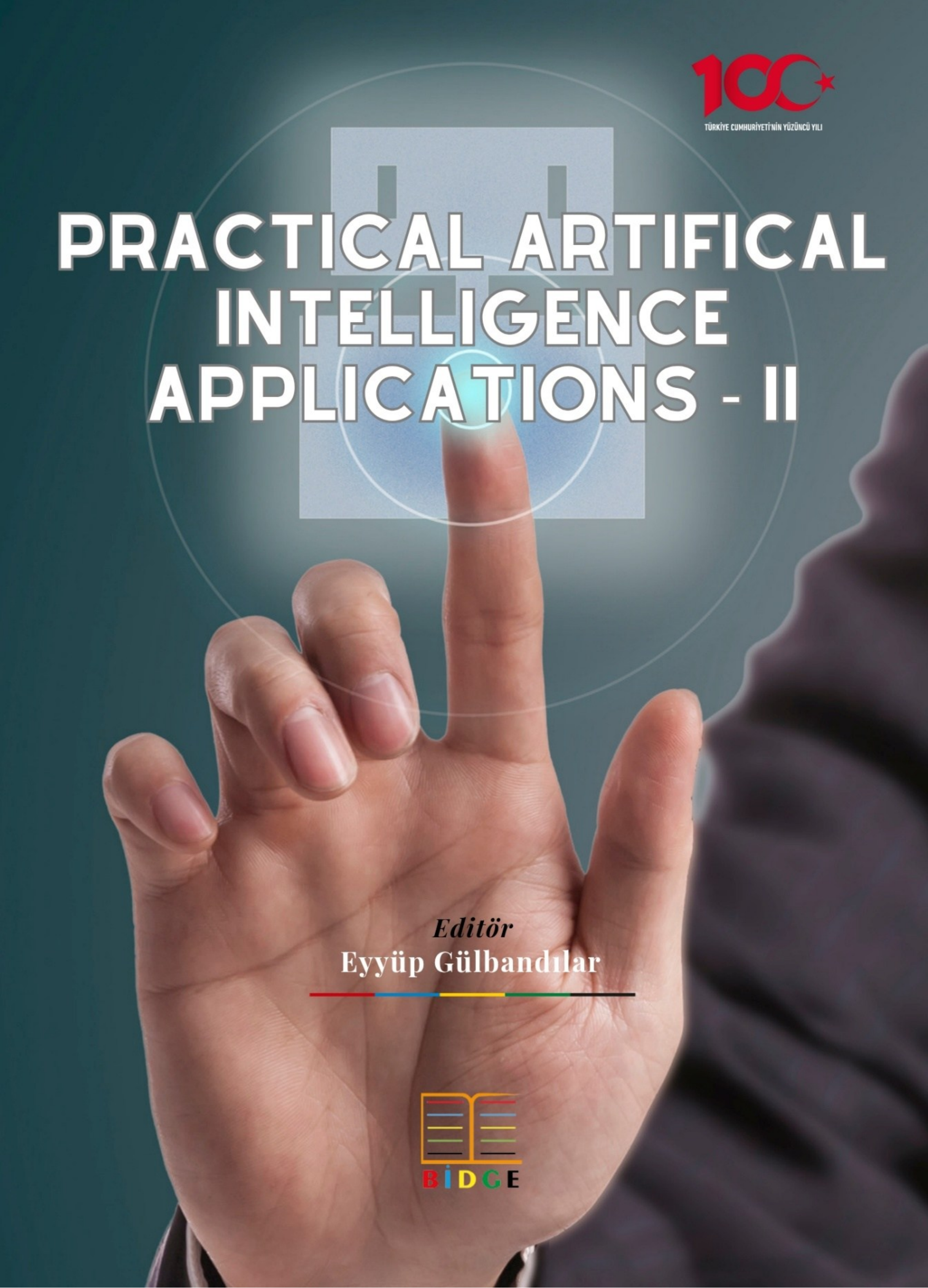


PRACTICAL ARTIFICAL INTELLIGENCE APPLICATIONS - II

Editör

Eyyüp Gülbandılar

BİDGE

BIDGE Publications

Practical Artificial Intelligence Applications 2

Editor: Prof.Dr.Eyyüp GÜLBANDILAR

ISBN: 978-625-372-394-1

Page Layout: Gözde YÜCEL

1st Edition:

Publication Date: 25.12.2024

BIDGE Publications,

Certificate No: 71374

www.bidgeyayinlari.com.tr - bidgeyayinlari@gmail.com

Krc Bilişim Ticaret ve Organizasyon Ltd. Şti.

Güzeltepe Mahallesi Abidin Daver Sokak Sefer Apartmanı No: 7/9 Çankaya / Ankara

Content

Whale Optimization Algorithm Bibliometric Analysis from 2016 to 2024 4

Ahmet AKKAYA 4

Cemil KÖZKURT 4

Deep Learning Approaches For Accurate And Efficient Classification Of Eye Diseases 39

Kamil ORMAN 39

Yıldız AYDIN 39

Artificial Intelligence Methods in Aquaculture: Review, Application and Challenges 49

Özge ZENCİR TANIR 49

Volkan KAYA 49

Top 1000 Books: Regressıon Analysıs And Recommendatıon System 91

Sena TINKIR 91

CHAPTER I

Whale Optimization Algorithm Bibliometric Analysis from 2016 to 2024

Ahmet AKKAYA[1]
Cemil KÖZKURT[2]

1. Introduction

Whale Optimization Algorithm (WOA), an optimization algorithm inspired by nature, was proposed by Seyedali Mirjalili in 2016 (Mirjalili & Lewis, 2016). This algorithm is designed based on whale hunting strategy and bubble-net feeding technique. It seeks solutions to global optimization problems by transforming the whales' behavior of creating bubbles in the water to catch their prey and trapping their prey in these bubbles into a mathematical model. Compared to other popular metaheuristic algorithms such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO) and Simulated Annealing (SA) algorithms, WOA stands out with its simple structure, fast convergence and efficient exploration of the solution space.

[1] *Bandırma Onyedi Eylül University, Bandırma, Turkey, aakkaya@bandirma.edu.tr, ORCID: 0000-0003-4836-2310*
[2] *Bandırma Onyedi Eylül University, Bandırma, Turkey, ckozkurt@bandirma.edu.tr, ORCID: 0000-0003-1407-9867*

The basic principle of the whale optimization algorithm is based on the whale hunting strategy and exploration and exploitation balancing. In the whaling strategy, whales encircle their prey using bubble nets, simulating the algorithm's circulation around the solution space. In exploration and exploitation balancing, WOA explores large areas for global optimization, while at the same time stabilizing by performing local searches around the best solution. In the Exploration phase, a wide search is performed in the solution space to randomly generate new solution points. In the Exploitation phase, there is no narrow search around the best solution, which allows for a deeper investigation in a narrower solution space.

