabling models to tackle multi-step logical challenges more effectively (Wang et al., 2025). These diverse approaches reflect the field's growing methodological sophistication and its recognition that different tasks may require fundamentally different prompting strategies. Despite the proliferation of prompt engineering applications across domains ranging from software development (Nguyen-Duc et al., 2025) to medical diagnostics (van Diessen et al., 2024), the field's rapid growth has outpaced comprehensive scholarly analysis. While individual studies have explored specific prompting techniques or evaluated model performance on particular tasks, the broader landscape of prompt engineering research remains incompletely mapped. Understanding the field's intellectual structure, key contributors, dominant themes, and evolutionary trajectory requires systematic bibliometric investigation.

This study addresses this gap through a comprehensive bibliometric analysis of prompt engineering literature spanning 2020 to 2025. By analyzing 4,890 publications indexed in Web of Science, we examine the field's exponential growth trajectory, identify leading contributors and institutions, map collaborative networks, and trace the evolution of research themes. Our analysis reveals an annual growth rate exceeding 125%, with publication volumes increasing from 30 papers in 2020 to over 2,000 in 2024, reflecting the field's explosive expansion. This quantitative assessment provides researchers, practitioners, and policymakers with empirical insights into prompt engineering's development as a scholarly discipline and its position within the broader artificial intelligence landscape. The following sections present our methodology for data collection and analysis, results encompassing publication trends, authorship patterns, geographic distribution, and thematic evolution, and discussion of implications for future research directions in this rapidly evolving field.

2. Materials and Methods

2.1. Research Design

This study employed a quantitative bibliometric analysis approach to systematically examine the scholarly landscape of prompt engineering research. Bibliometric analysis represents an established methodology for quantitative assessment of scientific literature through statistical and mathematical techniques (Aria and Cuccurullo, 2017). The research design encompasses data collection from a bibliographic database, application of bibliometric indicators, and visualization of research patterns through network analysis and performance metrics.

2.2. Data Source

Web of Science (WoS) Core Collection served as the sole data source for this analysis. WoS was selected based on its comprehensive coverage of high-impact publications in computer science and artificial intelligence disciplines, rigorous quality control through selective journal indexing, and robust citation indexing capabilities enabling citation-based analyses (Pranckutė, 2021). The use of a single database ensures methodological consistency and data compatibility across all bibliometric indicators.

2.3. Search Strategy and Query Development

A comprehensive search query was developed through iterative refinement to capture the multifaceted nature of prompt engineering research while maintaining domain specificity. The search was executed on November 27, 2025, using the following query structure:

((TS=("prompt engineering" OR "prompt design" OR "prompt optimization" OR ("prompt" AND ("tuning" OR "learning" OR "construction") AND ("AI" OR "LLM" OR "large language model*" OR "NLP" OR "machine learning")) OR "in-context learning" OR "few-shot prompting" OR "zero-shot prompting" OR "chain of thought prompting" OR "LLM prompting" OR "GPT prompting" OR "natural language prompting" OR "prompt-based learning" OR ("prompting" AND ("technique*" OR "method*") AND ("AI" OR "NLP")))) AND PY=(2020-2025)) AND WC=(Computer Science OR Computer Science, Artificial Intelligence)

The search terms were selected to encompass core dimensions of prompt engineering. The query includes primary field terms such as "prompt engineering", "prompt design", and "prompt optimization" to capture direct references to the field. Technical process terms including "prompt tuning", "prompt learning", and "prompt construction" represent key methodological approaches within the domain. Paradigm-level terms including "in-context learning", "few-shot prompting", "zero-shot prompting", and "chain of thought prompting" capture fundamental techniques and methodologies. System-level terms including "LLM prompting", "GPT prompting", and "natural language prompting" represent implementation perspectives. The wildcard operator (*) was applied to "large language model*", "technique*", and "method*" to retrieve morphological variations.

Boolean operators AND and OR were strategically employed to combine related concepts while maintaining query precision and avoiding retrieval of irrelevant publications. The temporal scope was restricted to 2020-2025 to focus on the contemporary emergence and rapid development of prompt engineering following the introduction of GPT-3 and subsequent large language models. The WoS category restriction to Computer Science and Computer Science, Artificial Intelligence ensures domain relevance while capturing interdisciplinary contributions at the intersection of natural language processing, machine learning, and human-computer interaction.

2.4. Inclusion and Exclusion Criteria

Publications were included if they met the following criteria: publication type classified as Article, Proceedings Paper, or Review Article as these represent peer-reviewed research contributions; publication year between 2020 and 2025 inclusive; indexed in Web of Science Core Collection with complete bibliographic metadata; and classification within Computer Science or Computer Science, Artificial Intelligence WoS categories. Publications were excluded if they met any of the following criteria: document type classified as Editorial Material, Letter, Book Chapter, Book Review, Correction, or Meeting Abstract as these do not represent original peer-reviewed research contributions; identification as duplicate records through DOI and title matching to avoid distortion of frequency-based metrics; or incomplete essential bibliographic fields compromising analysis reliability. No duplicate records were detected in the final dataset. The search query yielded 4,890 publications meeting all inclusion criteria.

2.5. Data Collection and Processing

Bibliographic data were exported from WoS on November 27, 2025, in plain text format. The export included complete records containing bibliographic information such as title, authors, source, publication year, volume, issue, pagination, and DOI; author information including names, affiliations, and corresponding author address; abstract and keywords comprising author keywords and Keywords Plus; citation data including times cited and citation count per year; and complete reference lists for each publication. Exported data underwent preprocessing to ensure analytical quality through four automated procedures. Author name normalization disambiguated author name variants through algorithmic matching based on affiliation and co-author patterns. Institutional affiliation standardization harmonized institutional name variants using the Bibliometrix institution cleaning algorithm. Keyword harmonization standardized British and American spelling variants, performed lowercase conversion, and removed punctuation inconsistencies. Duplicate detection identified and removed duplicate records based on DOI and title matching, though no duplicates were found in the final dataset. All preprocessing utilized automated Bibliometrix functions to ensure reproducibility without manual intervention.

2.6. Analytical Framework and Tools

Analysis was conducted using a Bibliometric Analysis and Paper Creator Agent system developed as part of the first author's Master's thesis. The system operates on R programming language version 4.5.1 (R Core Team, 2025) with Bibliometrix package version 4.3.5 (Aria and Cuccurullo, 2017) and supporting packages including dplyr version 1.1.4 for data manipulation (Wickham et al., 2023), ggplot2 version 3.5.0 for visualization (Wickham, 2016), and xtable version 1.8-4 for LaTeX table generation (Dahl et al., 2019). All analyses were executed in R environment ensuring reproducibility through documented code and version-controlled packages. The analysis applied five bibliometric techniques. Performance analysis evaluated productivity and impact of research actors (authors, institutions, countries, sources) through publication counts, citation metrics, and h-index values. Science mapping techniques visualized intellectual structure through keyword co-occurrence networks revealing thematic clusters and their interrelationships. Co-citation analysis identified intellectual linkages between publications frequently cited together, indicating shared conceptual foundations. Bibliographic coupling analysis grouped publications sharing common references, revealing thematic communities. Collaboration network analysis examined co-authorship patterns at author, institution, and country levels, quantifying international collaboration rates and identifying key collaborative actors. Network visualizations employed Fruchterman-Reingold force-directed layout algorithm with node size representing frequency, edge thickness representing co-occurrence strength, and color coding representing community clusters identified through modularity optimization. Keyword co-occurrence networks were constructed using author keywords and Keywords Plus terms with minimum frequency threshold of 5 occurrences to filter noise while preserving meaningful patterns. Citation-based metrics include total citations, average citations per document, and normalized citation scores accounting for publication age and field citation density. All metrics follow standard bibliometric definitions as established in Aria and Cuccurullo (2017).

2.7. Visualization

Network visualizations were generated to illustrate thematic and collaboration structures. Keyword co-occurrence networks employed node size to represent keyword frequency, edge thickness to represent co-occurrence strength, and force-directed layout algorithm for node positioning. Word clouds provided visual representation of keyword frequency through font size scaling. Temporal plots depicted publication trends over time through line graphs. All visualizations maintain minimum 600 dpi resolution for publication quality with network visualizations employing Fruchterman-Reingold layout algorithm for optimal node positioning.

2.8. Study Limitations

This study acknowledges six methodological limitations. The exclusive use of WoS Core Collection excludes publications indexed only in Scopus, IEEE Xplore, ACL Anthology, or other databases, prioritizing methodological consistency over absolute comprehensiveness. The English-only restriction may underrepresent research contributions from non-Anglophone regions, particularly Chinese research published in domestic journals. The temporal scope beginning in 2020 excludes earlier foundational works in natural language processing and machine learning that preceded the modern prompt engineering era, though these works' influence is acknowledged qualitatively. The WoS category restriction may exclude relevant interdisciplinary work published outside selected categories, particularly publications in computational linguistics, human-computer interaction, or domain-specific application venues. Despite comprehensive term selection, the search query may not capture all terminology variations used across different research communities, conferences, and time periods, particularly newly emerging terms in this rapidly evolving field. Recently published papers from 2024-2025 have limited citation accumulation time, potentially underestimating their eventual impact compared to earlier publications. These limitations are inherent to bibliometric studies and do not compromise the validity of findings within the defined scope.

3. Results

3.1. Dataset Overview and Descriptive Statistics

The search query retrieved 4,890 publications from Web of Science Core Collection spanning 2020 to 2025. Table 1 presents the main descriptive statistics of the dataset. The corpus comprised publications from 1,538 sources authored by 16,052 researchers. The average document age was 0.939 years with an average of 10.1 citations per document. The annual growth rate reached 125.1%, indicating explosive expansion of the field. Author collaboration averaged 4.9 co-authors per document with 26.44% of publications involving international collaboration.

Table 1. Main descriptive statistics of the dataset

Description	Results
Timespan	2020:2025
Sources (Journals, Books, etc)	1538
Documents	4890
Annual Growth Rate %	125.1
Document Average Age	0.939
Average citations per doc	10.1
Average citations per year per doc	3.544
Documents per Author	0.305
Co-Authors per Doc	4.9
International co-authorships %	26.44
International co-authorships %	23.22

3.2. Temporal Publication Trends

Figure 1 presents the temporal evolution of prompt engineering publications from 2020 to 2025. The field exhibited dramatic exponential growth characterized by three distinct phases: emergence phase (2020-2021) with 99 publications, acceleration phase (2022-2023) witnessing expansion from 145 to 819 publications coinciding with ChatGPT's release, and explosion phase (2024-2025) reaching 2,093 publications in 2024 and 1,734 through November 2025.

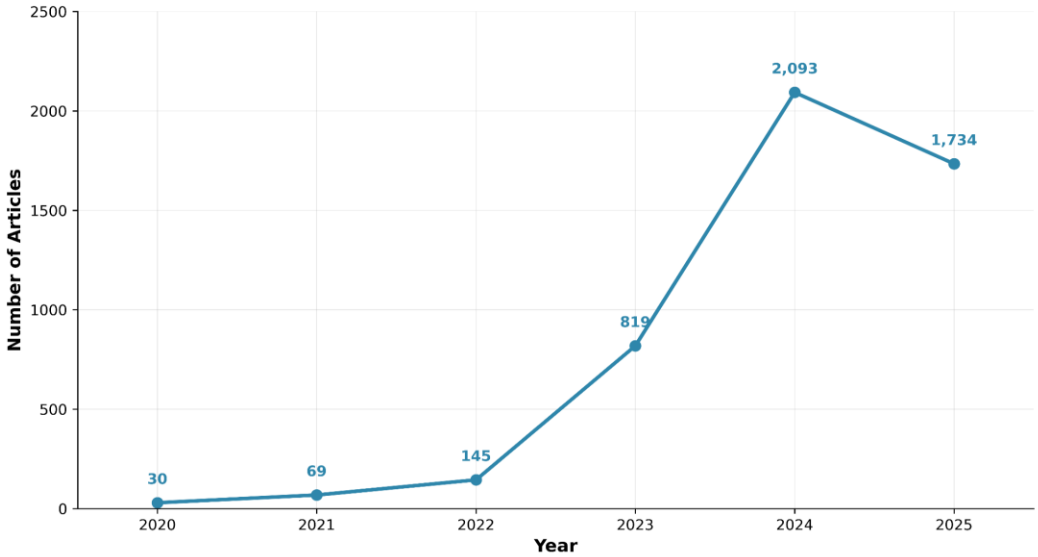


Figure 1. Annual publication output from 2000 to 2025.

3.3. Author Productivity and Impact

Table 2 presents the top 20 most productive authors measured by both absolute publication counts and fractionalized contributions. Zhang Y emerged as the most prolific author with 43 publications, followed by Liu Y (42 publications) and Wang Y (40 publications). When accounting for fractional contributions, Wang Y achieved the highest score (9.06), followed by Zhang Y (7.40) and Liu Y (7.26).

3.4. Institutional Productivity

Table 3 identifies the top 20 most productive institutions. The Chinese Academy of Sciences dominated institutional output with 345 publications, substantially exceeding the second-ranked University of California System with 171 publications. Chinese institutions occupied nine positions among the top 20, including Tsinghua University (149), Zhejiang University (133), and Peking University (125). Major technology corporations demonstrated substantial research contributions with Alphabet Inc. (125), Google Incorporated (112), and Microsoft (96) appearing among top producers, reflecting industry leadership in large language model development. United States institutions showed strong representation with University of California System, Carnegie Mellon University (85), University of Texas System (85), Harvard University (71), and Stanford University (69). Singapore maintained a notable presence through Nanyang Technological University (79) and National University of Singapore (74).

