Kafkas Universitesi Veteriner Fakultesi Dergisi
Journal Home-Page: http://vetdergikafkas.org
E-ISSN: 1309-2251

Kafkas Univ Vet Fak Derg
x (x): x-x, 2026
DOI: 10.9775/kvfd.2025.35867

Research Article

Thematic Mapping and Sustainability-Oriented Analysis of Veterinary Science Research in Türkiye: A Text Mining and Hybrid Hierarchical Clustering Approach (1980-2024)

Harun YONAR [1(*)], Furkan Çağrı BEŞOLUK [1], Aynur YONAR [2]

[1] Selcuk University, Faculty of Veterinary Medicine, Biostatistics Department, TR-42130 Konya - TÜRKİYE
[2] Selcuk University, Faculty of Science, Statistics Department, TR-42130 Konya - TÜRKİYE

(*) **Corresponding author:**
Harun Yonar
Phone: +90 332 223 3550
Cellular phone: +90 545 311 2080
E-mail: hyonar@selcuk.edu.tr

How to cite this article?

Yonar H, Beşoluk FÇ, Yonar A: Thematic Mapping and Sustainability-Oriented Analysis of Veterinary Science Research in Türkiye: A Text Mining and Hybrid Hierarchical Clustering Approach (1980-2024). *Kafkas Univ Vet Fak Derg*, 2026 (Article in Press).
DOI: 10.9775/kvfd.2025.35867

Article ID: KVFD-2025-35867
Received: 05.12.2025
Accepted: 11.04.2026
Published Online: 21.04.2026

Abstract

This study examines 45,769 veterinary medicine publication titles from Türkiye (1980–2024) using large-scale text mining, hybrid clustering, and Sustainable Development Goals (SDG) mapping. It aims to identify long-term thematic trends and assess alignment with global sustainability priorities. A two-stage hybrid clustering approach (k-means + hierarchical) revealed 11 thematic groups. SDG alignment was evaluated using a hybrid Aurora-Elsevier dictionary model enhanced with n-gram-based weighting and validation. Findings indicate a clear shift from traditional species- and production-focused research toward molecular, experimental, and data-driven domains. SDG mapping shows strong associations with Zero Hunger (SDG 2), Good Health and Well-being (SDG 3), Responsible Consumption and Production (SDG 12), and Life Below Water (SDG 14). In contrast, Climate Action (SDG 13) and Life on Land (SDG 15) remain underrepresented, highlighting critical gaps in environmental sustainability. Overall, veterinary research in Türkiye aligns with global production- and health-oriented trends but exhibits partial thematic divergence, particularly in aquatic systems. The comparatively lower emphasis on climate change, biodiversity, and ecosystem health suggests areas that could be further strengthened within the national research agenda. These results offer a data-driven basis for strengthening interdisciplinary research, advancing the environmental dimension of One Health, and improving alignment with global sustainability priorities.

Keywords: Hybrid clustering, One health, Sustainable development goals, Text mining, Türkiye, Veterinary medicine

Introduction

Veterinary medicine plays a critical role at the intersection of animal health, public health, food safety, environmental sustainability, and global health security. Beyond clinical practice, the discipline contributes significantly to the prevention of zoonotic diseases, the fight against antimicrobial resistance, sustainable livestock production systems, the conservation of biological diversity, and the support of ecosystem health within the One Health approach [1-3]. In this context, veterinary science is directly related to the Sustainable Development Goals (SDGs), particularly SDG 2 (Zero Hunger), SDG 3 (Good Health and Well-being), SDG 12 (Responsible Consumption and Production), SDG 13 (Climate Action), and SDG 15 (Life on Land). The strategic position of veterinary medicine in contributing to these global sustainability goals makes it important to understand the alignment of national research orientations with global priorities [4-6].

Türkiye is a strategic country in regional and global veterinary management due to its geopolitical position, animal production capacity, biological diversity, and wildlife corridors. This position makes veterinary research critical for both national animal husbandry and public health, as well as global disease surveillance, control of zoonotic risks, and sustainable food systems. Strengthening veterinary research capacity in Türkiye directly supports regional biosecurity and international collaborations such as FAO, WOAH, and WHO, in line with One Health initiatives [7-9].

In recent years, veterinary education and research infrastructure in Türkiye have developed significantly, accelerating international collaborations and integration with digital epidemiology, genomics, precision livestock farming, and sustainable agriculture policies [10,11].

Despite the growing use of text-based approaches in medicine, agriculture, and environmental sciences, the application of these methods in veterinary research within an SDG framework remains an emerging area.

This study examines 45,769 veterinary science publication titles from Türkiye (1980–2024) using text mining, hybrid hierarchical clustering, and SDG-based classification to reveal (i) the transformation of Turkish veterinary research trends over time, (ii) thematic clustering structures among departments, and (iii) the alignment of publication content with the SDGs – contributing to sustainability-focused scientific mapping in the international literature.

Material and Method

Ethical Statement

This study did not involve any procedures requiring ethical approval.

Material

The data used in this study includes the title, year, and division/department information of scientific articles published in the field of veterinary science in Türkiye between 1980 and 2024 [12]. The analyses were performed on article titles; the total data set consists of 45,769 unique titles obtained after cleaning.

Method

To ensure methodological transparency and reproducibility, the analysis was structured into four stages: (i) data preprocessing, (ii) TF-IDF-based text representation with SVD-based dimensionality reduction and scaling, (iii) hybrid clustering (k-means + hierarchical), and (iv) SDG mapping and validation using a hybrid dictionary approach with ranking-based evaluation. This workflow ensures a coherent and reproducible analytical process.

Data Preprocessing and Text Standardization

The texts were first subjected to a comprehensive preprocessing process. In the first stage, all titles were run through a language filter to identify those in Turkish and were translated into English to ensure linguistic consistency. Subsequently, the texts underwent cleaning, noise filtering, and tokenization processes.

Text Representation and Dimension Reduction

In the first stage, article titles were converted into high-dimensional feature vectors based on TF-IDF using a text mining approach. TF-IDF representations derived from text data typically have a high-dimensional and sparse feature space. This situation can negatively affect the performance and computational cost of clustering algorithms. Therefore, to reduce the dimensionality problem and preserve the fundamental variation components in the data structure, a dimension reduction process was applied using the Singular Value Decomposition (SVD) method. To ensure comparability among variables in the clustering analysis, the resulting vectors were subsequently scaled.

Hybrid Clustering (k-means + hierarchical)

Cluster analysis is an unsupervised learning approach for identifying homogeneous groups based on similarity patterns. In text mining, clustering reveals latent thematic structures; performance depends on appropriate feature representation, dimensionality reduction, and distance metric selection.

Due to the large-scale nature of the dataset in this study, a two-stage hybrid clustering (k-means + hierarchical) was chosen for the clustering analysis. Hybrid clustering strategies are widely recommended in the literature, particularly for large datasets, both to reduce computational costs and to enhance the interpretability of the resulting cluster structure. This approach combines the computational efficiency of the k-means algorithm with the structural interpretability provided by hierarchical clustering, offering an effective solution for revealing the thematic structures of large-scale datasets [13–18].

Pre-Clustering and Determination of Representative Centers

Given the complexity of hierarchical clustering, applying it directly to 45,769 observations would require approximately 2.1 billion pairwise distance calculations, which is computationally infeasible. By reducing the dataset to M=500 representative centroids via k-means pre-clustering, this number decreases to about 250,000. Thus, M=500 was selected as a balance between computational feasibility and preserving the global structure of the data.

Hierarchical Clustering Analysis

In the second stage, hierarchical clustering analysis was performed on the representative centers obtained using the Ward.D2 linkage method. The Ward method is an agglomerative approach aimed at minimizing intra-cluster variance and is widely used, particularly in text-based datasets. Similarities were calculated using the Euclidean distance during the clustering process.

Euclidean distance was selected as the similarity measure given the continuous, standardized feature space. Feature scaling ensured equal contribution of all dimensions. SVD-based dimensionality reduction addressed the high-dimensional, sparse nature of TF–IDF representations, reducing noise and preserving informative latent structures.

Determination of the Optimal Number of Clusters

Candidate cluster numbers were evaluated in the range k=2-30 using four internal validation indices: Silhouette (higher=better intra-cluster cohesion [19]), Calinski-

Harabasz (higher = stronger separation), Dunn (higher = better structure), and Davies-Bouldin (lower = better performance). As different indices may suggest varying optimal k values [20,21], a consensus across multiple indices was applied [22-24], supplemented by cluster size distribution and inter-cluster variance ratio to assess over-fragmentation.