WOA uses two main equations, given by Equation (1) and Equation (2), to simulate the movement of whales in the solution space.

$$X_{t+1} = X_t + A \cdot |C \cdot X^* - X_t| \qquad \text{Equation (1)}$$

Equation (1) mathematically models the movement in the solution space (encircling prey). Here, X_t is the current solution position, X^* is the best solution position, and A and C are the weight parameters used to calculate whale movements.

$$X_{t+1} = X^* - A \cdot |C \cdot X^* - X_t| \qquad \text{Equation (2)}$$

Equation (2) mathematically models bubble formation (bubble-net feeding). Equation (2) simulates whales moving in a narrow circle around the best solution.

A review of the literature shows that WOA has become an optimization tool that is effectively used in a wide range of applications thanks to its strengths such as its simple structure, fast convergence capability, and efficient exploration of the solution space. WOA is frequently used for machine learning problems, especially for its successful application in hyperparameter optimization and model selection. (Mirjalili & Lewis, 2016; Zhou et al., 2022) have shown that WOA can be used for hyperparameter optimization of models such as support vector machines (SVM) and artificial neural networks (ANN). The global optimization capability of WOA is essential for accurate parameter tuning. WOA has also been used in the optimization of renewable energy systems such as solar energy (Paul et al., 2023), wind energy (Al-Quraan et al., 2023), and energy storage systems (Peddakapu et al., 2024). (Oliva et al., 2017) reported that WOA provides robust results in energy systems optimization such as power generation capacity and distribution management. WOA has been used in energy systems for the integration of renewable energy sources (Yahya et al., 2024) and energy efficiency (Chandrasekaran & Rajasekaran, 2024). WOA has also been successfully used in engineering fields such as control systems design and system optimization. In such problems, WOA is often used to control dynamic systems and solve engineering problems (Nadimi-Shahraki et al., 2023). The ability of WOA to avoid local minima leads to more accurate and efficient solutions in engineering systems. The optimization power of WOA is also used in application areas such as distributed systems (Li & Chen, 2024) and logistics optimization (Chen, 2024). It is especially effective in supply chain management (Pham et al., 2024), vehicle routing problems (Dewi & Utama, 2021) and coverage optimization (Deepa & Venkataraman, 2021).

WOA offers some important advantages compared to evolutionary algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Simulated Annealing (SA). Although WOA has fewer parameters than PSO, it can converge faster. (Mirjalili & Lewis, 2016) stated that WOA's mobility in the solution space is more efficient than PSO. While PSO is generally effective in continuous solution spaces, WOA provides extensive exploration and better exploitation properties. While Genetic Algorithms (GA) generally search for solutions with a large population in large solution spaces, WOA produces more efficient solutions with smaller populations. (Mirjalili & Lewis, 2016) showed that WOA achieves faster and more accurate results than GA. Moreover, WOA requires fewer parameter adjustments, resulting in lower computational cost compared to GA.

In order to improve the efficiency of WOA, many researchers have developed hybrid algorithms. Hybrid algorithms usually combine the exploration capabilities of WOA with the exploitation capabilities of other algorithms. PSO-WOA: By combining the exploration power of PSO with the exploitation capabilities of WOA, it is stated to achieve effective optimization results in larger solution spaces. GA-WOA: Genetic Algorithms (GA) are combined with the hunting strategy of WOA to provide higher accuracy and faster solutions (Dai et al., 2023).

A. Methodology of the study

Bibliometric analysis is a method for the quantitative study of scientific publications. It is often applied to assess research trends, author networks, citation relationships and scientific impact. The main areas of bibliometric analysis include citation analysis, which is frequently used in academic performance evaluations, co-citation and bibliographic linkage to identify the scientific links of publications, author and collaboration networks to analyze authors' academic collaborations and international networks, and journal impact factor, which measures the prestige of a journal based on the citation frequency of articles published in a particular journal (Cooper, 2015). Within the scope of the research, bibliometric analysis was conducted with the VOSviewer program using the data obtained from the Web of Science (WoS) database.

B. Purpose of the study

The aim of this research is to draw the attention of researchers by revealing a total of 3873 studies in the WoS database of the whale optimization algorithm in the last nine years (2016-2024) using the search phrase (“whale optimization algorithm”) or (“whales optimization algorithm”) or (“whale algorithm”) or (“whales algorithm”) or (“whales algorithm”) with various quantitative data using bibliometric analysis method.

C. Data and analysis

In order to ensure the reliability of the study, the Web of Science database, which allows access to many academic data, was used. There are programs such as CiteSpace, Bibliometrix, Gephi, HistCite, EndNote and VOSviewer that enable the data to be analyzed by bibliometric analysis method. In this study, the data obtained from the Web of Science database were analyzed using the

VOSviewer program. VOSviewer is a tool specially designed for visualization of data obtained in scientific research. It maps scientific networks and interactions of publications. Using this program, researchers in areas such as literature analysis, network analysis and visual data exploration are carried out. VOSviewer is especially useful for understanding, visualizing and analyzing complex data sets such as collections of articles or bibliometric analyses that contain large datasets. VOSviewer is often preferred for its visualization capabilities, network analysis, bibliometric analysis, interactive analysis, free, user-friendly and data content analysis. These advantages of VOSviewer make it preferable for scientific researchers, academics and analysts in areas such as literature analysis, network analysis and visual data exploration.

In this study, using the sentence ("whale optimization algorithm") OR ("whales optimization algorithm") OR ("whale algorithm") OR ("whales algorithm") OR ("whales algorithm"), a bibliometric analysis on the subject was carried out with the data obtained from 3873 publications from the WoS database as of November 2024 and using the VOSviewer program.

2. Findings

In this section, the distribution of WoS categories, distribution of publications over the years, distribution of publications by document type, distribution of publications by Web of Since index, research profiles, distribution of publication titles by research category, distribution of publications by country, distribution of publications by publisher, distribution of publications by research field, meso distribution of publications by citation subject, micro distribution of publications by citation subject, The findings related to the distribution of publication languages of publications, analysis of researchers with co-authors, analysis of institutions with co-authors, analysis of countries with co-authors, citation analysis of publications on the basis of the published work, the link between the citations found in the published sources, citation analysis of authors, citation analysis of countries, citation analysis of institutions, keyword analysis concurrency author keyword analysis, bibliographic match analysis of texts are included.

Distribution of WoS categories

As a result of the analysis, Figure 1 shows the distribution of publications according to WoS categories.