Table 2. Top 20 most productive authors

Authors	Articles	Authors	Articles Fractionalized
Zhang Y	43	Wang Y	9.06
Liu Y	42	Zhang Y	7.4
Wang Y	40	Liu Y	7.26
Chen X	33	Kim J	6.65
Kim J	31	Chen X	6.27
Wang X	30	Lee S	6.06
Li X	29	Anonymous	6
Zhang C	29	Liu H	5.69
Li J	28	Li J	5.32
Liu H	28	Wang X	5.21
Li Y	27	Zhang C	5.12
Chen L	26	Li L	5.04
Li L	26	Li Y	4.84
Lee S	25	Li X	4.84
Liu X	24	Chen L	4.8
Kim S	23	Wang L	4.71
Wang C	23	Wang H	4.64
Wang Xy	23	Wang Xy	4.35
Wang H	22	Wang C	4.29
Zhang J	22	Zhang J	4.25

Table 3. Top 20 most productive institutions

Institution	Frequency
Chinese Academy of Sciences	345
University of California System	171
Tsinghua University	149
Zhejiang University	133
Alphabet Inc.	125
Peking University	125
Google Incorporated	112
University of Chinese Academy of Sciences, Cas	106
Microsoft	96
Carnegie Mellon University	85
University of Texas System	85
Harbin Institute of Technology	81
University of Science and Technology of China, Cas	81
Nanyang Technological University	79
National University of Singapore	74
Harvard University	71
University of London	70
Stanford University	69
University of Illinois System	68
University System of Georgia	66

3.5. Country-Level Productivity and Collaboration

Table 4 presents country-level publication statistics including single-country publications (SCP), multiple-country publications (MCP), and international collaboration ratios. China produced the most publications with 1,584 documents representing 32.67% of the corpus. The United States ranked second with 1,162 publications (23.96%). India, Germany, and the United Kingdom completed the top five contributors.

Table 4. Top 20 countries by productivity and collaboration patterns

Country	Articles	Freq	SCP	MCP	MCP Ratio
China	1584	0.32667	1209	375	0.237
USA	1162	0.23964	940	222	0.191
India	199	0.04104	160	39	0.196
Germany	183	0.03774	139	44	0.24
United Kingdom	168	0.03465	96	72	0.429
Korea	153	0.03155	114	39	0.255
Italy	142	0.02928	111	31	0.218
Canada	130	0.02681	72	58	0.446
Japan	111	0.02289	83	28	0.252
Australia	102	0.02104	52	50	0.49
Singapore	89	0.01835	47	42	0.472
Spain	58	0.01196	45	13	0.224
France	49	0.01011	33	16	0.327
Switzerland	47	0.00969	33	14	0.298
Netherlands	39	0.00804	26	13	0.333
Brazil	35	0.00722	26	9	0.257
Saudi Arabia	34	0.00701	25	9	0.265
U Arab Emirates	29	0.00598	15	14	0.483
Israel	25	0.00516	15	10	0.4
Greece	24	0.00495	18	6	0.25

SCP: single country publications, MCP: multi-country publications, Freq: frequency ratio.

Geographic distribution analysis reveals distinct regional research ecosystems. East Asian countries (China, Korea, Japan, Singapore) collectively contributed 1,937 publications (39.6% of corpus), establishing the region as a major prompt engineering research hub. European nations demonstrated substantial but more distributed contributions across multiple countries, with no single European nation matching the output concentration of China or the United States. Middle Eastern countries (Saudi Arabia, United Arab Emirates, Israel) showed emerging research activity, while Latin American and African contributions remained limited. The Single Country Publications (SCP) metric identifies research conducted entirely within national boundaries. China produced 1,209 single-country publications, indicating robust domestic research capacity and potentially limited international research mobility. Similarly, the United States generated 940 domestic publications, reflecting its large-scale independent research infrastructure. Smaller nations like Singapore (47 SCP) and Switzerland (33 SCP) relied more heavily on international collaboration, consistent with their research strategies emphasizing global partnerships to overcome limited domestic researcher populations.

3.6. Country-Level Citation Impact

Table 5 presents citation impact across countries, revealing distinct patterns between productivity and influence. The United States achieved the highest total citations (25,176) and substantial average citations per article (21.67), demonstrating both volume and quality leadership. Singapore, despite modest output (89 articles), attained the highest average citations per article (37.37), indicating exceptional research impact. This pattern suggests that research quality does not necessarily correlate with quantity, with smaller nations producing highly influential work through focused research programs and international collaborations

3.7. Most Influential Publications

Table 6 presents the top 20 most cited publications in prompt engineering research. Ouyang L et al. (2022) achieved the highest citation count (4,859 citations, 1,214.8 citations per year). Wei et al. (2022) ranked second (4,123 citations, 1,030.8 per year). Liu PF (2023) achieved the highest normalized citation score (119.57 NTC) with 2,124 total citations.

3.8. Source Productivity

Table 7 presents the top 20 publication venues. IEEE Access dominated with 176 publications, followed by major natural language processing conference proceedings including ACL 2024 Findings (86 publications) and NeurIPS 2023 (76 publications).

3.9. Keyword Analysis and Thematic Structure

Table 8 presents the most frequent keywords. "Large language models" emerged as the dominant author keyword with 946 occurrences, followed by "prompt engineering" (733 occurrences) and "large language model" (474 occurrences).

Figure 2 presents a word cloud visualization of author keywords. "Large language models" and "prompt engineering" dominated the visual representation, consistent with frequency analysis. Terms such as "machine learning", "generative AI", "ChatGPT", "in-context learning", and "natural language processing" appeared prominently. Visualization confirmed the centrality of large language model concepts while highlighting the increasing importance of specific prompting techniques and application domains.

Table 5. Top 20 countries by citation impact

Country	Total Citations	Average Article Citations
USA	25176	21.666
China	8560	5.404
Singapore	3326	37.371
Japan	1595	14.369
Canada	1267	9.746
India	893	4.487
Germany	846	4.623
Korea	835	5.458
United Kingdom	752	4.476
Israel	745	29.8
Italy	704	4.958
Australia	698	6.843
U Arab Emirates	585	20.172
New Zealand	321	16.05
France	221	4.51
Greece	187	7.792
Spain	172	2.966
Denmark	156	15.6
Armenia	153	153
Macedonia	145	145

Table 6. Most Influential Publications

Paper	DOI	TC	TCperYear	NTC
Ouyang L, 2022	-	4859	1214.8	37.46
Wei JS, 2022	-	4123	1030.8	31.79
Liu PF, 2023	10.1145/3560815	2124	708	119.57
Zhou KY, 2022	10.1007/s11263-022-01653-1	1565	391.2	12.07
Kojima T, 2022	-	1310	327.5	10.1
Ruiz N, 2023	10.1109/CVPR52729.2023.02155	1085	361.7	61.08
Jia ML, 2022	10.1007/978-3-031-19827-4_41	1014	253.5	7.82
ZHOU KY, 2022, -A	10.1109/CVPR52688.2022.01631	957	239.2	7.38
Gao TY, 2021	-	765	153	20.51
Min BN, 2024	10.1145/3605943	633	316.5	122.3
Zamfrescu-Pereira JD, 2023	10.1145/3544548.3581388	470	156.7	26.46
Khattak MU, 2023	10.1109/CVPR52729.2023.01832	456	152	25.67
Reynolds L, 2021	10.1145/3411763.3451760	417	83.4	11.18
Gao P, 2024	10.1007/s11263-023-01891-x	384	192	74.19
Liu VV, 2022	10.1145/3491102.3501825	357	89.2	2.75
Liu JC, 2022	-	350	87.5	2.7
Liang J, 2023	10.1109/ICRA48891.2023.10160591	302	100.7	17
Tumanyan N, 2023	10.1109/CVPR52729.2023.00191	296	98.7	16.66
Wang ZF, 2022	10.1007/978-3-031-19809-0_36	256	64	1.97
Abid A, 2021	10.1145/3461702.3462624	227	45.4	6.09

TC: total citations, TCperYear: citations per year, NTC: normalized total citations.

Table 7. Top 20 most productive sources

Sources	Articles
IEEE ACCESS	176
FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: ACL 2024	86
ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 36 (NEURIPS 2023)	76
PROCEEDINGS OF THE 62ND ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS VOL 1: LONG PAPERS	68
ELECTRONICS	63
PROCEEDINGS OF THE 2024 CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES VOL 1: LONG PAPERS	53
FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS - EMNLP 2023	52
2024 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING EMNLP 2024	48
2023 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING EMNLP 2023	47
KNOWLEDGE-BASED SYSTEMS	44
2023 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP 2023)	41
FRONTIERS IN ARTIFICIAL INTELLIGENCE	41
FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (EMNLP 2023)	40
EXPERT SYSTEMS WITH APPLICATIONS	38
FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: NAACL 2024	38
2024 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR)	37
NEUROCOMPUTING	36
2024 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS IJCNN 2024	35
PROCEEDINGS OF THE 18TH INTERNATIONAL WORKSHOP ON SEMANTIC EVALUATION SEMEVAL-2024	32
PROCEEDINGS OF THE 33RD ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT CIKM 2024	30

Table 8. Top 20 most frequent author keywords

Author Keywords (DE)	Articles	Keywords-Plus (ID)	Articles
Large Language Models	946	Classification	51
Prompt Engineering	733	Model	35
Large Language Model	474	Network	31
Machine Learning	278	System	28
Generative AI	222	Networks	23
In-Context Learning	211	Artificial-Intelligence	22
Chatgpt	210	Algorithm	21
Natural Language Processing	206	Diagnosis	20
Artificial Intelligence	192	Models	20
LLM	165	Prediction	20
Deep Learning	163	Challenges	19
Prompt Learning	141	Language	19
LLMS	123	Design	18
Training	111	Framework	18
Few-Shot Learning	98	Neural-Networks	17
Large Language Models (LLMS)	93	Information	16
Data Models	66	Recognition	16
Fine-Tuning	65	Risk	15
Accuracy	62	Internet	14
Transformers	62	Management	14

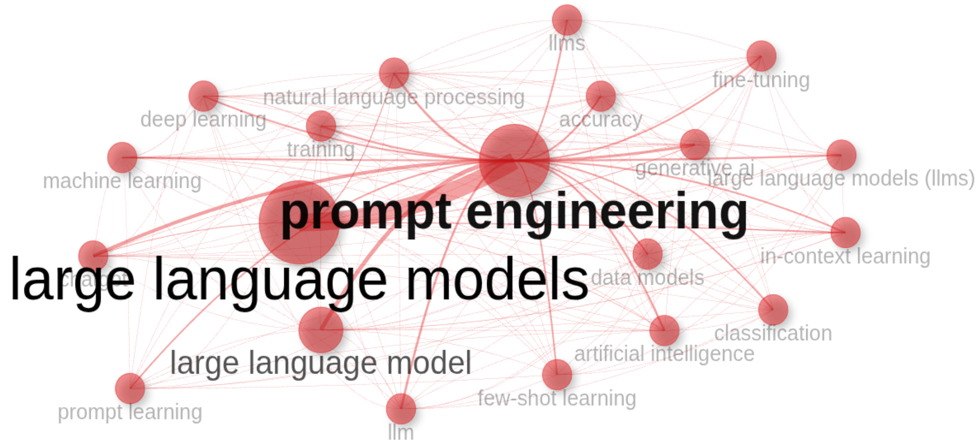


Figure 2. Keyword co-occurrence network based on keywords analysis.

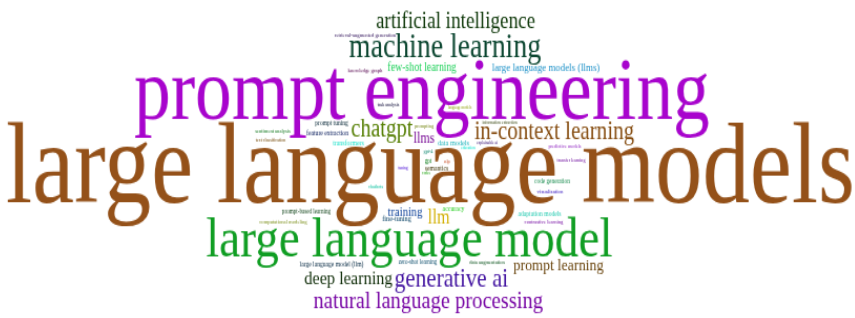


Figure 3. Word cloud of title-based keywords.

Figure 3 visualizes the keyword co-occurrence network based on author keywords. The network revealed three primary thematic clusters. The largest cluster centered on "large language models" and "prompt engineering" representing foundational concepts. A second cluster emphasized machine learning approaches including "deep learning", "neural networks", and "transformers". A third cluster focused on prompting techniques including "few-shot learning", "in-context learning", and "chain-of-thought prompting". The network structure indicated strong interconnections between theoretical prompting concepts and applied machine learning techniques.

4. Discussion and Conclusion

4.1. Publication Growth and Field Maturation

The 125.1% annual growth rate observed in prompt engineering research substantially exceeds typical expansion patterns in computer science subfields, indicating the domain's explosive emergence following widespread adoption of large language models. The temporal trajectory reveals three distinct developmental phases corresponding to major technological milestones. The emergence phase (2020-2021) produced limited output (99 publications) reflecting nascent exploration following GPT-3's introduction. The acceleration phase (2022-2023) witnessed exponential expansion from 145 to 819 publications, coinciding with ChatGPT's November 2022 release and subsequent mainstream recognition. The explosion phase (2024-2025) sustained unprecedented growth exceeding 2,000 annual publications, establishing prompt engineering as a permanent research domain within artificial intelligence. This growth pattern differs markedly from context engineering's 14.22% annual growth rate, reflecting prompt engineering's more concentrated temporal emergence within a compressed five-year window rather than gradual evolution over two decades. The field's rapid maturation compressed typical domain development cycles, with foundational methodologies, application domains, and theoretical frameworks emerging nearly simultaneously rather than through sequential accumulation. This accelerated development trajectory creates both opportunities and challenges for knowledge consolidation, quality assurance, and theoretical integration.

4.2. Geographic Distribution and Research Leadership

China's dominance with 32.67% of publications and the United States' 23.96% contribution collectively represent 56.63% of global prompt engineering research output. This concentration mirrors patterns observed in broader artificial intelligence research while exceeding context engineering's geographic distribution where China contributed 28.4% and USA 8.0%. The increased concentration in prompt engineering reflects these nations' strategic investments in large language model development and deployment, with major corporations

(OpenAI, Google, Meta, Baidu, Alibaba) driving both technological advancement and research priorities. The quality-quantity disparity between high-productivity and high-impact nations reveals distinct research strategies. Singapore's exceptional average citation impact (37.37 citations per article) from modest output demonstrates focused research programs emphasizing breakthrough contributions over volume production. Conversely, China's high productivity coupled with modest average citations (5.40) suggests rapid scaling potentially outpacing citation accumulation, consistent with national strategies prioritizing technological sovereignty and research output volume. International collaboration patterns reveal regional research ecosystems. Australia (49.0% MCP ratio), Canada (44.6%), and Singapore (47.2%) demonstrate high international integration, while China (23.7%) and USA (19.1%) maintain lower collaboration rates despite substantial productivity. These patterns reflect infrastructure capacity, policy incentives, and research culture differences, with smaller nations leveraging international partnerships to access resources and expertise while larger nations sustain independent research programs.

