

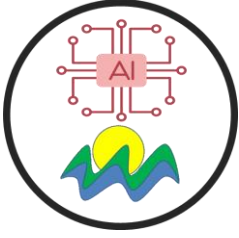
Black Sea Journal of Artificial Intelligence

Volume 2 | Issue 1



ISSN: 3108-4281


KAR-YAY



BLACK SEA JOURNAL OF ARTIFICIAL INTELLIGENCE
(BSJ AI)



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: June 15, 2026- Samsun, TÜRKİYE

Publishing kind: Electronically

OWNER

KARYAY Karadeniz Yayımcılık ve Organizasyon Ticaret Limited Company

DIRECTOR IN CHARGE

Prof. Dr. Hasan ÖNDER

EDITOR BOARDS

EDITOR IN CHIEF

Prof. Dr. Mehmet Serhat ODABAŞ, Ondokuz Mayıs University, TÜRKİYE

SECTION EDITORS*

Assoc. Prof. Dr. Cem TIRINK, Iğdır University, TÜRKİYE

Assoc. Prof. Dr. Thobela Louis TYASI, University of Limpopo, SOUTH AFRICA

Dr. Alper GÜLBE, Iğdır University, TÜRKİYE

Dr. Hakan Can ALTUNAY, Ondokuz Mayıs University, TÜRKİYE

* The ranking is arranged alphabetically within the academic title

STATISTIC EDITOR

Prof. Dr. Mehmet TOPAL, Kastamonu University, TÜRKİYE

ENGLISH EDITOR

Assist. Prof. Dr. Betül ÖZCAN DOST, Ondokuz Mayıs University, TÜRKİYE

TURKISH EDITOR

Prof. Dr. Serkan ŞEN, Ondokuz Mayıs University, TÜRKİYE

REVIEWERS OF THE ISSUE*

Prof. Dr. ıgdem TAKMA, Ege University, Department of Animal Science, Artificial Intelligence Algorithms, TRKİYE

Prof. Dr. Fatih NER, Ordu University, Department of Field Crops, Artificial Intelligence Algorithms, TRKİYE

Prof. Dr. Mustafa ŐAHİN, Kahramanmaraş St İmam University, Department of Agricultural Biotechnology, Artificial Intelligence Algorithms, TRKİYE

Assist. Prof. Dr. Esra YAVUZ, Şırnak University, Department of Accounting and Tax, Artificial Intelligence, TRKİYE

Assist. Prof. Dr. Hatice Kbra SEVİN, Karabk University, Department of Software Application Development and Analysis, Machine Learning Algorithms, TRKİYE

Dr. Asim Faraz, Bahauddin Zakariya University, Department of Livestock and Poultry Production, Artificial Intelligence Algorithms, PAKİSTAN

* The ranking is arranged alphabetically within the academic title

Table of Contents

Research Articles

1. EVALUATING BASELINE MACHINE LEARNING MODELS FOR EARTHQUAKE MAGNITUDE PREDICTION IN TÜRKİYE

Ayhan AYDIN.....1-6

2. MACHINE LEARNING IN ENVIRONMENTAL METAGENOMIC APPROACHES: A BIBLIOMETRIC AND THEMATIC ANALYSIS

Sevtap TIRINK.....7-17

3. FROM BIOACOUSTICS TO ARTIFICIAL INTELLIGENCE: A BIBLIOMETRIC STUDY OF STRESS AND WELFARE MONITORING IN POULTRY

Berkay ŞAHNAR, Cem TIRINK.....18-25

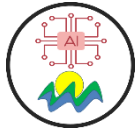
Reviews

4. TARIMDA YAPAY ZEKA KULLANIMI: MEVCUT UYGULAMALAR, ZOOTEKNİ BOYUTU VE GELECEK PERSPEKTİFLERİ

Hasan ÖNDER.....26-29

5. HAYVANSAL ÜRETİM VE BİYOTEKNOLOJİDE YAPAY ZEKÂ VE MAKİNE ÖĞRENMESİ UYGULAMALARI: PROF. DR. UĞUR ŞEN'İN ARAŞTIRMALARI ÜZERİNDEN BİR DEĞERLENDİRME VE GELECEK VİZYONU

Uğur ŞEN.....30-33



EVALUATING BASELINE MACHINE LEARNING MODELS FOR EARTHQUAKE MAGNITUDE PREDICTION IN TÜRKİYE

Ayhan AYDIN^{1*}

¹Ondokuz Mayıs University, Faculty of Engineering, 55200, Samsun, Türkiye

Abstract: Due to its tectonic framework, Türkiye faces a constant and severe threat from devastating earthquakes. To mitigate seismic risks, predictive modeling utilizing Artificial Intelligence (AI) methodologies has become increasingly vital. In this study, baseline machine learning models—including Decision Tree, Artificial Neural Networks (ANN), Bayesian Linear Regression, and Linear Regression—were evaluated on the Microsoft Azure Machine Learning Studio platform using historical seismic catalogs (magnitude ≥ 3.5) from the Kandilli Observatory. The performance results demonstrated that models relying solely on historical spatial and temporal parameters yielded insufficient predictive accuracy, with maximum R-squared (R^2) scores reaching only 0.249. These baseline findings confirm that classical catalog-based predictions fail to capture the chaotic, non-linear dynamics of seismicity. The study concludes that to achieve reliable forecasting and life-saving early warning mechanisms, future models must transition from simple historical regressions to multidimensional Deep Learning architectures (such as LSTM or XGBoost) that integrate real-time physical constraints, including GNSS strain data and active fault mechanics.

Keywords: Earthquake prediction, Machine learning, Deep learning, Big data, Seismotectonics, Azure

*Corresponding author: Ondokuz Mayıs University, Institute of Graduate, Department of Smart Systems Engineering, Samsun, Türkiye

E-mail: ayhan.aydin@omu.edu.tr (A. AYDIN)

Ayhan AYDIN



<https://orcid.org/0000-0001-9127-0951>

Received: April 10, 2025

Accepted: June 05, 2025

Published: June 15, 2026

Cite as: Aydın, A. (2026). Evaluating baseline machine learning models for earthquake magnitude prediction in Türkiye. *Black Sea Journal of Artificial Intelligence*, 2(1), 1-5.

1. Introduction

Throughout history, humanity has continuously struggled with natural disasters, with earthquakes being among the most destructive. Positioned within the Alpide-Himalayan orogenic belt, the Anatolian Plate is continuously squeezed between the Eurasian, African, and Arabian plates, creating highly active tectonic systems such as the North Anatolian Fault Zone (NAFZ) and the East Anatolian Fault Zone (EAFZ). These active fault networks have triggered devastating events, such as the 1939 Erzincan, 1999 Gölçük-Düzce, 2020 Samos (Mw 6.9), 2020 Elazığ (Mw 6.8), and the catastrophic 2023 Kahramanmaraş doublet (Mw 7.7 and 7.6). As populations and industrial centers expand near these tectonically volatile zones, the need for robust earthquake forecasting models has become paramount to prevent severe socioeconomic and human losses (Gencer et al., 2025; Khurshudov and Arslan, 2022).

Historically, classical seismology and statistics relied heavily on empirical models or straightforward statistical approaches, such as ARIMA, to identify linear trends in seismic catalogs. Early prediction attempts even observed animal behavior or electromagnetic anomalies. However, the mechanical processes underlying earthquakes—strain accumulation, friction, and viscoelastic relaxation—are highly non-linear, chaotic, and spatio-temporally heterogeneous. Consequently, traditional methods often

fail to capture the complex dynamics of fault ruptures.

1.1. Literature Review and Analysis

In recent years, the exponential growth of Big Data analytics and Artificial Intelligence (AI) has driven a paradigm shift in seismology, transitioning from deterministic constraints to flexible, probabilistic machine learning (ML) models (Khurshudov and Arslan, 2022). AI algorithms can uncover hidden, non-linear patterns across massive, multidimensional datasets that human analysts or classical statistics cannot process.

A critical analysis of the current literature reveals three distinct evolutionary phases in AI-based earthquake prediction in Türkiye:

1.2. Catalog-Based Spatial and Temporal Forecasting

Initial ML studies predominantly utilized historical catalogs. For instance, Khurshudov and Arslan (2026) analyzed over 66,000 seismic events in Türkiye using Random Forest (RF) and Gated Recurrent Unit (GRU) models, demonstrating that AI could effectively simulate small-to-medium earthquake magnitudes based solely on historical coordinates and magnitudes. Similarly, Doğan (2023) utilized algorithms such as XGBoost, Support Vector Machines (SVM), and Bayesian Ridge to successfully predict the spatial focal coordinates (latitude/longitude) and depths of future $M \geq 6.0$ earthquakes in Northwestern Türkiye (Bülteni et al., 2023). However, models relying solely on catalogs suffer



from a fundamental limitation: they record past energy releases but fail to account for the physical, real-time stress currently accumulating in the crust.

1.3. Physics-Informed Geodetic Integration: To address this limitation, researchers began integrating physical constraints into ML datasets. A landmark framework by Demirelli et al. (2023) demonstrated that integrating active geological fault geometries (fault length,

proximity) and GNSS-derived geodetic strain rates with historical catalogs drastically reduced prediction errors. By feeding strain-rate data into the Random Forest and XGBoost algorithms, they achieved an outstanding Mean Squared Error (MSE) of 0.09, demonstrating that combining crustal physics with big data yields vastly superior forecasts. Samiloglu et al. (2025) further validated this by using.



Figure 1. Devastation caused by the 7.7 and 7.6 magnitude twin earthquakes centered in Kahramanmaraş, Türkiye, on February 6, 2023 (*Kahramanmaraş Depreminin Kendine Has Özellikleri | EkoAvrasya Vakfı*, 2026).

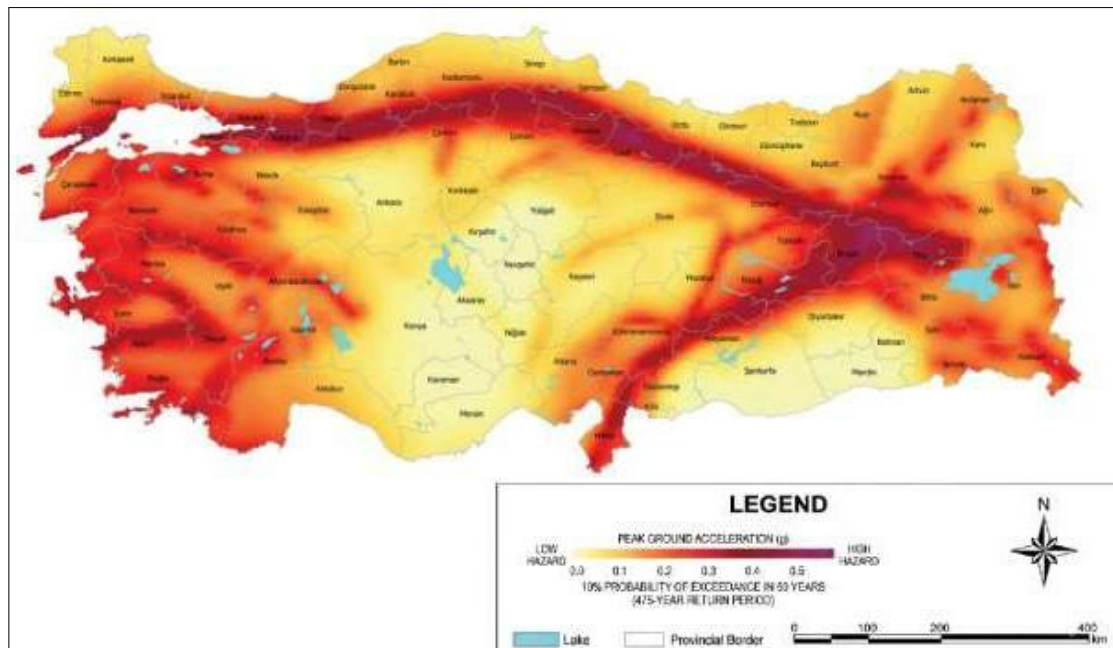


Figure 2. Seismic Hazard Map of Türkiye, illustrating peak ground acceleration (PGA) values with a 10% probability of exceedance in 50 years (475-year return period) (Sahin, 2019).

1.4. The Shift to Deep Learning and Hybrid Architectures

Most recently, researchers have adopted deep learning to capture both temporal memory and spatial relationships. Kaftan (2025) compared five models (LSTM, ANFIS, Decision Tree, Random Forest, CNN) in Western Türkiye and found that Long Short-Term Memory (LSTM) networks outperformed others due to their ability to remember long-term seismic dependencies, achieving an RMSE of 0.1391. Gencer and Cizmeci (2025) introduced even more advanced hybrid architectures, such as CNN+LSTM and Transformer models integrated with Gaussian Processes, to quantify prediction uncertainty, which is highly critical for early warning systems.

Interpretation: The literature strongly indicates that basic regression models are no longer sufficient. The future of seismic forecasting lies in multidimensional, hybrid Deep Learning models that synthesize real-time GNSS strain data, fault mechanics, and historical catalogs (CLAWS Journal Previous Issue, Claws Journal Issue - Centre for Land Warfare Studies (CLAWS), 2026.). The primary objective of the current study is to construct an earthquake magnitude prediction framework utilizing Azure Machine Learning Studio, while critically acknowledging the necessity of transitioning towards these advanced, multidimensional inputs.

2. Materials and Methods

2.1. Computational Platform and Architecture

Initial model development was conducted using Microsoft Azure Machine Learning Studio, a cloud-based platform

that provides a drag-and-drop interface for rapid algorithm deployment and performance evaluation. For advanced deep learning tasks (such as LSTM or CNNs), frameworks utilizing Python environments, including libraries like Pandas (for data manipulation), GeoPandas (for geodetic mapping), and SciKit-Learn, are established as the industry standard for handling high-dimensional seismic computations (Gencer and Cizmeci, 2025).

2.2. Multidimensional Dataset Construction

Based on modern literature best practices, a robust prediction model requires a multi-tiered dataset rather than just a simple catalog:

Historical Seismic Catalogs: The foundational data were acquired from the Boğaziçi University Kandilli Observatory and Earthquake Research Institute (KOERI) and verified against AFAD catalogs. The dataset includes events in Türkiye with magnitudes from 1900 to 2024. Core features include Date, Time, Latitude, Longitude, Depth, and Magnitude.

Geological Fault Parameters: To inject physical constraints, active fault parameters from MTA (General Directorate of Mineral Research and Exploration) must be integrated. These include the total length of the fault, the shortest distance from the earthquake epicenter to the fault line, and the fault mechanism (e.g., normal, strike-slip).

Geodetic GNSS Strain Data: Relative station movements derived from Global Navigation Satellite Systems (e.g., TUSAGA-Aktif) are incorporated to calculate crustal strain and shear stress using Hooke's Law. Incorporating GNSS time series enables algorithms to monitor active tectonic loading and stress drops.

Table 1. Data preview: Recent seismic activity in Türkiye and surrounding regions

Occurrence Date	Occurrence Time	Latitude	Longitude	Depth (km)	Mw	Location
2024-03-31 00:00:00	2024-07- 07T09:33:13.57	403.095	270.207	6.5	3.7	DISBUDAK-LAPSEKI (CANAKKALE) [South East 1.6 km]
2024-03-30 00:00:00	2024-07- 07T05:09:38.65	356.662	35.241	5.3	3.6	MEDITERRANEAN SEA
2024-03-30 00:00:00	2024-07- 07T00:25:03.23	384.867	39.35	5.0	4.7	GUNEYKOY-SIVRICE (ELAZIG) [South East 2.9 km]
2024-03-29 00:00:00	2024-07- 07T20:06:18.97	380.632	369.053	5.3	3.6	ERICEK-GOKSUN (KAHRAMANMARAS) [South East 3.0 km]
2024-03-25 00:00:00	2024-07-	353.425	276.872	16.5	3.6	MEDITERRANEAN SEA
2024-03-25 00:00:00	2024-07- 07T10:10:45.87	378.607	36.267	4.2	3.9	TOPALLAR- SAIMBEYLI (ADANA)

2.3. Data Preprocessing Pipeline

Data preprocessing is an absolute prerequisite for optimizing machine learning model convergence:

Data Cleaning: Rows containing a magnitude of 0 or incomplete geographic coordinates were removed.

Declustering (Noise Reduction): Seismic swarms and

aftershocks heavily bias machine learning models. Algorithms were applied to filter out aftershocks within specific spatial and temporal windows (e.g., removing subsequent events within 15 km and 2 months following a mainshock) to isolate independent tectonic ruptures.

Data Splitting and Normalization: The cleaned dataset was

partitioned into 70% for training and 30% for testing. Min-Max Normalization was subsequently applied to scale all inputs to a common range, preventing features with large numerical ranges (such as fault length) from dominating the learning process.

2.4. Machine Learning and Deep Learning Algorithms

Various algorithms were evaluated to understand their capacity to capture non-linear seismic behavior:

Linear and Bayesian Linear Regression: Foundational statistical techniques that model relationships between independent features and magnitude. Bayesian regression is particularly useful as it incorporates uncertainty in parameters, though linear models generally fail to capture the chaotic nature of seismicity (Pattern Recognition and Machine Learning | Springer Nature Link, 2006).

Decision Trees (DT) and Random Forest (RF): While single Decision Trees are prone to overfitting, Random Forest is an ensemble method that combines multiple trees to improve generalization. RF has proven highly effective when fed multidimensional spatial data, achieving MSEs as low as 0.09 when combining catalogs with fault and GNSS data (Breiman, 2001).

Extreme Gradient Boosting (XGBoost): A highly optimized, sequential ensemble algorithm that aggressively minimizes errors from previous trees. It excels in handling complex spatial boundaries and fault geometries (Chen and Guestrin, 2016).

Long Short-Term Memory (LSTM): A specialized Recurrent Neural Network (RNN) equipped with "forget, input, and output gates" designed to remember long-term temporal dependencies. LSTM is superior for analyzing GNSS time series, as it can track the slow accumulation of stress over decades and correlate it with sudden drops in seismic magnitude (Hochreiter and Schmidhuber, 1997).

2.5. Performance Evaluation Metrics

Model efficacy was evaluated using several standard regression metrics:

MAE (Mean Absolute Error): Measures the average magnitude of errors without considering their direction. A

value closer to 0 indicates high prediction accuracy.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

y_i = Actual values

$\hat{y}_i$ = Predicted values

n = Number of observations

MSE and RMSE (Mean Squared Error and Root Mean Squared Error): Because these metrics square the errors before averaging, they strongly penalize large deviations. In seismology, a low RMSE is critical because missing a high-magnitude earthquake carries catastrophic real-world consequences.