While validation indices serve as an important guide in determining the final number of clusters, the thematic consistency and interpretability of the clusters played a decisive role in the decision-making process. This approach is of critical importance in ensuring that the resulting clusters provide not only statistically valid but also meaningful and interpretable thematic structures [25-27].

SDG Mapping Using the Hybrid Dictionary Approach

A hybrid dictionary content and text mining approach was applied to determine the relationship between the articles in the study and the United Nations' 17 Sustainable Development Goals (SDGs). The analysis process was built on a two-stage design by combining the conceptual structures obtained from the Aurora [28] and Elsevier [29] SDG dictionaries [30-32]. The analysis was conducted in two consecutive design steps. In Design 1, the XML-based SDG query files of the Aurora platform were parsed, and the terms derived from this were converted into a systematic lexicon structure under the name "Aurora SDG Lexicon." In Design 2, the "Elsevier SDG Lexicon" was created using terms extracted from Elsevier's 2025 Sustainable Development Goals (SDGs) Mapping query files; then, it was integrated with the Aurora lexicon to develop a hybrid SDG assignment model based on source-based weighting. In Elsevier queries, expressions in double quotation marks were parsed as "phrases," while single words were parsed as "unigrams", and each was matched with the relevant SDG number. These terms were compared with article titles tokenized at the unigram, bigram, and trigram levels. Matching scores were calculated considering n-gram length (tri > bi > uni) and term type (phrase > unigram) weights. This hybrid structure enabled multi-source information integration at both the conceptual and linguistic levels; the most probable SDG label was identified for ranking and validation purposes based on the total score, number of matches, and maximum n-gram level, while multiple SDG matches were retained under the multi-label classification framework. Thus, a hybrid text-matching approach that integrates the strengths of dictionary-based approaches and n-gram matching techniques was developed [33,34].

Validation of the SDG Model

To assess the reliability of the SDG assignment model, a manual validation procedure was conducted using a stratified sample of article titles based on the highest detected n-gram level. This sampling strategy was adopted to separately evaluate the performance of the n-gram-based matching mechanism, which represents a fundamental component of the proposed hybrid system. In this context, 100 article titles were randomly selected from each n-gram category (trigram, bigram, and unigram) to ensure a representative evaluation. As a result, SDG labels were manually assigned to a total of 300 article titles, and these labels were compared with the SDG candidates generated by the model.

Manual verification was performed independently by two authors, and disagreements were resolved through discussion until consensus was reached. In the verification analysis, the manually assigned SDG labels were compared with the ranked SDG candidates generated by the model. Model performance was evaluated using ranking-based metrics, including coverage and Top-k accuracy (Top-1, Top-3, and Top-5). Coverage indicates the proportion of cases where the manually assigned SDG is present among the candidate SDGs generated by the model, while Top-k accuracy indicates whether the correct SDG is among the top k ranked positions. These evaluation metrics are widely used in SDG mapping and classification studies because such systems typically generate ranked lists of candidate SDGs rather than single-label predictions [35,36].

Multi-label Approach

SDG assignment was performed using a multi-label matching approach. A single article title may contain expressions related to multiple SDGs and can therefore be associated with more than one SDG. Consequently, the values derived from this process represent association counts, defined as the number of article titles matched with each SDG, rather than mutually exclusive category assignments. Percentages reported in the results are calculated based on these association counts; therefore, the total percentage may exceed 100%, which is expected in multi-label classification analyses.

Analysis of Cluster-SDG Relationships (Heatmap)

A heatmap visualization was created to examine the relationships between the identified thematic clusters and the Sustainable Development Goals (SDGs) more clearly. First, using the hybrid SDG lexicon approach, it was determined which SDGs each article title matched. Then, the SDG match frequencies were calculated for each cluster. To determine the relative weight of each SDG within a cluster, the number of SDG matches for that SDG was divided by the total number of SDG matches within the same cluster to obtain percentage values. These percentage values were visualized on the heatmap. In the heatmap, rows represent clusters, while columns represent the SDGs (SDG01–SDG17). Color intensity indicates the relative density of SDG matches within the respective cluster.

Software and Packages

All analyses were conducted in R (v4.5.1). Data handling used readxl, readr, dplyr, tidyr, stringr, and tidyverse; tokenization and n-gram generation used tidytext; lemmatization used textstem; TF–IDF construction used text2vec; stop-word filtering used stopwords and tokenizers. Dimensionality reduction used irlba; clustering used fastcluster, proxy, and RANN; validation used cluster and igraph. Visualization used dendextend, factoextra, ggplot2, gridExtra, and forcats.

Results

Publication Trends by Year (1980-2024)

Annual publication output in veterinary medicine from Türkiye increased steadily from the 1980s, accelerating after 2000 and exceeding 1,000 per year; output first exceeded 3,000 in 2015, followed by a slight decline in 2019 and 2024 *(Fig. 1)*.

Publication Distribution by Division and Department

Fig. 2 shows the publication volume in the veterinary field in Türkiye by division and department between 1980 and 2024.

Preclinical Sciences accounted for the highest publication share (29.9%), followed by Clinical Sciences (29.0%), Basic Sciences (19.3%), Zootechnics and Animal Nutrition (15.8%), and Food Hygiene and Technology (6.1%) *(Fig. 2)*.

Division and Department Collaboration Network

Fig. 3 shows the network structure of joint publication relationships at the division and department levels in the field of veterinary science. Each node represents a department, the edge thickness represents the frequency of joint publications between two departments, and the coloring represents the divisions.

The strongest inter-departmental connections occur between Preclinical and Clinical Sciences, with Basic Sciences (Biochemistry, Physiology, Histology) bridging both divisions. Zootechnics and Animal Nutrition has multiple connections, while Food Hygiene and Technology shows more limited network links *(Fig. 3)*.

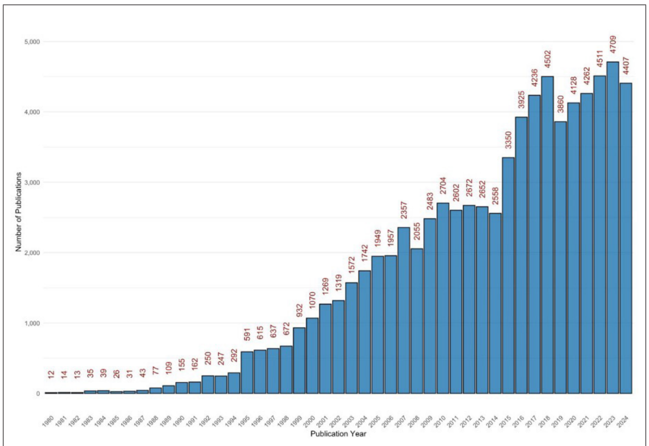

Fig 1. Annual publication numbers in the field of veterinary medicine between 1980 and 2024

Unigram and Bigram Frequency Analyses

Fig. 4 shows the distribution of the 20 most frequently occurring single words (unigrams) and two-word expressions (bigrams) in the titles of veterinary research in Türkiye.

The most frequently occurring terms in the unigram analysis were dog, rat, turkey, sheep, cattle, cow, and goat. Among methodological terms, investigation, parameter, evaluation, treatment, and performance showed high frequency. The most frequently used bigram in the bigram analysis was oxidative stress. This was followed by dairy cows, biochemical parameters, blood parameters, escherichia coli, rainbow trout, and broiler chickens.

Trend Topics Diagram

Fig. 5 and *Fig. 6* display, for the 30 most frequent unigrams and bigrams respectively, the first appearance, peak usage, and last appearance year as horizontal trend lines.

Species-based terms (dog, sheep, cattle, goat, turkey) dominated the 1980-2000 period. After 2000, molecular-focused terms such as oxidative, stress, biochemical, and antioxidant increased rapidly, reaching peak frequencies between 2020–2024 alongside production-related terms (broiler, dairy, meat) *(Fig. 5)*.

Production-related bigrams (broiler chickens, dairy cows, rainbow trout) appeared consistently from the 1980s. Post-2000, mechanistic bigrams (oxidative stress, biochemical parameters, lipid peroxidation, heat stress) became increasingly frequent, with marked increases in oxidative stress and biochemical parameters after 2020 *(Fig. 6)*.