4.3. Author Productivity Patterns and Collaboration

The dominance of researchers with East Asian names (Zhang, Liu, Wang, Chen, Li, Kim, Lee) among top-ranked authors reflects China and South Korea's concentrated research activity in artificial intelligence and natural language processing. The relatively modest fractionalized scores even among leading authors (highest 9.06) indicate that prompt engineering research occurs predominantly through collaborative teams rather than individual efforts. This collaborative pattern aligns with the field's interdisciplinary character spanning machine learning, natural language processing, human-computer interaction, and domain-specific applications. The average of 4.9 co-authors per document exceeds context engineering's 3.83, suggesting increased collaborative complexity in prompt engineering research. This pattern may reflect the domain's requirement for diverse expertise including model architecture understanding, application domain knowledge, human factors considerations, and evaluation methodology design. The 26.44% international co-authorship rate, while substantial, remains below levels observed in mature interdisciplinary domains, suggesting continued potential for expanded global research networks.

4.4. Thematic Structure and Research Focus

Keyword analysis reveals prompt engineering's intellectual structure organized around three thematic clusters. The core cluster centers on large language models, prompt engineering methodologies, and specific techniques (few-shot learning, zero-shot learning, chain-of-thought prompting), representing the field's technical foundation. The secondary cluster emphasizes machine learning and natural language processing foundations (transformers, neural networks, training), situating prompt engineering within broader artificial intelligence contexts. The tertiary cluster focuses on application domains (classification, sentiment analysis, question answering), demonstrating practical deployment across diverse tasks. The dominance of "large language models" (946 occurrences) and "prompt engineering" (733 occurrences) as primary keywords confirms the field's clear conceptual identity despite rapid emergence. The prominence of methodological terms (in-context learning, few-shot learning, prompt learning) indicates research emphasis on fundamental techniques rather than purely application-driven investigation. However, the presence of model-specific keywords (ChatGPT with 210 occurrences) suggests potential fragmentation around particular platforms rather than unified theoretical frameworks. Trend analysis reveals temporal concentration in 2024-2025 for high-frequency terms, reflecting the field's recent maturation. Earlier topics (neural networks, IoT, edge computing) from 2021-2022 represent conceptual antecedents predating modern prompt engineering's crystallization. This pattern suggests the field synthesized existing concepts rather than emerging entirely de novo, building upon prior work in neural language models, few-shot learning, and human-AI interaction.

4.5. Source Concentration and Publication Venues

The dominance of conference proceedings over traditional journals reflects computer science publishing culture where conferences serve as primary venues for rapid research dissemination. IEEE Access's leadership (176 publications) combined with strong representation from ACL, EMNLP, and NeurIPS proceedings establishes natural language processing and machine learning conferences as principal dissemination channels. This pattern differs from context engineering where journal publications maintained a stronger presence, suggesting prompt engineering's faster development cycle favoring conference submission timelines. The prevalence of "Findings" tracks from major conferences (ACL, EMNLP, NAACL) indicates substantial submission volumes exceeding main conference acceptance capacity. These supplementary tracks provide additional publication opportunities while maintaining peer review standards, accommodating the field's explosive growth without compromising quality control. The presence of computer vision venues (CVPR 2024) reflects prompt engineering's expansion into multimodal models and vision-language applications, demonstrating disciplinary boundary crossing. Journal representation concentrates in applied artificial intelligence venues (Knowledge-Based Systems, Expert Systems with Applications, Neurocomputing) rather than pure theory publications, suggesting research emphasis on practical implementation alongside methodological innovation. Open-access journals (Electronics, Frontiers in Artificial Intelligence) provide rapid dissemination channels consistent with the field's fast-paced development and community preference for accessibility over prestige-based gatekeeping.

4.6. Citation Patterns and Knowledge Base

The most influential publications reveal both contemporary breakthrough contributions and recognition of foundational works predating prompt engineering's formalization. Ouyang et al. (2022) and Wei et al. (2022) achieved exceptional citation counts (4,859 and 4,123 respectively) reflecting their roles establishing instruction following and chain-of-thought prompting as core techniques. Liu PF (2023) attained the highest normalized citation score (119.57 NTC), indicating exceptional impact relative to publication age and field norms, likely representing a comprehensive synthesis that codified early knowledge. The temporal concentration of highly-cited papers in 2022-2023 marks the field's formative period when foundational frameworks were established. The inclusion of computer vision conference papers among top-cited works demonstrates prompt engineering's expansion beyond pure natural language processing into multimodal applications. The presence of human-computer interaction venues reflects recognition that effective prompting requires understanding human communication patterns alongside technical model capabilities. The citation patterns reveal rapid knowledge turnover characteristic of fast-developing technological domains. Recent papers from 2024 achieving substantial citations despite limited accumulation time indicate continued methodological innovation and community recognition of emerging contributions. This pattern suggests the field has not yet stabilized around canonical works, with ongoing competition among frameworks and techniques for establishing dominant paradigms.

4.7. Implications for Theory and Practice

For researchers, the bibliometric patterns identify productive institutions, influential authors, and high-impact publication venues that can inform collaboration strategies and manuscript targeting decisions. The thematic structure revealed through keyword analysis provides a conceptual map guiding literature review scope and identifying underexplored intersections between established clusters. The geographic distribution highlights regions where resources, expertise, and collaborative opportunities concentrate, informing international partnership development. For practitioners, the analysis reveals dominant techniques (few-shot learning, chain-of-thought prompting, in-context learning) that have achieved research consensus regarding effectiveness. The application-oriented keyword cluster identifies domains where prompt engineering has demonstrated practical value, guiding technology adoption decisions. The source concentration in accessible conferences and open-access journals facilitates practitioner engagement with cutting-edge research without institutional subscription barriers. The rapid growth trajectory and thematic breadth indicate prompt engineering's establishment as a distinct research domain rather than transient technological trend. However, the concentration around specific models (ChatGPT) and platforms suggests potential fragmentation risk if research emphasizes model-specific optimization over generalizable principles. Future work should prioritize theoretical frameworks transcending particular implementations while maintaining practical relevance for diverse large language model architectures.

4.8. Limitations

This study acknowledges several methodological limitations inherent to bibliometric approaches. The exclusive reliance on Web of Science Core Collection excludes publications indexed solely in Scopus, IEEE Xplore, ACL Anthology, or other specialized databases. While this choice ensures methodological consistency, it may underrepresent contributions from conference proceedings and preprint repositories that play particularly important roles in rapidly evolving technological domains. The English-language restriction potentially underrepresents research from non-Anglophone regions, particularly Chinese publications in domestic journals that may not appear in international databases despite substantial research activity. The temporal scope beginning in 2020 excludes earlier foundational work in natural language processing and machine learning that preceded modern prompt engineering's crystallization. While these antecedent contributions are acknowledged qualitatively, they do not appear in quantitative metrics, potentially obscuring intellectual lineages and conceptual continuities. The limited citation accumulation time for recent publications (2024-2025) may underestimate their eventual impact compared to earlier works with longer citation windows. The keyword analysis relies on author-provided terms and algorithmically-generated Keywords Plus, both subject to terminology inconsistencies, disciplinary variations, and strategic keyword selection for discoverability rather than accurate conceptual representation. Network analysis reveals co-occurrence patterns but cannot definitively establish causal relationships, conceptual dependencies, or qualitative assessment of thematic integration depth. Institutional affiliation data exhibited inconsistencies preventing reliable productivity analysis, reflecting metadata quality challenges in rapidly evolving fields with substantial preprint and conference proceeding publications.

4.9. Future Research Directions

The bibliometric patterns suggest several productive research directions. First, the thematic analysis revealed limited integration between technical methodology development and domain-specific application research. Future work should investigate how domain characteristics influence optimal prompting strategies, developing taxonomies that map application requirements to appropriate techniques. Second, the modest international collaboration rate indicates opportunities for establishing multinational research networks addressing prompt engineering challenges requiring diverse linguistic, cultural, and application perspectives. Third, the keyword analysis identified emerging topics including multimodal prompting, automated prompt optimization, and adversarial robustness that warrant systematic investigation (Hsieh and Lee, 2025). The integration of classical artificial intelligence techniques with neural language models represents promising directions for developing more robust and interpretable systems (Lizarraga et al., 2025). Fourth, the concentration around specific models suggests need for research emphasis on generalizable principles transcending particular implementations, developing theoretical frameworks applicable across diverse large language model architectures. Fifth, addressing challenges identified in recent literature including data bias, hallucination mitigation, few-shot generalization, and real-time responsiveness requires continued methodological innovation (Jovanovic and Voss, 2025). The creative variability introduced by automated prompt engineering raises questions about balancing innovation with adherence to task specifications and evaluation rubrics (Xue et al., 2025). Finally, the field would benefit from comprehensive synthesis works integrating findings across technical development, application deployment, and human factors research to advance theoretical coherence and practical impact.

Author Contributions

The percentages of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	A.K.	N.Ş.
C	50	50
D	50	50
S	50	50
DCP	50	50
DAI	50	50
L	50	50
W	50	50
CR	50	50
SR	50	50
PM	50	50
FA	50	50

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Cheong, M. (2025). ChatGPT's performance evaluation in spreadsheets modelling to inform assessments redesign. *Journal of Computer Assisted Learning*, *41*(3), e70035. https://doi.org/10.1111/jcal.70035

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML* (Version 1.8-4) [R package]. https://CRAN.R-project.org/package=xtable

Hsieh, P., & Lee, W. (2025). A particle swarm optimization-based approach coupled with large language models for prompt optimization. *Expert Systems*, *42*(6), e70049. https://doi.org/10.1111/exsy.70049

Jovanovic, M., & Voss, P. (2025). Towards incremental learning in large language models: A critical review. *Expert Systems*, *42*(10), e70127. https://doi.org/10.1111/exsy.70127

Lizarraga, A., Honig, E., & Wu, Y. N. (2025). From stochastic parrots to digital intelligence: The evolution of language models and their cognitive capabilities. *WIREs Computational Statistics*, *17*(3), e70035. https://doi.org/10.1002/wics.70035

Nguyen-Duc, A., Cabrero-Daniel, B., Przybylek, A., Arora, C., Khanna, D., Herda, T., Rafiq, U., Melegati, J., Guerra, E., Kemell, K.-K., Saari, M., Nguyen, C., Yousuf, O., Matthies, C., Rafiq, A., Becker, S., Muñoz, R. F., Chanin, R., Sales, A., . . . Abrahamsson, P. (2025). Generative artificial intelligence for software engineering—A research agenda. *Software: Practice and Experience*, *55*(11), 1806–1843. https://doi.org/10.1002/spe.70005

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, *9*(1), 12. https://doi.org/10.3390/publications9010012

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Sasson Lazovsky, G., Raz, T., & Kenett, Y. N. (2024). The art of creative inquiry—From question asking to prompt engineering. *The Journal of Creative Behavior*, *59*(1), e671. https://doi.org/10.1002/jocb.671

Vaira, L. A., Lechien, J. R., Abbate, V., Allevi, F., Audino, G., Beltramini, G. A., Bergonzani, M., Bernini, L., Boccaletti, R., Bonvin, N., Bossi, P., Bruschini, L., Calvanese, L., Canzi, P., Carobbio, A. L. C., Cazzador, D., Cocuzza, S., Colombo, G., D'Aguanno, V., . . . De Riu, G. (2023). Accuracy of ChatGPT-generated information on head and neck and oromaxillofacial surgery: A multicenter collaborative analysis. *Otolaryngology–Head and Neck Surgery*, *170*(6), 1492–1503. https://doi.org/10.1002/ohn.489

van Diessen, E., Amerongen, R. A., Zijlmans, M., & Otte, W. M. (2024). Potential merits and flaws of large language models in epilepsy care: A critical review. *Epilepsia*, *65*(4), 873–886. https://doi.org/10.1111/epi.17907

Wang, Y., Li, L., & Chen, L. (2025). Recent advances in finetuning multimodal large language models. *AI Magazine*, *46*(3), e70025. https://doi.org/10.1002/aaai.70025

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A grammar of data manipulation* (Version 1.1.4) [R package]. https://CRAN.R-project.org/package=dplyr

Xue, M., Liu, Y., Xiao, X., & Wilson, M. (2025). Automatic prompt engineering for automatic scoring. *Journal of Educational Measurement*. Advance online publication. https://doi.org/10.1111/jedm.70002

Black Sea Journal of Artificial Intelligence
doi: 10.5281/zenodo.18213838

Open Access Journal
e-ISSN: 3108-4281

Research Article
Volume 1 - Issue 2: 53-56 / December 2025

ARTIFICIAL INTELLIGENCE APPLICATIONS IN SCIENCE EDUCATION: OPPORTUNITIES, CHALLENGES, AND FUTURE PERSPECTIVES

Hatice DİLAVER[1]*, Kâmil Fatih DİLAVER[2]

[1]Niğde Ömer Halisdemir University, Department of Eurasia Studies, 51200, Niğde, Türkiye
[2]Niğde Ömer Halisdemir University, Faculty of Engineering, Department of Electric and Electronics, 51200 Niğde, Türkiye

Abstract: Artificial Intelligence (AI) has become a transformative driver of innovation in education, reshaping teaching and learning processes across science disciplines. AI-based instructional tools such as intelligent tutoring systems, virtual simulations, adaptive platforms, and predictive analytics enable students to engage with scientific phenomena through data-driven and interactive learning environments. These technologies support science inquiry, facilitate real-time feedback, and personalize instructional pathways by continuously analyzing student performance. This study aims to address this gap by developing conceptual and mathematical frameworks for AI integration in science classrooms, supported by simulation-based evaluations. Specifically, the research analyzes the effectiveness of intelligent tutoring systems, adaptive platforms, and virtual laboratories through model-driven feedback mechanisms and learning analytics. By linking AI prediction models with pedagogical outcomes, the study proposes a structured and scalable framework for responsible AI adoption in science education.