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

R-squared (R^2): Determines the proportion of variance in earthquake magnitude that can be explained by the input features. A score close to 1 implies that the model successfully learned the underlying physical patterns, whereas a low score indicates that the model's predictions are unreliable.

$$R^2 = \frac{SSR}{SST} \quad (3)$$

3. Results

Using the historical catalog of earthquakes ≥ 3.5 magnitude, prediction mechanisms were tested across the four chosen algorithms in Azure. The performance results are summarized below in Table 2. While the MAE and RMSE values appear reasonably low in isolation, they are severely low across all models (ranging from 20% to 25%). Bayesian Linear Regression achieved the best relative result, but due to the low scores, the baseline models failed to capture sufficient variance, rendering these predictions unreliable.

Table 2. Performance evaluation metrics of machine learning models for earthquake magnitude prediction

Model	MAE	RMSE	R^2
Artificial Neural Networks	0.528119	0.891551	0.200725
Bayesian Linear Regression	0.565008	0.882595	0.216703
Linear Regression	0.537651	0.864171	0.249064
	0.538475	0.876844	0.226878

3.1. Comparison with Advanced Literature Findings

The limitations observed in our baseline results are echoed in the advanced literature, which proves that relying strictly on historical catalogs is insufficient. For instance, Demirelli et al. (2023) demonstrated that when historical catalogs are augmented with GNSS strain rates and active fault line geometries, ensemble methods like Random Forest and XGBoost achieve an outstanding MSE of 0.09. Furthermore, Deep Learning methods drastically outperform traditional algorithms for time-series seismic

data. Kaftan (2025) utilized a Long Short-Term Memory (LSTM) network for Western Türkiye, achieving highly accurate predictions with an RMSE of 0.1391 and an MAE of 0.1046. Similarly, Samiloglu et al. (2025) integrated GNSS-derived stress accumulation data with an LSTM architecture along the North Anatolian Fault Zone, successfully modeling the non-linear build-up and drop of physical stress that correlates directly with earthquakes.

4. Discussion and Conclusion

In this study, an earthquake magnitude prediction framework was evaluated using historical data and Big Data methods on Azure Machine Learning Studio. The experimental results demonstrated that baseline algorithms—Decision Trees, ANNs, Bayesian Linear Regression, and Linear Regression—produced inadequate scores (max 0.249) when trained solely on basic historical parameters (coordinates, depth, time).

The poor performance of the baseline model emphasizes a crucial scientific reality: earthquake forecasting cannot be reduced to simple historical coordinate regression. Earthquakes are driven by chaotic, non-linear mechanisms involving active crustal deformation and stress accumulation (Kaftan, 2025). Our model lacked critical physical input parameters, which severely limited its predictive power.

The integration of current literature reveals the necessary path forward. To accurately predict complex seismic events, prediction models must transition to Multidimensional Deep Learning Architectures. By incorporating real-time GNSS geodetic strain measurements and proximity to geological faults into algorithms such as LSTM, XGBoost, or Random Forest, models can physically track active tectonic loading of the crust.

For future work, it is essential to optimize the dataset by consulting domain experts to incorporate physical constraints (e.g., fault types, continuous GNSS time series, and geodetic stress drops) and to deploy advanced recurrent neural networks (such as LSTM) that can capture long-term seismic dependencies. Transitioning to these physics-informed AI models will yield highly reliable earthquake early-warning mechanisms, ultimately saving lives and mitigating disaster risks.

Author Contributions

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

	A.A.
C	60
D	50
S	40
DCP	50
DAI	30
L	
W	40
CR	50
SR	
PM	
FA	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.

Data availability

For verification and permission requests, please contact Ayhan AYDIN (<https://github.com/AyhanAydinPhD>). In accordance with the Regulation on the Protection of Data, researchers are prohibited from sharing their data with third parties. The data may be used in scientific studies provided that permission is obtained from the relevant institutions. The authors have shared the code generated in this study after the project was completed. These can be requested via Ayhan AYDIN's GitHub page or by email.

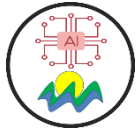
Declaration of generative AI and AI-assisted technologies

In the manuscript preparation process, the authors used Gemini (Google) and ChatGPT (OpenAI) to improve the clarity and readability of the English text. This involved using specific prompts such as: "Please improve the academic tone and grammatical flow of the following section while preserving the technical meaning." After using these tools, the authors carefully reviewed and edited the content as needed and take full responsibility for the integrity of the published article.

References

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- Bülteni, M., Doğan, A., Üniversitesi, H., Osb, B., Meslek, T. B., Okulu, Y., & Ankara, T. (2023). Makine öğrenimi yöntemleri kullanılarak Türkiye'nin kuzeybatısı için deprem tahmini. *Yerbilimleri*, 44(2), 166–178. <https://doi.org/10.17824/YERBILIMLERI.1325321>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785;CSUBTYPE:STRING:CONFERENCE>
- CLAWS Journal Previous Issue, *Claws Journal Issue - Centre for Land Warfare Studies (CLAWS)*. (n.d.). Retrieved June 28, 2026, from <https://archive.claws.co.in/journal/journal-claws-journal/index.html>
- Gencer, K., & Cizmeci, İ. H. (2025). Enhanced Earthquake Magnitude Prediction Using Hybrid Machine Learning and Deep Learning Models. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 16(2), 369–376. <https://doi.org/10.24012/DUMF.1663473>
- Gencer, K., Fakültesi, İ. C.-D. Ü. M., & 2025, undefined. (2025). Enhanced Earthquake Magnitude Prediction Using Hybrid Machine Learning and Deep Learning Models. *Dergipark.Org.TrK Gencer, İH CizmeciDicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 2025•dergipark.Org.Tr, 16(2), 369–376*. <https://doi.org/10.24012/dumf.1663473>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/NECO.1997.9.8.1735>

- Kahramanmaraş Depreminin Kendine Has Özellikleri / EkoAvrasya Vakfı.* (2026). Retrieved June 28, 2026, from <https://ekoavrasya.net/kahramanmaraş-depreminin-kendine-has-ozellikleri/>
- Khurshudov, E., & Arslan, A. (2022). *From Historical to AI-Based Approaches in Earthquake Forecasting: The Case of Türkiye.* <https://doi.org/10.36937/cebel.2026.11114>
- Pattern Recognition and Machine Learning / Springer Nature Link.* (n.d.). Retrieved June 28, 2026, from <https://link.springer.com/in/book/9780387310732>
- Sahin, S. (2019). Türkiye’de Afet Yönetimi ve 2023 Hedefleri. *Türk Deprem Araştırma Dergisi*, 1(2), 180–196. <https://doi.org/10.46464/TDAD.600455>
- Şamiloğlu, F. B., Işık, V., Aktuğ, B., & Şamiloğlu, A. T. (2025). Bolu İli Çevresinde Geçmiş Deprem Kataloğu ve GNSS Zaman Serilerinin Derin Öğrenme ile Deprem Tahmininde Kullanımı. *Turkish Journal of Earthquake Research*, 7(3), 571–589. <https://doi.org/10.46464/TDAD.1706258>
- Sciences, I. K.-A., & 2025, undefined. (n.d.). Machine Learning Applications for Earthquake Magnitude Prediction in Western Turkey. *Mdpi.ComI KaftanApplied Sciences*, 2025•*mdpi.Com*. Retrieved June 28, 2026, from <https://www.mdpi.com/2076-3417/15/20/10909>



MACHINE LEARNING IN ENVIRONMENTAL METAGENOMIC APPROACHES: A BIBLIOMETRIC AND THEMATIC ANALYSIS

Sevtap TIRINK^{1*}

¹Iğdır University, Health Services Vocational School, Department of Medical Services and Techniques, Environmental Health Program, Iğdır, Türkiye

Abstract: Machine learning is increasingly integrated with metagenomic analysis, yet the development, conceptual organization, environmental applications, and recent thematic directions of this research area require systematic characterization. This study examined English-language article and review records indexed in the Web of Science Core Collection from 2016 to 2025 using a predefined search strategy. In this study, bibliometric and science-mapping analyses were conducted to examine annual scientific production, source productivity, keyword co-occurrence, thematic structure, trend topics, and geographic and collaboration patterns. The analysis covered 286 documents published across 139 sources, with an annual growth rate of 54.79%, and publication output reached 102 documents in 2025. Metagenomics and machine learning formed major conceptual foundations, with thematic structure linking computational methods to microbial ecology, diversity, soil, and bacterial communities. Keyword co-occurrence further distinguished interconnected groupings centered on machine learning–metagenomics–bioinformatics, soil–rhizosphere and nutrient-related research, bioremediation–biodegradation, and resistome–wastewater surveillance. Recent trend topics included artificial intelligence, prediction, and degradation. When considered together with the thematic-map position of degradation–bioremediation–artificial intelligence, these patterns support a cautious interpretation that this research direction is recently gaining visibility rather than representing a definitively emerging or dominant research front. Scientific production was geographically concentrated, particularly in China and the United States, while international collaboration was also evident. Overall, the field shows increasing methodological diversity, with reference-based approaches being complemented by predictive and explainable methods

Keywords: Environmental metagenomics, Machine learning, Science mapping, Microbial ecology, Bibliometrics

*Corresponding author: Iğdır University, Health Services Vocational School, Department of Medical Services and Techniques, Environmental Health Program, Iğdır, Türkiye

E mail: sevtaptirink@gmail.com (S. TIRINK)

Sevtap TIRINK  <https://orcid.org/0000-0003-0123-0054>

Received: April 21, 2026

Accepted: June 06, 2026

Published: June 15, 2026

Cite as: Tırink, S. (2026). Machine learning in environmental metagenomic approaches: a bibliometric and thematic analysis. *Black Sea Journal of Artificial Intelligence*, 2(1), 7-17.

1. Introduction

Environmental metagenomics provides a culture-independent framework for characterizing microbial communities directly from environmental samples through the analysis of their genetic material (Zhang et al., 2021). It has been applied across diverse matrices, including soils, surface and groundwater, aquatic sediments, air and bioaerosols, and sewage-associated environments, allowing microbial communities to be investigated across a wide range of terrestrial and aquatic ecosystems (Liu et al., 2025). Unlike cultivation-dependent approaches, metagenomic analysis does not require microorganisms to be isolated and grown under laboratory conditions before their genetic characteristics are examined. This is particularly relevant in environmental microbiology, where many microorganisms remain difficult to cultivate using conventional laboratory methods. By reducing this dependence on cultivation, metagenomic approaches extend access to microbial diversity that may be poorly represented in culture-based studies, including difficult-

to-culture and low-abundance members of microbial communities (Chettri et al., 2024).

Building on these capabilities, environmental metagenomics are widely used to characterize microbial diversity and taxonomic structure across natural ecosystems (Ghannam and Techtmann, 2021). Beyond taxonomic characterization, genome-resolved metagenomic analyses can link uncultivated microbial lineages with metabolic functions involved in major biogeochemical processes, including carbon, nitrogen, and sulfur transformations that contribute to ecosystem functioning (Mirete et al., 2025). Metagenomic information is also used in environmental monitoring, where changes in microbial community composition and source-associated microbial markers can support the assessment and tracing of contamination within environmental systems (Böke Özkoç et al., 2025; Chettri et al., 2024). In remediation contexts, metagenomic analysis can further identify microbial functions associated with pollutant transformation and degradation, thereby informing the evaluation of strategies such as



biostimulation and bioaugmentation (Böke Özkoç et al., 2025; Chettri et al., 2024). The increasing use of metagenomics across these ecological and applied contexts, however, also produces datasets that place substantial demands on data processing, interpretation, and integration. High-throughput metagenomic sequencing generates large datasets that require considerable computational resources for storage, preprocessing, and downstream analysis (Sharma et al., 2025). In shotgun metagenomics, these challenges are further intensified by heterogeneous DNA mixtures, uneven organismal abundance and sequence coverage, and complex community composition, all of which can complicate genome reconstruction and downstream analysis (Mathieu et al., 2022). Metagenomic datasets are also frequently high-dimensional, with the number of taxa, genes, or sequence-derived features often exceeding the number of available samples. Their analysis is further complicated by sparsity and compositional constraints, which can affect conventional statistical interpretation if not appropriately addressed (Sharma et al., 2025; Roy et al., 2024). Incomplete taxonomic and functional reference databases present an additional limitation, particularly for sequences derived from poorly characterized or uncultivated microorganisms (Pallavi and Uday Bhaskar, 2025). Moreover, differences in sample processing, DNA extraction, sequencing platforms, and related methodological choices can introduce technical heterogeneity and affect reproducibility across studies (Papoutsoglou et al., 2023). Addressing these challenges requires analytical approaches capable of handling high-dimensional, heterogeneous, and structurally complex metagenomic data. Machine-learning approaches have consequently been applied across several stages of metagenomic analysis. Sequence-derived features can be used by supervised learning methods to classify metagenomic reads or contigs into taxonomic groups, supporting taxonomic assignment in large and heterogeneous sequence datasets (Tonkovic et al., 2020). Machine-learning methods have also been applied to metagenomic binning and genome reconstruction, where neural-network and representation-learning approaches can assist in grouping contigs and recovering metagenome-assembled genomes from complex sequence data (Herazo-Álvarez et al., 2025). Metagenomic features can likewise be incorporated into predictive models to support the classification or inference of functional characteristics (Wassan et al., 2019). In environmental research, machine learning has been applied to microbial community data for environmental assessment and prediction, including the identification of patterns associated with ecosystem conditions and contamination status (Ghannam and Techtmann, 2021). Environmental microbial profiles may also support geographic or source classification of samples (Ghannam and Techtmann, 2021). The breadth of these applications makes it increasingly relevant to examine how research linking machine learning with metagenomics has developed and

been organized within the scientific literature.

Previous reviews and bibliometric studies have approached this intersection from several related perspectives. Tonkovic et al. (2020) provided the closest synthesis to the present topic through a scoping review of applied machine learning in metagenomic classification; however, its scope is broadly metagenomic and only partially environmental. Marcos-Zambrano et al. (2021), by contrast, reviewed statistical and machine-learning approaches primarily in human and clinical microbiome research, including disease-related applications. At a broader disciplinary level, Xu et al. (2022) mapped artificial intelligence research across biotechnology and applied microbiology rather than focusing specifically on environmental metagenomics. Abdelwahab et al. (2024) similarly examined the development of AI-microbiome research, with much of the literature centered on gut microbiota, disease prediction, and other human-health contexts. Velasco Cardona et al. (2025) analyzed AI applications in bacteriology, particularly bacterial identification, diagnostics, and antimicrobial resistance. Thus, although these studies demonstrate substantial interest in AI- and machine-learning-related microbial research, their primary scopes address domains that only partially overlap with machine learning in environmental metagenomics.

Taken together, these studies provide valuable perspectives on related areas, but they only partially capture the bibliometric and thematic development of research at the intersection of machine learning and environmental metagenomics. The present study therefore characterizes the literature at this intersection during 2016–2025 by examining publication dynamics and productivity, influential sources, geographical and collaborative patterns, citation impact, and the thematic structure and temporal development of research topics

2. Material and Methods

2.1. Data Source and Search Strategy

The scientific dataset used in this study was obtained from the Web of Science (WoS) Core Collection, one of the most widely recognized and rigorously curated multidisciplinary citation databases. WoS is commonly used in scientometric studies due to its transparent indexing criteria, comprehensive metadata, and robust citation structure (Donthu et al., 2021). The search was designed to identify publications at the intersection of metagenomics, machine learning or artificial intelligence, and environmental applications. Search terms were applied to the title (TI), abstract (AB), and author keyword (AK) fields. The search expression used in WoSCC was retained without modification:

(TI=(metagenom*) OR AB=(metagenom*) OR AK=(metagenom*)) AND (TI=("machine learning" OR "deep learning" OR "artificial intelligence" OR "random forest*" OR "support vector machine*" OR "neural network*" OR "convolutional neural network*" OR "graph neural network*" OR "gradient boosting" OR XGBoost OR

LightGBM) OR AB=("machine learning" OR "deep learning" OR "artificial intelligence" OR "random forest*" OR "support vector machine*" OR "neural network*" OR "convolutional neural network*" OR "graph neural network*" OR "gradient boosting" OR XGBoost OR LightGBM) OR AK=("machine learning" OR "deep learning" OR "artificial intelligence" OR "random forest*" OR "support vector machine*" OR "neural network*" OR "convolutional neural network*" OR "graph neural network*" OR "gradient boosting" OR XGBoost OR LightGBM)) AND (TI=("environmental metagenom*" OR "environmental microbiome*" OR soil* OR rhizosphere* OR sediment* OR freshwater* OR "surface water*" OR river* OR stream* OR lake* OR estuar* OR marine OR ocean* OR seawater OR groundwater OR wastewater OR sewage OR sludge OR "activated sludge" OR wetland* OR landfill* OR compost* OR "built environment*" OR airborne OR "air microbiome*" OR bioremediat*) OR AB=("environmental metagenom*" OR "environmental microbiome*" OR soil* OR rhizosphere* OR sediment* OR freshwater* OR "surface water*" OR river* OR stream* OR lake* OR estuar* OR marine OR ocean* OR seawater OR groundwater OR wastewater OR sewage OR sludge OR "activated sludge" OR wetland* OR landfill* OR compost* OR "built environment*" OR airborne OR "air microbiome*" OR bioremediat*) OR AK=("environmental metagenom*" OR "environmental microbiome*" OR soil* OR rhizosphere* OR sediment* OR freshwater* OR "surface water*" OR river* OR stream* OR lake* OR estuar* OR marine OR ocean* OR seawater OR groundwater OR wastewater OR sewage OR sludge OR "activated sludge" OR wetland* OR landfill* OR compost* OR "built environment*" OR airborne OR "air microbiome*" OR bioremediat*))

The search results were restricted to the 2016–2025 publication period using the recorded WoS search-history expression #2 AND (PY=(2016–2020) OR PY=(2021–2025)). Records assigned to 2026 were subsequently excluded (#3 NOT PY=(2026)).

The search results were restricted to the 2016–2025 period and to English-language Article and Review Article records. Articles and Reviews were retained because they represent the principal peer-reviewed research and synthesis outputs used for bibliometric comparison, while Early Access records were excluded to avoid duplicate or unstable year assignments and to maintain a consistent closed publication window through 2025. These database-level restrictions resulted in 286 documents. No manual relevance screening was performed beyond these restrictions.

2.2. Bibliometric and Temporal Analysis

Bibliometric and science-mapping analyses were conducted in R version 4.6.0 using the bibliometrix package and the Biblioshiny interface (Aria and Cuccurullo, 2017; Aria et al., 2026; R Core Team, 2026). Performance analyses included annual scientific production, source productivity, institutional and geographic contributions, and citation indicators,

following established bibliometric principles (Donthu et al., 2021). Geographic and collaboration patterns were examined through affiliation and corresponding-author country analyses, including single-country publications (SCP), multiple-country publications (MCP), country-level citation patterns, and international collaboration networks. Conceptual and thematic structure was evaluated using keyword-frequency analysis, keyword co-occurrence networks, thematic mapping, and Trend Topics (Aria and Cuccurullo, 2026). Source concentration was assessed using Bradford's Law and the Kolmogorov–Smirnov goodness-of-fit test.