Examining SDG Associations Using a Hybrid Dictionary-Based Approach

Fig. 7 shows the pattern of SDG associations obtained by applying a hybrid dictionary-based approach to all article titles published between 1980 and 2024.

The results indicate that academic publications are most strongly associated with SDG 2 (36.3%), followed by SDG 14 (19.0%), SDG 3 (18.1%), and SDG 12 (15.8%). Goals with moderate association levels include SDG 4 (12.3%), SDG 6 (10.6%), SDG 8 (11.6%), SDG 11 (11.9%), and SDG 17 (9.3%). Lower association levels are observed for SDG 1 (8.6%), SDG 5 (7.6%), SDG 7 (8.5%), SDG 9 (5.7%), SDG 10 (7.0%), SDG 13 (6.4%), SDG 15 (8.0%), and SDG 16 (5.2%).

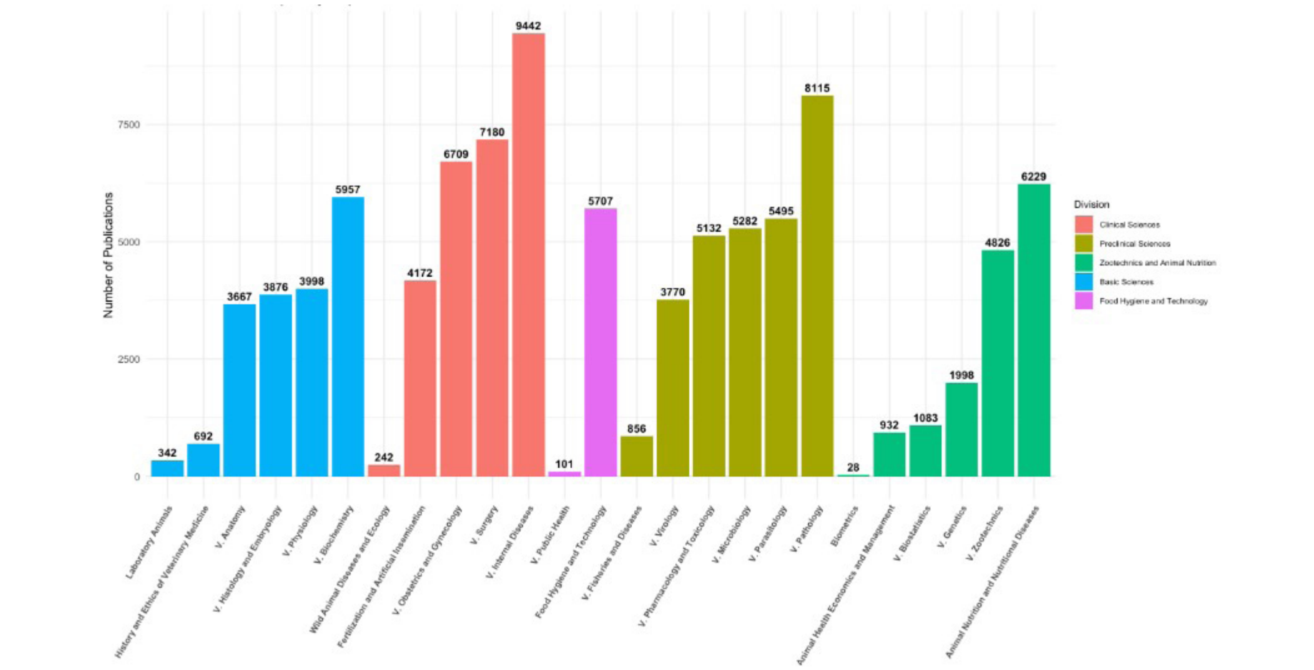

Fig 2. Distribution of publications by veterinary departments according to divisions

Validation of the SDG Classification Model

Comparative validation results for the Aurora, Elsevier, and Hybrid SDG models across different n-gram levels and for the overall validation dataset are presented in *Table 1*.

The hybrid model outperformed both Aurora and Elsevier models across all n-gram levels (*Table 1*). For the overall validation set (n = 300), hybrid coverage was 0.755 vs. Aurora 0.631 and Elsevier 0.359; Top-3 accuracy was 0.645 vs. 0.510 and 0.355, respectively, confirming that the hybrid approach provides more reliable identification of SDG-related themes.

These results indicate that the hybrid SDG assignment approach provides a more reliable identification of SDG-related themes in article titles compared with the single-dictionary approaches.

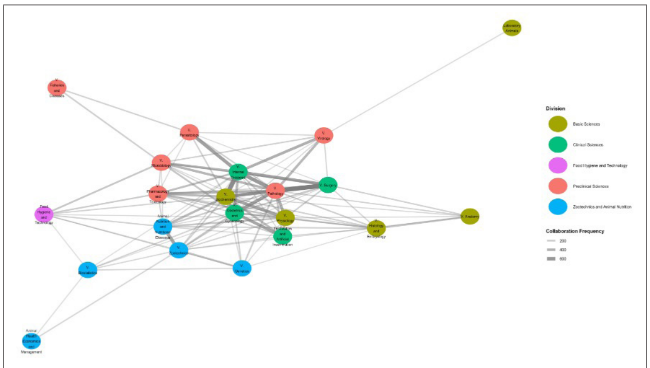

Fig 3. Collaborative publication network among veterinary departments

SDG and Division Relationship

Fig. 8 shows the top 5 veterinary divisions contributing most to each SDG, revealing the distribution of research production in these divisions from a sustainability perspective.

Clinical Sciences and Preclinical Sciences lead across the majority of SDGs. Basic Sciences ranks among the top three for nearly all SDGs, while Zootechnics and Animal Nutrition stands out particularly in SDG 2, 8, and 17. Overall, the findings suggest that interdisciplinary contributions within veterinary sciences exhibit a thematically differentiated yet balanced distribution across the SDG portfolio.

Cluster Validation and Dendrogram

Cluster validation indices for k=2-30 indicated the Silhouette maximum at k=9 (0.288), with k=10–12 forming a stable region (0.186-0.220). Calinski-Harabasz

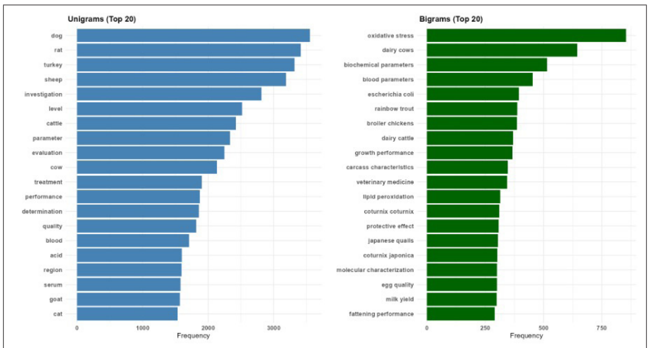

Fig 4. Top 20 unigrams and bigrams from article titles

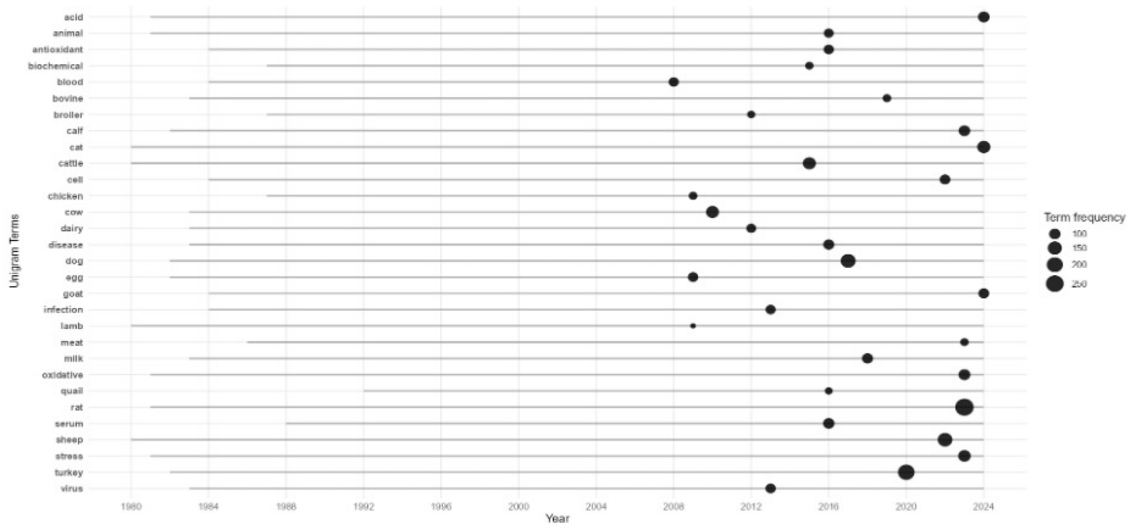

Fig 5. Trend analysis of unigram terms by year

increased gradually through this range (10: 12.30; 11: 12.56; 12: 12.94) while Davies-Bouldin reached its minimum at k=11 (2.17). The Dunn index peaked locally at k=12 (0.145) and k=15 (0.150), but increases at higher k were attributed to fragmentation effects. At k≥12, minimum cluster size dropped substantially; the k=11 solution yielded a minimum size of 7 and a between-cluster variance ratio of 0.197, representing the optimal balance of separation, interpretability, and balance.