Keywords: Artificial intelligence, Science education, Educational technology, Personalized learning, Simulation

***Corresponding author:** Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye
E mail: haticedilaver509@gmail.com (H. DİLAVER)
Hatice DİLAVER https://orcid.org/0000-0002-4484-5297
Kâmil Fatih DİLAVER https://orcid.org/0000-0001-7557-9238
Received: November 18, 2025
Accepted: December 10, 2025
Published: December 15, 2025
Cite as: Dilaver, H., & Dilaver, K. F. (2025). Artificial intelligence applications in science education: Opportunities, challenges, and future perspectives. *Black Sea Journal of Artificial Intelligence*, *1*(2), 53–56.

1. Introduction

Recent studies highlight that the pedagogical impact of AI in science education extends beyond personalized learning, improving students' conceptual understanding, experimentation skills, and metacognitive awareness (Tatli and Ayas, 2013; Holmes et al., 2019). Moreover, adaptive AI models have been shown to reduce misconceptions in physics, chemistry, and biology by providing targeted interventions based on real-time analytics (Johnson, 2016; Luckin and Holmes, 2016). Despite these developments, empirical and model-based evaluations that integrate both pedagogical and technical dimensions remain limited in the literature.

Artificial Intelligence (AI) is increasingly embedded in science education, offering advanced computational capabilities that reshape instructional design and learning processes. This study investigates the pedagogical and technical dimensions of AI integration in science classrooms through an extensive literature review, data-driven analyses, and model-based evaluations. Findings indicate that AI supports improved conceptual understanding, adaptive learning pathways, automated assessment, and real-time analytic feedback, thereby strengthening evidence-based instructional decision making. Moreover, AI-supported simulations and intelligent tutoring systems enhance students' inquiry skills and facilitate the development of higher-order scientific reasoning. Despite these advantages, the study identifies critical constraints associated with data governance, algorithmic transparency, ethical compliance, and infrastructural disparities. Overall, the research provides a systematic and methodologically grounded framework for the responsible, scalable, and pedagogically aligned adoption of AI technologies in science education (Chen et al., 2020).

The integration of Artificial Intelligence (AI) in education is reshaping teaching and learning practices. In science education, AI offers innovative methods for knowledge acquisition, simulation experiments, and interactive learning environments. AI tools such as intelligent tutoring systems, virtual labs, and predictive analytics provide both teachers and students with personalized experiences, improving engagement and learning efficiency. Furthermore, AI enables data-driven decision-making for educators, enhancing instructional strategies and assessment methods.

Recent studies indicate a growing adoption of AI in science classrooms across different educational levels, highlighting its potential to transform pedagogical approaches (Albacete, 1999; Luan and Tsai, 2021). This paper examines AI applications in science education, their impact on learning outcomes, challenges encountered, and future perspectives for integrating AI-driven technologies effectively. Additionally, it proposes conceptual and mathematical models to optimize AI integration in science learning.

Theoretical Background of this issue:

AI in science education involves the integration of machine learning models, adaptive analytics, and simulation technologies to enhance instructional outcomes. Major applications include:

- Intelligent Tutoring Systems (ITS): Personalized instruction through adaptive task sequencing and feedback mechanisms.
- Adaptive Learning Platforms: Dynamic content recommendations based on learners' knowledge states.
- Virtual and Augmented Reality Simulations: Modeling complex scientific phenomena and providing inquiry-based laboratory experiences.
- Predictive Analytics for Student Performance: Identifying misconceptions and forecasting achievement trends using data-driven models.

Recent research emphasizes that science learning environments supported by AI provide enhanced inquiry skills, reduced cognitive load during complex simulations, and more precise scaffolding of problem-solving behaviors (Baker et al., 2016; Holmes et al., 2019). However, the effectiveness of these systems depends on how accurately predictive and adaptive models represent real classroom learning processes.

AI in science education involves the use of algorithms, machine learning models, and data-driven techniques to enhance instructional methods. Key applications include:

- Intelligent Tutoring Systems (ITS): Systems that adapt to the individual learner's pace, providing personalized guidance.
- Adaptive Learning Platforms: Platforms that tailor learning content based on student performance.
- Virtual and Augmented Reality Simulations: Immersive environments for conducting science experiments and visualizing complex phenomena.
- Predictive Analytics for Student Performance: Tools that forecast student success and identify learning gaps.

Recent AI research includes simulation-based learning, reinforcement learning for adaptive content, and predictive models to personalize instruction.

2. Materials and Methods

This study employs a literature review methodology, data-driven simulations, and model development, focusing on peer-reviewed articles published between 2020 and 2025. Data were collected from databases including Google Scholar, ScienceDirect, and ERIC. Quantitative and qualitative analyses were applied to identify trends, evaluate AI tools' effectiveness, and measure limitations in science classrooms.

Mathematical frameworks were used to construct predictive and adaptive learning models, supported by simulated classroom datasets. These models were designed to estimate student performance and optimize feedback mechanisms in AI-supported instructional environments.

To support the analytical models, simulation outputs were included to demonstrate AI-driven improvements across science domains. Table 1 presents the simulation-based performance outcomes for commonly used AI tools, including intelligent tutoring systems, adaptive platforms, and virtual laboratories. This study employs a literature review methodology, data-driven simulations, and model development, focusing on peer-reviewed articles published from 2020 to 2025. Data were collected from databases including Google Scholar, ScienceDirect, and ERIC. Quantitative and qualitative analyses were applied to identify trends,

2.1. Common AI Tools

AI tools' effectiveness, and limitations within science classrooms. Overall, simulation-intensive AI tools appear to yield the strongest improvements in science learning outcomes (Table 2).

2.2. Predictive Model for Student Performance

The mathematical explanation was given in Equation 1:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Where; TP= true positives, TN= true negatives, FP= false positives, FN= false negatives.

2.3. Student Competency Scoring Model

Student competency scoring model was given in Equation 2:

$$Score_{student} = \sum_{i=1}^{n} w_i \times R_i \quad (2)$$

here; w_i= weight of topic I, R_i= student performance in topic I, and n = number of topics covered.

2.4. AI Workflows

2.4.1. Intelligent tutoring system workflow

1. Student inputs answer
2. AI evaluates correctness
3. Adaptive feedback is generated
4. Learning path updated based on performance.

2.4.2. Adaptive learning feedback loop

1. Data collection on student activity
2. AI predicts knowledge gaps
3. Personalized recommendations provided

Continuous monitoring and update.

Table 1. Simulation-based improvement outcomes of AI applications in science education

AI Model	Domain	Improvement (%)	Notes
ITS	Physics	18	Personalized feedback
Adaptive Platform	Biology	15	Dynamic content adaptation
Virtual Labs	Chemistry	22	Simulated experiments

Table 2. Common AI tools

AI Tool	Purpose	Example
Intelligent Tutoring System	Personalized instruction	Carnegie Learning
Virtual Labs	Simulated experiments	Labster
Adaptive Platforms	Customized learning paths	Smart Sparrow
Predictive Analytics	Student performance prediction	Knewton

Table 3. Simulation outcomes

AI model	Domain	Improvement (%)	Notes
ITS	Physics	18	Personalized feedback
Adaptive platform	Biology	15	Dynamic content adaptation
Virtual labs	Chemistry	22	Simulated experiments

Table 4. Simulation result

Simulation Scenario	Number of Students	Pre-Test Avg	Post-Test Avg	Improvement %
Virtual Chemistry Lab	100	60	72	20
Adaptive Physics Platform	80	55	65	18
Biology ITS	90	62	71	15

3. Results and Discussion

AI tools in science education significantly improve learning outcomes by providing tailored content, immediate feedback, and interactive experiences. The simulation outcomes presented in Table 3 indicate that AI-supported learning environments significantly enhance students' scientific competencies. Among the evaluated tools, virtual laboratories yielded the highest improvement (22%) in chemistry learning outcomes, primarily due to their capacity to visualize molecular interactions and allow repeated experimentation without safety constraints (Tatli and Ayas, 2013). Intelligent tutoring systems demonstrated notable gains in physics learning (18%), reflecting the effectiveness of adaptive feedback in addressing individual misconceptions (Albacete, 1999). Adaptive platforms showed measurable yet comparatively lower improvement (15%), suggesting that personalized learning gains increase when real-time experimentation is combined with simulation-based interactivity.

These results collectively reveal that AI-driven science environments are most impactful when they provide interactive, adaptive, and experimental learning opportunities simultaneously. The findings align with recent evidence indicating that performance analytics and intelligent scaffolding substantially improve students' inquiry skills and conceptual reasoning in data-intensive simulations (Baker et al., 2016; Johnson, 2016). A virtual chemistry lab experiment simulated 100 students using adaptive learning; results showed an average improvement of 20% in concept understanding.

These models and simulations provide a framework to systematically integrate AI into science education, ensuring improved learning outcomes and enhanced teacher efficiency (Table 4).

The findings of this study indicate that AI-supported science instruction is most effective when adaptive feedback, data-driven assessment, and simulation-based experimentation are integrated within the same learning ecosystem. This interpretation aligns with international evidence demonstrating that multimodal AI systems reduce conceptual errors by offering personalized scaffolding and inquiry-based engagement (Zawacki-Richter et al., 2019; Holmes et al., 2019). In particular, intelligent tutoring systems and virtual laboratories enhance cognitive processing by transforming abstract concepts into interactive representations, while providing corrective feedback during experimentation.

Moreover, the simulation results in this study confirm emerging trends in AI-driven science learning, where predictive analytics play a critical role in diagnosing misconceptions and tailoring instruction (Baker et al., 2016). These results support claims that AI models grounded in performance analytics can help educators make evidence-based decisions, reducing ambiguity in instructional planning (Johnson, 2016). Accordingly, the effectiveness of AI in science classrooms depends on how accurately adaptive models are aligned with real learner behaviors, making the integration of data mining and pedagogy a central requirement for future science education research.

4. Conclusion

AI has transformative potential in science education, shaping new paradigms for personalized learning, data-driven instruction, and simulation-based inquiry. The findings of this study demonstrate that intelligent tutoring systems, adaptive learning platforms, and virtual laboratories collectively enhance students' scientific competencies by providing targeted feedback, reducing misconceptions, and offering repeatable experimentation opportunities without laboratory constraints. Simulation results further revealed that performance gains are maximized when adaptive analytics and interactive experimentation are integrated within a unified instructional framework.

These results highlight the importance of designing science learning environments that combine pedagogical modeling with technical AI capabilities. The integration of predictive analytics into instructional decision-making enables more accurate diagnosis of misconceptions and supports scalable feedback systems across science domains. Such advancements indicate that AI is not merely a supplementary tool but a strategic component in future science education models.

Future research should focus on developing hybrid AI ecosystems that align classroom pedagogy with real-time learning analytics, incorporate transparent and ethical data governance, and evaluate long-term learning impacts across diverse educational contexts. By linking adaptive algorithms with curriculum-based science learning, AI can contribute to more equitable, accessible, and empirically grounded learning environments that enhance both teacher practice and student achievement.AI-driven science education will be most impactful when pedagogy and analytics evolve together.

Emerging AI trends in science education include reinforcement learning models, natural language tutoring systems, and augmented reality labs.

Future research should explore long-term outcomes and cross-disciplinary applications of AI in science education.

Author Contributions

The percentages of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	H.D.	K.F.D.
C	50	50
D	50	50
S	50	50
DCP	50	50
DAI	50	50
L	50	50
W	50	50
CR	50	50
SR	50	50

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

Albacete, P. L. (1999). *An intelligent tutoring system for teaching fundamental physics concepts* [Doctoral dissertation, University of Pittsburgh]. ProQuest Dissertations and Theses Global.

Baker, R. S., Martin, T., & Rossi, L. M. (2016). Educational data mining and learning analytics. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 379–396). John Wiley & Sons. https://doi.org/10.1002/9781118956588.ch16

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264–75278.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Johnson, C. (2016). *Adaptive learning platforms: Creating a path for success*. ProQuest. https://search.proquest.com/openview/8f52ac336c6016db45b6de908f7721f6/1

Luan, H., & Tsai, C.-C. (2021). A review of using machine learning approaches for precision education. *Educational Technology & Society*, *24*(1), 250–266.

Luckin, R., & Holmes, W. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education. https://discovery.ucl.ac.uk/id/eprint/1475756

Tatli, Z., & Ayas, A. (2013). Effect of a virtual chemistry laboratory on students' achievement. *Educational Technology & Society*, *16*(1), 159–170.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Black Sea Journal of Artificial Intelligence
doi: 10.5281/zenodo.18213862

Open Access Journal
e-ISSN: 3108-4281

Research Article
Volume 1 - Issue 2: 57-61 / December 2025

BIBLIOMETRIC ANALYSIS ON USING ARTIFICIAL INTELLIGENCE IN DAIRY SCIENCE

Hasan ÖNDER[1]*

[1]*Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye*

Abstract: Interest in artificial intelligence, which began in the 1950s, has gradually increased and has caused the same interest to renew and increase day by day and spread many of the scientific fields including dairy science. In this study, making the bibliometric analysis of the usage of the artificial intelligence methods on the dairy sciences between the years of 2001 and 2025 was aimed. For the artificial intelligence methods used on the dairy sciences the annual percentage growth rate which was calculated as 9.05 showed that the artificial intelligence methods used on the dairy sciences will continue to increase. This increasing trend is also depend on the increasing studies on precision in livestock farming.

Keywords: Dairy, Bibliometric study, Precision in livestock farming, Artificial intelligence

***Corresponding author:** Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye
E mail: hasanonder@gmail.com (H. ÖNDER)
Hasan ÖNDER https://orcid.org/0000-0002-8404-8700

Received: August 26, 2025
Accepted: December 12, 2025
Published: December 15, 2025
Cite as: Önder, H. (2025). Bibliometric analysis on using artificial intelligence in dairy science. *Black Sea Journal of Artificial Intelligence*, *1*(2), 57–61.

1. Introduction

The act of learning, which has existed since the dawn of humanity, has revealed a tendency to imitate nature, and natural imitation has become a means of acquiring much knowledge. The work done by scientists on this ability to imitate is one of the reasons why the scientific world has advanced in this field. In this sense, the book written by Arf (1959) called "Can a machine think and how can it think" on the production and functional structure, considering the working principles of the human brain, and many studies have been carried out, the foundations of the soldiers made in the field of artificial intelligence today have been formed (Fetzer, 1990; Chatterjee and Laudato, 1995). The main purpose is to solve problems by modeling the intelligence structure of human beings and by using a set of algorithms and artificial intelligence on computers and machines controlled by computers (Baş, 2006).