Temporal publication dynamics were additionally evaluated using the Life Cycle of Scientific Production analysis based on a logistic growth model. Model fit was summarized using R^2 and RMSE, while model-derived indicators included the estimated saturation level, inflection point, peak annual publication output, and the interval between 10% and 90% of the estimated saturation level.

3. Results

3.1. General Characteristics and Temporal Development of Literature

The bibliometric dataset comprised 286 documents published across 139 sources between 2016 and 2025. These publications involved 1,854 authors, 17,638 cited references, and 944 distinct author keywords. Scientific production showed an annual growth rate of 54.79% over the study period. The literature was relatively recent, with a mean document age of 2.74 years, while the average citation count was 29.37 per document. Collaborative authorship was also prominent, with an average of 7.29 co-authors per document and international co-authorship accounting for 34.27% of the publications (Table 1).

Annual publication output increased substantially between 2016 and 2025 (Figure 1a). Publication output rose from two documents in 2016 to 13 in 2020 and 16 in 2021. The increase became more pronounced thereafter, reaching 31 documents in 2022, 39 in 2023, 66 in 2024, and 102 in 2025. Thus, more than one-third of all documents in the dataset were published in the final year of the study period. The Life Cycle of Scientific Production analysis provided a model-based characterization of the publication trajectory (Figure 1b). The fitted logistic model showed a close correspondence with the observed data ($R^2 = 0.995$; RMSE = 2.13). The model estimated a saturation level of 2,860 publications and an inflection point at approximately 2030.1, corresponding to a modeled maximum annual output of 339 publications. The estimated interval between 10% and 90% of the saturation level was 9.3 years. By 2025, the cumulative output of 286 documents corresponded to 10.0% of the modeled saturation level. Because the inflection point and modeled peak occur beyond the observed 2016–2025 period, these estimates represent model-based extrapolations rather than observed publication counts.

Table 1. General bibliometric characteristics of the dataset

Indicator	Value
Timespan	2016–2025
Sources	139
Documents	286
Annual growth rate	54.79%
Authors	1,854
Authors of single-authored documents	1
Co-authors per document	7.29
International co-authorship	34.27%
Author keywords	944
References	17,638
Document average age	2.74 years
Average citations per document	29.37

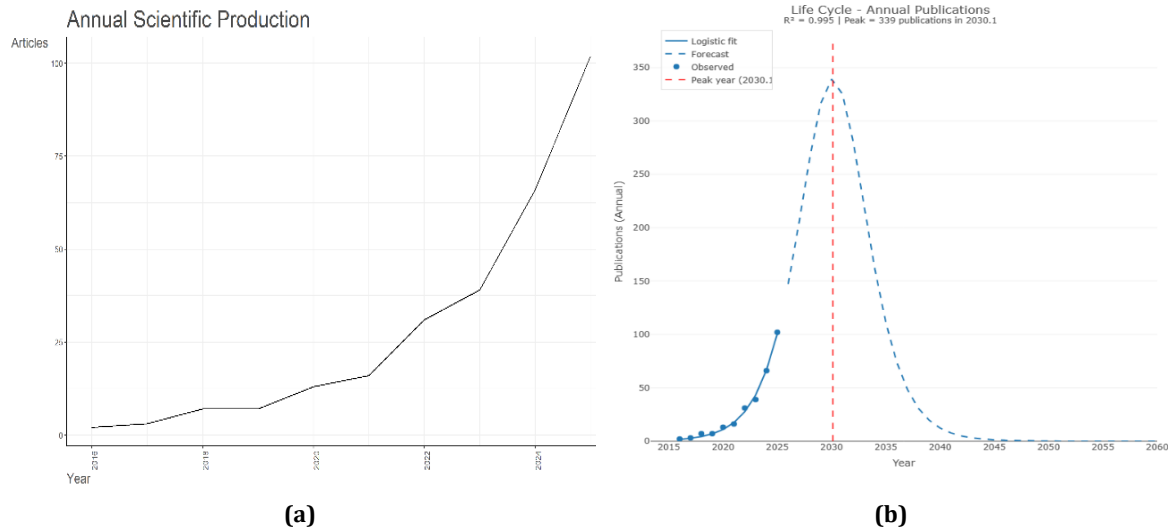


Figure 1. Temporal development of research on machine learning in environmental metagenomics: (a) annual scientific production during 2016–2025 and (b) Life Cycle of Scientific Production analysis showing the model-derived annual publication trajectory.

3.2. Scientific Sources and Citation Structure

Water Research was the most productive source with 14 documents, followed by Frontiers in Microbiology and Microbiome with 12 documents each. Water Research also had the highest local h-index ($h = 11$), whereas Frontiers in Microbiology received the highest number of local citations (599). The Bradford model showed a high coefficient of determination ($R^2 = 0.9766$) and a multiplier of 2.94. The three Bradford zones contained 11, 33, and 95 sources, respectively, with each zone contributing approximately one-third of the total publications. The Kolmogorov–Smirnov test did not provide sufficient evidence to reject the specified Bradford distribution ($D = 0.1367$, $P = 0.149$); this non-significant result should not be interpreted as proof of exact conformity.

Citation analysis identified two documents with substantially higher global citation counts than the other highly cited publications in the dataset. Here, global citations refer to citations received by a document across the full Web of Science citation database, whereas local citations refer to citations received from documents within the 286-record collection analyzed in this study.

The most globally cited document was a 2022 article published in Nature Communications, with 865 citations, followed by a 2018 article published in Microbiome, with 699 citations. Other highly cited documents included articles published

Other highly cited documents included articles published in Environmental Science and Technology in 2017 (158 citations), FEMS Microbiology Ecology in 2018 (153), Nature Communications in 2022 (152), and Nature Microbiology in 2023 (151). Within the analyzed collection, the 2018 Microbiome article with 699 global citations and the 2022 Nature Communications article with 865 global citations each received 14 local citations.

3.3. Institutional, Geographic, and International Collaboration Patterns

Scientific production was unevenly distributed across institutions and countries. The Chinese Academy of Sciences was the most frequently represented affiliation, with 48 publications. It was followed by the University of California System with 20, Cornell University with 18, and the United States Department of Energy with 16. East China Normal University and Weill Cornell Medicine each

contributed 14 publications, while the Centre National de la Recherche Scientifique contributed 13.

The corresponding-author analysis placed China first in publication output, followed by the United States and India (Figure 2a). Germany and the United Kingdom formed the next group of contributing countries, followed by France, Brazil, Spain, Australia, Italy, and several other countries with smaller publication counts. Both single-country publications (SCP) and multiple-country publications (MCP) were represented among the leading countries.

Country-level citation patterns showed a similar concentration. China accumulated the highest number of citations (4,241), followed by the United States (2,009). India and Germany received 388 and 367 citations,

respectively, while Brazil and Italy each recorded approximately 200 citations.

International co-authorship accounted for 34.27% of the complete dataset. This value is a dataset-level co-authorship indicator and is distinct from the collaboration-network visualization, which depicts the pattern and strength of cross-country relationships rather than a collaboration rate. The country collaboration map showed multiple cross-border connections among the major contributing countries, including links involving China and the United States as well as countries across Europe, Asia, and Australia (Figure 2b). Thus, the geographical structure combined a concentration of publication output in a limited number of countries with a broader network of international research links.

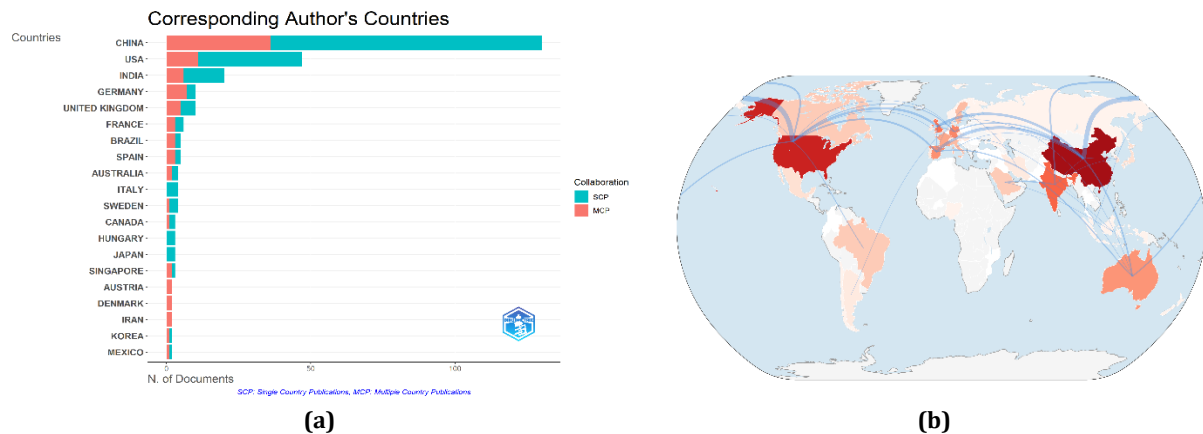


Figure 2. Geographic distribution and international collaboration in the analyzed literature: (a) corresponding authors' countries according to SCP and MCP, and (b) international country collaboration network.

3.4. Conceptual and Thematic Structure

Author-keyword analysis identified metagenomics as the most frequent term, occurring 78 times, followed by machine learning with 71 occurrences. Other prominent terms were diversity (57), bacteria (41), bacterial (27),

microbiome (25), soil (22), metagenome (21), community (18), and microbial community (18). The word cloud (Figure 3) is included only as a descriptive visual summary of these frequency patterns and is not used as an independent analytical result.



Figure 3. Word cloud showing the most frequent author keywords in the dataset.

The keyword co-occurrence network identified four interconnected conceptual groupings (Figure 4). The first was centered on machine learning, metagenomics, microbiome, and related bioinformatics terms. The second comprised diversity, soil, rhizosphere, carbon, and nitrogen, reflecting a cluster of soil- and ecosystem-related concepts. A third grouping linked bacteria with bioremediation, biodegradation, and identification, whereas the fourth was characterized by resistome, antibiotic resistance genes, and wastewater. Connections between these groupings indicated that the identified conceptual domains were not completely isolated from one another.

The thematic map provided a complementary representation based on centrality and density (Figure 5). The bacteria–community–microbial community cluster occupied the motor-theme quadrant. Two groups were positioned among the basic themes: metagenomics–machine learning–microbiome and diversity–bacterial–soil. The nitrogen–carbon–rhizosphere and random forest–metabolism–denitrification groups occurred in the niche-theme quadrant. In contrast, degradation–bioremediation–artificial intelligence was located in the quadrant conventionally labeled emerging or declining themes.

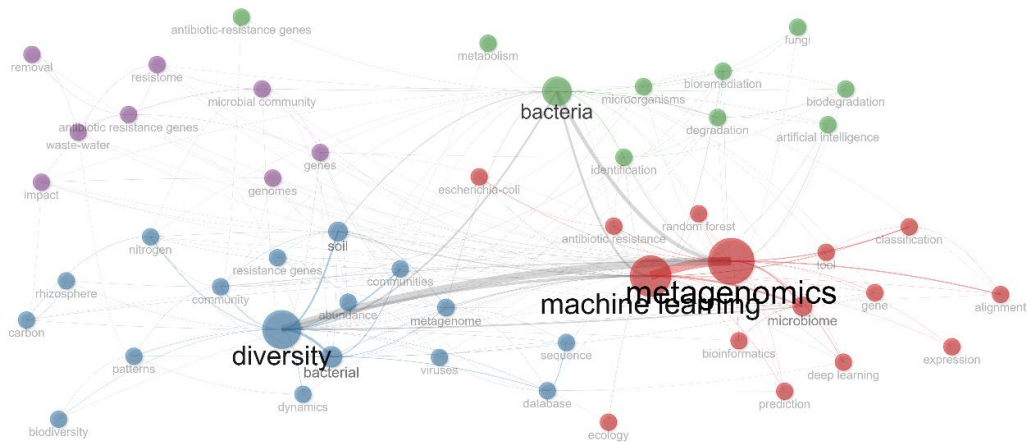


Figure 4. Author-keyword co-occurrence network showing the principal conceptual groupings in research on machine learning and environmental metagenomics.

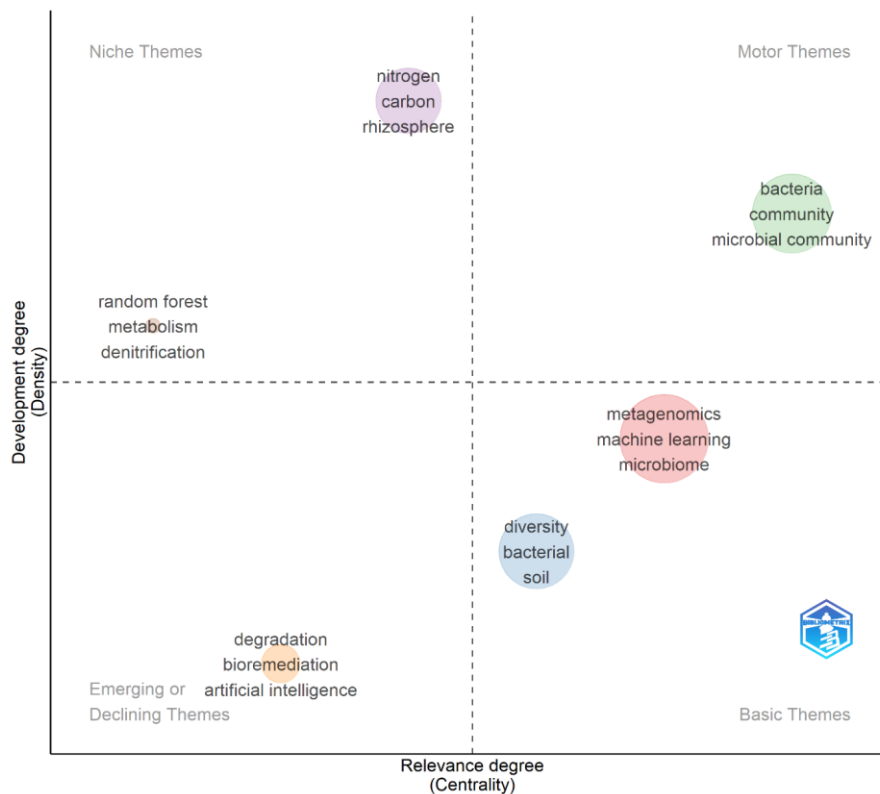


Figure 5. Thematic map of author keywords according to centrality and density.

3.5. Temporal Evolution of Research Topics

The Trend Topics analysis showed differences in the temporal positioning of frequently used terms (Figure 6). Alignment, resistome, and *Escherichia coli* were centered around 2022, whereas metagenomics, bacteria, and microbiome were positioned around 2023. Machine learning, diversity, and bacterial were centered around 2024 and were represented by comparatively larger

frequency markers than several of the more recent terms. The most recent group was positioned in 2025 and included artificial intelligence, prediction, and degradation. These terms had smaller frequency markers than the dominant terms such as machine learning and metagenomics, but their temporal placement distinguished them from earlier terminology.

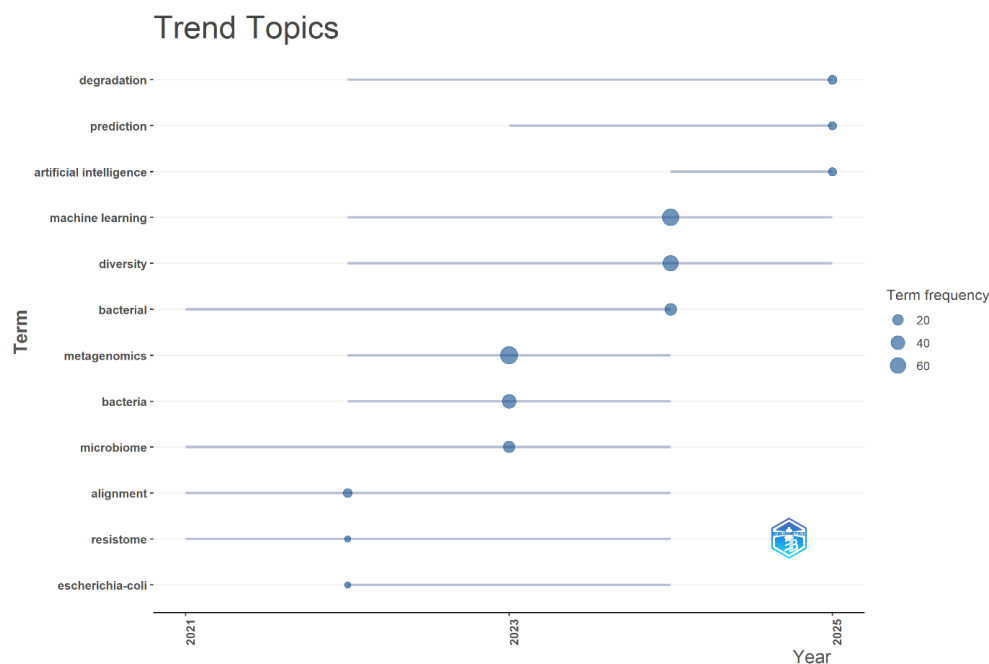


Figure 6. Trend Topics analysis showing the temporal positioning and relative frequency of selected author keywords. Bubble size represents term frequency; temporal recency indicates the location of recent use within the selected analysis settings and does not by itself indicate dominance or growth rate.

4. Discussion

4.1. Rapid Growth and Developmental Trajectory of Machine Learning in Environmental Metagenomics

The bibliometric results indicate a marked recent expansion of research at the intersection of machine learning and environmental metagenomics. Across 286 documents identified from 139 sources, the annual growth rate was 54.79%, and publication output reached 102 documents in 2025. The mean document age of 2.74 years indicates that much of the literature has accumulated relatively recently, while the average of 29.37 citations per document suggests that this literature has already generated a visible citation footprint. This publication trajectory is consistent with broader methodological developments in metagenomics. Mathieu et al. (2022) described analytical limitations associated with conventional alignment- and reference-dependent approaches in metagenomic analysis, including computational demands and incomplete representation of microbial diversity, providing an important context for complementary machine-learning and deep-learning strategies. Roy et al. (2024) likewise documented an expanding range of deep-learning applications across metagenomic tasks and increasing diversity in model

architectures and analytical purposes, although their review addressed metagenomics broadly rather than environmental applications alone.