When all validation indices were considered together, the range of k = 10-12 emerged as an appropriate solution region in terms of clustering quality. Within this range, the k = 11 solution was selected as the optimal cluster structure, as it (i) ensured a balanced distribution of cluster sizes, (ii) provided adequate cluster separation, and (iii) prevented over-fragmentation. Therefore, k = 11 was determined as the final number of clusters that best represents the thematic structure of the dataset from both statistical and interpretability perspectives.

Eleven thematic clusters were obtained as a result of two-stage hybrid clustering. The dendrogram generated using the hybrid structure (k-means & Ward) shows distinct clustering between clusters.

Fig. 9 shows the final dendrogram resulting from the Ward. D2 hierarchical clustering applied to the 500 centroids obtained in the k-means preprocessing stage.

The dendrogram (*Fig. 9*) shows five main groupings: Cluster 3 as a dominant independent block; Clusters 2-6-9 on a shared main branch; Clusters 5-8 in the same region; Clusters 7-10 under a common branch; and Clusters 1-4-

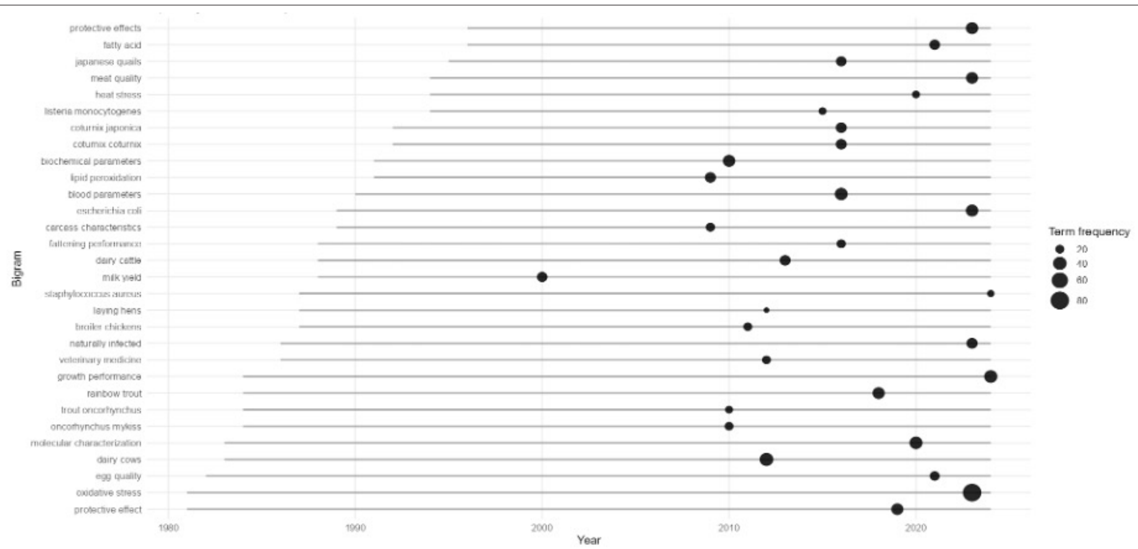

Fig 6. Trend analysis of bigram terms by year

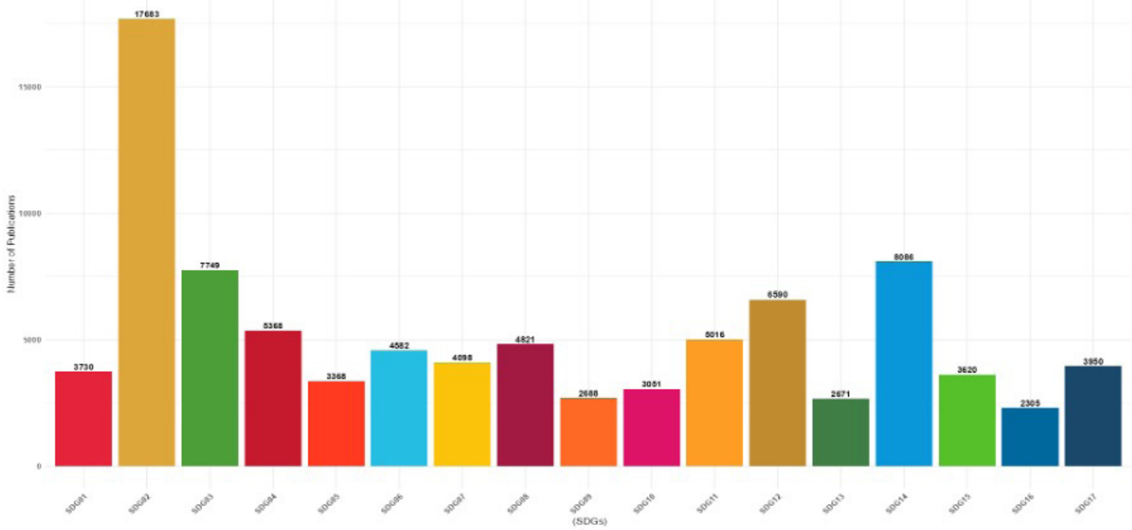

Fig 7. Article-SDG associations identified using a hybrid dictionary-based multi-label matching approach (1980-2024). A single article title may be associated with more than one SDG because the SDG assignment follows a multi-label matching approach. Therefore, the total counts across SDGs may exceed the total number of articles. The values shown in the figure represent the number of article titles matched with each SDG, while percentages reported in the text are calculated based on these counts

11 within the same large block. Long top-level branches indicate high semantic distances, while Clusters 7, 10, 9, and 6 form compact, closely related structures.

Content Characteristics of Clusters

Table 2 shows the terms with the highest TF-IDF weight and cluster sizes for each of the 11 thematic clusters obtained using the hybrid clustering method.

Cluster 3 (n=27,081) is the largest and most heterogeneous structure, dominated by species terms. Clusters 2, 6, and 9 form a production-biochemical-molecular branch; Clusters 5 and 8 a pathogen-dairy branch; Clusters 7 and 10 represent aquaculture and poultry-egg specializations; Clusters 1, 4, and 11 form a clinical-institutional-public health block *(Table 2, Fig. 9)*.

The Evolution of Clusters Over the Years

Fig. 10 shows the annual number of articles in the thematic cluster for the period 1980-2024. In each panel, the black line represents the number of publications in the relevant cluster in the relevant year.