This interest in artificial intelligence, which began in the 1950s, continues to grow day by day. Thus, it has provided the emergence of a field different from the classical algorithmic calculation methods of today's computers (Çanakçı and Hosoz, 2006). It is possible to list the areas where artificial intelligence technology is widely used as genetic algorithm, expert systems, artificial neural networks and fuzzy logic (Elmas, 2003).

Increasing the production and quality of animal derived foods, which play a critical role in human nutrition, requires improvements in both the environmental conditions and genetic structures of farm animals (Kolenda et al., 2025). Dairy farming is one of the oldest areas of animal production and is widespread throughout the world (Dagtekin et al., 2023). According to the farm's milk or meat production purpose, it is necessary to raise animals of qualified breeds with quality feeds in a modern way under hygienic care feeding conditions. In intensive animal breeding, which is done by providing the environmental conditions needed by animals with good genetic characteristics and feeding them with quality feed sources to obtain higher yields (Ermetin and Erkan Can, 2023; Okuyucu et al., 2023). Important environmental factors that do not have a genetic effect on milk yield are lactation period, calving age, calving season and calving stage (Elahi Torshizi, 2016). In addition, the effects of environmental factors such as calving year, calving interval, calving season, number of births, herd and milking frequency on milk yield should be investigated and therefore, to study these relationships, machine learning methodologies from traditional statistical methods have been increasingly adopted (Çanga Boğa et al., 2024).

To understand the usage of the artificial intelligence methods on the dairy sciences, the bibliometric analysis seems to be a valuable tool, which is used many areas of science (Özlü, 2022; Önder and Tırınk, 2022; Önder, 2025). Bibliometrics, which refers to the application of mathematical and statistical methods to analyze scientific publications on a specific topic, serves to provide quantitative information on bibliographic properties, such as authors, journals, citation scores, and countries of distribution. Many different techniques such as citation analysis, co-citation analysis, and bibliometric matching analysis, co-asset analysis and bibliometric mapping can be used together in bibliometric analysis methods (Özlü, 2022).

In this study, bibliometric analysis of the usage of the artificial intelligence methods on the dairy sciences was aimed to evaluate the literature related to artificial intelligence methods on the dairy sciences since it started to be worked was examined to understand the evaluation and spreading of it.

2. Materials and Methods

In this study, studies related to artificial intelligence methods used on the dairy sciences between the years 2001-2025 were taken into account. In this context, the "artificial intelligence", "machine learning", and "dairy" expression were used for searching on the Web of Science (WoS) database. The bibliographic information under the heading "artificial intelligence", "machine learning", and "dairy" of 49 studies from 2001 to 2025 was used as material

In this study, the bibliometric analysis for artificial intelligence in dairy science was performed with R software (R Core Team, 2020). For this aim, the bibliometrix package were used (Aria and Cuccurullo, 2017). The bibliographic data were obtained from the WoS system in Plain text format. Further, the data was changed as the data frame by using "convert2pdf" function. The biblioAnalysis function was used for performing the bibliometric analysis.

3. Results and Discussion

The most productive authors were given in Figure 1, where Goli (4 articles) can be recognized as the first productive author over time even if Goli started the publication in the year of 2019.

In the year of 2001 and 2002 only one articles has been published. Until 2018 there was no more publication about artificial intelligence methods on the dairy sciences. From the year of 2018 the number of articles proceed on the use of artificial intelligence methods on the dairy sciences started to increase. In 2024 the number of articles reached to 16 which is eight time more than 2018. Interestingly, the first article in 2001 has been cited 26 times. The maximum number of average total citations per year was observed as 65 in 2019 (Figure 2). From the year of 2018 the growth of number of article has been increased, the main case of this should be recognizing the artificial intelligence as a precision in livestock farming. The annual percentage growth rate was calculated as 9.05, which is a great growth rate. The document average age was observed as 3.08, average citations per document was 14.24, and average citations per year per document was 3.458.

The top manuscripts per citations was recognized to belong to Alireza Goli from University of Tehran with the article "Hybrid artificial intelligence and robust optimization for a multi-objective product portfolio problem Case study: The dairy products industry" which achieved to take 113 citations (Table 1).

The most article producer country was determined as USA and the follower was Spain. These top ten producer countries were produce the 83.33% of the total production. The rate of publications produces by single country was calculated as 87.50%. The rate of publications produces by multiple country was the highest for the USA (Table 2).

As Iran took the first place for the total citations because 33.28% of total production of top 10 countries was belong to Iran. Even the Zimbabwe, Poland, and Hungary not listed in top ten countries, they took place in the most cited publisher countries. The case of this situation could be the publications of Golia' publications (Table 1 and Table 3). The most relevant journal was determined as Journal of Dairy Science and the follower was Applied Sciences-Basel. In the first five ranked journals only one of them were the journals that it' aim of scope is dairy science (Table 4).

When the collaboration matrix (Figure 3) was examined, it was understood that some working groups existed on the studies of the artificial intelligence methods used on the dairy sciences. The co-citations couldn't be centeralized (Figure 4). The conceptual structure map (Figure 5) showed that "body temperature", "ruminal acidosis", and "heat stress" were generally used together.

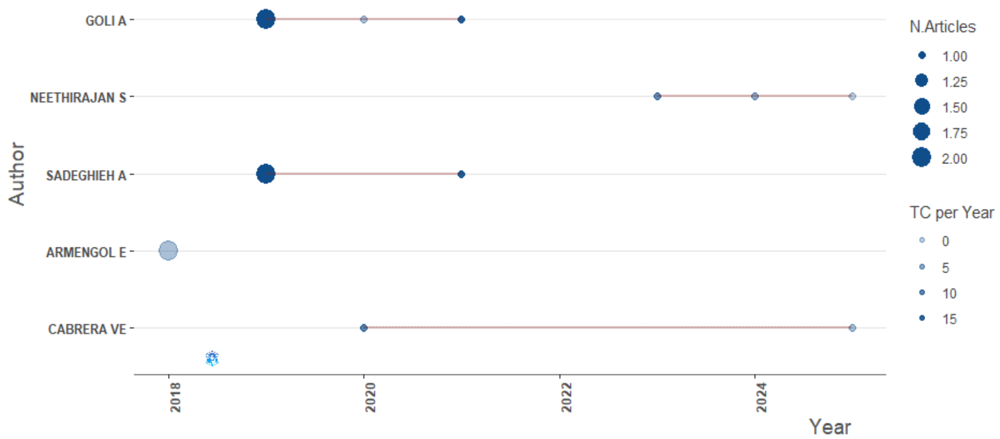


Figure 1. Authors' production over time (the first 20).

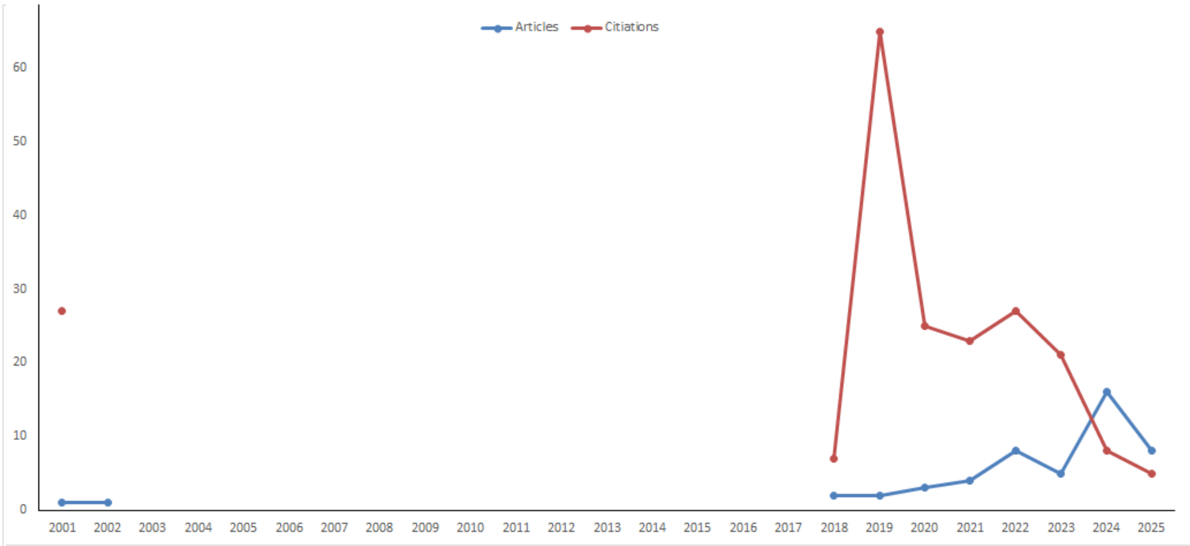


Figure 2. Number of articles and citations published from 2001 to 2025.

Table 1. Top manuscripts per citations

Paper	DOI	TC	TC/Y
GOLI A. 2019	10.1016/j.cie.2019.106090	113	16.14
BAO J. 2022	10.1016/j.jclepro.2021.129956	102	25.50
GOLI A. 2021	10.1080/0954898X.2020.1849841	85	17.00
CABRERA VE. 2020	10.3168/jds.2019-17145	63	10.50
ADDANKI M. 2022	10.1016/j.afres.2022.100126	49	12.25
KUTYAURIPO I. 2023	10.1016/j.jafr.2023.100502	46	15.33
NEETHIRAJAN S. 2023	10.3390/s23167045	28	9.33
GOYACHE F. 2001	10.1017/S1357729800058045	26	1.04
DE VRIES A. 2023	10.15232/aas.2022-02345	21	7.00
GOLI A. 2019	10.9781/ijimai.2019.03.003	20	2.86

Table 2. Corresponding author's countries

	Country	Articles	SCP	MCP	MCP Ratio
1	USA	7	5	2	0.286
2	Spain	6	5	1	0.167
3	Brazil	5	4	1	0.200
4	Canada	5	5	0	0.000
5	China	4	4	0	0.000
6	India	4	3	1	0.250
7	Iran	4	4	0	0.000
8	Japan	2	2	0	0.000
9	UK	2	2	0	0.000
10	Ecuador	1	1	0	0.000

SCP= Single Country Publications, MCP= Multiple Country Publications.

Table 3. Total citations per country

	Country	Total Citations	Average Article Citations
1	Iran	227	56.75
2	China	123	30.75
3	USA	102	14.57
4	India	70	17.50
5	Zimbabwe	46	46.00
6	Spain	41	6.83
7	Canada	37	7.40
8	Poland	18	18.00
9	Hungary	13	13.00
10	Brazil	5	1.00

Table 4. The most relevant sources for the publications about the artificial intelligence methods used on the dairy sciences

Rank	Journal title	Number of articles
1	JOURNAL OF DAIRY SCIENCE	4
2	APPLIED SCIENCES-BASEL	2
3	JOURNAL OF AGRICULTURE AND FOOD RESEARCH	2
4	JOURNAL OF FOOD COMPOSITION AND ANALYSIS	2
5	SENSORS	2

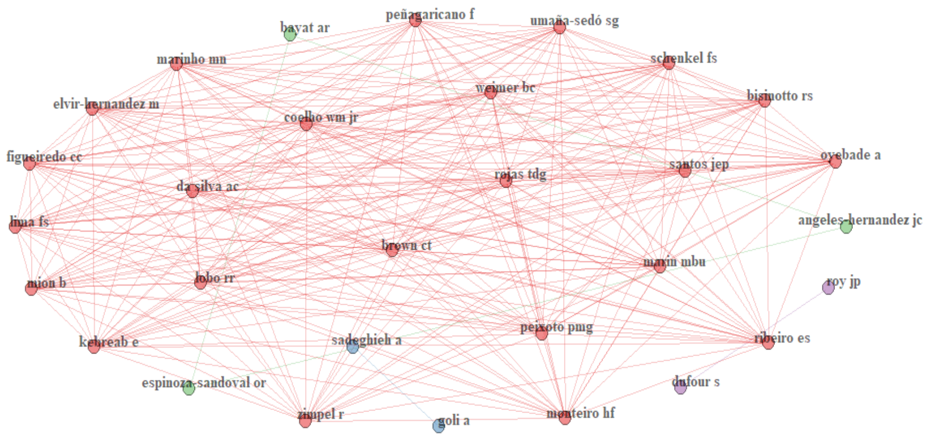


Figure 3. Collaboration matrix.

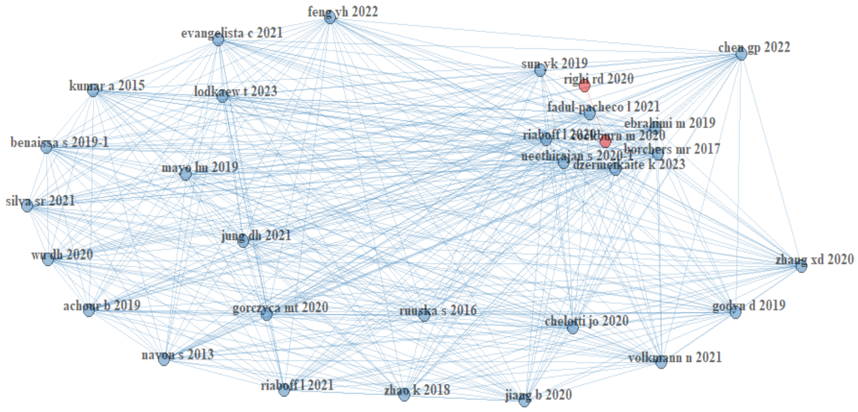


Figure 4. Co-citations.