The increasing availability of sequencing data and the analytical demands imposed by large, complex datasets may therefore help explain the growing interest in scalable computational approaches. Herazo-Álvarez et al. (2025), for example, documented increasing use of neural-network architectures for metagenomic binning, including applications involving environmental datasets and complex microbial communities. Such methodological diversification is consistent with the widening computational scope accompanying the publication growth observed here. The fitted life-cycle model ($R^2 = 0.995$) indicated that cumulative publication output in 2025 corresponded to approximately 10% of the modeled saturation level. This suggests that the field remains in an expansion phase, although the model-derived estimates should be interpreted cautiously.

4.2. Conceptual Foundations and Thematic Organization of the Field

Beyond publication growth, the conceptual structure of literature was dominated by terms linking metagenomic methodology with microbial ecology. Metagenomics (78

occurrences) and machine learning (71) were the two most frequent keywords, accompanied by diversity (57), bacteria (41), bacterial (27), microbiome (25), and terms associated with soil and microbial communities. Consistent with this pattern, metagenomics–machine learning–microbiome and diversity–bacterial–soil was positioned as basic themes, whereas bacteria–community–microbial community occupied the motor-theme quadrant. Together, these results indicate that computational analysis and community-level ecological characterization form closely connected components of the field.

The basic-theme position of metagenomics–machine learning–microbiome is consistent with the integration of machine learning into metagenomic analysis. Mathieu et al. (2022) described ML and DL applications in taxonomic classification and functional annotation, while Ghannam and Techtman (2021) showed how machine-learning approaches can relate high-dimensional microbial-community data to environmental conditions and predictive outcomes. Within the thematic-map framework, the basic-theme position therefore reflects broad connectivity to the field rather than limited relevance.

The diversity–bacterial–soil theme represents a similarly broad ecological foundation. Bahram et al. (2018) demonstrated relationships among bacterial diversity, community composition, functional gene repertoires, and environmental conditions in global topsoil microbiomes. Jansson and Hofmockel (2018) emphasized the diversity and spatial complexity of soil microbiomes and the role of metagenomics in characterizing microbial identities and functional potential. These findings provide an ecological context for the prominence of soil-associated bacterial diversity in the thematic structure identified here.

Within the same mapping framework, the motor-theme position of bacteria–community–microbial community reflects both strong connectivity with the wider network and comparatively high internal development. Grossart et al. (2020), for example, described how metagenomic information can link microbial community composition and interactions with ecological functions and biogeochemical processes in aquatic systems. By contrast, nitrogen–carbon–rhizosphere and random forest–metabolism–denitrification occupied niche-theme positions, indicating more specialized combinations of ecological processes, settings, and analytical approaches. Trend Topics further showed that artificial intelligence, prediction, and degradation were among the most recently positioned terms. Their recent temporal position, together with the placement of degradation–bioremediation–artificial intelligence in the thematic map, indicates increasing attention to this research direction, although it has not yet reached the prominence of the dominant themes. Overall, the thematic structure depicts a field organized around a broad metagenomic–computational foundation, a well-developed community–ecology axis, and several more specialized process- and

method-oriented streams.

4.3. Environmental Application Domains of Machine Learning in Metagenomic Research

This conceptual organization is further resolved by the keyword co-occurrence network, which indicates several differentiated but interconnected environmental application domains rather than a single homogeneous research stream. Four principal structures were evident: a machine learning–metagenomics–microbiome/bioinformatics domain; a diversity–soil–rhizosphere–carbon/nitrogen domain; bacteria–bioremediation–biodegradation–identification domain; and a resistome–antibiotic resistance genes–wastewater domain. These groupings suggest that computational approaches are associated with distinct environmental research contexts involving ecological prediction, contamination assessment, remediation, and environmental surveillance.

Within the soil-related domain, Gardiner et al. (2024) used shotgun metagenomic data to combine taxonomic and genetically encoded functional information in models of soil organic carbon, including graph-based approaches representing relationships between microbial taxa and their associated functions. This illustrates how environmental metagenomic information can contribute to predictive models of ecosystem properties. However, model importance or predictive association should not, by itself, be interpreted as evidence of causal microbial control over carbon dynamics.

Aquatic monitoring and source tracking provide another application. Zhu et al. (2024) applied machine learning to riverine nitrogen-source classification using environmental information that included shotgun-metagenomic taxonomic and functional profiles as well as 16S rRNA amplicon data. The metagenomic component demonstrates how functional information can complement community composition in source-attribution analyses. This is consistent with broader applications of microbial-community profiles to environmental prediction and monitoring described by Ghannam and Techtman (2021) and with the supervised-learning and de novo bioindicator framework discussed by Cordier et al. (2021). However, microbiome profiling, 16S rRNA amplicon sequencing, metabarcoding, and shotgun metagenomics represent distinct molecular and analytical approaches; examples from these categories are therefore discussed as related evidence rather than treated as methodologically equivalent.

The bioremediation–biodegradation domain similarly connects metagenomic characterization with predictive analysis. Wijaya et al. (2024) combined taxonomic and metabolic profiles from groundwater metagenomes with machine-learning and explainable-AI approaches to diagnose petroleum contamination and identify degradation-associated predictors. Moreover, Chettri et al. (2024) described the use of metagenomics to identify pollutant-degradation genes and pathways, trace contamination sources, and characterize microbial

responses relevant to bioremediation. Similarly, Böke Özkoç et al. (2025) highlighted the integration of AI and machine learning with metagenomic data in contaminated environments, including applications for predicting pollutant-degradation pathways and supporting bioremediation strategies in polluted soils, waters, and sediments. In the context of potentially toxic element (PTE)-contaminated aquatic environments, Farooq Tayeb Alsaadawi et al. (2026) further highlighted metagenomic functional profiling as a complementary approach for identifying microbial signatures related to metal resistance, detoxification, oxidative-stress response, and redox transformations, thereby extending sediment assessment beyond concentration-based measurements alone.

Finally, the resistome-antibiotic resistance genes-wastewater cluster identifies a distinct environmental-surveillance stream. Metagenomic analysis can support resistome characterization and contamination-source assessment, although the evidence considered here supports the specific integration of machine learning with wastewater shotgun metagenomics less directly than the soil, aquatic, and contaminated-groundwater applications. Detection of resistance genes also indicates genomic presence or potential rather than expression or ecological activity. Collectively, the co-occurrence structure therefore portrays machine learning as supporting multiple environmental metagenomic tasks rather than a single uniform analytical application.

5. Conclusion

The literature linking machine learning with environmental metagenomics expanded rapidly between 2016 and 2025, indicating a comparatively young research area. The dataset comprised 286 documents and showed an annual growth rate of 54.79%, with publication output reaching 102 documents in 2025. This recent acceleration indicates that the integration of computational learning approaches with environmental metagenomic research has developed into a broader and increasingly visible body of literature.

The conceptual structure of the field is organized around metagenomics and machine learning as broad conceptual foundations, with microbial diversity, bacterial communities, soil, and related ecological processes forming major thematic components. The co-occurrence network further indicated differentiated but interconnected research domains involving a machine learning-metagenomics-microbiome/bioinformatics core, soil-rhizosphere and nutrient-related research, bioremediation and biodegradation, and resistome and wastewater surveillance. Trend Topics indicated increasing visibility of machine learning, artificial intelligence, prediction, and degradation-related terminology, although the *degradation-bioremediation-artificial intelligence* direction is more appropriately interpreted as recently gaining visibility than as a definitively established emerging field. Methodologically,

predictive and explainable approaches are broadening the analytical repertoire of environmental metagenomic research.

The field is internationally connected but geographically uneven, with scientific production concentrated particularly in China and the United States. These geographical comparisons are descriptive and reflect publication and citation counts within the Web of Science-indexed dataset; they should not be interpreted as direct measures of research quality because they are influenced by database coverage, publication volume, collaboration patterns, and publication age. Future research may benefit from broader geographic participation, greater integration across diverse environmental systems, and rigorous validation of predictive AI/ML outputs against experimentally or independently measured environmental and biological processes.

Author Contributions

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

	S.T.
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 declares 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

- Abdelwahab, S. I., Taha, M. M. E., Jerah, A. A., Farasani, A., Abdullah, S. M., Aljahdali, I. A., Ibrahim, R., Oraibi, O., Oraibi, B., Alfaifi, H. A., Alzahrani, A. H., & Babiker, Y. O. H. (2024). Artificial intelligence and microbiome research: Evolution of hotspots, research trends, and thematic-based narrative review. *Cellular and Molecular Biology*, 70(10), 182-192. <https://doi.org/10.14715/cmb/2024.70.10.24>
- Aria, M. & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*, 11(4), pp 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Aria, M., & Cuccurullo, C. (2026). *Science Mapping Analysis - A*

- primer with *Biblioshiny*. McGraw-Hill, ISBN: 978-88-386-2297-7.
- Aria, M., Cuccurullo, C., D'Aniello, L., & Spano, M. (2026). Biblioshiny and the SAAS Workflow: An integrated framework for transparent and reproducible science mapping. A demonstration through the replication of a study. *Journal of Informetrics*, 20(3), 101837 <https://doi.org/10.1016/j.joi.2026.101837>
- Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., Bengtsson-Palme, J., Anslan, S., Coelho, L. P., Harend, H., Huerta-Cepas, J., Medema, M. H., Maltz, M. R., Mundra, S., Olsson, P. A., Pent, M., Pölme, S., Sunagawa, S., Ryberg, M., ... Bork, P. (2018). *Structure and function of the global topsoil microbiome*. *Nature*, 560(7717), 233–237. <https://doi.org/10.1038/s41586-018-0386-6>
- Böke Özkoç, H., Tırink, S., & Farooq Tayeb Alsaadawi, S. (2025). Metagenomic approaches to monitoring and management of potential toxic element pollution. *UBAK Publishing House*. <https://doi.org/10.5281/zenodo.17939033>
- Chettri, D., Verma, A. K., Chirania, M., & Verma, A. K. (2024). Metagenomic approaches in bioremediation of environmental pollutants. *Environmental Pollution*, 363, 125297. <https://doi.org/10.1016/j.envpol.2024.125297>
- Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D. A., Bouchez, A., Chariton, A. A., Creer, S., Frühe, L., Keck, F., Keeley, N., Laroche, O., Leese, F., Pochon, X., Stoeck, T., Pawlowski, J., & Lanzén, A. (2021). Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. *Molecular Ecology*, 30(13), 2937–2958. <https://doi.org/10.1111/mec.15472>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Farooq Tayeb Alsaadawi, S., Tırink, S., & Böke Özkoç, H. (2026). Integrating metagenomic functional profiling to evaluate public health risks derived from PTE-contaminated river sediments. In *International Congress on Natural Sciences and Applied Mathematics* (pp. 259–275). <https://doi.org/10.5281/zenodo.19459760>
- Gardiner, L., Marshall, M., Reusch, K., Dearden, C., Birmingham, M., Carrieri, A. P., Pyzer-Knapp, E. O., Krishna, R., Neal, Andy (2024) DGCNN approach links metagenome-derived taxon and functional information providing insight into global soil organic carbon. *NPJ Biofilms and Microbiomes*, 10. p. 113. [10.1038/s41522-024-00583-9](https://doi.org/10.1038/s41522-024-00583-9)
- Ghannam, R. B., & Techtmann, S. M. (2021). Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. *Computational and Structural Biotechnology Journal*, 19, 1092–1107. <https://doi.org/10.1016/j.csbj.2021.01.028>
- Grossart, H. P., Massana, R., McMahon, K. D., & Walsh, D. A. (2020). Linking metagenomics to aquatic microbial ecology and biogeochemical cycles. *Limnology and Oceanography*, 65(S1), S2–S20. <https://doi.org/10.1002/lno.11382>
- Herazo-Álvarez, J., Mora, M., Cuadros-Orellana, S., Vilches-Ponce, K., & Hernández-García, R. (2025). A review of neural networks for metagenomic binning. *Briefings in Bioinformatics*, 26(2), bbaf065. <https://doi.org/10.1093/bib/bbaf065>
- Jansson, J. K., & Hofmockel, K. S. (2018). The soil microbiome—from metagenomics to metaphenomics. *Current Opinion in Microbiology*, 43, 162–168. <https://doi.org/10.1016/j.mib.2018.01.013>
- Liu, S., Rodriguez, J. S., Munteanu, V., Ronkowski, C., Sharma, N. K., Alser, M., ... & Mangul, S. (2025). Analysis of metagenomic data. *Nature Reviews Methods Primers*, 5, 5. <https://doi.org/10.1038/s43586-024-00376-6>
- Marcos-Zambrano, L. J., Karaduzovic-Hadziabdic, K., Loncar Turukalo, T., Przymus, P., Trajkovic, V., Aasmets, O., Berland, M., Gruca, A., Hasic, J., Hron, K., Klammsteiner, T., Kolev, M., Lahti, L., Lopes, M. B., Moreno, V., Naskinova, I., Org, E., Paciência, I., Papoutsoglou, G., ... Truu, J. (2021). Applications of machine learning in human microbiome studies: A review on feature selection, biomarker identification, disease prediction and treatment. *Frontiers in Microbiology*, 12, 634511. <https://doi.org/10.3389/fmicb.2021.634511>
- Mathieu, A., Leclercq, M., Sanabria, M., Perin, O., & Droit, A. (2022). Machine learning and deep learning applications in metagenomic taxonomy and functional annotation. *Frontiers in Microbiology*, 13, 811495. <https://doi.org/10.3389/fmicb.2022.811495>
- Mirete, S., Sánchez-Costa, M., González-de-Figueras, C., Díaz-Rullo, J., Martínez-Rodríguez, P., & González-Pastor, J. E. (2025). Metagenome-assembled genomes (MAGs): Advances, challenges, and ecological insights. *Microorganisms*, 13(5), 985. <https://doi.org/10.3390/microorganisms13050985>
- Pallavi, H., & Uday Bhaskar, N. (2025). Computational and machine learning frameworks for microbial data analysis: A systematic review. *International Journal of Advanced Research in Computer Science*, 16(6), 60–69. <https://doi.org/10.26483/ijarcs.v16i6.7377>
- Papoutsoglou, G., Tarazona, S., Lopes, M. B., Klammsteiner, T., Ibrahim, E., Eckenberger, J., Novelli, P., Tonda, A., Simeon, A., Shigdel, R., Béreux, S., Vitali, G., Tangaro, S., Lahti, L., Temko, A., Claesson, M. J., & Berland, M. (2023). Machine learning approaches in microbiome research: challenges and best practices. *Frontiers in Microbiology*, 14, 1261889. <https://doi.org/10.3389/fmicb.2023.1261889>
- R Core Team (2026). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL: <https://www.R-project.org/>.
- Roy, G., Prifti, E., Belda, E., & Zucker, J.-D. (2024). Deep learning methods in metagenomics: A review. *Microbial Genomics*, 10(4), 001231. <https://doi.org/10.1099/mgen.0.001231>
- Sharma, S., Narahari, H. P., & Raman, K. (2025). Harnessing machine learning for metagenomic data analysis: Trends and applications. *mSystems*, 10(11), e01642-24. <https://doi.org/10.1128/msystems.01642-24>
- Tonkovic, P., Kalajdziski, S., Zdravevski, E., Lameski, P., Corizzo, R., Pires, I. M., Garcia, N. M., Loncar-Turukalo, T., & Trajkovic, V. (2020). Literature on applied machine learning in metagenomic classification: A scoping review. *Biology*, 9(12), 453. <https://doi.org/10.3390/biology9120453>
- Velasco Cardona, D. C., Cardona-Acevedo, S., Valencia-Arias, A., Pérez-Delgado, O., & Pinella Vega, M. (2025). Bibliometric analysis of the application of artificial intelligence techniques in bacteriology: A PRISMA-guided research agenda. *Frontiers in Microbiology*, 16, 1641967. <https://doi.org/10.3389/fmicb.2025.1641967>
- Wassan, J. T., Wang, H., Browne, F., & Zheng, H. (2019). A comprehensive study on predicting functional role of metagenomes using machine learning methods. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(3), 751–763. <https://doi.org/10.1109/TCBB.2018.2858808>
- Wijaya, J., Park, J., Yang, Y., Siddiqui, S. I., & Oh, S. (2024). A metagenome-derived artificial intelligence modeling framework advances the predictive diagnosis and

- interpretation of petroleum-polluted groundwater. *Journal of Hazardous Materials*, 472, 134513. <https://doi.org/10.1016/j.jhazmat.2024.134513>
- Xu, D., Liu, B., Wang, J., & Zhang, Z. (2022). Bibliometric analysis of artificial intelligence for biotechnology and applied microbiology: Exploring research hotspots and frontiers. *Frontiers in Bioengineering and Biotechnology*, 10, 998298. <https://doi.org/10.3389/fbioe.2022.998298>
- Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y., Hao, H., Yi, W., Li, M., & Xie, Y. (2021). Advances in metagenomics and its application in environmental microorganisms. *Frontiers in Microbiology*, 12, 766364. <https://doi.org/10.3389/fmicb.2021.766364>
- Zhu, Z., Ding, J., Du, R., Zhang, Z., Guo, J., Li, X., Jiang, L., Chen, G., Bu, Q., Tang, N., Lu, L., Gao, X., Li, W., Li, S., Zeng, G., & Liang, J. (2024). Systematic tracking of nitrogen sources in complex river catchments: Machine learning approach based on microbial metagenomics. *Water Research*, 253, 121255. <https://doi.org/10.1016/j.watres.2024.121255>



FROM BIOACOUSTICS TO ARTIFICIAL INTELLIGENCE: A BIBLIOMETRIC STUDY OF STRESS AND WELFARE MONITORING IN POULTRY

Berkay ŞAHNAR^{1*}, Cem TIRINK¹

¹Iğdır University, Faculty of Agriculture, Department of Animal Science, TR76000, Iğdır, Türkiye

Abstract: In this study, the scientific literature regarding the acoustic monitoring of stress and welfare in poultry was examined from both bibliometric and thematic perspectives. Publications covering the period from 1987 to 2025 were retrieved from the Web of Science Core Collection database; only research articles and reviews in English were included in the evaluation. Bibliometric analyses were conducted using the Bibliometrix package and the Biblioshiny interface within the R environment. Annual scientific output, sources, authors, countries, citation structure, keyword co-occurrence networks, and thematic structure were analyzed. The findings revealed significant growth in the research field, particularly after 2020, with an annual scientific output growth rate of 8.47%. While "Poultry Science", "Applied Animal Behaviour Science", and "Animals" emerged as the most productive publication sources, keyword and thematic analyses highlighted concepts such as animal welfare, behaviour, poultry, broiler, bioacoustics, precision livestock farming, and machine learning. Notably, the inclusion of "poultry-machine learning" and "animal welfare-behaviour" clusters among the motor themes indicates an evolution in the field from classical bioacoustics approaches toward AI-supported, automated welfare monitoring systems. In conclusion, the study finds that research on acoustic monitoring in poultry is becoming increasingly interdisciplinary and that AI-based methods could play a pivotal role in the future development of objective, continuous, and non-invasive welfare assessment systems.