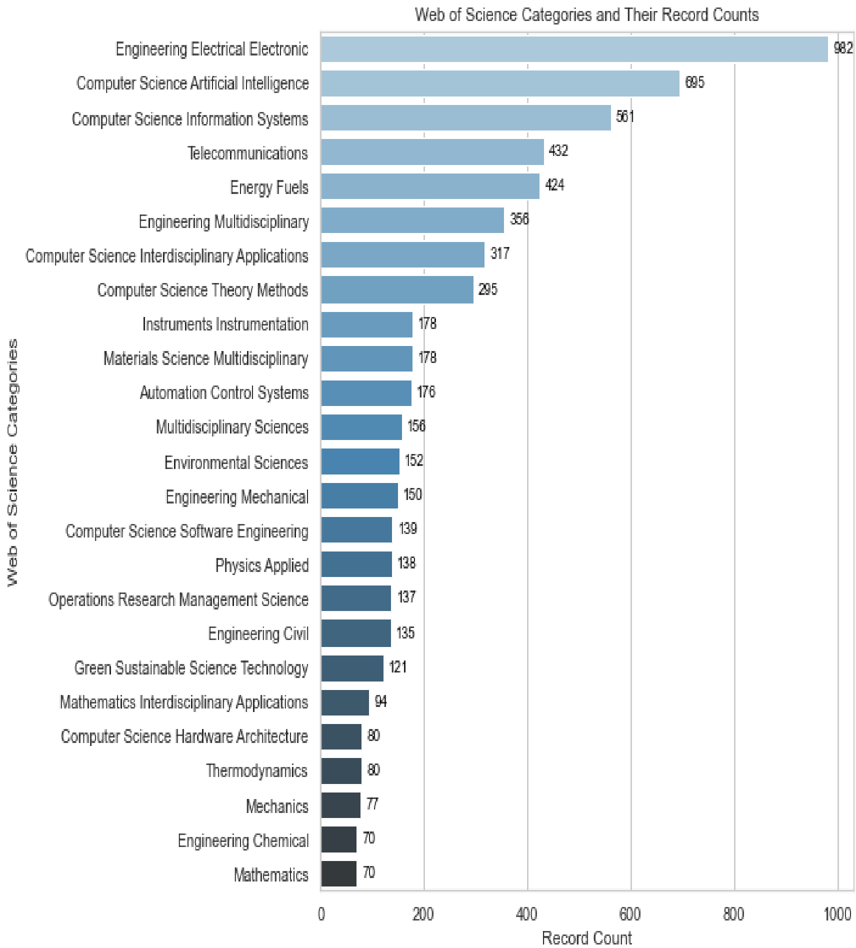


Figure 1: Distribution of publications according to WoS categories

When Figure 1 is examined, it is seen that the Engineering Electrical Electronic category (982 publications) has the highest number of publications in all data. Accordingly, it is seen that the studies conducted in the field of electrical and electronics engineering have a very wide scope. Computer Science Artificial

Intelligence (695 publications) ranks second. Today, research on artificial intelligence has a large place in the field of computer science. It is seen that this category also hosts a lot of studies. Computer Science Information Systems (561 publications) and information systems in computer science constitute a very large research area and attract attention with the high number of publications. Categories in the field of engineering and energy such as Telecommunications (432 publications) and Energy Fuels (424 publications) also stand out with their high number of publications. Multidisciplinary categories such as Engineering Multidisciplinary (356 publications) and Computer Science Interdisciplinary Applications (317 publications) show that important studies are being conducted in areas where research from different disciplines comes together. Lower publication numbers are observed in some areas such as Mathematics (70 publications) and Engineering Chemical (70 publications). This means that these areas are relatively less researched or the data are more limited.

Distribution of publications according to the year they were published

As a result of the analysis, the distribution of publications according to the years they were published is given in Figure 2.

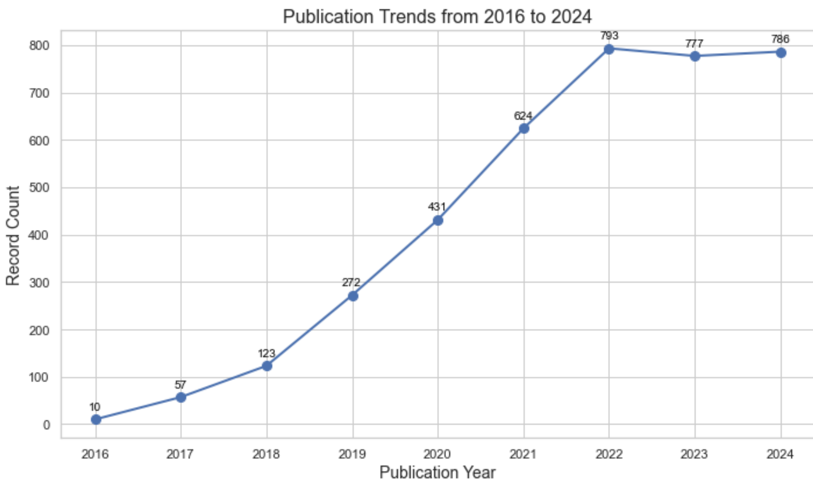


Figure 2: Distribution of publications according to the years they were published

When Figure 2 is examined, the years 2022, 2023, and 2024 stand out as the years with the most intense publications. There were 793 publications in 2022, 777 in 2023, and 786 in 2024. The fact that these years stand out with high numbers indicates that the research conducted in this period increased or more articles were published. An increase is observed especially after 2022 compared to 2021 (it reaches 793), which indicates that there was a concentration in scientific research in these years. In 2020 and before, the number of publications is clearly low. In 2020, there were 431 publications, and this number decreased to 272 in 2019. The number of publications was quite low in the years before 2018 (in 2016 and 2017). There were 57 publications in 2017 and only 10 in 2016, which shows that fewer articles were published at that time because the algorithm was new, and the algorithm was not popular.

Document type distribution of publications

As a result of the analysis, the distribution of publications according to document type is given in Figure 3.

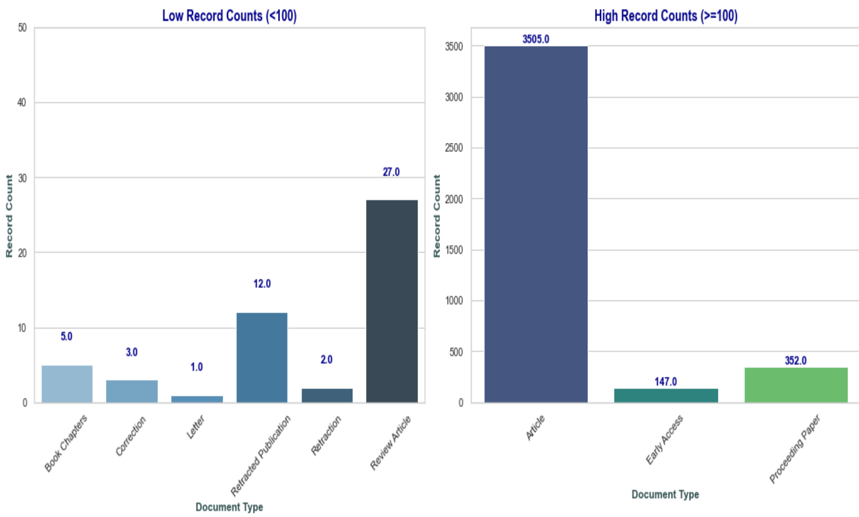


Figure 3: Distribution of publications by document type

When Figure 3 is examined, it is seen that Article is the most common document type with 3505 publications. This shows that most of the research is published in article format. Proceeding Paper has 352 publications, conference proceedings also have an important place. These types of documents usually include research results presented in conferences. Early Access 147 publications, early access documents are articles that can be accessed before the official publication of research results. Review Article 27 publications, compilation articles are also a part of research but are seen to be used less commonly. Retraction and Retracted Publication These types of documents indicate that published content has been retracted or is incorrect. There are 14 publications in total (12 retracted articles and 2 retraction notices). Other Types, on the other hand, show that the number of rarer document types such as Book Chapters, Correction and Letter is very low. This situation indicates that these publication types are published less.