Cluster 3 shows the largest and most steadily increasing trend from the 1980s; Cluster 2 rose markedly after 2000; Clusters 5 and 6 show steady growth. Smaller clusters (1,

Table 1. *Comparative validation performance of Aurora, Elsevier, and Hybrid SDG models across different n-gram levels*

N-gram Level	Model	n	Coverage	Top-1	Top-3	Top-5
Trigram	Hybrid	100	**0.833**	**0.456**	**0.711**	**0.744**
	Aurora	100	0.689	0.311	0.556	0.611
	Elsevier	100	0.511	0.467	0.511	0.511
Bigram	Hybrid	100	**0.780**	**0.420**	**0.650**	**0.750**
	Aurora	100	0.590	0.330	0.430	0.540
	Elsevier	100	0.490	0.390	0.480	0.490
Unigram	Hybrid	100	**0.660**	**0.300**	**0.580**	**0.620**
	Aurora	100	0.620	0.280	0.550	0.590
	Elsevier	100	0.090	0.080	0.090	0.090
Overall	Hybrid	300	**0.755**	**0.390**	**0.645**	**0.703**
	Aurora	300	0.631	0.307	0.510	0.579
	Elsevier	300	0.359	0.307	0.355	0.359

Coverage and Top-k accuracy values represent proportions of correct matches. Top-k accuracy indicates the proportion of cases in which the manually assigned SDG appears within the first k ranked SDG candidates generated by the model

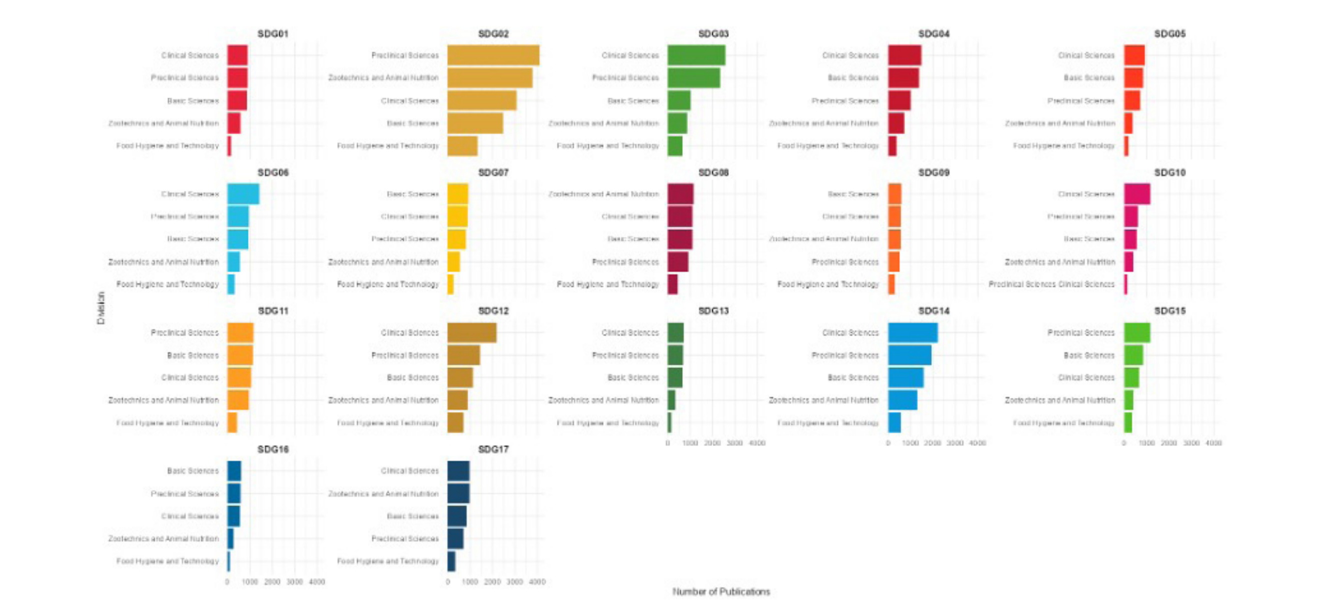

Fig 8. Comparative analysis of publication contributions by veterinary divisions according to SDGs

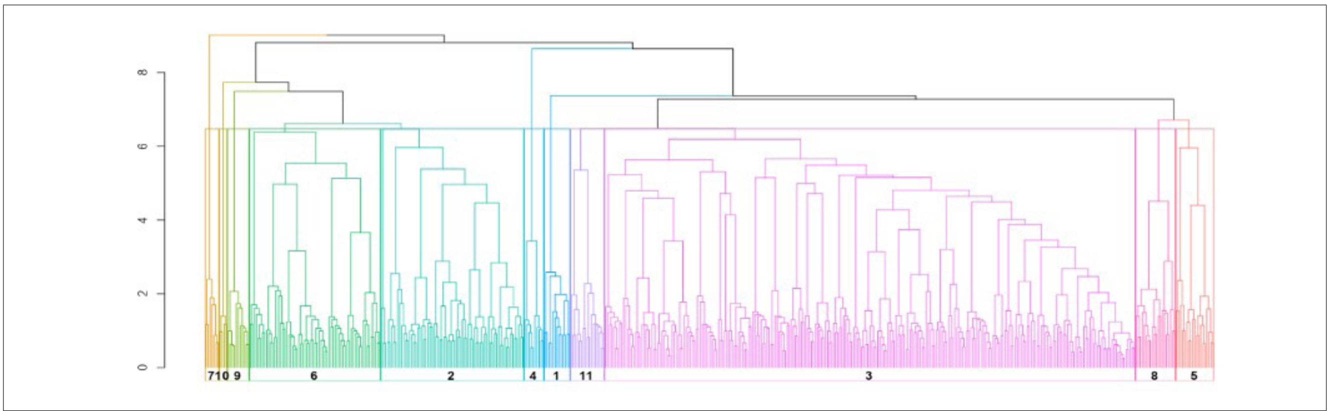

Fig 9. Dendrogram obtained using the hybrid clustering method (k=11)

4, 7, 8, 9, 10, 11) show lower publication volumes and no discernible trend *(Fig. 10)*.

The Relationship Between Thematic Clusters and the Sustainable Development Goals (Heatmap Analysis)

Fig. 11 presents a heatmap of within-cluster relative SDG association densities, with rows representing thematic clusters and columns representing SDGs (SDG01-SDG17); color intensity reflects the percentage of SDG matches within each cluster.

SDG02 exhibits relatively higher density in Clusters 2, 5, 8, 10, and 11; SDG03 is more prominent in Cluster 5; SDG14 in Cluster 7; and SDG06 in Cluster 9. Other SDGs show a more balanced distribution across clusters *(Fig. 11)*.

Discussion

This study provides a comprehensive analysis of the thematic evolution of veterinary science in Türkiye (1980–2024) based on 45,769 publications, integrating hybrid clustering and SDG alignment within a unified analytical framework [37-43]. The hybrid clustering approach reduced the computational load on the large dataset [14,44], while the n-gram-weighted combination of Aurora and Elsevier SDG dictionaries substantially reduced the risk of context-disconnected matching inherent in single-dictionary approaches [32,45].

International scientometric research indicates that sustainability research is unevenly distributed across SDGs, with environmental goals relatively underrepresented [46], a pattern reflected in veterinary and animal sciences where production, health, and clinical fields dominate over environmental topics [42].

Global veterinary research capacity remains geographically skewed toward Western Europe and North America, while emerging countries like Türkiye have concentrated thematically on specific areas [4]. Methodological

***Table 2.** Dominant terms and publication count of the cluster*

Cluster	Representative Terms (Top-TF-IDF)	n
1	case, case_report, report, dog, calf, cat, dog_case, congenital, cause, two, myiasis, atresia, treatment, case_congenital, calf_case	1122
2	effect, performance, parameter, egg, broiler, different, lamb, characteristic, biochemical, blood, hen, in vitro, diet, quail, carcass	5642
3	dog, turkey, sheep, effect, study, cattle, treatment, investigation, use, cat, evaluation, disease, region, goat, infection	27081
4	veterinary, medicine, veterinary_medicine, faculty, university, faculty_veterinary, university_faculty, journal, clinic, veterinary_faculty, education, animal, student, turkish_veterinary, history	690
5	isolate, resistance, coli, escherichia, escherichia coli, antibiotic, staphylococcus, listeria, monocytogenes, listeria monocytogenes, o, aureus, staphylococcus aureus, antibiotic resistance, strain	1582
6	rat, effect, acid, activity, sperm, protective, antioxidant, vitamin, protective_effect, ram, damage, model, injury, lipid, e	5347
7	trout, rainbow, rainbow_trout, mykiss, oncorhynchus, oncorhynchus_mykiss, trout_oncorhynchus, walbaum, mykiss_ walbaum, effect, isolate_rainbow, isolate, w, mykiss_w, farm	389
8	milk, cow, dairy_cow, dairy, milk_yield, yield, mastitis, subclinical, subclinical_mastitis, effect, lactation, somatic_cell, somatic, relationship, holstein	1480
9	oxidative stress, oxidative, stress, rat, effect, apoptosis, inflammation, stress inflammation, stress parameter, stress rat, parameter, damage, protective, inflammation apoptosis, antioxidant	755
10	Coturnix, Coturnix coturnix, japonica, Coturnix japonica, quail coturnix, quail, Japanese, Japanese quail, egg, effect, performance, weight, hatch, egg weight, egg quality	319
11	animal, food, health, importance, public_health, public, nutrition, use, safety, domestic_animal, welfare, food_safety, animal_ nutrition, animal_welfare, human	1362

variability across SDG-based analyses limits direct quantitative comparisons; findings are therefore evaluated at the contextual and conceptual level.