Keywords: Bioacoustics, Animal welfare, Poultry, Vocalization, Artificial intelligence, Bibliometrics

*Corresponding author: Iğdır University, Faculty of Agriculture, Department of Animal Science, TR76000, Iğdır, Türkiye

E mail: berkay76000@gmail.com (B. ŞAHNAR)

Berkay ŞAHNAR



<https://orcid.org/0009-0005-4814-4833>

Cem TIRINK



<https://orcid.org/0000-0001-6902-5837>

Received: April 21, 2026

Accepted: June 07, 2026

Published: June 15, 2026

Cite as: Şahnar, B., & Tırink, C. (2026). From bioacoustics to artificial intelligence: A bibliometric study of stress and welfare monitoring in poultry. *Black Sea Journal of Artificial Intelligence*, 2(1), 18-25.

1. Introduction

The reliable and continuous monitoring of animal welfare in poultry is becoming increasingly important in modern production systems, both in terms of ethical requirements and sustainable production. Traditional welfare assessments rely largely on methods requiring direct observation or intervention, such as behavioral observations, physiological measurements, and periodic clinical checks. In contrast, the development of precision livestock farming technologies has brought to the fore sensor-based approaches that enable the monitoring of animal health, behavior, and welfare status in a contactless, continuous, and real-time manner. Among these approaches, bioacoustics and acoustic monitoring are particularly notable for their ability to provide low-cost, non-invasive, and continuous data. It has long been emphasized that animal vocalizations can convey information regarding emotional, physiological, and individual characteristics, and that automated sound analysis can be utilized for large-scale monitoring of animal welfare (McLoughlin et al., 2019). More recent studies also demonstrate that bioacoustics methods hold

significant potential for monitoring various components of welfare, such as physiology, health, behavior, and affective states (Coutant et al., 2024).

Animal vocalizations can be viewed not only as means of communication but also as indicators of the biological and environmental conditions the animals are experiencing. Characteristics of vocalizations, such as frequency, duration, fundamental frequency, energy distribution, spectral structure, and call type, can be linked to stress, health status, behavior, and emotional valence. Research on farm animals indicates that vocal characteristics can be used to assess positive welfare states, not just negative ones. For instance, it has been reported that the duration and fundamental frequency of vocalizations associated with positive states can vary in certain species, and that specific call types in chickens can be linked to positive states (Laurijs et al., 2021). Therefore, using vocalizations in conjunction with other behavioral and physiological indicators is considered a promising approach for the non-invasive assessment of welfare under farm conditions.

The importance of sound-based assessments in poultry has become increasingly evident through studies



conducted on broilers and laying hens. Current research demonstrates that distinct types of vocalizations can be automatically identified and utilized to monitor welfare-related conditions. Soster et al. (2025a) developed a system based on a convolutional neural network to distinguish between four types of broiler vocalizations, such as distress calls, pleasure notes, short peeps, and warbles, achieving a balanced accuracy of 91.1%. The study relied on a dataset comprising over 2,000 labeled vocalization samples, and the results highlighted the potential of automated acoustic monitoring for assessing broiler welfare. Similarly, Lev-ron et al. (2025) classified the vocal responses of broilers to environmental stressors such as cold, heat, and wind using acoustic signal processing and a transformer-based artificial neural network, reporting that the developed approach achieved a mean average precision of 0.97. These findings indicate that vocalizations could serve as a potential digital welfare indicator for the early detection of environmental stress. Alongside classical acoustic techniques, such as Mel-frequency cepstral coefficients, spectral features, and spectrogram analyses, convolutional neural networks, long short-term memory networks, and attention-based models are increasingly being employed for the automated interpretation of poultry vocalizations. A comprehensive systematic review from 2025 demonstrated that artificial intelligence-assisted (AI-assisted) vocalization analysis in poultry is advancing rapidly, particularly in the monitoring of health, behavior, and welfare, with the majority of the examined studies published after 2020 (Manikandan and Neethirajan 2025a; 2025b). Furthermore, another recent systematic review reveals that automated acoustic analyses in broiler and laying hens are beginning to be linked to the "Five Domains" welfare framework.

However, the bulk of existing literature focuses on the performance of specific algorithms, particular vocalization types, or specific welfare indicators. There is a scarcity of bibliometric studies that provide a holistic overview of the field's scientific evolution over time, prominent research themes, collaborations between countries and authors, and, crucially, the transition from traditional bioacoustics approaches to AI-driven monitoring systems. Therefore, the aim of this study is to examine scientific literature on the acoustic monitoring of stress and welfare in poultry using bibliometric and thematic methods, and to elucidate publication trends, influential sources, collaboration structures, key research themes, and the field's conceptual shift from bioacoustics to artificial intelligence.

2. Material and Methods

In this study, the Web of Science Core Collection (WoSCC) database was used to evaluate the bibliometric structure of the scientific literature regarding the monitoring of stress and welfare in poultry using acoustic methods. Web of Science was selected because it provides the detailed bibliographic metadata—such as author, source, citation,

keyword, and institutional information—necessary for bibliometric analysis. The literature search was designed to encompass the concepts of poultry, vocalization/acoustic monitoring, and stress/welfare. The "Topic Search" query employed was defined as follows:

TS=(broiler* OR "broiler chicken" OR poultry) AND (vocalization* OR vocalisation* OR acoustic* OR sound*) AND (stress OR welfare)

Bibliometric analyses were conducted using the R statistical programming environment (R Core Team, 2026) and the RStudio (version 2026.09.0+174) integrated development environment (Posit Software, PBC). The "Bibliometrix" R package was utilized for processing bibliographic data, calculating performance indicators, and performing science mapping analyses. Bibliometrix is a comprehensive open-source tool developed for bibliometric data analysis and science mapping (Aria and Cuccurullo, 2017). Biblioshiny, which is the web-based user interface of the Bibliometrix package, was used to generate the visual outputs of the analyses (Aria and Cuccurullo, 2026; Aria et al., 2026). The process of identifying, screening, and ultimately including studies in the analysis is presented in accordance with the PRISMA 2020 reporting framework (Page et al., 2021). Through this interface, analyses were performed regarding annual scientific output, source and author performance, citation metrics, keyword co-occurrence networks, thematic maps, and other aspects of conceptual structure.

The search initially yielded 211 records. The PRISMA 2020 approach was utilized to ensure transparent reporting of the selection process, and the process was illustrated using a flow diagram adapted for bibliometric studies. PRISMA flow diagrams enable clear reporting of the movement of identified records through the screening, eligibility assessment, and final inclusion stages, as well as the number of excluded records (Figure 1).

In the first stage, since 2026 was an incomplete publication year, 19 records from that year were excluded to avoid introducing bias into the time-series and thematic trend analyses; this left 192 records covering the 1987–2025 period. Subsequently, only "research article" and "review article" types were included, while 11 records belonging to other document types were removed, reducing the total count to 181. In the final stage, only studies published in English were retained; 9 non-English records were eliminated, resulting in a final dataset of 172 publications for bibliometric analysis. Due to the interdisciplinary nature of the subject, no Web of Science subject category filter was applied, thereby preventing the exclusion of studies published in fields such as animal science, veterinary medicine, behavioral science, agricultural engineering, computer science, electronic engineering, and related disciplines.

Bibliographic records for the 172 finally included publications were exported from Web of Science in a format containing the "Full Record and Cited References." This format enables comprehensive bibliometric analyses,

as it encompasses publication years, authors, journals, institutions, countries, author keywords, Keywords Plus, citation data, and reference lists.

During the analysis process, the general bibliometric characteristics of the dataset were evaluated first. In this context, the total number of publications, annual scientific output, annual growth trends, number of sources, number of authors, co-authorship structure, and citation indicators were examined. Additionally, the most productive journals, most productive authors, countries with the highest publication output, and studies receiving the highest global citations were identified. During the analysis process, the general bibliometric characteristics of the dataset were evaluated first. In this context, the total

number of publications, annual scientific output, annual growth trends, number of sources, number of authors, co-authorship structure, and citation indicators were examined. Additionally, the most productive journals, most productive authors, countries with the highest publication output, and studies receiving the highest global citations were identified.

To reveal the scientific collaboration structure of the field, co-authorship and international collaboration patterns at the country level were assessed. Bibliometrix/Biblioshiny enables the analysis of co-authorship networks at the author, institution, and country levels, alongside key indicators of scientific output performance.

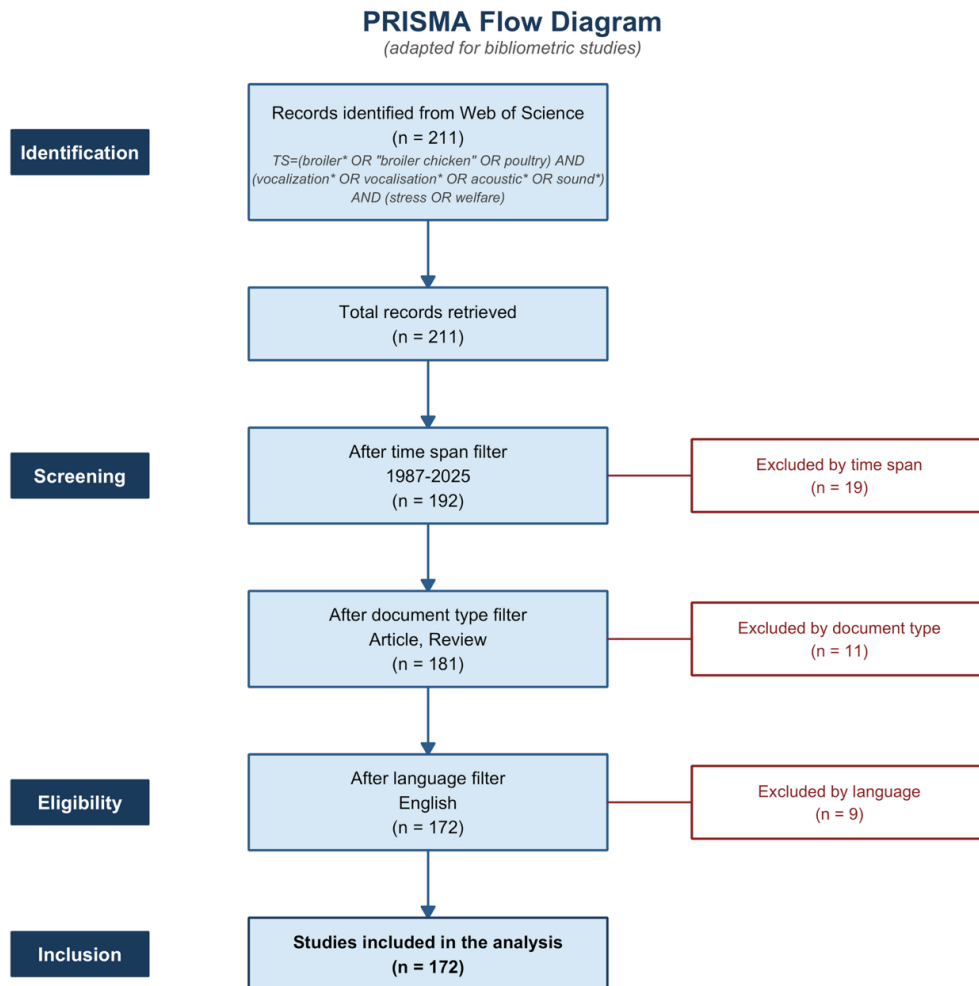


Figure 1. PRISMA flow diagram.

3. Results

The final dataset, obtained from the Web of Science Core Collection and analyzed using Biblioshiny, comprises 171 documents published between 1987 and 2025 (Figure 2). These publications appeared in a total of 70 different sources and were authored by 666 individuals. The fact that only 11 of the studies were single authored indicates that the research field is characterized by a high degree of collaboration. The average number of authors per

document was found to be 4.68, while the international co-authorship rate stood at 23.98%. The dataset contains a total of 544 author keywords and 7,225 bibliographic records. The average age of the documents was calculated as 9.64 years, with an average of 34.13 citations per document. An average annual growth rate of 8.47% in scientific output demonstrates a long-term increase in research interest regarding the sound-based monitoring of stress and welfare in poultry.

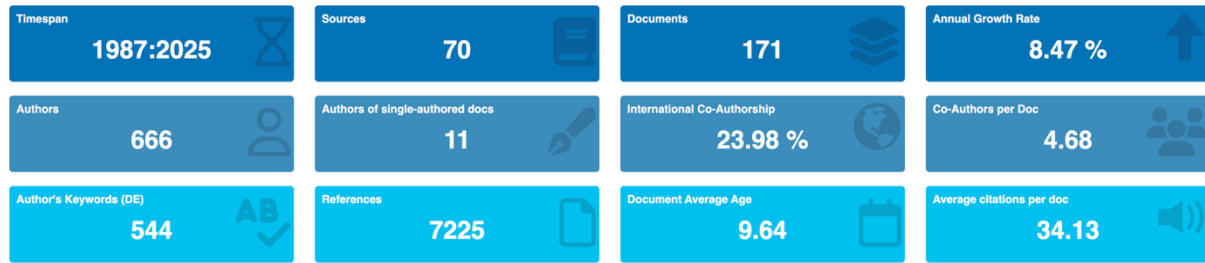


Figure 2. Main information.

An examination of annual scientific production reveals that the field initially had a very limited volume of publications (Figure 3). While the number of annual publications remained largely confined to a few studies from the late 1980s through the early 2000s, the first significant increase was observed between 2008 and 2010. Nevertheless, the primary acceleration occurred

after 2020. Output reached approximately 16 publications in 2022; following a temporary dip in 2023, it rose to 20 publications in 2024 and approximately 22 in 2025, marking the highest level recorded in the dataset. This pattern suggests that the field has expanded in recent years from classical animal behavior and welfare studies toward sensor-based and automated monitoring systems.

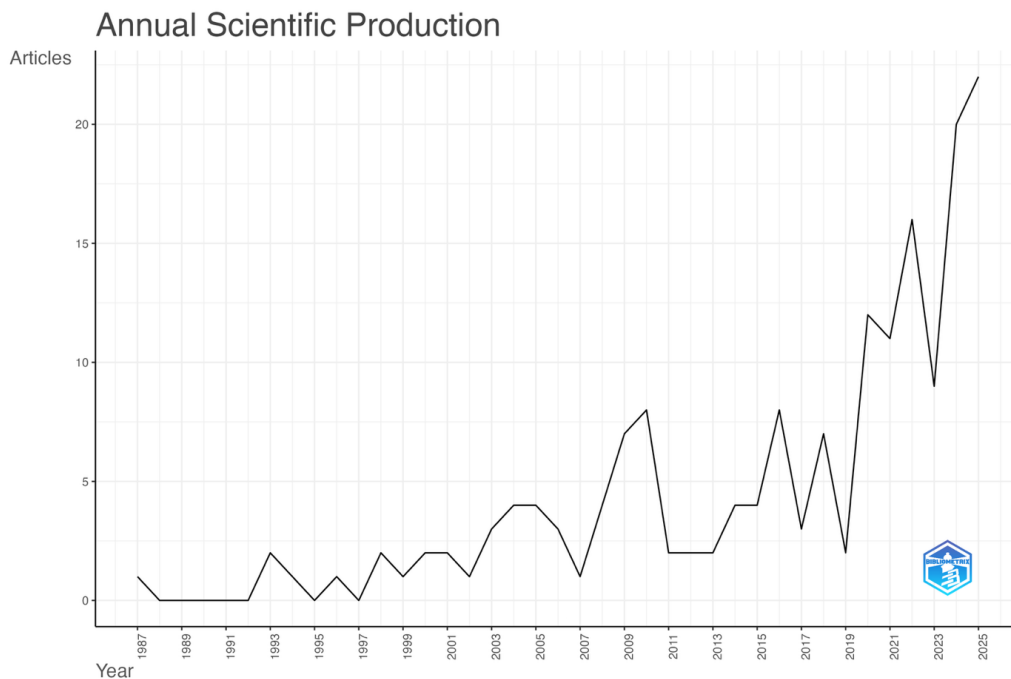


Figure 3. Annual scientific production on acoustic monitoring of stress and welfare in poultry between 1987 and 2025.

This acceleration is consistent with current literature. It is reported that acoustic monitoring in poultry has increasingly integrated with artificial intelligence, deep learning, and automated vocalization classification in recent years. For instance, the ability to classify different vocalization types in broiler chickens with a balanced accuracy of 91.1% using a convolutional neural network has demonstrated that acoustic data can be transformed into practical welfare monitoring systems (Soster et al., 2025a). Similarly, mean average precision values reaching 0.97 have been reported in the classification of broiler vocalizations associated with environmental stressors using artificial neural networks (Lev-Ron et al., 2025). Consequently, the increase in publications since 2020 reflects not only numerical growth but also a

methodological transformation in which acoustic data is integrated with AI-based decision support systems. An analysis of the sources indicates that the field is developing across journals covering both animal science/welfare and engineering-technology (Figure 4). "Poultry Science" holds the highest number of publications, ranking first with 21 documents. It is followed by "Applied Animal Behaviour Science" with 17 publications and "Animals" with 15. Other notable sources include "Computers and Electronics in Agriculture" (eight publications), "Animal" (seven), and "Frontiers in Veterinary Science" (six). Additionally, "British Poultry Science" and "Sensors" (five publications each), along with "Animal Welfare" and "Applied Sciences-Basel" (four publications each), are among the top ten sources.