Meso distribution of publications' citation topics

As a result of the analysis, the meso distributions of the citation topics of the publications are given in Table 1.

Table 1: Citation topics of publications meso distributions

Citation Topics Meso	Record Count
4.84 Supply Chain & Logistics	896
4.18 Power Systems & Electric Vehicles	675
4.13 Telecommunications	241
4.61 Artificial Intelligence & Machine Learning	157
7.215 Friction & Vibration	114
4.17 Computer Vision & Graphics	97
4.29 Automation & Control Systems	88
7.133 Geotechnical Engineering	83
4.46 Distributed & Real Time Computing	78
2.62 Electrochemistry	61
8.19 Oceanography, Meteorology & Atmospheric Sciences	55
7.192 Testing & Maintenance	51
4.58 Wireless Technology	48
4.101 Security, Encryption & Encoding	47
4.47 Software Engineering	43
4.48 Knowledge Engineering & Representation	39
7.227 Manufacturing	34
4.116 Robotics	33
4.183 Transportation	31
4.187 Security Systems	30
6.115 Sustainability Science	30
6.153 Climate Change	30
7.70 Thermodynamics	30
4.224 Design & Manufacturing	26
7.12 Metallurgical Engineering	26

When Table 1 is examined, Supply Chain & Logistics (896 citations) covers studies on the management of global supply chains and logistics processes. The high number of citations shows the importance of research in this field and its practical impact. Global crises such as the COVID-19 pandemic are thought to have increased research on supply chains. Power Systems & Electric Vehicles (675 citations) is a major research area in terms of electrical systems and electric vehicles, energy efficiency and sustainability. As the popularity of electric vehicles increases, the number of studies in this field is also increasing rapidly. This is an important research topic in the fields of energy production, storage and transportation. Telecommunications (241 citations) includes research on the development of telecommunication internet connections and communication technologies. There is also a significant number of citations in this field because research on new generation communication technologies such as 5G continues intensively. Artificial Intelligence & Machine Learning (157 citations): Artificial intelligence and machine learning have begun to revolutionize many sectors today. However, this area is a relatively new research topic and has still achieved an impressive number of citations. AI and ML have applications in many industries, increasing the interest in the work done in this area. Friction & Vibration (114 citations), this field focuses on the study of friction and vibration in engineering, especially in materials science and mechanical systems. Solving such problems in high-tech systems plays a critical role in the design of more efficient and safer machines.

Micro distribution of publications' citation topics

As a result of the analysis, the micro distribution of the citation topics of the publications is given in Table 2.

Table 2: Micro distribution of publications' citation topics

Citation Topics Micro	Record Count
4.84.169 Particle Swarm Optimization	786
4.18.296 Unit Commitment	186
4.18.204 Distributed Generation	160
4.18.472 Voltage Stability	125
7.215.818 Fault Diagnosis	109
4.18.575 Mppt	89
4.13.43 Wireless Sensor Networks	78
4.46.85 Cloud Computing	65
4.13.807 Internet Of Things	56
4.61.145 Feature Selection	52
8.19.7 Evapotranspiration	52
4.84.401 Scheduling	45
4.61.1302 Intrusion Detection	42
7.133.359 Rock Mechanics	42
2.62.138 Lithium-ion Battery	40
4.17.128 Deep Learning	33
4.18.101 Power Quality	29
4.29.435 Multi Agent Systems	27
7.192.650 Damage Detection	27
7.227.355 Tool Wear	27
4.84.471 Vehicle Routing Problem	26
4.18.754 Doubly Fed Induction Generator	23
4.84.260 Supply Chain	22
4.48.672 Natural Language Processing	21
2.62.571 Proton Conductivity	20

When Table 2 is examined, the publications' citation topics are as follows, which have over 100 citations in their micro distributions. Particle Swarm Optimization (786 citations), Particle Swarm Optimization is among the evolutionary algorithms and is used especially in machine learning, artificial intelligence and optimization problems. The high number of citations shows that this technique is popular and widely used in many different fields. Unit

Commitment (186 citations), a "unit commitment" problem in energy systems, deals with deciding how power plants will come into operation and go out. Studies in this area are studies carried out to provide more efficient energy management and planning in electrical grids. Distributed Generation (160 citations), distributed energy production, the use of energy resources in a decentralized, local production form. It can be said that this subject is increasingly being researched with the increasing use of renewable energy sources. Voltage Stability (125 citations), voltage stability is of critical importance to ensure a safe and stable power distribution in electrical grids. Studies in this area aim to increase the security of the energy infrastructure. Fault Diagnosis (109 citations), fault detection is a critical area in engineering and automation systems. The high number of citations indicates the interest in solving such problems, especially in electrical systems and robotic applications.

Web of Science index distributions

As a result of the analysis, the distribution of publications according to the Web of Science index is given in Figure 4.

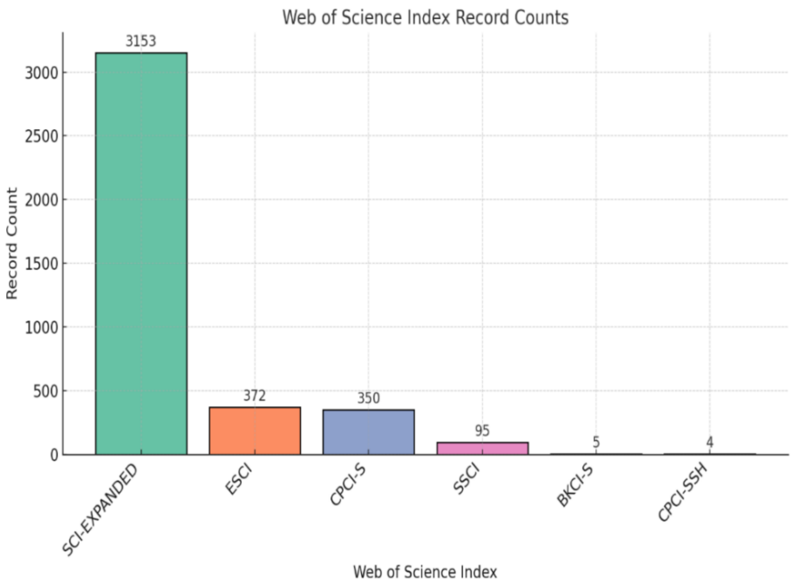


Figure 4: Distribution of publications according to Web of Science index

When Figure 4 is examined, it is seen that the highest number of publications is Science Citation Index Expanded (SCI-EXPANDED) – 3153 publications, this index includes the most prestigious and comprehensive scientific journals. The high number of publications shows that the studies published in these journals address a wide research area and include a large number of articles. This index covers the most important journals in the fields of basic sciences and engineering. Other indexes are Emerging Sources Citation Index (ESCI) – 372 publications, Conference Proceedings Citation Index – Science (CPCI-S) – 350 publications, Social Sciences Citation Index (SSCI) – 95 publications, Book Citation Index – Science (BKCI-S) – 5 publications, Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH) – 4 publications.

Journal name distribution according to publication titles

As a result of the analysis, the distribution of journal names according to the publication titles is given in Table 3.