The dominance of SDG 2 (Zero Hunger) and SDG 3 (Good Health and Well-being) reflects veterinary science's long-standing focus on animal production, food safety, and zoonotic disease control - a structure aligned with global veterinary orientation rather than a national deviation [11,47-49].

The cluster evolution data corroborate the global shift toward molecular and data-driven veterinary research [41,50-53]: species- and production-focused studies dominated before 2000, after which molecular-biochemical terms (oxidative stress, antioxidant activity, lipid peroxidation) rapidly gained prominence, consistent with methodological diversification in the global literature [4,54,55].

The relatively high representation of SDG 14 (19%) is notable, as this goal is predominantly associated with marine sciences globally [46]. Türkiye's prominence likely reflects its aquaculture production capacity and geographical advantages, evidenced by the concentration of rainbow trout and aquaculture terms in Cluster 7,

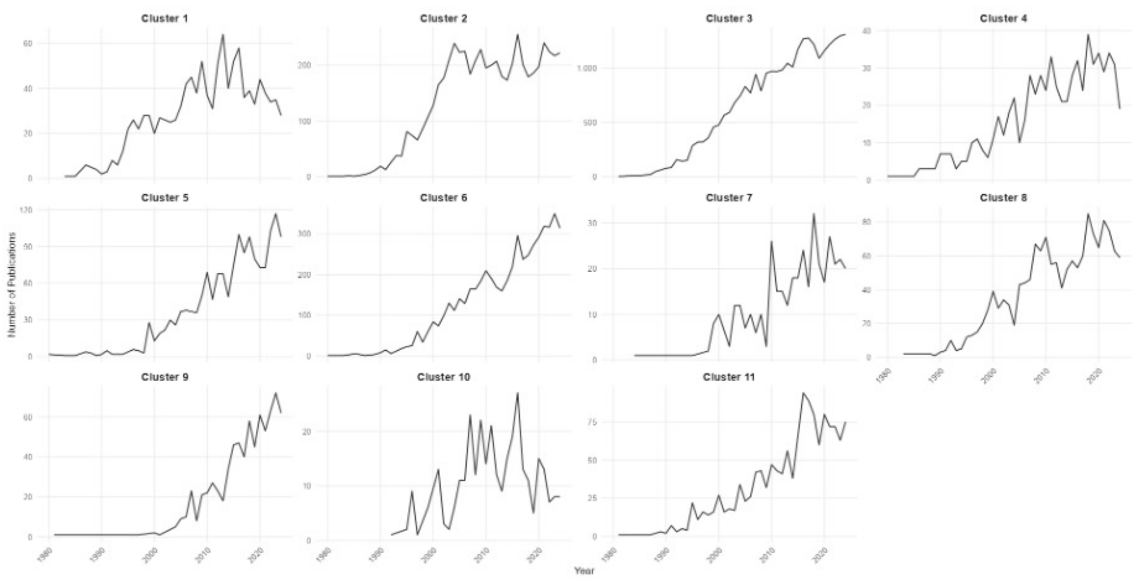

Fig 10. Development profiles of thematic clusters by year

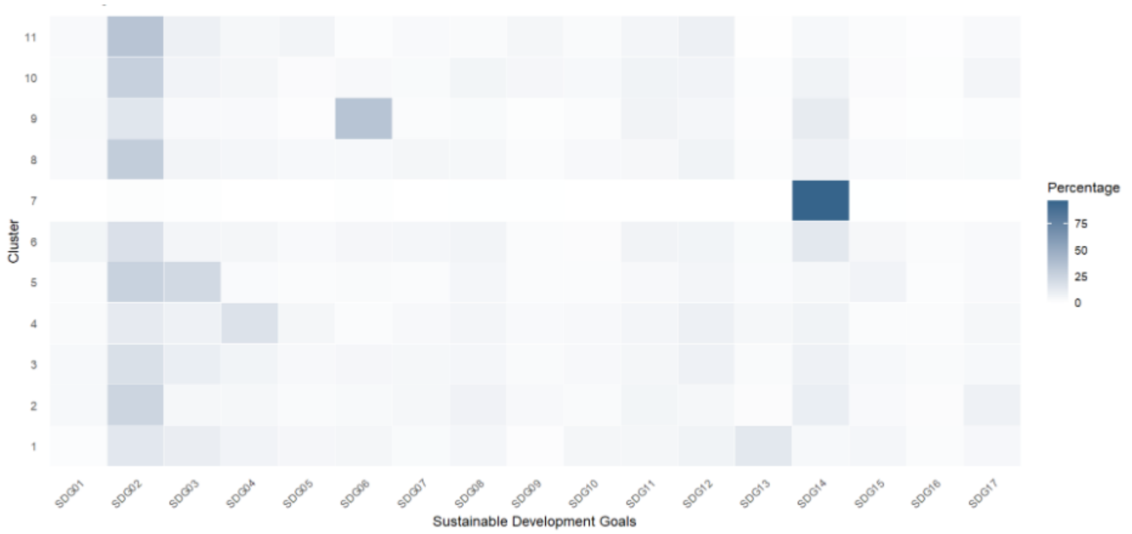

Fig 11. Heatmap showing the within-cluster relative associations between thematic clusters and the Sustainable Development Goals (SDGs)

suggesting a distinct national specialization partially diverging from the global veterinary literature [2].

The comparatively lower representation of SDG 13 (Climate Action) and SDG 15 (Life on Land) represents an area warranting further attention. Global veterinary and livestock literature increasingly emphasizes climate change, heat stress, methane emissions, vector dynamics, and biodiversity loss [2,11,52,56-59], indicating a partial misalignment between national veterinary research priorities and these emerging global trends.

Several structural factors may explain this pattern: veterinary curricula have historically centered on clinical sciences and production efficiency; research funding has favored directly observable outcomes such as disease control and production gains; and climate/biodiversity topics inherently require interdisciplinary integration with ecology, environmental sciences, and public health that remains underdeveloped [2,11].

From a One Health perspective, the strong associations with SDG 2, 3, 12, and 14 confirm that veterinary research robustly addresses human-animal-production system linkages [2,6,58]. However, weaker representation of SDG 13 and SDG 15 indicates that the environmental dimension of One Health has not been sufficiently addressed; a stronger research focus on climate change, ecosystem degradation, and biodiversity is needed [6,59].

The low publication volume in Food Hygiene and Technology highlights the need for strengthening research in food safety surveillance and sustainable food chains. The literature indicates that the relationship between meat inspection, surveillance, food safety, and animal welfare is multifaceted; however, data gaps and insufficient systematic utilization are reported in many areas [47]. Whether this reflects a Türkiye-specific pattern or a broader global trend warrants comparative investigation.

Academic output in Türkiye is primarily concentrated in animal health and production, consistent with existing bibliometric analyses [11,56]. Emerging global priorities -climate change, ecosystem health, and biodiversity- have not yet been fully integrated into the national veterinary research agenda.

The spread of zoonotic diseases emerging due to climate change, changes in vector distributions, habitat transformation, and the impacts of extreme weather events on animal health and production systems are risks that are difficult to manage without adequate scientific infrastructure [6,59]. Furthermore, the strengthening of relationships between ecosystem degradation, pathogen circulation, interactions between wild and domestic animals, and agroecological vulnerabilities makes the environmental dimension of veterinary research even more critical [2]. In this context, comparatively lower levels of knowledge production in relevant fields may influence Türkiye's preparedness for multidimensional environmental and biological challenges. Global scientific trends increasingly shift toward climate change, environmental adaptation, and ecosystem-based research, comparatively lower academic output in these areas may influence Türkiye's international scientific competitiveness [11].

This study has certain limitations. Since analyses rely solely on publication titles, context-sensitive themes -particularly those associated with SDG 13 and SDG 15- that tend to be elaborated in abstracts or full texts may not be fully captured. Manual entry inconsistencies in the YÖK Academic dataset and the English-based nature of the SDG dictionaries may have led to partial underrepresentation of Turkish-titled studies. While the cluster number k was determined through multiple validation criteria, the final selection inevitably involves a degree of interpretive judgment. Additionally, the multi-label SDG approach may have contributed to a relative overrepresentation of structurally overlapping goals (e.g., SDG 2 and SDG 14 in aquaculture-related studies), though this is an inherent feature of multi-label classification frameworks rather than a methodological flaw.