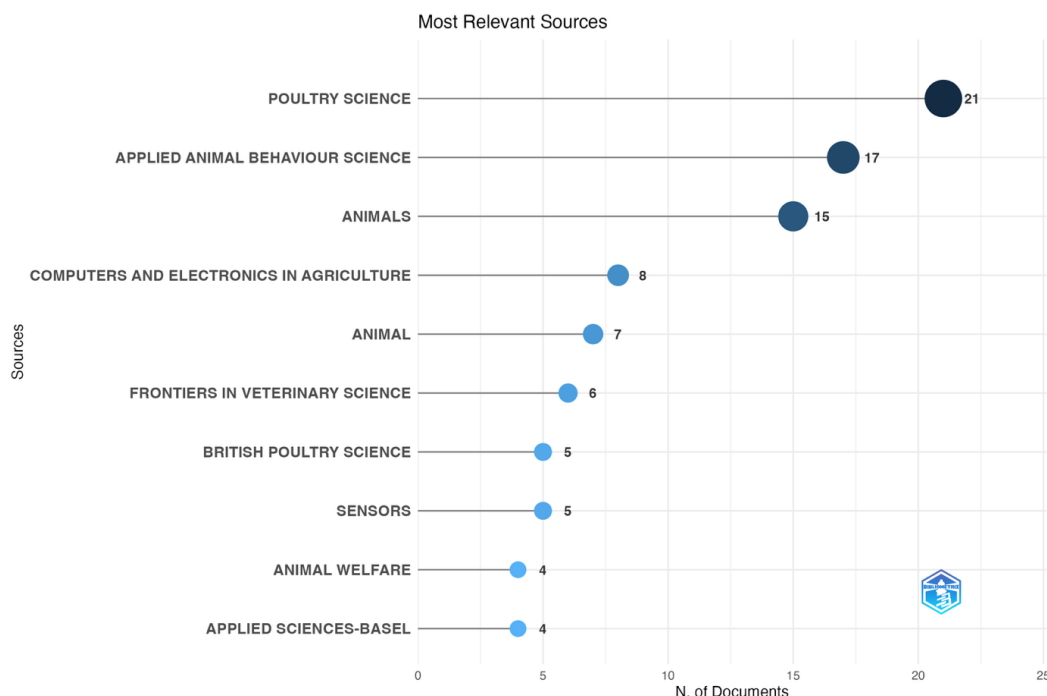


Figure 4. Most relevant sources, ranked by the number of publications in the analyzed dataset.

This distribution clearly demonstrates the interdisciplinary nature of the field. The prominence of welfare and behavior-focused journals, such as “Applied Animal Behaviour Science” and “Animal Welfare”, alongside species-specific journals like “Poultry Science” and “British Poultry Science” highlights the biological foundations of the literature. Conversely, the high ranking of technology-oriented sources such as “Computers and Electronics in Agriculture”, “Sensors”, and “Applied Sciences” indicates that acoustic welfare monitoring is increasingly integrating sensor technologies, signal processing, and artificial intelligence applications. Current literature also notes that the application of machine learning in poultry farming is expanding beyond the assessment of behavior, health, and welfare to include

sound- and image-based automated diagnostic systems (Naeem et al., 2025). Therefore, resource allocation can be considered a bibliometric indicator supporting the transition from “bioacoustics to artificial intelligence” mentioned in the study’s title.

The co-occurrence network based on author keywords demonstrates that the research field has evolved not around a single conceptual axis, but through multiple interconnected thematic clusters (Figure 5). The network reveals links between concepts related to welfare and behavior and those focused on poultry production, stress, acoustic/sound analysis, and technology. The fact that certain major nodes are connected to multiple clusters indicates that these terms serve as conceptual bridges within the research field.

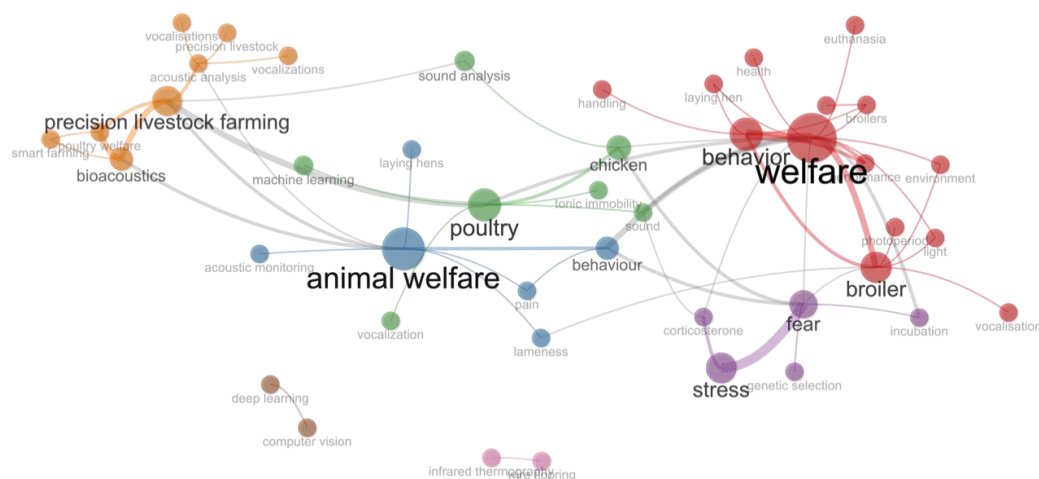


Figure 5. Keyword co-occurrence network illustrating the conceptual structure of research on acoustic monitoring of poultry stress and welfare.

The positioning of concepts such as animal welfare, behavior, poultry, machine learning, broilers, stress, bioacoustics, and precision livestock farming on the thematic map further supports this network structure. This result suggests that acoustic monitoring studies are no longer focused solely on identifying vocalizations but have evolved into a broader research framework integrated with behavioral welfare indicators, precision livestock farming systems, and automated classification models. Indeed, current studies demonstrate that stress, behavior, and health status can be assessed non-invasively by utilizing acoustic features in conjunction with machine learning and deep learning models (Manikandan and Neethirajan, 2025b).

The thematic map categorizes clusters formed from the author's keywords into four regions based on centrality and density metrics (Figure 6). In the bibliometric approach, clusters with high centrality and high density values are interpreted as "motor themes"; those with low centrality and high density as "niche themes"; those with low centrality and low density as "emerging or declining themes"; and those with high centrality and low density as "basic themes."

In this context, there are two distinct clusters in the "Motor Theme" region. The first cluster is shaped around the terms animal welfare, behavior, and lameness. This indicates that research on animal welfare and behavior constitutes a fundamental axis of the field—one that is well-developed and strongly linked to other themes. The second motor theme centers on the axis of poultry, machine learning, and laying hens. The inclusion of "machine learning" as a motor theme is particularly significant; this finding demonstrates that AI-based analyses have evolved from a peripheral innovation into a central, advanced component of the current research landscape. This result aligns with the methodological shift—reported in a 2025 systematic review—from traditional MFCC- and spectrogram-based methods toward CNNs, LSTMs, attention mechanisms, and self-supervised audio models (Manikandan and Neethirajan, 2025a).

The presence of the "welfare-broiler-behavior" cluster within the "Basic Themes" region indicates that broiler welfare and behavior constitute one of the most fundamentals and interconnected conceptual backbones of the research field. The high centrality yet relatively low density of this cluster suggests that, while the topic is linked to numerous sub-fields of the literature, it continues to evolve internally around diverse methods

and indicators. This situation aligns with existing methodological gaps regarding standardized audio recording protocols, shared datasets, and models generalizable across species, ages, and environmental conditions. Current literature shows ongoing research into the automated detection of broiler vocalizations and the identification of new call clusters, demonstrating that this core theme is still actively expanding (Soster et al., 2025a; 2025b).

The cluster comprising precision livestock farming, bioacoustics, and poultry welfare is also situated near the core themes. This finding indicates that bioacoustics is increasingly being integrated into precision livestock farming systems, evolving from a purely laboratory-focused research topic into a technological application for real-time farm management. The non-invasive nature of acoustic monitoring and its capacity to generate continuous data can be considered key drivers of this transformation. Recent machine learning-based studies also treat sound signals as digital biomarkers that can be directly utilized for welfare assessment (Manikandan and Neethirajan, 2025b).

The "niche themes" region comprises the clusters of environmental enrichment-nutrition-cattle and disease-euthanasia. While these themes exhibit a high degree of internal development, they show more limited connections to the field's core conceptual structure. The emergence of the term "cattle" as a niche theme within a poultry-focused dataset may stem from the use of keywords in certain publications within the broader context of farm animal welfare or comparative animal welfare. Consequently, such terms should be regarded not as central components of the primary research axis, but rather as peripheral themes representing the field's connection to the wider animal welfare literature.

In light of all this information, the findings indicate that research on the acoustic monitoring of stress and welfare in poultry has progressed through three fundamental stages. In the first stage, studies focused primarily on behavior, stress, and the biological significance of vocalizations. In the second stage, approaches involving acoustic feature extraction, sound analysis, and precision livestock farming came to the forefront. In the current stage, machine learning, deep learning, and automated welfare classification have emerged as key driving themes in the field. This transformation demonstrates a strong parallel between bibliometric findings and the current experimental literature.

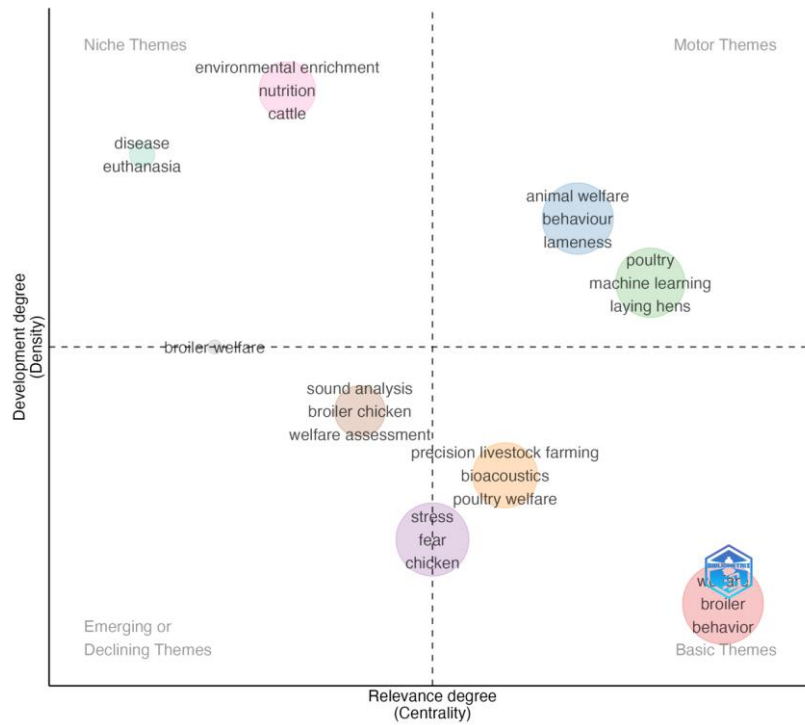


Figure 6. Thematic map based on author keywords.

Consequently, the existing bibliometric structure reveals that the field is not only expanding in terms of publication volume but is also undergoing a conceptual and methodological shift, moving from classical bioacoustics observations toward automated, real-time, AI-driven welfare monitoring systems. However, the low-density values of core themes and the positioning of certain stress/sound analysis clusters in the "emerging/declining" zone indicate that the field has not yet fully matured; issues such as data standardization, model validation under field conditions, explainable AI, and generalizability across different herd, breed, and age groups remain critical for future research. Systematic reviews identify field generalizability, data standardization, and the integration of next-generation AI architectures as key research priorities for the future (Manikandan and Neethirajan, 2025a).

4. Conclusion

This study presents the scientific evolution of literature regarding the acoustic monitoring of stress and welfare in poultry from 1987 to 2025, analyzing it from both bibliometric and thematic perspectives. The findings reveal that while the field maintained a limited publication volume for many years, it has demonstrated significant growth, particularly after 2020. The increase in annual scientific output indicates that acoustic monitoring technologies are gaining increasing importance in animal welfare research.

Analyses of sources and keywords demonstrate that the research field is not confined solely to animal behavior and welfare; rather, it possesses an interdisciplinary structure that integrates agricultural engineering, sensor

technologies, precision livestock farming, and artificial intelligence applications. The prominence of concepts such as "animal welfare," "behavior," "poultry," "broiler," "bioacoustics," "precision livestock farming," and "machine learning" suggests that the field is evolving from traditional bioacoustics approaches toward automated, data-driven monitoring systems.

The thematic map results, which position the "animal welfare–behavior" and "poultry–machine learning" clusters among the motor themes, indicate that AI-based approaches have transitioned from being secondary components to becoming core research axes within the field. Conversely, the placement of themes such as "sound analysis," "stress," and "welfare assessment" in the "emerging/declining" zone suggests not that these areas are losing importance, but rather that they are being reshaped by new technological approaches. Overall, the scientific evolution of stress and welfare monitoring in poultry shows a shift from behavioral observation and classical bioacoustic analysis toward sensor-based monitoring and AI-driven automated assessment systems. However, issues such as dataset standardization, model validation across diverse production conditions, generalizability in real-world farm environments, explainable AI, and multimodal sensor integration remain critical for future research. Therefore, integrating acoustic data with artificial intelligence holds significant potential for developing more objective, continuous, and non-invasive welfare monitoring systems for poultry.

Author Contributions

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

	B.Ş.	CT
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 declare 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>
- Aria, M., Cuccurullo, C., D'Aniello, L., & Spano, M. (2026). Biblioshiny and the SAAS Workflow: An integrated framework for transparent and reproducible science mapping. A demonstration through the replication of a study, *Journal of Informetrics*, DOI: 10.1016/j.joi.2026.101837.
- Aria, M., & Cuccurullo, C. (2026), Science Mapping Analysis - A primer with Biblioshiny, McGraw-Hill, ISBN: 978-88-386-2297-7.
- Coutant, M. P., Villain, A. S., & Briefer, E. F. (2024). A scoping review of the use of bioacoustics to assess various components of farm animal welfare. *Applied Animal Behaviour Science*, 275, 106286. doi:10.1016/j.applanim.2024.106286.

- Laurijs, K. A., Briefer, E. F., Reimert, I., & Webb, L. E. (2021). Vocalisations in farm animals: A step towards positive welfare assessment. *Applied Animal Behaviour Science*, 236, 105264. doi:10.1016/j.applanim.2021.105264.
- Lev-Ron, T., Yitzhaky, Y., Halachmi, I., & Druyan, S. (2025). Classifying vocal responses of broilers to environmental stressors via artificial neural network. *Animal*, 19(1), 101. doi:10.1016/j.animal.2024.101378.
- Manikandan, V., & Neethirajan, S. (2025a). AI-Powered Vocalization Analysis in Poultry: Systematic Review of Health, Behavior, and Welfare Monitoring. *Sensors* (Basel, Switzerland), 25(13), 4058. doi:10.3390/s25134058.
- Manikandan, V., & Neethirajan, S. (2025b). Decoding poultry welfare from sound—a machine learning framework for non-invasive acoustic monitoring. *Sensors*, 25(9), 2912.
- McLoughlin, M. P., Stewart, R., & McElligott, A. G. (2019). Automated bioacoustics: Methods in ecology and conservation and their potential for animal welfare monitoring. *Journal of the Royal Society Interface*, 16(155), 20190225. doi:10.1098/rsif.2019.0225.
- Naeem, M., Jia, Z., Wang, J., Poudel, S., Manjankattil, S., Adhikari, Y., ... & Bourassa, D. (2025). Advancements in machine learning applications in poultry farming: a literature review. *Journal of Applied Poultry Research*, 100602.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Posit Software, PBC. (2026). RStudio: Integrated development environment for R. Posit Software, PBC, Boston, MA.
- R Core Team (2026). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Soster, P. D. C., Grzywalski, T., Hou, Y., Thomas, P., Dedeurwaerder, A., De Gussem, M., Tuytens, F., Devos, P., Botteldooren, D., & Antonissen, G. (2025a). Automated detection of broiler vocalizations: A machine learning approach for broiler chicken vocalization monitoring. *Poultry Science*, 104(5), 104962. doi:10.1016/j.psj.2025.104962.
- Soster, P., Devos, P., Smets, R., Tuytens, F. A., Andretta, I., Buyse, K., De Gussem, M., & Antonissen, G. (2025b). A systematic review linking welfare and automated analysis of acoustic signals in broiler chickens and layers. *World's Poultry Science Journal*, 81(4), 1049–1080. doi:10.1080/00439339.2025.2543042.



TARIMDA YAPAY ZEKA KULLANIMI: MEVCUT UYGULAMALAR, ZOOTEKNİ BOYUTU VE GELECEK PERSPEKTİFLERİ

Hasan ÖNDER^{1*}

¹Ondokuz Mayıs University, Faculty of Agriculture, Department of Animal Science, 55139, Samsun, Türkiye

Özet: Birleşmiş Milletler nüfus projeksiyonlarına göre dünya nüfusunun 2050 yılına kadar 9,7 milyara ulaşacak olması, sınırlı tarım arazileri ve su kaynakları üzerindeki baskıyı artırmış, gıda güvenliğini stratejik bir mesele haline getirmiştir. Geleneksel tarımsal üretim metodolojilerinin karşıya kaldığı iş gücü yetersizliği, iklim değişikliği ve verim kayıpları gibi kronik sorunların çözümünde Yapay Zeka (YZ) teknolojilerinin entegrasyonu hayati role sahiptir. Bu derleme çalışmada, veri toplama mekanizmalarından makine öğrenmesi algoritmalarına, bitkisel üretim ve toprak izlemeye kadar geniş bir yelpazede tarımsal YZ uygulamaları bütüncül bir yaklaşımla incelenmiştir. Özellikle zootekni alanında termal kamera destekli Yapay Sinir Ağları (YSA) ile mastitis teşhisi, otonom yemleme sistemleri ve davranış analizi değerlendirilmiştir. Ayrıca AB'deki agritech girişim sermayesi yatırımları, teknik iş gücü dönüşümü ve geleceğin otonom dikey tarım projeksiyonları tartışılmıştır.

Anahtar kelimeler: Yapay zekâ, Hassas tarım, Sürdürülebilirlik, Akıllı tarım, IoT

AI Usage in Agriculture: Current Applications, Animal Science Perspective, and Future Trends

Abstract: Global population projections indicating a reach of 9.7 billion by 2050 have significantly magnified the pressure on limited arable land and freshwater resources, turning food security into a paramount strategic objective. Artificial Intelligence (AI) integration offers innovative solutions to chronic challenges faced by traditional agricultural practices, such as severe labor shortages, climate variability, and yield inefficiencies. This review paper synthesizes AI applications in modern agriculture, ranging from multi-source data collection and predictive machine learning models to crop monitoring, precision irrigation, and supply chain optimization. Special emphasis is dedicated to Animal Science (Zootechnics), showcasing thermal imaging-based Artificial Neural Network (ANN) algorithms for mastitis detection in dairy cattle, automated feeding technologies, and continuous welfare monitoring. Furthermore, venture capital investments in the EU agritech sector, labor force transitions, structural implementation barriers, and future projections are thoroughly analyzed.

Keywords: Artificial intelligence, Precision agriculture, Sustainability, Smart agriculture, IoT

*Sorumlu yazar (Corresponding author): Ondokuz Mayıs University, Faculty of Agriculture, Department of Animal Science, 55139, Samsun, Türkiye

E mail: hasanonder@gmail.com (H.ÖNDER)

Hasan ÖNDER <https://orcid.org/0000-0002-8404-8700>

Gönderi: 06 Şubat 2025

Kabul: 20 Nisan 2025

Yayınlanma: 15 Haziran 2026

Received: February 06, 2025

Accepted: April 20, 2025

Published: June 15, 2026

Cite as: Önder, H. (2026). AI usage in agriculture: current applications, animal science perspective, and future trends. *Black Sea Journal of Artificial Intelligence*, 2(1), 26-29.