Table 3: Journal name distribution according to publication titles

Publication Titles	Record Count
IEEE Access	179
Expert Systems with Applications	75
Energies	63
Applied Sciences Basel	55
Cmc Computers Materials Continua	54
Neural Computing Applications	53
Applied Soft Computing	49
Mathematics	48
Scientific Reports	45
Sensors	45
Soft Computing	45
Sustainability	41
Energy	40
Electronics	39
Energy Reports	39
Multimedia Tools and Applications	36
Journal of Intelligent Fuzzy Systems	34
Wireless Personal Communications	30
Journal of Supercomputing	26
Applied Intelligence	25
Knowledge Based Systems	24
Concurrency and Computation Practice Experience	23
Symmetry Basel	22
Arabian Journal for Science and Engineering	21
International Journal of Communication Systems	21

When Table 3 is examined, it is seen that IEEE Access has the highest number of publications with 179 publications. This journal offers a wide range of research in multidisciplinary engineering and technology fields. Medium Publication Numbers Expert Systems with Applications (75 publications) and ENERGIES (63 publications) are journals that publish important studies on artificial intelligence applications and energy, Applied Sciences

Basel (55 publications) and Cmc Computers Materials Continua (54 publications). These journals focus on technical areas such as applied sciences and materials engineering. Medium Low Publication Numbers, journals such as Neural Computing Applications, Applied Soft Computing, and Soft Computing publish on artificial intelligence and computational techniques. These journals have an important place with 53, 49, and 45 publications, respectively. Journals such as Sustainability, Energy, and Electronics cover a narrower range of topics with 41, 40, and 39 publications, respectively. Other journals have contributions ranging from 30 to 21 publications.

Language distribution of publications

As a result of the analysis, the distribution of publication languages is given in Table 4.

Table 4: Publication language distribution of publications

Languages	**Record Count**	**% of 3.873**
English	3851	99.432
Chinese	12	0.31
Turkish	9	0.232
Portuguese	1	0.026

When Table 4 is examined, English constitutes 99.432% of the total publications with 3851 publications. This shows that the vast majority of the literature is in English. Chinese constitutes 12 publications, 0.31% of the total publications. Although Chinese is a growing language in the academic field, it can be said that it has a very low share in this data. Turkish constitutes 9 publications, 0.232% of the total publications. Publications in Turkish are also limited, but it is seen as a language with the potential to lead to more publications in the academic field. Portuguese constitutes 1 publication, 0.026% of the total publications.

Country distribution of publications

As a result of the analysis, the distribution of publications according to the number of countries is given in Figure 5.

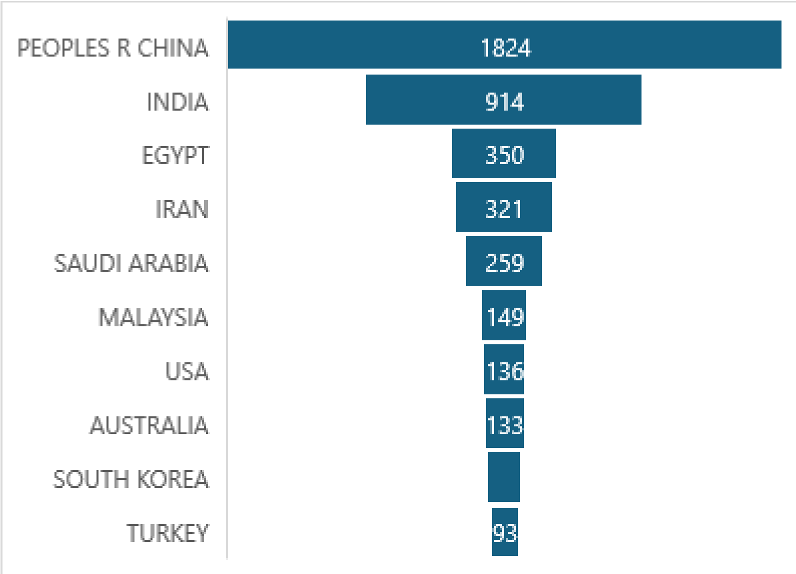


Figure 5: Country distribution of publications

When Figure 5 is examined, the highest number of publications is China, 1824 publications, which shows China's great influence in academic and research fields and its extensive literature production. India, 914 publications reflect that India has a significant share in scientific and technical research, especially academic contributions in developing economies. Countries such as Egypt, Iran, Saudi Arabia are listed with 350, 321, and 259 publications, respectively. These countries generally conduct active academic research in the Middle East and North Africa region. Countries such as Malaysia and USA are listed with 149 and 136 publications, respectively. The number of academic publications in Malaysia shows that it contributes to research and development in the region. The USA, on the other hand, has a high number of publications worldwide, but it can be said that the data provided here is related to a more limited database or source. Countries such as Turkey, Taiwan, Pakistan are in the middle level with 93, 83 and 81 publications respectively. These countries also make scientific contributions but have fewer publications than other major research countries. Countries such as France and South Africa are also in the lower part of the table with around 30 publications. Japan 27 publications, countries such as Mexico, South Africa, France are also included in the list but they have low publication numbers.

Publisher distribution of publications

As a result of the analysis, the publisher distributions of the publications are given in Table 5.

Table 5: Publisher distribution of publications

Publishers	**Record Count**	**% of 3.873**
Elsevier	893	23.057
Springer Nature	716	18.487
IEEE	504	13.013
Mdpi	465	12.006
Wiley	173	4.467
Taylor & Francis	128	3.305
Tech Science Press	74	1.911
Hindawi Publishing Group	71	1.833
Sage	62	1.601
Ios Press	48	1.239
NATURE PORTFOLIO	45	1.162
World Scientific	39	1.007
Wiley-Hindawi	37	0.955
Frontiers Media Sa	36	0.93
Emerald Group Publishing	30	0.775
Igi Global	30	0.775
Amer Inst Mathematical Sciences-Aims	23	0.594
Iop Publishing Ltd	23	0.594
Assoc Computing Machinery	18	0.465
Oxford Univ Press	18	0.465
Inst Engineering Technology-Iet	17	0.439
Public Library Science	16	0.413
Science & Information Sai Organization Ltd	15	0.387
Walter De Gruyter	14	0.361
AIP Publishing	12	0.31

When Table 5 is examined, it is seen that Elsevier (893 publications) is the publisher with the highest number of publications. Elsevier is one of the largest publishers in the academic world and has a wide impact with scientific journals, books, and conference proceedings. Other publishers are Springer Nature (716 publications), IEEE (504 publications), Mdpi (465 publications), Wiley (173 publications), Taylor & Francis (128 publications), Tech Science Press (74 publications), Hindawi Publishing Group (71 publications), Sage (62 publications), Ios Press (48 publications), Nature Portfolio (45 publications), World Scientific (39 publications), Wiley-Hindawi (37 publications), Frontiers Media Sa (36 publications), Emerald Group Publishing (30 publications) Igi Global (30 publications), Amer Inst Mathematical Sciences-Aims (23 publications), Iop Publishing Ltd (23 publications), Oxford Univ Press (18 publications), Inst Engineering Technology-Iet (17 publications), Public Library Science (16 publications), Science & Information Sai Organization Ltd (15 publications), Walter De Gruyter (14 publications), AIP Publishing (12 publications).