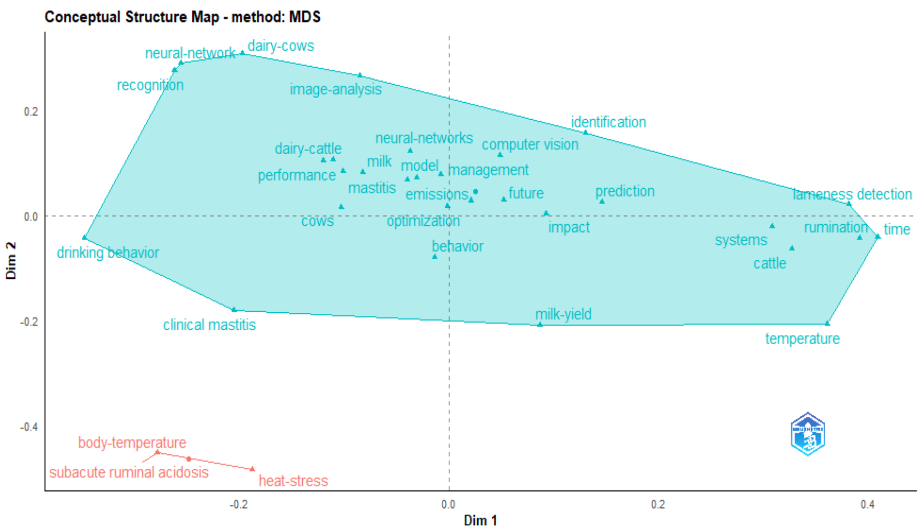


Figure 5. Conceptual structure map.

4. Conclusion

For the artificial intelligence methods used on the dairy sciences the annual percentage growth rate which was calculated as 9.05 showed that the artificial intelligence methods used on the dairy sciences will continue to increase. This increasing trend is also depend on the increasing studies on precision in livestock farming.

Author Contributions

The percentages of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	H.Ö.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

Acknowledgements

This study was presented as an oral presentation in the proceeding book of the International Congress on Domestic Animal Breeding, Genetics and Husbandry-2025 (ICABGEH-25), was held on September 02 - 04, 2025, at the University of Szeged, Faculty of Agriculture, Hódmezővásárhely - Hungary.

References

Arf, C. (1959). *Makine düşünebilir mi ve nasıl düşünebilir?* Atatürk Üniversitesi Yayınları.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Baş, N. (2006). *Artificial neural networks approach and an application* [Yüksek lisans tezi, Mimar Sinan Güzel Sanatlar Üniversitesi]. Ulusal Tez Merkezi.

Çanakçı, M., & Hosoz, M. (2006). Energy and exergy analyses of a diesel engine fuelled with various biodiesels. *Energy Sources, Part B: Economics, Planning, and Policy*, *1*(4), 379–394. https://doi.org/10.1080/15567240500400713

Çanakçı Boğa, D., Boğa, M., & Bulut, M. (2024). Forecasting seasonal milk production using MARS algorithm for multiple continuous responses in Holstein dairy cattles. *Black Sea Journal of Agriculture*, *7*(3), 203–214. https://doi.org/10.47115/bsagriculture.1383832

Chatterjee, S., & Laudato, M. (1995). Gender and performance in athletics. *Social Biology*, *42*(3-4), 397–412. https://doi.org/10.1080/19485565.1995.9988907

Dagtekin, M., Kic, P., Demirel, B., & Gurdil, G. A. K. (2023). Analyses of operational and economic conditions in selected dairy farms. *Black Sea Journal of Agriculture*, *6*(1), 32–41. https://doi.org/10.47115/bsagriculture.1157832

Elahi Torshizi, M. (2016). Effects of season and age at first calving on genetic and phenotypic characteristics of lactation curve parameters in Holstein cows. *Journal of Animal Science and Technology*, *58*(1), 8. https://doi.org/10.1186/s40781-016-0089-1

Elmas, Ç. (2003). *Bulanık mantık denetleyicileri*. Seçkin Yayıncılık.

Ermetin, O., & Erkan Can, M. (2023). Determination of structural characteristics of large-scale dairy farms: Example of Yozgat province. *Black Sea Journal of Agriculture*, *6*(5), 444–451. https://doi.org/10.47115/bsagriculture.1288564

Fetzer, J. H. (1990). What is artificial intelligence? In J. H. Fetzer (Ed.), *Artificial intelligence: Its scope and limits* (ss. 3–27). Springer. https://doi.org/10.1007/978-94-009-1900-6_1

Kolenda, M., Önder, H., Piwczyński, D., Kurnaz, B., & Sitkowska, B. (2025). Comparative analysis of Bayesian methods and GBLUP for genomic evaluations in dairy cattle. *Folia Biologica (Kraków)*, *73*(1), 21–29. https://doi.org/10.3409/fb_73-1.03

Okuyucu, İ. C., Akdağ, A., Erdem, H., & Ocak, N. (2023). Physiological and physical responses of dairy cattle to heat stress. *Black Sea Journal of Agriculture*, *6*(1), 95–103. https://doi.org/10.47115/bsagriculture.1177117

Önder, H., & Tırınk, C. (2022). Bibliometric analysis for genomic selection studies in animal science. *Journal of the Institute of Science and Technology*, *12*(3), 1849–1856. https://doi.org/10.21597/jist.1133397

Önder, O. K. (2025). Analysis of published manuscripts on Napoleon Bonaparte. *Black Sea Journal of Public and Social Science*, *8*(1), 1–6. https://doi.org/10.52704/bssocialscience.1541621

Özlü, A. (2022). Bibliometric analysis of publications on pulmonary rehabilitation. *Black Sea Journal of Health Science*, *5*(2), 219–225. https://doi.org/10.19127/bshealthscience.1032380

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Black Sea Journal of Artificial Intelligence
doi: 10.5281/zenodo.18213880

Open Access Journal
E-ISSN: 3108 – 4281

Araştırma Makalesi (Research Article)
Cilt 1 - Sayı 2: 62-67 / Aralık 2025
(Volume 1 - Issue 2: 62-67 / December 2025)

YÜRÜME BOZUKLUĞU OLAN HASTALARDA ETKİLENEN ANATOMİK SİSTEMİN YAPAY ZEKA İLE TESPİTİ

Kübra ASLAN KOCA[1]*, Murat TERZİ[2]

[1]*Ondokuz Mayıs University, School of Information and Technology, Department of Software, Application Development and Analysis, 55200, Samsun, Türkiye*
[1]*Ondokuz Mayıs University, Faculty of Medicine, Department of Neurology, 55200, Samsun, Türkiye*

Özet: Yürüme bozuklukları, santral ve periferik sinir sisteminden ekstrapiramidal yapılara, kas-iskelet sistemine ve eklem-kemik patolojilerine uzanan geniş bir anatomik spektrumun etkilenmesi sonucu ortaya çıkan multidisipliner bir klinik sorundur. Bu çalışma, yüksek çözünürlüklü klinik yürüme videolarının 3B Evrişimsel Sinir Ağı (3D-CNN) mimarisi ile analiz edilerek etkilenen anatomik sistemin otomatik olarak sınıflandırılmasını amaçlamaktadır. Ondokuz Mayıs Üniversitesi Nöroloji Kliniği'nde toplanan video tabanlı veri seti, santral sinir sistemi, periferik sinir sistemi, ekstrapiramidal sistem ve kas-kemik-eklem patolojileri olmak üzere dört kategori altında etiketlenmiştir. Videolar OpenCV tabanlı bir ön-işleme hattından geçirilmiş, mekansal ve zamansal standardizasyon uygulanmış ve model %80 eğitim – %20 test ayrımı ile optimize edilmiştir. 3D-CNN modeli 50 epoch sonunda %96,20 doğruluk düzeyine ulaşarak klinik genelleme açısından yüksek performans göstermiştir. Bulgular, video-temelli derin öğrenme yaklaşımlarının nörolojik ve hareket sistemi hastalıklarına ilişkin karar destek mekanizmalarında erken tanı, risk azaltma ve kişiselleştirilmiş tedavi planlamasını güçlendirebileceğini göstermektedir. Literatürde çoğunlukla tek hastalık odaklı modellere kıyasla, bu çalışma çoklu sınıf ayrımıyla kapsamlı bir klinik sınıflandırma sunmakta ve yapay zeka temelli gait analitiğinin gerçek yaşam uygulamalarına entegrasyonuna stratejik bir temel oluşturmaktadır.

Anahtar kelimeler: Yürüme bozuklukları, 3D-CNN, Yapay zeka, Video tabanlı analiz, Nörolojik sınıflandırma

Detection of Affected Anatomical Systems in Patients with Gait Disorders Using Artificial Intelligence

Abstract: Gait disorders are a multidisciplinary clinical problem resulting from the involvement of a wide anatomical spectrum ranging from the central and peripheral nervous systems to extrapyramidal structures, the musculoskeletal system, and joint-bone pathologies. This study aims to automatically classify the affected anatomical system by analyzing high-resolution clinical gait videos using a 3D Convolutional Neural Network (3D-CNN) architecture. The video-based dataset collected at the Neurology Clinic of Ondokuz Mayıs University was labeled under four categories: central nervous system, peripheral nervous system, extrapyramidal system, and musculoskeletal-joint pathologies. The videos were processed through an OpenCV-based pre-processing pipeline, spatial and temporal standardization was applied, and the model was optimized with an 80% training – 20% test split. The 3D-CNN model achieved an accuracy level of 96.20% after 50 epochs, demonstrating high performance in terms of clinical generalization. The findings demonstrate that video-based deep learning approaches can enhance early diagnosis, risk reduction, and personalized treatment planning in decision support mechanisms related to neurological and musculoskeletal disorders. Unlike the mostly single-disease focused models in the literature, this study offers a comprehensive clinical classification with multiple class distinctions and provides a strategic foundation for the integration of AI-based gait analytics into real-world applications.

Keywords: Gait disorders, 3D-CNN, Artificial intelligence, Video-based analysis, Neurological classification

***Sorumlu yazar (Corresponding author):** Ondokuz Mayıs University, School of Information and Technology, Department of Software, Application Development and Analysis, 55200, Samsun, Türkiye
E mail: kubra.aslankoca@omu.edu.tr (K. ASLAN KOCA)
Kübra ASLAN KOCA https://orcid.org/0000-0002-2828-3239
Murat TERZİ https://orcid.org/0000-0002-3586-9115

Gönderi: 05 Kasım 2025 **Received:** November 05, 2025
Kabul: 15 Aralık 2025 **Accepted:** December 15, 2025
Yayınlanma: 15 Aralık 2025 **Published:** December 15, 2025

Cite as: Aslan Koca, K., & Terzi, M. (2025). Detection of affected anatomical systems in patients with gait disorders using artificial intelligence. *Black Sea Journal of Artificial Intelligence*, *1*(2), 62–67.

1. Giriş

Yürümenin fizyolojik bütünlüğü, merkezi ve periferik sinir sisteminden kas-iskelet yapısına uzanan çok katmanlı bir nöromüsküler organizasyonun senkronizasyonunu gerektirmektedir (Vun vd., 2024). 21. yüzyılda hızlanan nöroteknolojik inovasyonlar ve yapay zeka temelli hesaplama yaklaşımları, hareket bozukluklarının analizi ve sınıflandırılmasına ilişkin klinik değerlendirme paradigmalarını yeniden tanımlamaya başlamıştır. Yürüme, motor planlama, ritmik koordinasyon, denge ve hız kontrolü gibi birbirini tamamlayan fonksiyonların çıktısı olduğundan, bu yapıların herhangi birindeki anatomik veya fonksiyonel düzensizlik, gait paterninde ölçülebilir sapmalar üretmektedir (Liang, 2024). Bu durum, yalnızca nörolojik hastalıkların değil kas-iskelet sistemi patolojileri, ortopedik deformiteler ve metabolik süreçlerle ilişkili çoklu etiyolojilerin klinik pratiğe yansıyan ortak sonuçlarından biridir.

Klasik yaklaşımlarda yürüme bozukluklarının değerlendirilmesi, gözleme dayalı subjektif analizlere, klinisyenin deneyimine ve sınırlı kantitatif parametrelerin yorumuna bağlıdır. Bu yaklaşım, tanısal belirsizlik, geciken müdahale kararları ve kişiselleştirilmiş rehabilitasyon planlarının eksik kalması gibi operasyonel kırılganlıklar üretmektedir. Güncel çalışmalar, santral sinir sistemi lezyonlarından ekstrapiramidal disfonksiyona kadar geniş bir spektrumda, gait dinamiklerinin erken tanısal biyobelirteç niteliği taşıyabileceğini göstermektedir (Ali vd., 2025). Buna rağmen mevcut çalışmalar çoğunlukla tek hastalık odağında sınırlı veri setleriyle ilerlemekte ve çoklu anatomik sistemleri aynı model altında ayrıştıran, klinik genellenebilirliği yüksek yapılar henüz yeterince geliştirilememektedir.

Yapay zeka, özellikle de derin öğrenme tabanlı video analizi, karmaşık motor örüntülerin otomatik çıkarımı, çerçeve bazlı uzamsal-zamansal özelliklerin modellenmesi ve yüksek doğrulukla sınıflandırılması için güçlü bir mimari sunmaktadır. 3B evrişimsel sinir ağları (3D-CNN), video temelli gait verisini mekansal ve zamansal boyutlarıyla birlikte yorumlayarak, klasik gözlemsel yaklaşımların ötesine geçen kantitatif, tekrarlanabilir ve objektif bir karar destek yetkinliği sağlamaktadır.

Bu çalışma, klinik ortamda kaydedilmiş yürüme videolarından elde edilen yüksek çözünürlüklü görsel verileri işleyerek santral sinir sistemi, periferik sinir sistemi, ekstrapiramidal sistem ve kas-kemik-eklem yapılarının etkilenmesine bağlı yürüme bozukluklarını aynı model üzerinde ayrıştırmayı hedeflemektedir. Amaç, erken tanıyı hızlandıran, tedavi planlamasında belirsizliği azaltan ve hastaya özel müdahaleleri destekleyen bir yapay zeka mimarisi geliştirmektir. Modelin klinik entegrasyona aday bir yapıda olması düşme riski yönetimi, rehabilitasyon çıktılarının izlenmesi ve uzun dönem nöromüsküler hastalık takibinde yeni bir operasyonel standart yaratma potansiyeli sunmaktadır. Bu perspektifte çalışma, nörolojik ve kas iskelet temelli gait bozukluklarının değerlendirilmesine yapay zeka aracılığıyla sistematik, çok sınıflı ve yüksek doğruluk oranlı bir yaklaşım getirerek literatürdeki metodolojik boşluğu doldurmayı ve klinisyenler, araştırmacılar ve karar vericiler için uygulanabilir bir teknolojik yol haritası üretmeyi amaçlamaktadır.