In conclusion, the field of veterinary sciences in Türkiye demonstrates a strong thematic concentration, particularly within the context of SDG 2 (Zero Hunger) and SDG 3 (Good Health and Well-being), and in this regard, aligns significantly with global trends. Conversely, the relatively high representation of SDG 14 suggests that Türkiye may have developed a distinct specialization in aquatic systems and aquaculture, whereas the comparatively lower representation of SDG 13 and SDG 15 indicates areas that may benefit from further research attention in climate change, biodiversity, and ecosystem health.

Strengthening interdisciplinary collaborations and directing funding toward environmental sustainability, food security, and ecosystem health -and expanding the One Health framework to encompass its environmental dimension- will enhance Türkiye's sustainability-focused scientific output and global alignment [58]. In this context, the findings reveal that veterinary sciences are not merely a discipline focused on production and health, but a strategic field that must be re-examined and repositioned within the context of global environmental crises.

Declarations

Availability of Data and Materials: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. Due to the inclusion of publication records obtained from the YÖK Academic platform, data sharing is restricted to aggregated or anonymized formats to comply with platform usage policies.

Acknowledgements: The authors would like to thank the institutions and researchers whose publicly available publication records contributed to the construction of the dataset analyzed in this study.

Financial Support: This research received no specific grant from any funding agency, commercial organization or not-for-profit body.

Competing Interests: The authors declared that there is no competing of interests.

Declaration of Generative Artificial Intelligence (AI): The authors declare that the article, tables and figures were not written/created by AI and AI-assisted Technologies. All analyses, visualizations, and textual content were produced solely by the authors using only R (v4.5.1) and its associated packages, with no other software or tools employed.

Author Contributions: H.Y. designed the study, developed the text-mining workflow, performed the hybrid clustering and SDG-matching analyses, and interpreted the results. F.Ç.B. contributed to the literature review, data collection and curation, and assisted in the implementation and validation of analytical procedures. A.Y. supported the analytical evaluation and contributed to the review and critical revision of the manuscript. All authors read and approved the final manuscript.

References

1. **Akello W:** Climate change and veterinary medicine: A call to action for a healthier planet. *F1000Research,* 13:1360, 2024. DOI: 10.12688/f1000research.158307.1
2. **Destoumieux-Garzón D, Bonnet P, Teplitsky C, Criscuolo F, Henry PY, Mazurais D, Prunet P, Salvat G, Usseglio-Polatera P, Verrier E, Friggens NC:** Animal board invited review: OneARK: Strengthening the links between animal production science and animal ecology. *Animal,* 15 (1):100053, 2020. DOI: 10.1016/j.animal.2020.100053
3. **Kiran D, Sander W, Duncan C:** Empowering veterinarians to be planetary health stewards through policy and practice. *Front Vet Sci,* 9:775411, 2022. DOI: 10.3389/fvets.2022.775411
4. **Christopher MM:** A new decade of veterinary research: societal relevance, global collaboration, and translational medicine. *Front Vet Sci,* 2:1, 2015. DOI: 10.3389/fvets.2015.00001
5. **Olmos G, Tunón H, de Oliveira D, Jones M, Wallenbeck A, Swanson J, Blokhuis H, Keeling L:** Animal welfare and the United Nations' sustainable development goals - Broadening students' perspectives. *Sustainability*, 13 (6):3328, 2021. DOI: 10.3390/su13063328
6. **Pappaioanou M, Kane TR:** Addressing the urgent health challenges of climate change and ecosystem degradation from a One Health perspective: What can veterinarians contribute? *J Am Vet Med Assoc,* 261 (1): 49-55, 2022. DOI: 10.2460/javma.22.07.0315
7. **Küçükaslan Ö, Küçükaslan İ:** Bibliometric profile of postgraduate theses in veterinary obstetrics and gynecology in Turkey. *J Turk Vet Med Assoc,* 93 (1): 52-64, 2021. DOI: 10.33188/vetheder.1016877
8. **Sacchini S, Castro-Alonso A:** Veterinary Medical Education: Challenges and Perspectives. MDPI Books, Basel, 2024.
9. **Yıldırım Ö, Yaralı C, Danyer E:** The importance of veterinary medicine in the provision of sufficient and healthy food and the role of Veterinary Control Research Institutes. *Etlik Vet Microbiol J,* 31 (1): 101-109, 2020. DOI: 10.35864/evmd.746011
10. **Baker D, Jackson E, Cook SJ:** Perspectives of digital agriculture in diverse types of livestock supply chain systems. Making sense of uses and benefits. *Front Vet Sci,* 9:992882, 2022. DOI: 10.3389/fvets.2022.992882
11. **Manyike JZ, Taruvinga A, Akinyemi BE:** Mapping the research landscape of livestock adaptation to climate change: A bibliometric review using Scopus database (1994-2023). *Front Clim,* 7:1567674, 2025. DOI: 10.3389/fclim.2025.1567674
12. **Yüksek Öğretim Kurulu:** YÖK Akademik. Academic Search System. https://akademik.yok.gov.tr; *Accessed:* 05.02.2025.
13. **Akoğul S, Erişoğlu M:** An approach for determining the number of clusters in a model-based cluster analysis. *Entropy,* 19 (9): 452, 2017. DOI: 10.3390/e19090452
14. **Alasali T, Ortakcı Y:** Clustering techniques in data mining: A survey of methods, challenges, and applications. *Computer Sci,* 9 (1): 32-50, 2024. DOI: 10.53070/bbd.1421527
15. **Hennig C:** Clustering strategy and method selection. **In**, Hennig C, Meila M, Murtagh F, Rocci R: Handbook of Cluster Analysis, 724, Chapman and Hall/CRC eBooks, Informa, London, 2015.
16. **Lizenberger A, Pfeifer F, Polewka B:** Rethinking Recommender Systems: Cluster-based Algorithm Selection. *arXiv,* 2024:1-16, 2024. DOI: 10.48550/arxiv.2405.18011
17. **Sangam RS, Om H:** Hybrid data labeling algorithm for clustering large mixed type data. *J Intell Inf Syst,* 45 (2): 273-294, 2014. DOI: 10.1007/s10844-014-0348-x

18. **Yang Y, Nan F, Yang P:** Effective multilayer hybrid classification approach for automatic bridge health assessment on large-scale uncertain data. *J Ind Inf Integr,* 24, 100234, 2021. DOI: 10.1016/j.jii.2021.100234

19. **Rousseeuw PJ:** Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math,* 20, 53-65, 1987. DOI: 10.1016/0377-0427(87)90125-7

20. **Milligan GW, Cooper MC:** An examination of procedures for determining the number of clusters in a data set. *Psychometrika,* 50 (2): 159-179, 1985. DOI: 10.1007/BF02294245

21. **Halkidi M, Batistakis Y, Vazirgiannis M:** On clustering validation techniques. *J Intell Inf Syst,* 17 (2-3): 107-145, 2001. DOI: 10.1023/A:1012801612483

22. **Kaufman L, Rousseeuw PJ:** Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York, 1990.

23. **Han J, Kamber M, Pei J:** Data Mining: Concepts and Techniques. 3rd ed., Morgan Kaufmann, Waltham, 2012.

24. **Tan PN, Steinbach M, Kumar V:** Introduction to Data Mining. 2nd ed., Pearson, New York, 2019.

25. **Chen C:** CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inf Sci Technol,* 57 (3): 359-377, 2006. DOI: 10.1002/asi.20317

26. **Waltman L, Van Eck NJ:** A smart local moving algorithm for large-scale modularity-based community detection. *Eur Phys J B,* 86 (11):471, 2013. DOI: 10.1140/epjb/e2013-40829-0

27. **Šubelj L, Van Eck NJ, Waltman L:** Clustering scientific publications based on citation relations. *J Informetr,* 10 (4): 1020-1032, 2016. DOI: 10.1371/journal.pone.0154404

28. **Vanderfeesten M, Otten R, Spielberg E:** Search Queries for Mapping Research Output to the Sustainable Development Goals (SDGs) v5.0.2 (5.0.2). Zenodo, 2020.