Co-authorship works by authors

As a result of the analysis, the connection between the co-authorship of the publications is given in Figure 7.

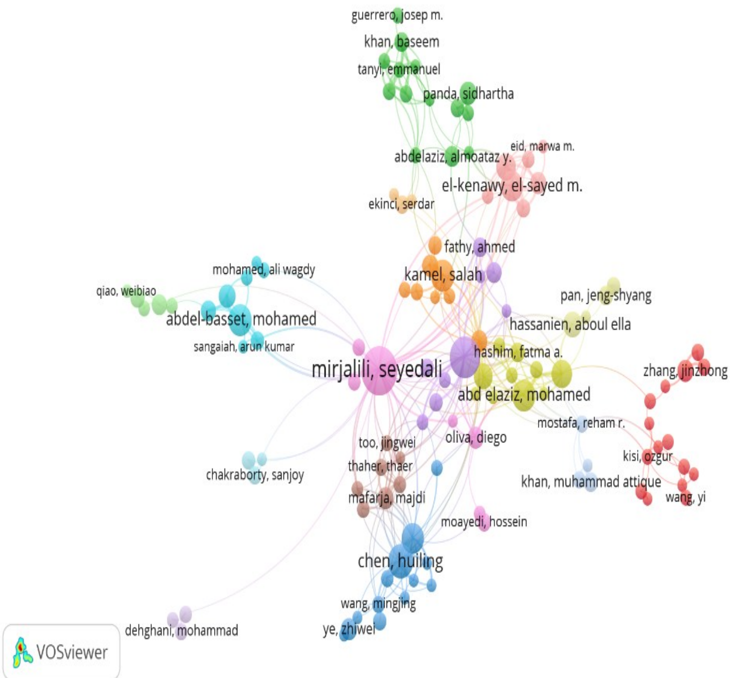


Figure 6: Connection between co-authored authors of publications

Figure 7 summarizes the research performances of the authors with the most co-authored work in terms of the number of publications, number of citations, and total link strength. Link strength refers to the collaborations the researcher has established with other authors and the strength of the network. The most active author, Mohamed Houssein (41 publications, 3859 citations, 76 link strength), has the highest link strength. Seyedali Mirjalili (58 publications, 15450 citations, 75 link strength) is by far the most cited author. Names such as Amir H. Gandomi and Hashim Fatma A. are seen to have high values in citation numbers. El-Sayed M. El-Kenawy (59 link strength) and Mohamed Houssein (76 link strength) are seen to be authors with high collaboration potential.

The connection between organizations' citation analysis

As a result of the analysis, the connection between organizations and citation amounts is given in Figure 8.

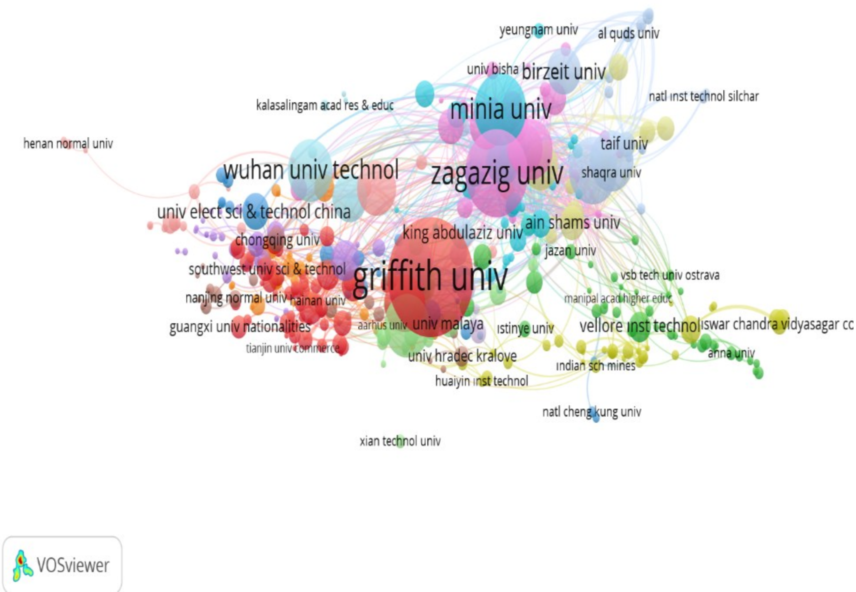


Figure 7: Network visualization graph according to the citation counts of institutions

According to Figure 7, Griffith University has the highest number of citations with 11,160, followed by Zagazig University with 5,926 citations, Minia University with 4,133 citations, Torrens University with 3,527 citations and Islamic Azad University with 3,174 citations. Griffith University stands out as a university with high research impact by receiving the highest number of citations with 20 published documents. Zagazig University, despite having published more documents, lags Griffith in terms of the number of citations but draws attention with its strong network of connections

(212). Minia University has similarly received many citations (127 link strength) but has relatively few published documents (53). Islamic Azad University, on the other hand, lags the others in terms of the number of citations and link strength, despite being the university that has published the most documents. This ranking shows that the number of citations reflects the research impact of universities, but the strength of links and the number of documents are also important factors.

Co-authorship country analysis

As a result of the analysis, the connection between the countries where co-authored publications work is given in Figure 8.

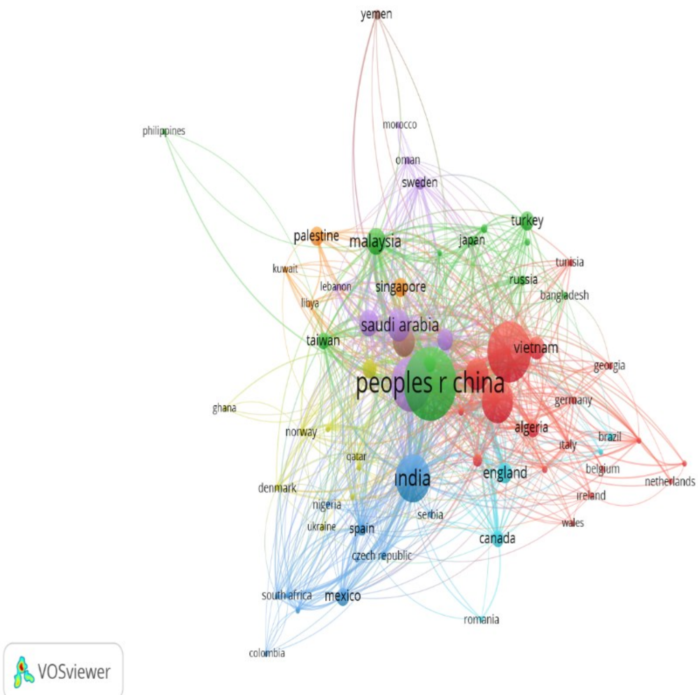


Figure 8: The connection between the countries in which co-authored publications work

According to the data presented in Figure 9, China has the highest number of publications with 1,824 documents and the highest impact with 26,881 citations, and also has the strongest collaborative network with a total link strength of 650. Egypt, despite publishing relatively few documents with 350 documents, shows a remarkable impact with 14,965 citations and 513 link strength. India is productive and influential with 914 documents and 11,432 citations, and also exhibits strong collaboration with 394 link strength. Australia, despite publishing fewer documents with 133 documents, has a high impact with 18,260 citations and 293 link strength. Iran, although having lower document and citation number, have a significant collaboration capacity with 9,144 citations and 384 link strength. Türkiye, on the other hand, appears to have a more limited area of influence with 1,860 citations and 51 link strengths.