1. Giriş

Küresel ölçekte yaşanan hızlı nüfus artışı ve kentleşme, gıda maddelerine olan talebi hızla artırırken; küresel iklim değişikliği, kuraklık, erozyon ve tarım arazilerinin amaç dışı kullanımı gibi faktörler kullanılabilir kaynakları ciddi oranda daraltmaktadır. Birleşmiş Milletler verilerine göre 2050 yılında 9,7 milyara ulaşması beklenen dünya nüfusunu besleyebilmek için küresel gıda üretiminin en az %60 oranında artırılması gerekmektedir (FAO, 2022). Buna karşılık, kırsal nüfusun yaşlanması ve kentlere göç nedeniyle tarım sektöründe akut bir iş gücü açığı baş göstermektedir.

Yapay Zeka (YZ) kavramı ilk kez 1955 yılında gerçekleştirilen tarihi Dartmouth Konferansı'nda John McCarthy vd. (2006) tarafından tanımlanmıştır ve "öğrenmenin veya zekanın diğer tüm özelliklerinin bir makine tarafından simüle edilebilecek kadar hassas

tanımlanabileceği" hipotezinden yola çıkmıştır. Geçen zaman zarfında bilgisayar bilimlerinin en dinamik alanlarından biri haline gelen YZ; eğitim, sağlık, finans ve imalattan sonar günümüzde tarım ve çevre yönetimi sektörlerinin de merkezine yerleşmiştir.

Yapay Zeka, modern tarım ve çevre yönetiminde giderek daha önemli bir temel taşı haline geliyor. Nüfus artışı, iklim değişikliği ve gıda güvensizliği gibi acil sorunlarla karşı karşıya olan dünyada, YZ, veri analizi, makine öğrenimi ve otomasyondan yararlanarak tarımsal uygulamaları dönüştürebilecek ve sürdürülebilirliği teşvik edebilecek yenilikçi çözümler sunmaktadır.

Modern tarımda YZ; veri analizi, makine öğrenmesi, bilgisayarlı görü (computer vision) ve otomasyon teknolojilerini birleştirerek, geleneksel sezgisel tarım yöntemlerinden veri tabanlı "Hassas Tarım" (Precision Agriculture) modeline geçişi sağlamaktadır. Yapay zekâ, tarımın geleneksel kalıplarını ve sınırlarını yeniden



tanımlayacak gelişmiş yaklaşımlarla kapsamlı bir dönüşüm sunmaktadır (Liu, 2020).

Hassas tarımda yapay zekâ, sadece teknolojik bir evrim değil, daha sürdürülebilir ve verimli bir geleceğe giden bir yolu olarak tanımlanabilir. Paydaşların iş birliği, disiplinler arası araştırma ve sorumlu inovasyon, yapay zekânın tüm potansiyelini çiftçilerin, toplulukların ve küresel gıda sisteminin yararına kullanmada kilit rol oynaması beklenmektedir (Adewusi vd., 2024).

BU çalışmada tarımsal alanda yapay zeka kullanımı ile ilgili kısa bir derleme sunulması amaçlanmıştır.

2. Tarımsal Yapay Zeka Veri Kaynakları ve Veri İşleme Şeması

Tarımsal YZ sistemlerinin başarısı, sahadan toplanan verilerin çeşitliliği, kalitesi ve sürekli güncellenmesine bağlıdır. Veri kaynakları meteoroloji, toprak, tohum, sulama, verim geçmişi, gübreleme, işletme yönetimi, uydu görüntüleri, otonom robotlar ve İHA/dronlardan oluşmaktadır (Tablo 1). Ham verinin karara dönüştürülmesi aşamasında YZ Veri İşleme Şeması kritik rol oynamaktadır (Şekil 1).



Şekil 1. Tarımsal yapay zeka veri işleme ve karar döngüsü

Tablo 1. Tarımsal yapay zeka için veri kaynakları, teknolojileri ve uygulama alanları (Mohan vd., 2023)

Veri Kaynağı	Toplama Teknolojisi	Tarımsal Uygulama ve Karar Alanı
Meteoroloji	İstasyonlar, Mikro-klima Sensörleri	Don riski tahmini, ekim/hasat zamanlaması, sulama planlaması
Toprak	IoT Toprak Nemi ve pH Sensörleri	Besin elementi eksikliği tespiti, taban gübreleme dozajı
Tohum	Genetik Veritabanları, Çimlenme Analizörleri	Çeşit seçimi, iklime dayanıklı varyete geliştirme
Sulama	Debimetreler, Tansiyometreler	Otonom değişken oranlı sulama (VRI) kontrolü.
Geçmiş Verim	Hasat Makinesi Verim Haritalayıcıları	Rekolte tahmini, ekonomik risk analizi
Gübreleme	Spektrofotometreler, NDVI Sensörleri	Hassas yaprak ve toprak gübrelemesi
İşletme Yönetimi	Çiftlik Yönetim Bilgi Sistemleri	Maliyet analizi, iş gücü ve ekipman planlaması
Uydu Görüntüleri	Sentinel, Landsat Uyduları	Bölgesel rekolte ve kuraklık haritalama, bitki indeksi
Otonom Robotlar	LiDAR, RGB-D Kameralar	Otonom yabancı ot ayıklama, hassas meyve toplama
İHA / Dronlar	Multispektral ve Termal Kameralar	Erken hastalık tespiti, nokta atışı ilaçlama

3. Bitkisel Üretimde Yapay Zeka

3.1. Bitki Sağlığı ve Toprak İzleme

Geleneksel izleme yöntemleri zaman alıcıdır. Derin öğrenme tabanlı bilgisayarlı görü algoritmaları, multispektral görüntüler üzerindeki renk ve doku değişimlerini analiz ederek besin eksikliklerini ve fizyolojik stresi erken aşamada tespit eder.

3.2. Zararlı ve Yabancı Ot Yönetimi

Evrişimli Yapay Sinir Ağları (CNN), zararlı böcek türlerini sınıflandırır. Otonom yabancı ot robotları, kültür bitkisi ile yabancı otu milisaniyeler düzeyinde ayırt ederek tarlanın tamamına ilaç sıkmak yerine sadece hedef ota nokta atışı pestisit/herbisit uygular. Bu yaklaşım kimyasal kullanımını %80'e varan oranda azaltır.

3.3. Entegre Ürün Yönetimi ve Karar Destek Sistemleri

YZ tabanlı Karar Destek Sistemleri (DSS), mikro-klima verileri ile hastalık risk modellerini birleştirerek çiftçiye anlık müdahale tavsiyeleri üretir. Gübre, ilaç ve su kullanımını optimize ederek çevresel etkiyi azaltır.

3.4. Tedarik Zinciri Optimizasyonu

YZ algoritmaları; tüketici talebini önceden tahmin etmek, BSJ AI / Hasan ÖNDER

lojistik rotaları optimize etmek ve soğuk zincir sıcaklık kırılmalarını izlemek suretiyle gıda israfını en aza indirir.

4. Hayvansal Üretimde Yapay Zeka

4.1. Hassas Hayvan Besleme

YZ tabanlı otonom yemleme sistemleri (örneğin Real Appetite AI), bireysel yem tüketimini, yemlik başında geçirilen süreyi ve yem yeme frekansını sensörler ile takip eder. Mekanistik yazılımlar hayvanın ağırlığı, laktasyon evresi ve çevre sıcaklığını birleştirerek bireye özel rasyon formüle eder. Bu durum yem dönüşüm oranını (FCR) iyileştirirken metan emisyonunu düşürür.

4.2. Hayvan Sağlığı ve Termal Görüntüleme ile Mastitis Teshisi

Sürü sağlığında subklinik hastalıkların erken tespiti büyük ekonomik önem taşır. Şahin, Mikó, Önder ve Baccouri tarafından yürütülen ve MDPI Animals dergisinde yayınlanan çalışmada (Şahin vd., 2026), Holstein süt sığırlarında mastitis tespiti için Termal Kamera (Infrared Thermography - IRT) verileri ile Yapay Sinir Ağları entegre edilmiştir. Meme loblarının yüzeysel sıcaklık haritasındaki milimetrik değişimleri işleyen YSA

modeli, mastitisi henüz klinik semptomlar ortaya çıkmadan yüksek doğrulukla tespit etmiştir. Bu sayede antibiyotik kullanımı sınırlandırılmakta ve süt kalitesi korunmaktadır.

5. Ekonomik Etkiler ve AB Perspektifi

AB genelinde YZ odaklı tarım teknolojileri (Agri-Tech) girişimlerine yapılan girişim sermayesi yatırımları 2012-2022 döneminde ivmelenerek 2022'de 67 milyon USD ile zirve yapmıştır (European Commission, 2023). Tarımsal YZ otomasyonu beden gücüne dayalı istihdamı azaltırken, robotik sistemleri yönetebilen ve veri tabanlı kararlar alabilen uzman teknik iş gücüne ihtiyacı artırmaktadır. Yapay Zeka uygulamaları, tarım ve hayvancılıkta proaktif bir karar destek mekanizması olarak ekolojik sürdürülebilirlik ile ekonomik karlı arasındaki dengeyi kurmaktadır (Önder, 2026). Birim alandan ve hayvandan alınan verimi artırırken çevresel ayak izini en aza indirmektedir.

6. Zorluklar ve Gelecek Projeksiyonu

Yüksek kurulum maliyetleri, veri gizliliği standartları ve kırsal alandaki dijital okuryazarlık eksikliği temel engellerdir. Gelecekte tarımsal üretimin kapalı alan dikey tarım tesisleri (Vertical Farming) ve otonom robotik sistemler ile tamamen veri güdümlü bir yapıya evrileceği öngörülmektedir.

7. Sonuç

Yapay zeka, hassas tarım uygulamalarında toprak yönetimi, sulama optimizasyonu, yabancı ot kontrolü ve hastalık teşhisi gibi alanlarda devrim niteliğinde dönüşümler yaratmaktadır. Özellikle, Evrişimsel Sinir Ağları ve Nesnelerin İnterneti (IoT) tabanlı sensör ağlarının kullanımı, girdi maliyetlerini (su, gübre, pestisitler) en aza indirirken, birim alan başına verimi en üst düzeye çıkarmaktadır. Bununla birlikte, bu teknolojik dönüşümün önündeki en önemli engeller arasında yüksek kurulum maliyetleri, veri gizliliği standartları ve kırsal alanlarda teknik okuryazarlık eksikliği yer almaktadır.

Yapay zeka (YZ) ve Nesnelerin İnterneti (IoT) tabanlı dijital dönüşüm, modern tarım ve hayvancılık sektörlerinde geleneksel sezgisel yaklaşımlardan veri tabanlı "Hassas Tarım" ve "Hassas Hayvancılık" modellerine geçişi hızlandırmaktadır. Birleşmiş Milletler projeksiyonlarına göre 2050 yılında 9,7 milyara ulaşacak küresel nüfusun gıda ihtiyacını karşılamak; iklim değişikliği, azalan su kaynakları ve daralan tarım arazileri göz önüne alındığında ancak birim alandan ve hayvandan alınan verimin sürdürülebilir biçimde artırılmasıyla mümkündür.

Bu derleme çalışması kapsamında incelenen veriler, YZ'nin hem bitkisel hem de hayvansal üretimde devrim niteliğinde yenilikler sunduğunu ortaya koymaktadır:

1. Bitkisel Üretim ve Girdi Optimizasyonu: Evrişimsel Sinir Ağları (CNN) ve multispektral

görüntüleme teknolojileri, bitki sağlığı, zararlı tespiti ve yabancı ot kontrolünde nokta atışı müdahalelere imkan tanımaktadır. Bu durum pestisit ve herbisit kullanımını %80'e varan oranlarda azaltarak sulama ve gübreleme maliyetlerini düşürmekte, böylece hem ekonomik tasarruf hem de çevresel sürdürülebilirlik sağlamaktadır.

2. Hassas Hayvancılık ve Hayvan Refahı: Hayvansal üretim boyutunda YZ; otonom yemleme sistemleri ile bireye özgü rasyon formülasyonu sunarak Yem Dönüşüm Oranını (FCR) iyileştirmekte ve metan emisyonunu azaltmaktadır. Ayrıca, termal kamera (IRT) verilerinin Yapay Sinir Ağları (YSA) ile entegre edildiği araştırmalar (Şahin vd., 2026), subklinik mastitis gibi kritik hastalıkların henüz semptom göstermeden erken aşamada tespit edilebildiğini göstermiştir. Bu yaklaşım sürü sağlığını korurken, antibiyotik kullanımını sınırlandırmakta ve süt kalitesini güvenceye almaktadır.

3. Ekonomik Dönüşüm ve İş Gücü: AB genelinde Agri-Tech girişim sermayesi yatırımlarının (2022'de 67 milyon USD zirvesi ile) gösterdiği üzere, tarımsal YZ sadece operasyonel bir araç değil, stratejik bir yatırım alanıdır. Otomasyon beden gücüne dayalı geleneksel istihdamı azaltırken, dijital okuryazarlığı yüksek, YZ ve robotik sistemleri yönetebilen nitelikli teknik iş gücüne olan ihtiyacı artırmaktadır.

Bununla birlikte, teknolojinin yaygınlaşmasının önünde yüksek ilk kurulum maliyetleri, kırsal kesimdeki dijital okuryazarlık yetersizliği, veri mülkiyeti ve siber güvenlik gibi altyapısal engeller bulunmaktadır.

Sonuç olarak, Yapay zeka destekli sistemler, ekolojik sürdürülebilirlik ile ekonomik karlı arasındaki dengeyi kuran proaktif bir karar destek mekanizmasıdır. Gelecekte, veri akışının kesintisiz sağlandığı kapalı alan dikey tarım tesisleri (Vertical Farming), otonom tarım robotları ve gelişmiş derin öğrenme algoritmalarının entegrasyonu ile tarımsal üretimin tamamen uçtan uca veri güdümlü bir ekosisteme evrileceği öngörülmektedir.

Katkı Oranı Beyanı

Yazarın katkı yüzdeleri aşağıda verilmiştir. Yazar makaleyi incelemiş ve onaylamıştır.

%	H.Ö.
K	100
T	100
Y	100
VTI	100
VAY	100
KT	100
YZ	100
GR	100
PY	100

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.

Çatışma Beyanı

Yazar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Bilgilendirme ve Teşekkür Beyanı

Bu çalışma, 20 Mayıs 2026 tarihinde Szegedi Tudományegyetem Mezőgazdasági Kar (SZTE MGK - Hódmezővásárhely, Macaristan) ev sahipliğinde gerçekleştirilen 23rd Wellmann International Scientific Conference kapsamında sunulan "AI Usage in Agriculture" başlıklı davetli sunumun genişletilmiş metnidir.

Kaynaklar

Adewusi, A. O., Asuzu, O. F., Olorunsogo, T., Iwuanyanwu, C., Adaga, E., & Daraojimba, D. O. (2024). AI in precision agriculture: A review of technologies for sustainable farming practices. *World Journal of Advanced Research and Reviews*, 21(1), 2276-2285.

European Commission. (2023). Digitalisation in EU Agriculture: Trends, Venture Capital Investments and Labor Market Transformation. EU Agricultural Market Observatory.

FAO. (2022). The State of Food and Agriculture 2022: Leveraging automation in agriculture for transforming agrifood systems. Food and Agriculture Organization of the United Nations, Rome.

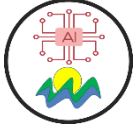
Liu, S. Y. (2020). Artificial intelligence (AI) in agriculture. *IT professional*, 22(3), 14-15.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine*, 27(4), 12-14.

Mohan, S.S., Venkat, R., Rahaman, S., Vinayak, M. & Hari Babu, B. (2023). Role of AI in agriculture: Applications, limitations and challenges: A review. *Agricultural Reviews*. 44(2): 231-237. <https://doi.org/10.18805/ag.R-2215>

Önder, H. (2026). AI Usage in Agriculture. *23rd Wellmann International Scientific Conference*. [Keynote Presentation].

Şahin, H. A., Mikó, E., Önder, H., & Baccouri, W. (2026). Thermal Image-Based Artificial Neural Network Approach to Determine Mastitis Detection in Holstein Dairy Cattle. *Animals*, 16(7), 1048. <https://doi.org/10.3390/ani16071048>



HAYVANSAL ÜRETİM VE BİYOTEKNOLOJİDE YAPAY ZEKÂ VE MAKİNE ÖĞRENMESİ UYGULAMALARI: PROF. DR. UĞUR ŞEN'İN ARAŞTIRMALARI ÜZERİNDEN BİR DEĞERLENDİRME VE GELECEK VİZYONU

Uğur ŞEN^{1*}

¹Ondokuz Mayıs University, Faculty of Agriculture, Department of Agricultural Biotechnology, 55139, Samsun, Türkiye

Özet: Günümüz hayvansal üretim ve tarımsal biyoteknoloji disiplinlerinde, karmaşık verilerin analizi, verimlilik tahmini, fizyolojik ve fenotipik modeller için yapay zekâ (YZ) ve makine öğrenmesi (MÖ) algoritmaları kritik bir role sahip hale gelmiştir. Bu çalışmada, Prof. Dr. Uğur Şen ve çalışma arkadaşlarının literatüre kazandırdığı makaleler YZ ve veri madenciliği perspektifinden sistematik olarak değerlendirilmiştir. Çalışmalarda özellikle XGBoost, Random Forest (RF), Multivariate Adaptive Regression Splines (MARS), Support Vector Machines (SVM) ve LightGBM gibi gelişmiş makine öğrenmesi algoritmalarının; farklı çiftlik hayvanı türlerinde (koyun, keçi, sığır, manda, domuz, tavşan, Pekin ördeği) canlı ağırlık tahmini, vücut ölçütlerinin sınıflandırılması, FAMACHA® skoru tespiti ve genomik tahminleme süreçlerinde yüksek doğrulukla kullanıldığı görülmüştür. Ayrıca, moleküler biyoloji, kas lifi fenotipleme, plasenta morfolojisi ve in vitro embriyo gelişimi odaklı biyoteknolojik verilerin derin öğrenme ve bilgisayarlı görü (computer vision) entegrasyonu gelecekte nasıl daha ileriye taşınabileceği tartışılmıştır. Bu değerlendirme, Hassas Hayvancılık (Precision Livestock Farming - PLF) alanında çalışan araştırmacılara disiplinler arası bir vizyon sunmayı amaçlamaktadır.