1.1. Literatür

Yürüme bozuklukları, santral sinir sistemi, periferik sinir sistemi, kas-iskelet yapıları ve ekstrapiramidal ağlar gibi farklı anatomik sistemlerin etkilenmesine bağlı olarak heterojen klinik paternler sergilemektedir. Son yıllarda yapay zeka tabanlı gait analiz yaklaşımları, bu farklı patofizyolojik kaynakları objektif ve kantitatif biçimde ayırt edebilen güçlü karar destek sistemleri olarak öne çıkmaktadır. Özellikle markerless motion capture ve video tabanlı derin öğrenme yöntemleri, klinik gözleme dayalı değerlendirmelerin ötesine geçerek, etkilenen anatomik sistemlerin dolaylı olarak tanımlanmasına olanak sağlamaktadır (Iseki vd., 2023; Vun vd., 2024). Derin öğrenme modelleri, yürüyüş sırasında ortaya çıkan karmaşık motor örüntüleri uzamsal ve zamansal boyutlarıyla birlikte ele alarak, santral ve periferik kaynaklı bozukluklar arasında ayrım yapabilmektedir. Parkinson hastalığında gerçekleştirilen çalışmalar, uzamsal-zamansal sinir ağları ve zaman serisine dönüştürülmüş video verileri aracılığıyla alt ekstremite, gövde ve üst ekstremite segmentlerinin etkilenme düzeylerinin ayırt edilebildiğini göstermektedir (Liang, 2024; Edison vd., 2025). Bu yaklaşımlar, ekstrapiramidal disfonksiyonun gait üzerindeki karakteristik etkilerini nicel olarak ortaya koymaktadır. Giyilebilir sensörler ve plantar basınç sistemleri ile entegre edilen derin öğrenme modelleri, hemiplejik, Parkinsonian ve sensori-ataxik yürüyüş paternlerini yüksek doğrulukla sınıflandırarak, yürüme bozukluğunun köken aldığı anatomik sistemi dolaylı biçimde tanımlayabilmektedir (Bradski vd., 2008). Benzer şekilde, spatiotemporal gait parametrelerine dayalı makine öğrenmesi yaklaşımları, Parkinson, Huntington ve ALS gibi nörodejeneratif hastalıklarda santral sinir sistemi tutulumuna özgü gait analizlerini başarıyla ortaya koymuştur (Erdaş vd., 2023). Multimodal derin öğrenme yaklaşımlarının kullanıldığı çalışmalarda, yalnızca tanısal ayrım değil, aynı zamanda hastalık şiddetinin ve fonksiyonel etkilenme düzeyinin de tahmin edilebildiği gösterilmiştir. Bu durum, etkilenen anatomik sistemin ilerleyici doğasının izlenmesine olanak sağlamaktadır (Faiem vd., 2024). Markerless video tabanlı sistemlerin LSTM ve autoencoder mimarileriyle desteklenmesi, gait anomalilerinin gürültüden bağımsız biçimde tespit edilmesini mümkün kılarak klinik uygulanabilirliği artırmaktadır (Yoon vd., 2025). Sistematik derlemeler ve literatür anketleri, yapay zeka destekli gait analizinin yalnızca tanısal sınıflandırma değil, aynı zamanda etkilenen sistemlere özgü rehabilitasyon stratejilerinin planlanmasında da önemli bir rol oynadığını vurgulamaktadır (A review of artificial intelligence-based gait evaluation…, 2024; A survey of artificial intelligence in gait-based…, 2024). Rehabilitasyon odaklı çalışmalarda, AI tabanlı gait değerlendirmelerinin tedaviye yanıtı izleme ve fonksiyonel iyileşmeyi nesnel olarak ölçme açısından güçlü bir araç sunduğu bildirilmektedir (Edison vd., 2025). Son olarak, büyük hasta kohortları üzerinde gerçekleştirilen çalışmalar, yapay zeka modellerinin gait verileri üzerinden yürüme kalitesini ve fonksiyonel prognozu tahmin edebildiğini göstermekte ve bu durum, etkilenen anatomik sistemlerin uzunlamasına takibinde yapay zekanın klinik karar destek potansiyelini güçlendirmektedir (Ben Chaabane vd., 2023).

2. Materyal ve Yöntem

2.1. Veri Seti

Çalışmada kullanılan veri seti, yürüme bozukluğu şikayeti ile Ondokuz Mayıs Üniversitesi Nöroloji Kliniği'ne başvuran ve santral sinir sistemi, periferik sinir sistemi, ekstrapiramidal sistem ve kas, kemik ve eklem problemleri kaynaklı farklı patolojilere sahip bireylerden elde edilen yüksek çözünürlüklü yürüme videolarından oluşturulmuştur. Veri seti, gerçek klinik ortamda, rutin hasta değerlendirme sürecine entegre edilmiş bir kayıt protokolü çerçevesinde toplanmış olup, bu yönüyle çalışmanın klinik geçerliliğini ve genellenebilirliğini desteklemektedir.

Veri toplama sürecinde katılımcılardan, belirli uzunlukta ve sabit kamera yerleşimine sahip bir koridorda, kendi doğal yürüme hızlarında ve herhangi bir yönlendirme olmaksızın yürümeleri istenmiştir. Bu yaklaşım, bilinçli yürüme modifikasyonlarını en aza indirerek, bireylerin spontan ve patolojiye özgü gait paternlerinin kaydedilmesini amaçlamaktadır. Tüm video kayıtları, benzer kamera açısı, yükseklik ve mesafe parametreleri korunarak elde edilmiş ve böylece görüntüleme koşullarına bağlı varyasyonlar kontrol altına alınmıştır. Kayıt sürecinde görüntü alanına giren hasta dışı bireylerin varlığı, derin öğrenme tabanlı otomatik kişi tespit ve izleme algoritmaları kullanılarak belirlenmiştir. Tespit edilen bu bireyler, görüntüden maskeleme ve çıkarma yöntemleri ile elimine edilmiş ve analiz sürecine yalnızca hedef bireyin yürüyüşüne ait hareket verileri dahil edilmiştir.

Ham video verileri, analiz öncesinde kapsamlı bir ön işleme sürecinden geçirilmiştir. Bu kapsamda, hareket bulanıklığını azaltmaya yönelik blind deconvolution tabanlı bulanıklık giderme algoritmaları kullanılmış, kontrast ve parlaklık dağılımını dengelemek amacıyla adaptif kontrast optimizasyonu ve sensör kaynaklı gürültüyü azaltmak için uzamsal filtreleme için Adaptif Histogram Eşitleme (CLAHE – Contrast Limited Adaptive Histogram Equalization) uygulanmıştır. Ayrıca, aşırı ışık parlamaları, gölgelenme ve yetersiz aydınlatma gibi çevresel faktörlerin neden olabileceği görsel bozulmalar, otomatik normalizasyon ve eşikleme yöntemleri ile minimize edilmiştir. Uygulanan bu çok aşamalı veri temizleme ve iyileştirme süreci sayesinde, yürüyüş sırasında eklem noktalarının, gövde hizalanmasının ve uzuv hareketlerinin daha doğru ve tutarlı biçimde çıkarılabilmesi hedeflenmiştir. Böylece, oluşturulan veri setinin yapay zeka tabanlı gait analizi ve etkilenen anatomik sistemlerin tespiti için güvenilir, tekrarlanabilir ve klinik açıdan anlamlı bir temel oluşturması sağlanmıştır. En son elde edilen veriler, etkilenen anatomik sisteme göre dört ana kategoriye ayrılmıştır: (1) Santral Sinir Sistemi, (2) Periferik Sinir Sistemi, (3) Ekstrapiramidal Sistem ve (4) Kas, Kemik ve Eklem Problemleri. Her kategoriye ait hasta sayıları Tablo 1'de verilmiştir.

2.2. Yöntem

Nörolojik hastalıklar, merkezi ve periferik sinir sisteminde oluşturdukları yapısal ve fonksiyonel bozulmalar nedeniyle bireylerde çeşitli hareket bozukluklarına yol açmakta ve bunun bir çıktısı olarak yürüme yeteneği, denge kontrolü ve motor koordinasyon gibi temel mobilite yetileri ciddi ölçüde etkilenebilmektedir. Bu durum yalnızca adım uzunluğu, adım hızı veya ritmiklik gibi temel gait parametrelerinde sapmalara neden olmakla kalmamakta, aynı zamanda bireyin günlük yaşam aktivitelerine katılımını, düşme riskini, yaşam kalitesini ve klinik iyileşme sürecinin bütünsel yönetimini doğrudan şekillendiren kritik bir operasyonel problem alanı yaratmaktadır.

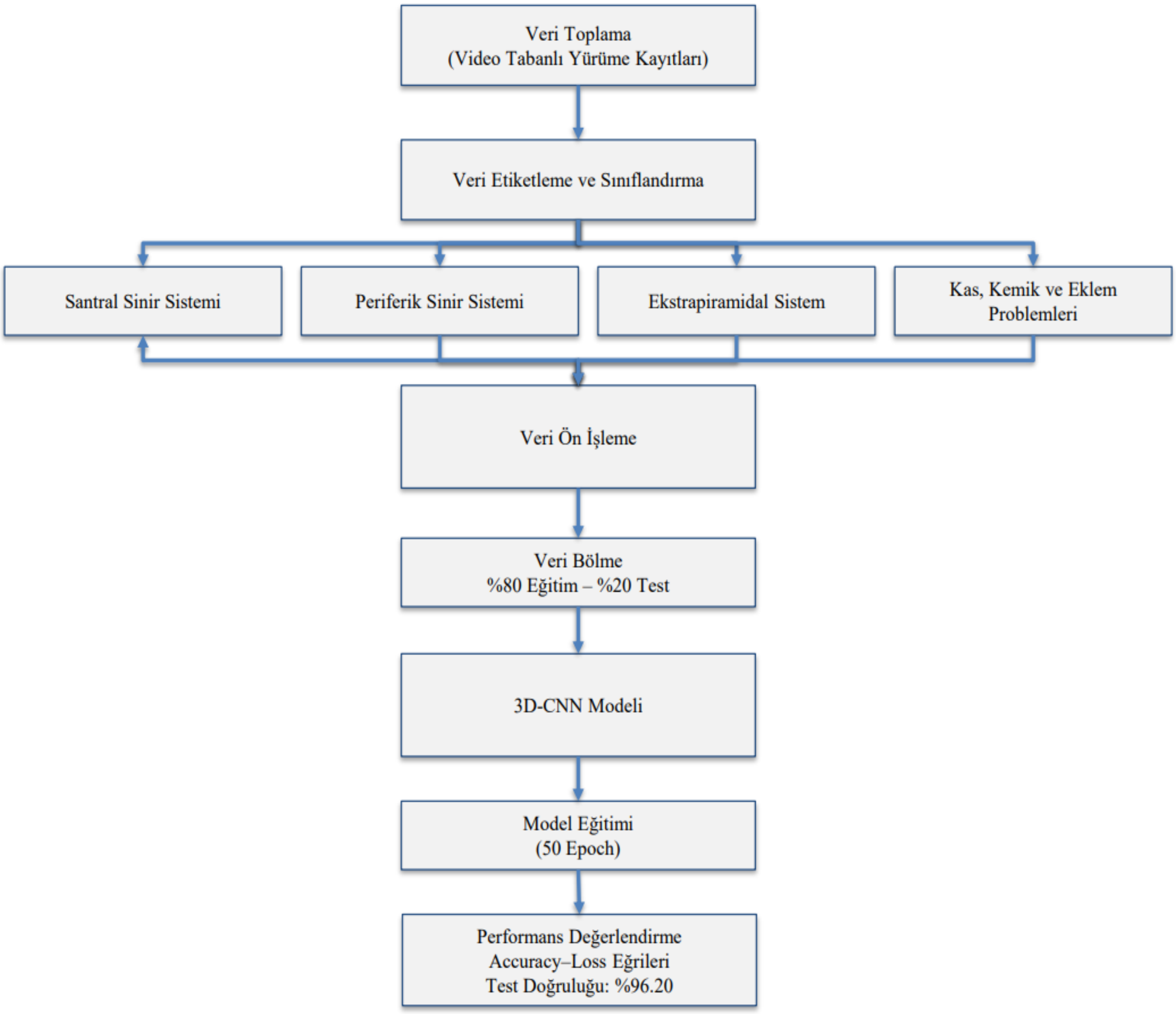
Tablo 1. Veri setinde yer alan hasta grupları ve sayıları

Hasta grupları	n
Santral sinir sistemi	286
Periferik sinir sistemi	24
Ekstrapiramidal sistem	124
Kas, kemik ve eklem problemleri	47

Bir Evrişimsel Sinir Ağı (CNN), bir veya daha fazla Evrişimsel katmandan ve ardından standart birçok katmanlı sinir ağında olduğu gibi bir veya daha fazla tamamen bağlı katmandan oluşmaktadır (Erdaş ve Sümer, 2020). CNN, görüntü ve video gibi görsel verilerin analizi için özel olarak geliştirilmiş, derin öğrenme mimarilerinin önemli bir bileşenidir. CNN, verileri işlemek için katmanlı yapılar kullanarak girdilerin belirli özelliklerini otomatik olarak çıkarabilen evrişimsel filtrelere dayanmaktadır. Bu filtreler, verilerin uzamsal hiyerarşisini keşfederek kenarlar, dokular, şekiller gibi özellikleri öğrenir ve yüksek düzeyde soyutlamalar oluştururlar. Evrişim katmanları ve bu katmanların ardından gelen havuzlama katmanları, yüksek boyutlu görsel verileri daha kompakt hale getirir ve böylece veri işlemeyi hızlandırırken veri üzerindeki gereksiz bilgileri de filtrelerler. Bir CNN modelinde, ilk katmanlar genellikle kenar ve dokular gibi düşük seviyeli özellikleri öğrenirken, daha derin katmanlar nesneler ve kavramlar gibi daha soyut ve yüksek seviyeli özellikleri öğrenirler. Bu yapı, CNN'i özellikle görüntü sınıflandırma, nesne tanıma ve segmentasyon gibi bilgisayarla görme görevlerinde güçlü kılmaktadır. CNN mimarileri, geleneksel yapay sinir ağlarından farklı olarak, ağırlık paylaşımı ve evrişim filtrelerinin kullanımı sayesinde parametre sayısını azaltmakta ve modelin öğrenme kapasitesini arttırmaktadır. Özellikle sınırlı veri ile eğitilme gereksinimini azaltan bu özellikler, CNN modellerini görsel veri ile çalışan pek çok uygulamada standart hale getirmiştir. OpenCV (Open Source Computer Vision Library), bilgisayarla görme ve görüntü işleme alanında kullanılan açık kaynaklı bir yazılım kütüphanesidir. 1999 yılında Intel tarafından geliştirilmiş ve 2000 yılında açık kaynak olarak yayımlanmıştır. OpenCV hem akademik hem de endüstriyel uygulamalarda sıkça kullanılan, görüntü işleme, nesne algılama, hareket takibi, yüz tanıma, derin öğrenme entegrasyonu gibi birçok görsel analitik görevi için zengin fonksiyonlara sahiptir. Kütüphane C++, Python, Java ve MATLAB gibi çeşitli dillerle uyumlu çalışmakta olup, çoklu platform desteği sunmaktadır. Görüntüleri filtreleme, dönüştürme, segmentasyon gibi temel işlemlerden karmaşık derin öğrenme modellerinin uygulanmasına kadar geniş bir yelpazede görevleri desteklemektedir (Bradski vd., 2008).