Co-occurrence analysis of all keywords

As a result of the analysis, the network visualization graph according to the number of keywords is given in Figure 9.

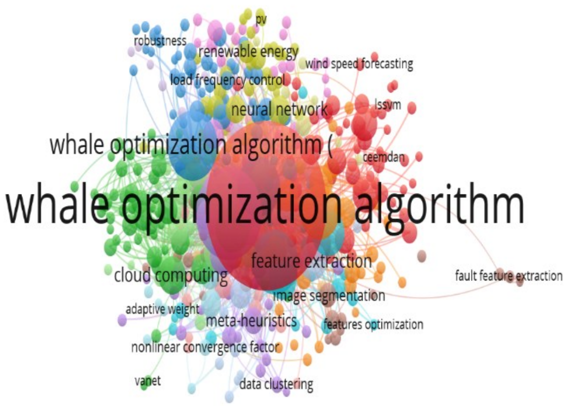


Figure 9: Network visualization graph based on keyword counts

The data obtained within the scope of the research conducted show that the keywords related to "Whale Optimization Algorithm" are used much more than others and have an effective collaboration network. "Whale Optimization Algorithm" is mentioned 1,214 times and has the highest total connection strength with 1,907. "Optimization" is seen to have an important place with 1,092 connection strength with 370 times. "Feature Selection" is mentioned 182 times and draws attention with 434 connection strength. "Machine Learning" and "Deep Learning" are used 126 and 90 times respectively and have reasonable connection strengths.

Analysis of publications according to citation count

As a result of the analysis, the connection between the responsible author and the publication year in terms of the number of citations is given in Figure 10.

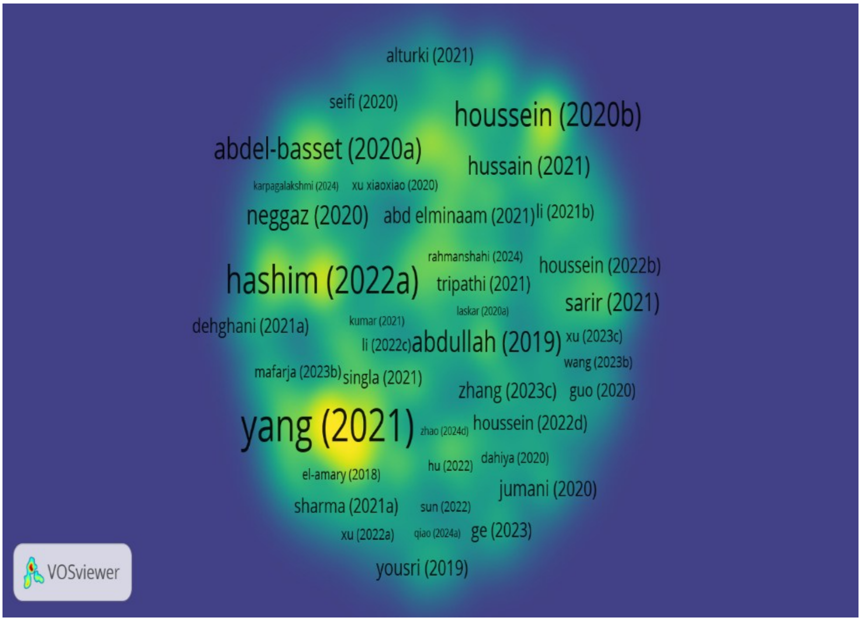


Figure 10: The relationship between the number of citations of publications based on the published work

According to the data obtained in the study, the most cited document is (Yang et al., 2021) and it is seen that it has the highest impact with 723 citations. (Hashim et al., 2022) ranks second with 582 citations, while (Hashim & Hussien, 2022), which belongs to the same author, has 467 citations, which shows that the author's work has a significant impact. (Houssein et al., 2020) ranks fourth with

300 citations, while (Agrawal et al., 2021) manages to enter the top five most cited publications with 251 citations.

The connection between the citations in the published sources

As a result of the analysis, the connection between Figure 11 and the number of citations in the published sources is given.

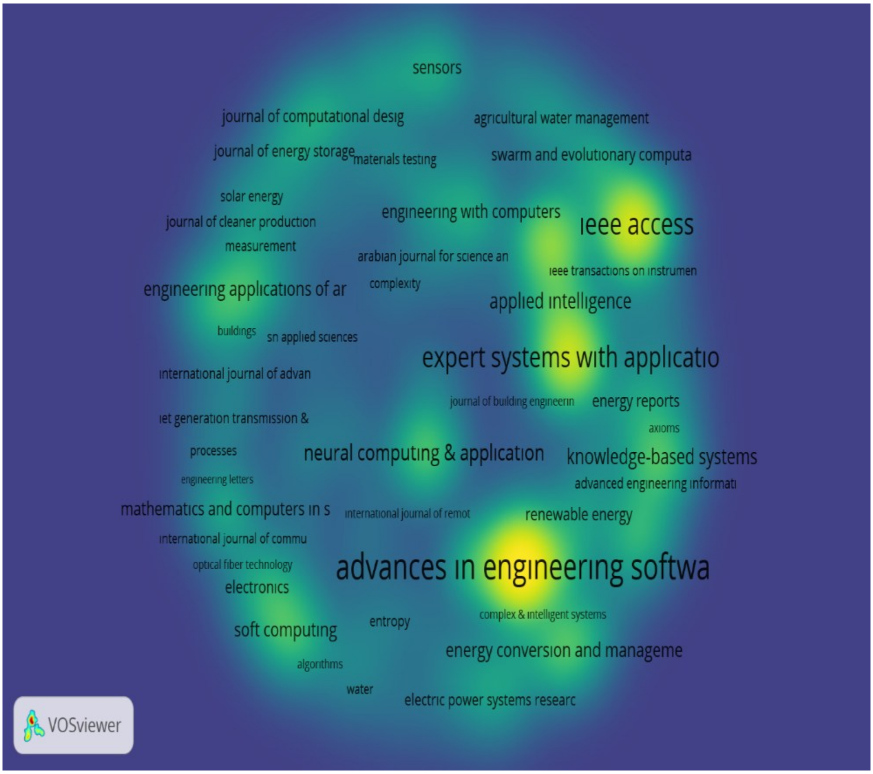


Figure 11: the relationship between the number of citations of publications in published sources

The data obtained through the study show the academic productivity and impact levels of the five sources mentioned. IEEE Access has the most documents (179) and has received 4799 citations, indicating a wide coverage and high impact potential. Expert Systems with Applications has received 4263 citations with 75 documents, providing a high citation rate per document. The most striking source is Advances in Engineering Software, which has received 8001 citations with only 10 documents, showing a very high impact rate. Applied Soft Computing and Neural Computing & Applications have relatively lower citation numbers but have managed to enter the top five sources.