29. **Roberge G, James C:** Elsevier 2025 Sustainable Development Goals (SDGs) Mapping (V1). Elsevier Data Repository, 2025.

30. **Bordignon F:** Dataset of search queries to map scientific publications to the UN sustainable development goals. *Data Brief,* 34:106731, 2021. DOI: 10.1016/j.dib.2021.106731

31. **Salazar MA, Loaiza MEB, Gómez M:** Natural language processing to analyze textual information from the SDGs. *International Meeting on Engineering Education,* 1-8, 2024. DOI: 10.26507/paper.3759

32. **Wulff DU, Meier DS, Mata R:** Using novel data and ensemble models to improve automated labeling of sustainable development goals. *Sustain Sci,* 19 (5): 1773-1786, 2024. DOI: 10.1007/s11625-024-01516-3

33. **Meier D, Mata R, Wulff DU:** text2sdg: An open-source solution for monitoring sustainable development goals from text. arXiv, Cornell University, 2021. http://arxiv.org/abs/2110.05856; *Accessed:* 08.09.2025.

34. **Pukelis L, Bautista-Puig N, Skrynik M, Stančiauskas V:** OSDG -- Open-source approach to classify text data by UN sustainable development goals (SDGs). *arXiv,* 2020:1-10, 2020. DOI: 10.48550/arxiv.2005.14569

35. **Tsoumakas G, Katakis I:** Multi-label classification: An overview. *Int J Data Warehous Min,* 3 (3): 1-13, 2007. DOI: 10.4018/jdwm.2007070101

36. **Schmidt F, Vanderfeesten M:** Evaluation of the accuracy of mapping science to the United Nations' Sustainable Development Goals (SDGs) using Aurora SDG queries. *Zenodo,* 2021.

37. **Adamakopoulou C, Benedetti B, Zappaterra M, Felici M, Masebo ND, Previti A, Passantino A, Paladino B:** Cats' and dogs' welfare: Text mining and topics modeling analysis of the scientific literature. *Front Vet Sci,* 10:1268821, 2023. DOI: 10.3389/fvets.2023.1268821

38. **Benedetti B, Felici M, Costa LN, Padalino B:** A review of horse welfare literature from 1980 to 2023 with a text mining and topic analysis approach. *Ital J Anim Sci,* 22 (1): 1095-1109, 2023. DOI: 10.1080/1828051x.2023.2271038

39. **Davies H, Nenadić G, Alfattni G, Castelerio AM, Moubayed NA, Farrell SO, Radford AD, Noble PM:** Text mining for disease surveillance in veterinary clinical data. Part one: The language of veterinary clinical records and searching for words. *Front Vet Sci,* 11:1352239, 2024. DOI: 10.3389/fvets.2024.1352239

40. **İstek Ö:** Bibliometric analysis of postgraduate theses on eye and their diseases in animals in Turkey. *Kocatepe Vet J,* 18 (3): 329-336, 2025.

41. **Ouyang Z, Sargeant JM, Thomas A, Wycherley K, Ma R, Esmaeilbeigi R, Versluis A, Deborah S, Stone E, Poljak Z, Bernardo TM:** A scoping review of 'big data', 'informatics', and 'bioinformatics' in the animal health and veterinary medical literature. *Anim Health Res Rev,* 20 (1): 1-16, 2019. DOI: 10.1017/s1466252319000136

42. **Vaitsi GA, Bourganou MV, Lianou DT, Kiouvrekis Y, Michael CC, Gougoulis DA, Fthenakis GC:** Scientometric analysis: An emerging tool in veterinary and animal scientific research. *Animals,* 14 (21): 3132, 2024. DOI: 10.3390/ani14213132

43. **Taçbaş E, Çakır Ş, Özkan S, Demir D:** Bibliometric analysis of articles produced from research projects supported by TAGEM according to scientific disciplines. *Etlik Vet Microbiol J,* 35 (2): 179-180, 2025. DOI: 10.35864/evmd.1248719

44. **Dörterler S, Dumlu H, Özdemir D, Temurtaş H:** Hybridization of meta-heuristic algorithms with K-means for clustering analysis: Case of medical datasets. *Gazi J Eng Sci,* 10 (1): 1-15, 2024. DOI: 10.30855/gmbd.0705n01

45. **Lemarchand P, MacMahon C, McKeever M, Owende P:** An evaluation of a computational technique for measuring the embeddedness of sustainability in the curriculum aligned to AASHE-STARS and the United Nations Sustainable Development Goals. *Front Sustain,* 4:997509, 2023. DOI: 10.3389/frsus.2023.997509

46. **Asatani K, Takeda H, Yamano H, Sakata I:** Scientific attention to sustainability and SDGs: Meta-analysis of academic papers. *Energies,* 13 (4):975, 2020. DOI: 10.3390/en13040975

47. **Stärk KDC, Alonso S, Dadios N, Dupuy C, Ellerbroek L, Georgiev M, Hardstaff J, Huneau-Salaün A, Laugier C, Mateus A, Nigsch A, Afonso A, Lindberg A:** Strengths and weaknesses of meat inspection as a contribution to animal health and welfare surveillance. *Food Control,* 39, 154-162, 2013. DOI: 10.1016/j.foodcont.2013.11.009

48. **Christopher MM, Marusic A:** Geographic trends in research output and citations in veterinary medicine: Insight into global research capacity, species specialization, and interdisciplinary relationships. *BMC Vet Res,* 9, 115, 2013. DOI: 10.1186/1746-6148-9-115

49. **Uzabacı E:** Evaluation of the 100 most-cited articles published in veterinary journals. *Vet Med Sci,* 11 (5):e70545, 2025. DOI: 10.1002/vms3.70545

50. **Topuz D, Tekgöz S:** Hybrid ensemble model for lactation milk yield prediction of Holstein cows. *Kafkas Univ Vet Fak Derg,* 31 (5): 603-611, 2025. DOI: 10.9775/kvfd.2025.34031

51. **Bağlıoğlu M, Özen A:** Veteriner Hekimliği Açısından İnformatik Devrimi. *Lokman Hekim Dergisi,* 48-48, 2014.

52. **Yardibi F, Fırat MZ, Teke EÇ:** Trend topics in animal science: A bibliometric analysis using CiteSpace. *Turk J Vet Anim Sci,* 45 (5): 833-841, 2021. DOI: 10.3906/vet-2001-103

53. **Savaş T, Yurtman İY:** Hayvan davranış bilimi ve zootekni: Tanım ve izlem. *Hay Üretim Derg,* 49 (2): 36-42, 2008.

54. **Yıldız Bİ:** Comparative performance of convolutional neural network models in wing morphometric classification of honey bee populations across Europe. *Kafkas Univ Vet Fak Derg,* 31 (5): 653-659, 2025. DOI: 10.9775/kvfd.2025.34544

55. **Akıllı A, Kezer G, Kul E:** Evaluation of microbiological properties in kefir production with fuzzy logic-based decision support system. *Kafkas Univ Vet Fak Derg,* 31 (5): 679-688, 2025. DOI: 10.9775/kvfd.2025.34643

56. **Diop S, İnci A, Kızgın AD, Düzlü Ö:** Understanding one health and zoonosis: A systematic review with a bibliometric analysis of Turkish research and global perspectives (1974-2023). *Kafkas Univ Vet Fak Derg,* 2025. DOI: 10.9775/kvfd.2025.34009

57. **Ramírez-Durán JA, Niebles W, Stojanovich-Morente Z, Gallego G, Guerra-Cogollo JA:** Use of technology for sustainable livestock processes: A bibliometric review. *Afr J Food Agric Nutr Dev,* 24 (7): 23934-23953, 2024. DOI: 10.18697/ajfand.132.23515

58. **Abbasi K, Ali P, Barbour V, Benfield T, Bibbins-Domingo K, Hancocks S, Horton R, Laybourn-Langton L, Mash R, Sahni P, Mohammad Sharief W, Yonga P, Zielinski C:** Time to treat the climate and nature crisis as one indivisible global health emergency. *Kafkas Univ Vet Fak Derg,* 30 (1): 1-3, 2024. DOI: 10.9775/kvfd.2023.editorial

59. **Alkheraije KA:** Climate change, vector-borne animal diseases: Impacts on livestock health: A narrative review. *Kafkas Univ Vet Fak Derg,* 2025. DOI: 10.9775/kvfd.2025.35482