Anahtar kelimeler: Yapay zekâ, Makine öğrenmesi, XGBoost, Hassas hayvancılık, Tarımsal biyoteknoloji, Genomik tahmini

Applications of Artificial Intelligence and Machine Learning in Animal Production and Biotechnology: An Evaluation and Future Perspective Based on the Research of Prof. Dr. Uğur Şen

Abstract: In modern animal production and agricultural biotechnology disciplines, artificial intelligence (AI) and machine learning (ML) algorithms have become crucial for analyzing complex data, predicting productivity, and modeling physiological and phenotypic traits. In this study, the articles contributed to the literature by Prof. Dr. Uğur Şen and colleagues were systematically evaluated from an AI and data mining perspective. The reviewed studies demonstrate that advanced machine learning algorithms such as XGBoost, Random Forest (RF), Multivariate Adaptive Regression Splines (MARS), Support Vector Machines (SVM), and LightGBM have been successfully applied across various livestock species (sheep, goats, cattle, buffaloes, pigs, rabbits, Pekin ducks) for body weight prediction, classification of biometric measurements, FAMACHA® scoring, and genomic predictions. Furthermore, the future integration of biotechnological data—focused on molecular biology, muscle fiber phenotyping, placental morphology, and in vitro embryo development—with deep learning and computer vision techniques is discussed. This evaluation aims to provide an interdisciplinary vision for researchers working in the field of Precision Livestock Farming (PLF).

Keywords: Artificial intelligence, Machine learning, XGBoost, Precision livestock Farming, agricultural Biotechnology, Genomic prediction

*Sorumlu yazar (Corresponding author): Ondokuz Mayıs University, Faculty of Agriculture, Department of Agricultural Biotechnology, 55139, Samsun, Türkiye

E mail: ugur.sen@omu.edu.tr (U. ŞEN)

Uğur ŞEN  <https://orcid.org/0000-0001-6131-137X>

Gönderi: 16 Şubat 2025

Kabul: 20 Nisan 2025

Yayınlanma: 15 Haziran 2026

Received: February 06, 2025

Accepted: April 20, 2025

Published: June 15, 2026

Cite as: Şen, U. (2026). Applications of artificial intelligence and machine learning in animal production and biotechnology: An evaluation and future perspective based on the research of Prof. Dr. Uğur Şen. *Black Sea Journal of Artificial Intelligence*, 2(1), 30-33.

1. Giriş

Dünya nüfusundaki hızlı artış, gıda güvenliği ihtiyacı ve iklim değişikliği gibi küresel faktörler; hayvansal üretimde sürdürülebilirliği, verimliliği ve hayvan refahını en üst düzeye çıkarmayı zorunlu kılmaktadır. Son yıllarda tarım ve hayvancılık alanında yaşanan dijital dönüşüm, Hassas Hayvancılık (Precision Livestock Farming - PLF) paradigmasını öne çıkarmıştır. Hassas hayvancılığın temel taşı ise büyük ve karmaşık verilerin

(Big Data) işlenmesini sağlayan Yapay Zekâ (YZ) ve Makine Öğrenmesi (MÖ) teknolojileridir.

Zootekni, veteriner hekimliği ve tarımsal biyoteknoloji alanlarında yürütülen çalışmalarda geleneksel doğrusal parametrik istatistiksel yöntemler (ANOVA, Regresyon vb.), değişkenler arasındaki lineer olmayan (non-linear) karmaşık ilişkileri ve etkileşimleri yakalamakta bazen yetersiz kalabilmektedir. Buna karşın, karar ağacı tabanlı toplu öğrenme (ensemble learning) yöntemleri (Random



Forest, XGBoost, LightGBM) ve esnek regresyon yaklaşımları (MARS), biyolojik verilerin analizinde yüksek doğruluk ve açıklanabilirlik sunmaktadır.

Ondokuz Mayıs Üniversitesi Ziraat Fakültesi Tarımsal Biyoteknoloji Bölümü öğretim üyesi Prof. Dr. Uğur Şen'in akademik külliyatı incelendiğinde; hem klasik ve moleküler biyoteknoloji (plasenta fizyolojisi, in vitro embriyo üretimi, kas lifi gelişimi, gen ifadesi analizleri) hem de veri bilimi ve yapay zekâ uygulamaları (canlı ağırlık tahmini, FAMACHA skoru sınıflandırması, tek adımlı genomik tahminleme) alanında son derece nitelikli ve uluslararası etkiye sahip eserler ürettiği görülmektedir.

Bu çalışmanın amacı, Prof. Dr. Uğur Şen ve çalışma arkadaşlarının yayınladığı araştırmalar üzerinden hayvansal üretim ve biyoteknolojide YZ kullanımını analiz etmek, kullanılan algoritmaların başarısını

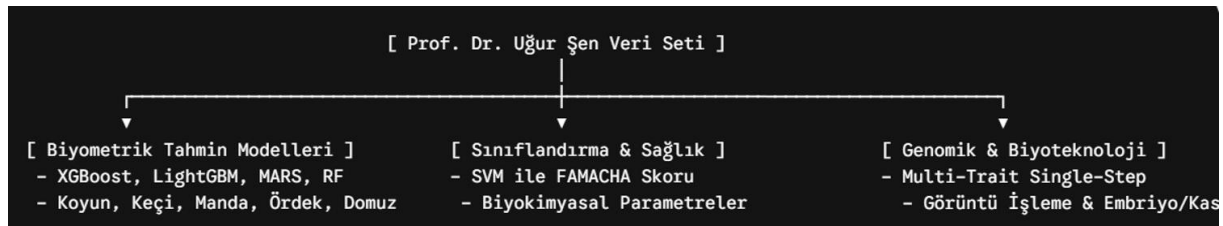
değerlendirmek ve geleceğe yönelik biyoteknoloji-YZ entegrasyonu vizyonu sunmaktır.

2. Metodoloji ve Yayınların Yapay Zekâ Persepektifinden Değerlendirilmesi

Prof. Dr. Uğur Şen'in AVESİS veritabanında yer alan 135'i aşkın bilimsel eseri incelendiğinde, yapay zekâ ve veri analitiği ile doğrudan bağlantılı yayınlar 3 ana kategoride toplanabilir (Şekil 1):

1. Biyometrik Ölçümlerden Canlı Ağırlık ve Performans Tahmini (Makine Öğrenmesi Algoritmaları)
2. Sağlık, Sınıflandırma ve Biyoistatistiksel Modellemeler (Destek Vektör Makineleri ve Karar Ağaçları)

Genomik Veri Analizi ve Çok Özellikli Tahminleme (Multi-Trait Single-Step Genomic Prediction).



Şekil 1. Prof. Dr. Uğur Şen'in AVESİS'te yer alan 135'i aşkın bilimsel eserinin yapay zekâ ve veri analitiği açısından üç ana kategoride incelenmesi.

2.1. Makine Öğrenmesi Algoritmaları ile Canlı Ağırlık ve Fenotip Tahmini

Çiftlik hayvanlarında canlı ağırlığın sahada doğru ve hızlı bir şekilde tahmini, besleme stratejileri, ilaç dozajı belirleme ve damızlık seçimi açısından kritik önem taşır. Prof. Dr. Uğur Şen ve araştırma grubunun bu alandaki çalışmaları, farklı makine öğrenmesi algoritmalarının kıyaslamalı performanslarını ortaya koymuştur:

- XGBoost ve Random Forest Uygulamaları: Delgado vd. (2026) ile Faraz vd. (2023) çalışmalarında, Blackbelly ve Kajli koyunlarında biyometrik ölçümler kullanılarak XGBoost ve Random Forest algoritmaları ile canlı ağırlık tahmini gerçekleştirilmiş; XGBoost'un yüksek R^2 ve düşük RMSE (Kök Ortalama Kare Hata) değerleriyle geleneksel regresyonlara göre üstün performans sergilediği gösterilmiştir.
- MARS ve XGBoost Algoritmalarının Karşılaştırılması: Tyasi vd. (2026) tarafından Kalahari Red keçilerinde yapılan araştırmada, MARS (Multivariate Adaptive Regression Splines) ve XGBoost modelleri sınanmış; doğrusal olmayan ve değişkenler arası etkileşim içeren veri setlerinde esnek regresyon yaklaşımlarının tahmin gücü doğrulanmıştır. Ağyar vd. (2022) ise Anadolu Mandalarında MARS algoritmasını başarıyla uygulamıştır.
- Hibrit ve Karşılaştırmalı ML Modelleri: Tırınk, Önder, Şen vd. (2023 - *PLOS ONE*), Romane koyunlarında veri madenciliği ve ML algoritmalarını karşılaştırarak canlı ağırlık tahmininde model seçim standartlarını

belirlemiştir. Benzer şekilde Önder, Tırınk, Şen vd. (2025), et tipi tavşan hatlarında LightGBM, XGBoost ve SVM algoritmalarını kıyaslayarak LightGBM'in büyük veri setlerinde hız ve doğruluk avantajını vurgulamıştır. Sierra-Vasquez vd. (2026) ise Meksika Yucatecan Criollo domuzlarında ML yaklaşımlarını başarıyla uygulamıştır.

2.2. Destek Vektör Makineleri (SVM) ve Sağlık/Sınıflandırma Modelleri

Parazit yükünün ve aneminin tespiti için küçükbaş hayvancılıkta yaygın kullanılan FAMACHA® skorlaması, göz mukozası rengine dayanan ve uzman deneyimi gerektiren görsel bir yöntemdir. Torres-Chable, Tırınk, Şen vd. (2025 - *Animals*), Pelibuey koyunlarında vücut kondisyon skoru (BCS) ve hematolojik parametrelerden yararlanarak Destek Vektör Makineleri (SVM) algoritması ile FAMACHA® skorlarını yüksek doğrulukla sınıflandırmıştır. Bu çalışma, yapay zekânın sürülerde parazit yönetiminde dijital karar destek sistemi olarak kullanılabileceğini kanıtlamaktadır.

2.3. Genomik Veri Analizi ve İleri Biyoistatistiksel Yaklaşımlar

Yapay zekânın bir diğer boyutu olan karmaşık genomik veri analitiğinde; Önder, Sitkowska, Kurnaz, Piwczyński, Kolenda, Şen vd. (2023 - *Animals*), Polonya Holstein popülasyonunda süt verimi ve süt bileşenleri için Çok Özellikli Tek Adımlı Genomik Tahmin (Multi-Trait Single-Step Genomic Prediction - ssGBLUP) yöntemini kullanmıştır. Bu yaklaşım, fenotipik ve soykütüğü verileri ile SNP marker verilerini birleştirerek genetik değer tahmininde (EBV) doğruluğu artırmıştır.

3. Biyoteknoloji ve Yapay Zekâ Entegrasyonu: Prof. Dr. Uğur Şen'in Çalışmalarının Gelecek Vizyonu

Prof. Dr. Uğur Şen'in temel uzmanlık alanlarından biri de plasenta fizyolojisi, myogenesis (kas lifi tipleri ve Myf5, Myf6, MyoD gen ifadeleri) ve in vitro embriyo biyolojisi üzerine odaklanmaktadır (Şen vd., 2016, 2021, 2025; Şirin, Şen vd., 2017, 2024). Bu biyoteknolojik ve histolojik veriler, YZ ile entegre edildiğinde gelecekte şu inovatif uygulama alanlarını doğurma potansiyeline sahiptir

[Histoloji / Biyoteknoloji Verisi]	[Yapay Zekâ Teknolojisi]	[Uygulama Alanı]
- Kas Lifi Tipolojisi	→ Bilgisayarlı Görü (Computer Vision)	→ Et Kalitesi Otomasyonu
- Plasenta Vaskülarizasyonu	→ Derin Öğrenme (CNN / U-Net)	→ Doğum Ağır. Tahmini
- In Vitro Embriyo Kalitesi	→ Derin Öğrenme / Görüntü Analizi	→ Non-Invasive Embriyo Seçimi

Şekil 2. Plasenta fizyolojisi, myogenesis ve in vitro embriyo biyolojisi verilerinin yapay zekâ ile entegrasyonuna dayalı potansiyel inovatif uygulama alanları.

3.2. Görüntü İşleme ile Plasental Etkinlik ve Doğum Ağırlığı Tahmini:

Şen ve Önder'in (2016, 2022) yerli koyun ve keçi ırklarında incelediği plasenta cotyledon sayıları, yüzey alanı ve vaskülarizasyon verileri, ultrasonik görüntüleme ve YZ görüntü işleme algoritmaları birleştirilerek, gebelik esnasında yavru doğum ağırlığı ve yaşama gücü önceden tahmin edilebilir.

3.3. In Vitro Embriyo Gelişiminde Yapay Zekâ Tabanlı Kalite Derecelendirmesi (Non-Invasive Embryo Scoring)

Şen vd. (2018, 2022) tarafından yürütülen sığır oositlerinin in vitro olgunlaşması ve embriyo gelişimi çalışmalarında, zaman atlamalı (time-lapse) görüntüleme sistemleri YZ algoritmaları ile birleştirilerek, en yüksek gebelik potansiyeline sahip blastokistlerin tahribatsız (non-invasive) olarak seçimi sağlanabilir.

4. Tartışma ve Sonuç

Prof. Dr. Uğur Şen'in yayınları, zootehni ve tarımsal biyoteknolojide klasik deneysel yöntemler ile modern yapay zekâ teknolojilerinin sentezlendiği öncü bir çizgiyi temsil etmektedir.

Araştırmalarda elde edilen bulgular göstermektedir ki:

- XGBoost, LightGBM ve MARS gibi veri madenciliği ve makine öğrenmesi yaklaşımları, hayvansal üretimde canlı ağırlık ve verim parametrelerinin tahmininde klasik istatistiksel modellerden belirgin şekilde daha başarılıdır.
 - SVM ve Sınıflandırma Algoritmaları, parazit yükü (FAMACHA) ve sağlık durumlarının erken teşhisinde güvenilir dijital araçlardır.
 - Biyoteknoloji alanındaki (embriyo, gen ifadesi, kas histolojisi) verilerin yapay zekâ ve derin öğrenme modellerine entegre edilmesi, önümüzdeki dönemde Hassas Hayvancılık (PLF) ve Akıllı Biyoteknoloji uygulamalarına ivme kazandıracaktır.
- Sonuç olarak, Prof. Dr. Uğur Şen ve ekibinin

(Şekil 2).

3.1. Bilgisayarlı Görü (Computer Vision) ile Kas Lifi ve Et Kalitesi Tespiti

Şen ve Şirin'in (2017, 2021) yerli keçi ve koyun ırklarında immunhistokimya ve histolojik boyama ile analiz ettiği kas lifi çapı, alanı ve tipleri (Type I, Type IIA, Type IIX); derin öğrenme tabanlı Convolutional Neural Networks (CNN) ve U-Net mimarileri ile otomatik olarak segmentlere ayrılabilir. Bu sayede manuel sayım hataları sıfırlanabilir.

gerçekleştirdiği çalışmalar, *Black Sea Journal of Artificial Intelligence* yayın odağında yer alan "Yapay Zekânın Biyolojik ve Tarımsal Bilimlere Uygulanması" ilkesiyle tam uyum içerisindedir. Gelecekte biyoteknolojik veri üretimi ile derin öğrenme (Deep Learning) entegrasyonunun artırılması, sektörün dijital dönüşümüne büyük katkı sunacaktır.

Katkı Oranı Beyanı

Yazarın katkı yüzdeleri aşağıda verilmiştir. Yazar makaleyi incelemiş ve onaylamıştır.

	U.Ş.
K	100
T	100
Y	100
VTI	100
VAY	100
KT	100
YZ	100
GR	100
PY	100

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

Yazar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Kaynaklar

- Ağyar, O., Tırınk, C., Önder, H., Şen, U., Piwczynski, D., & Yavuz, E. (2022). Use of Multivariate Adaptive Regression Splines Algorithm to Predict Body Weight from Body Measurements of Anatolian buffaloes in Türkiye. *Animals*, 12(21), 2910.
- Delgado, M. Á. G., Tırınk, C., Parra-Cortés, R. I., Martínez, I. V., Gomez-Vazquez, A., Cruz-Hernandez, A., & Şen, U. (2026). Application of XGBoost and Random Forest algorithms for body weight prediction in Blackbelly sheep using biometric measurements. *BMC Veterinary Research*, 22(1), 15-27.

- Faraz, A., Tırınk, C., Önder, H., Şen, U., Ishaq, H. M., & Tauqir, N. A. (2023). Usage of the XGBoost and MARS algorithms for predicting body weight in Kajli sheep breed. *Tropical Animal Health and Production*, 55(4), 231.
- Önder, H., Sitskowska, B., Kurnaz, B., Piwczynski, D., Kolenda, M., & Şen, U. (2023). Multi-Trait Single-Step Genomic Prediction for Milk Yield and Milk Components for Polish Holstein Population. *Animals*, 13(19), 3050.
- Önder, H., Tırınk, C., Yakubets, T., Getya, A., Matvieiev, M., Kononenko, R., & Şen, U. (2025). Predicting Live Weight for Female Rabbits of Meat Crosses From Body Measurements Using LightGBM, XGBoost and Support Vector Machine Algorithms. *Veterinary Medicine and Science*, 11(1), e1420.
- Sierra-Vasquez, A. C., Tırınk, C., Mezo-Solis, J. A., Önder, H., Cih-Angulo, N., Şen, U., et al. (2026). Application of Machine Learning Algorithms in Estimating Live Weight of Yucatecan Criollo Pigs Through Biometric Measurements. *Animals*, 16(8), 1102.
- Şen, U., & Önder, H. (2022). Determination of Effect of Placental Characteristics on Lamb Birth Weight in Bafra Sheep by Path Analysis. *Anadolu Tarım Bilimleri Dergisi*, 37(3), 479-486.
- Şen, U., Şirin, E., Filik, A. G., Önder, H., Piwczynski, D., & Kolenda, M. (2021). Placental Characteristics Classification of Various Native Turkish Sheep Breeds. *Animals*, 11(4), 1012.
- Tırınk, C., Önder, H., Francois, D., Marcon, D., Şen, U., Shaikenova, K., et al. (2023). Comparison of the data mining and machine learning algorithms for predicting the final body weight for Romane sheep breed. *PLoS ONE*, 18(8), e0289110.
- Torres-Chable, O. M., Tırınk, C., Parra-Cortes, R. I., Delgado, M. A. G., Martinez, I. V., Gomez-Vazquez, A., & Şen, U. (2025). Classification of FAMACHA© Scores with Support Vector Machine Algorithm from Body Condition Score and Hematological Parameters in Pelibuey Sheep. *Animals*, 15(5), 682.
- Tyasi, L. T., Tırınk, C., Mokoena, K., Önder, H., Şen, U., & Boga, D. C. (2026). Comparison of the MARS and XGBoost algorithms for predicting body weight in Kalahari Red goats. *Turkish Journal of Veterinary & Animal Sciences*, 50(1), 45-53.