Bu çalışmada, yürüme bozukluklarına neden olan anatomik sistemlerin ayrıştırılmasını hedefleyen video tabanlı bir sınıflandırma modeli geliştirilmiş ve bu amaçla üç boyutlu evrişimsel sinir ağı (3D Convolutional Neural Network, 3D-CNN) mimarisi kullanılmıştır. Model, ardışık video kareleri üzerinden hem mekansal hem de zamansal özellikleri eş zamanlı olarak öğrenebilecek şekilde tasarlanmıştır. Mimari yapı, zamansal (t), yatay (x) ve dikey (y) boyutlarda evrişim işlemi gerçekleştiren ardışık Conv3D katmanlarından oluşmakta olup, her bir evrişim katmanını doğrusal olmayanlık kazandırmak amacıyla Rectified Linear Unit (ReLU) aktivasyon fonksiyonu takip etmiştir. Özellik haritalarının boyutunu kademeli olarak azaltmak, hesaplama yükünü düşürmek ve yerel zamansal-mekansal temsilleri özetlemek amacıyla belirli evrişim bloklarının ardından 3D max-pooling katmanları uygulanmıştır. Modelin genelleme kapasitesini artırmak ve aşırı uyum riskini sınırlamak amacıyla derin katmanlarda dropout düzenlileştirme tekniği kullanılmıştır. Evrişimsel katmanlardan elde edilen yüksek boyutlu özellik temsilleri flatten işlemi ile vektör formuna dönüştürülmüş ve bunu takiben tam bağlantılı (fully connected) katmanlar aracılığıyla sınıflandırma aşamasına aktarılmıştır. Çıkış katmanında, yürüme bozukluğuna yol açan anatomik sistem sınıflarına ait olasılık dağılımlarını üretmek amacıyla softmax aktivasyon fonksiyonu kullanılmıştır. Bu yapı ile video temelli yürüme verilerinde zamansal süreklilik içeren hareket paternlerinin etkin biçimde modellenmesini ve farklı anatomik sistem kaynaklı bozuklukların ayrıştırılmasını amaçlamaktadır.

Video tabanlı yürüme verilerinin toplanmasından başlayarak, veri etiketleme ve sınıflandırma, ön işleme, eğitim–test ayrımı (%80–%20), 3D-CNN modeli ile eğitim (50 epoch) ve performans değerlendirme aşamalarını içeren önerilen çalışma akış diyagramı Şekil 1'de verilmiştir.

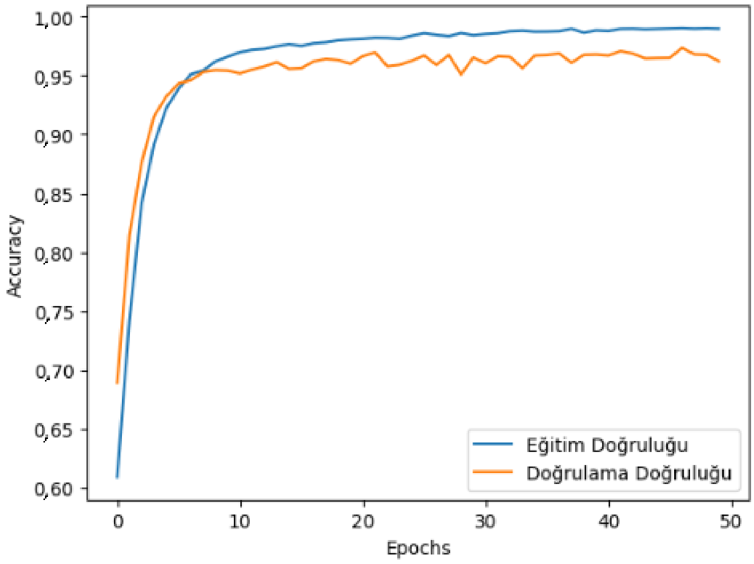


Şekil 1. Video tabanlı yürüme verilerinin toplanmasından başlayarak, veri etiketleme ve sınıflandırma, ön işleme, eğitim–test ayrımı (%80–%20), 3D-CNN modeli ile eğitim (50 epoch) ve performans değerlendirme aşamalarını içeren önerilen çalışma akış diyagramı verilmiştir.

3. Bulgular

Çalışmada önerilen video tabanlı 3D-CNN modelinin yürüme bozukluğuna yol açan anatomik sistemlerin ayrıştırılmasındaki performansı nicel ve görsel değerlendirme metrikleri üzerinden sunulmaktadır. Özellikle sınıf dağılımındaki dengesizlik ve bazı anatomik sistemlere ait örnek sayısının sınırlı olması göz önünde bulundurularak, modelin genelleme yeteneği ve aşırı öğrenme eğilimi ayrıntılı biçimde analiz edilmiştir. Bu kapsamda eğitim ve test aşamalarına ait doğruluk ve kayıp değerleri incelenmiş, modelin öğrenme süreci epoch bazında değerlendirilmiş ve elde edilen sonuçlar grafiksel olarak raporlanmıştır.

Ön işlemeden sonra özellikle Periferik Sinir Sistemi (24) ve Kas, Kemik ve Eklem Problemleri (47) gibi düşük örneklemli sınıflarda veri sayısının kısıtlı olması sebebiyle eğitim aşamasında veri seti yaklaşık %80 eğitim (train) ve %20 test olacak şekilde ikiye ayrılmış ve 3D CNN mimarisi eğitilecek biçimde yapılandırılmıştır. Model, eğitim verileri üzerinde 50 epoch boyunca eğitilmiştir. Bu epoch sayısı modelin örüntüleri yeterli düzeyde öğrenmesini sağlarken aşırı öğrenme riskini kontrol altında tutan denge noktası olduğu için tercih edilerek %96,20 doğruluk oranına ulaşılmıştır. Modelin performansını değerlendirmek amacıyla eğitim ve test sonuçları görselleştirilmiş, doğruluk ve kayıp (accuracy–loss) eğrileri analiz edilmiştir. Şekil 2'de yer alan grafikte hem eğitim hem doğrulama doğruluğunun yüksek seviyelere ulaştığını görülmektedir. Eğitim doğruluğu ile doğrulama doğruluğu arasındaki farkın düşük ve tutarlı olması, modelin aşırı öğrenme eğiliminden uzak olduğunu ve doğrulama verisi üzerinde başarılı bir genelleme sağladığını göstermektedir.



Şekil 2. Modelin eğitim ve doğruluma eğrileri.

4. Tartışma ve Sonuç

Bu çalışmada, yürüme videolarının analizi temel alınarak santral sinir sistemi, periferik sinir sistemi, kas ve kemik-eklem patolojileri ile ekstrapiramidal sistem hastalıklarının sınıflandırılmasına yönelik etkili bir yapay zeka modeli geliştirilmiştir. Elde edilen model, yüksek doğruluk oranı ile bu hastalık gruplarının ayrıştırılmasında güçlü bir performans sergilemiş ve klinik uygulamalara entegre edilebilecek nitelikte sonuçlar ortaya koymuştur. Model performansının ulaştığı seviye, özellikle erken tanı süreçlerinin desteklenmesi, hasta takibinin objektif ölçütlere dayandırılması ve kişiselleştirilmiş tedavi stratejilerinin geliştirilmesi açısından önemli bir potansiyel taşımaktadır. Bulgular, yürüme analizi temelli yapay zeka yaklaşımlarının nörolojik ve hareket sistemi hastalıklarının değerlendirilmesinde tamamlayıcı bir karar destek mekanizması olarak kullanılabileceğini göstermektedir.

Literatürdeki çalışmalar, özellikle Parkinson hastalığı ve belirli nörolojik bozukluklar üzerine odaklanmış olup, çoğunlukla tek bir hastalık grubunu hedefleyen modellerle sınırlı kalmıştır. Örneğin Pham vd. (2022) Parkinson hastalığında yürüyüş dinamiklerini kullanarak geliştirdiği model %85 doğruluk seviyesine ulaşmıştır. Ancak bu çalışmaların büyük bölümü çoklu hastalık sınıflarını aynı veri seti altında ele almamış ve kapsamlı bir klinik ayrıştırma sunmamıştır. Bu çalışmada ise farklı sinir ve kas sistemi patolojilerini aynı çerçevede analiz ederek literatüre çoklu sınıflandırma perspektifi kazandırılması amaçlanmış ve hem tanı sürecinin hızlanmasına hem de geniş bir hasta popülasyonunda uygulanabilirliğin artmasına katkı sağlanmıştır.

Sonuç olarak, önerilen yöntem literatürdeki benzer çalışmalara kıyasla daha kapsamlı, daha genellenebilir ve klinik uygulamaya daha yakın bir model sunmaktadır. Bu yönüyle çalışma, yürüme analizi tabanlı yapay zeka modellerine ilişkin bilgi birikimine anlamlı bir katkı sağlayarak gelecekte geliştirilecek tanısal yazılımlar için güçlü bir referans çerçevesi oluşturmayı hedeflemektedir.

Katkı Oranı Beyanı

Yazarların katkı yüzdeleri aşağıda verilmiştir. Yazarlar makaleyi incelemiş ve onaylamıştır.

%	K.A.K.	M.T.
K	50	50
T	50	50
Y	50	50
VTI	50	50
VAY	50	50
KT	50	50
YZ	50	50
GR	50	50
PY	50	50

K= kavram, T= tasarım, Y= yönetim, VTI= veri toplama ve/veya işleme, VAY= veri analizi ve/veya yorumlama, KT= kaynak tarama, YZ= Yazım, GR= gönderim ve revizyon, PY= proje yönetimi.

Çatışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Etik Onay Beyanı

Araştırmada yer alan bireylere bilgilendirilmiş onam formu doğrultusunda çalışmanın amacı ve hedefleri detaylı olarak anlatılmıştır. Katılımcılar onam formunu imzaladıktan sonra araştırmaya dahil edilmiştir.

Kaynaklar

Ali, F., Padilla, H., Blazek, A. M., Barnard, L., & Kaufman, K. R. (2025). Gait analysis in neurologic disorders: Methodology, applications, and clinical considerations. *Neurology*, *105*(8), Article e214154. https://doi.org/10.1212/WNL.0000000000214154

Ben Chaabane, N., Conze, P. H., Lempereur, M., Quellec, G., Rémy-Néris, O., Brochard, S., Cochener, B., & Lamard, M. (2023). Quantitative gait analysis and prediction using artificial intelligence for patients with gait disorders. *Scientific Reports*, *13*, Article 23099. https://doi.org/10.1038/s41598-023-49883-8

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media.

Edison, E., Awaiz, A. D., Sehr, S., Afzal, A., Tahir, H., Aslam, M., & Khan, M. W. (2025). Impact of artificial intelligence on clinical decision-making and support systems in hospital environments. *Insights in Health and Rehabilitation*, *3*(6), 622–628.

Erdaş, Ç. B., & Sümer, E. (2020, 19–20 Kasım). *Nörodejeneratif hastalıkların tespiti için derin öğrenme tabanlı bir yaklaşım* [Konferans bildirisi]. Tıp Teknolojileri Kongresi (TIPTEKNO), Online.

Erdaş, Ç. B., Sümer, E., & Kibaroğlu, S. (2023). Neurodegenerative diseases detection and grading using gait dynamics. *Multimedia Tools and Applications*, *82*(15), 22925–22942. https://doi.org/10.1007/s11042-023-14461-7

Faiem, N., Asuroglu, T., Acici, K., Kallonen, A., & van Gils, M. (2024). Assessment of Parkinson's disease severity using gait data: A deep learning-based multimodal approach. In M. Särestöniemi et al. (Eds.), *Digital Health and Wireless Solutions* (ss. 31–45). Springer. https://doi.org/10.1007/978-3-031-59091-7_3

Iseki, C., Hayasaka, T., Yanagawa, H., Komoriya, Y., Kondo, T., Hoshi, M., Fukami, T., Kobayashi, Y., Ueda, S., Kawamae, K., Ishikawa, M., Yamada, S., Aoyagi, Y., & Ohta, Y. (2023). Artificial intelligence distinguishes pathological gait: The analysis of markerless motion capture gait data acquired by an iOS application (TDPT-GT). *Sensors*, *23*(13), Article 6217. https://doi.org/10.3390/s23136217

Liang, A. (2024). Assessing gait dysfunction severity in Parkinson's disease using 2-stream spatial-temporal neural network. *Journal of Biomedical Informatics*, *157*, Article 104679. https://doi.org/10.1016/j.jbi.2024.104679

Pham, T. D. (2022, 6–8 Aralık). *Visual concurrent analysis of gait patterns among healthy young, old adults, and patients with Parkinson's disease* [Konferans bildirisi]. 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, ABD. https://doi.org/10.1109/BIBM55620.2022.9995627

Vun, D. S. Y., Bowers, R., & McGarry, A. (2024). Vision-based motion capture for the gait analysis of neurodegenerative diseases: A review. *Gait & Posture*, *112*, 95–107. https://doi.org/10.1016/j.gaitpost.2024.04.029

Yoon, H., Jo, E., Ryu, S., Yoo, J. I., Kim, M., & Kim, J. H. (2025). Noise-robust markerless video gait anomaly detection via two-stage acquisition and LSTM autoencoders. *Scientific Reports*, *15*, Article 42074. https://doi.org/10.1038/s41598-025-26169-9