Citation analysis of published authors

As a result of the analysis processes, the citation analysis of the published authors is given in Figure 12.

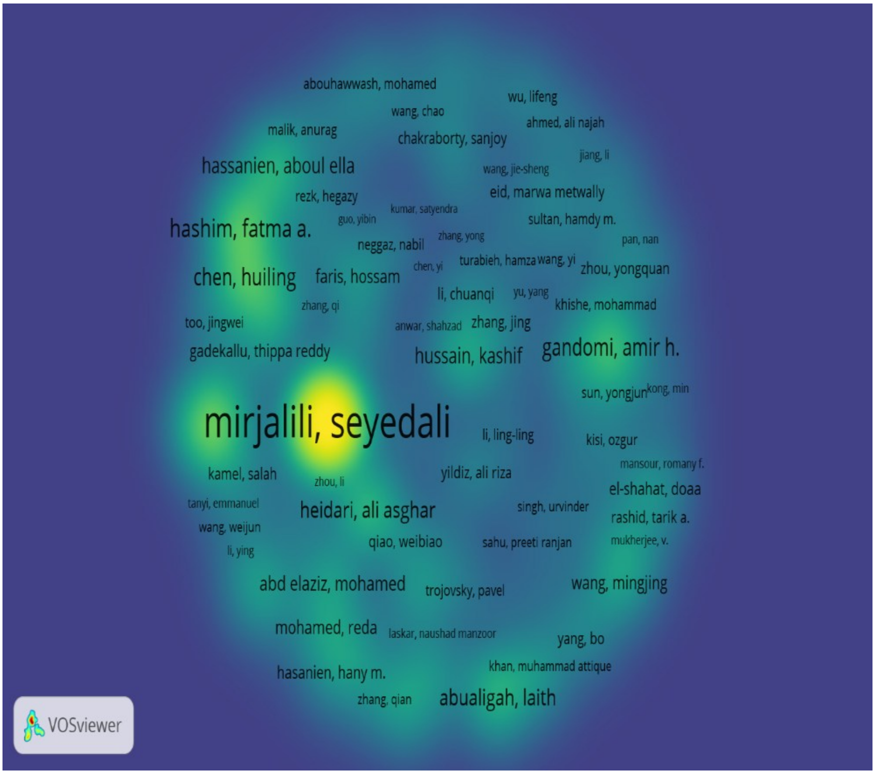


Figure 12: Citation analysis of published authors

Figure 12 shows the academic productivity and impact levels of the authors. Seyedali Mirjalili, has the most documents (58) and has 15,450 citations, which shows that they have made a significant impact. Essam H. Houssein, has 3859 citations in 41 documents, which shows a significant impact and productivity. Hashim, Fatma A. has 2748 citations with 10 documents, which shows that their research is effective but has a more limited productivity in terms of number. Huiling Chen, and Amir H. Gandomi, have 2356 and 2424 citations with 29 and 6 documents, respectively. Amir H. Gandomi, has a higher number of citations despite having fewer publications with 6 publications and 2424 citations, which shows that their research is heavily cited.

A term co-occurrence map based on text data

As a result of the analysis, a term association map based on text data is given in Figure 13.

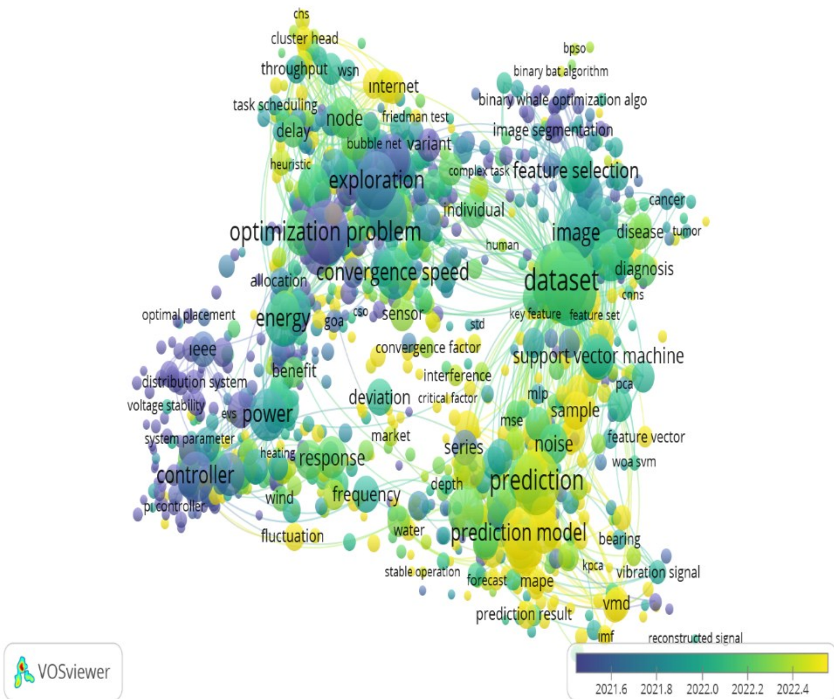


Figure 13: A term co-occurrence map based on text data

The map given in Figure 13 shows how frequently certain terms are used in the research and their importance in the relevant context. The term "prediction model" stands out as the most strongly associated term with the highest relevance score (2.2132), meaning that this term is considered to be very important for research. The terms "prediction" (1.6489) and "optimization problem" (1.5347) also have a very high relevance. Among the other terms, "dataset" (0.4931) and "classification" (0.4027) appear with lower relevance scores but are still terms with significant frequencies. "energy" (0.609) and "feature" (0.6187) have an average relevance.

3. Results

In this study, 3873 works published in the last nine years covering the years 2016 - 2024 were accessed in the WoS database using the keywords ("whale optimization algorithm") OR ("whales optimization algorithm") OR ("whale algorithm") OR ("whales algorithm"). Within the scope of the research, a bibliometric study was conducted in terms of various criteria and the following results were obtained.

According to WoS categories, fields such as Electrical, Electronics, Artificial Intelligence and Computer Science stand out as the fields with the most records in the Web of Science database. This shows that these fields have an important place in today's scientific research and offer a wide range of research. Multidisciplinary and Cross-Disciplinary Research (e.g. "Engineering Multidisciplinary" and "Computer Science Interdisciplinary Applications") also reflect an important trend. This shows that scientific research interacts more between different fields and tries to solve problems from broader perspectives. More Niche Fields (e.g. Mathematics and Engineering Chemical) may generally contain more specific and in-depth studies, but the number of studies in these fields appears to be lower.

When the data is examined according to the publication years, it is seen that there has been an increase in the number of publications especially after 2020 and that more research is published. This trend shows that scientific research and publications are increasing and becoming more visible. The fact that this increase continues to accelerate from 2021 may indicate an important period of scientific progress.

When we look at the distribution of document types, it is determined that studies in article format are preferred by a very large margin.

When the citation distributions of publications are examined according to the meso criterion, it is seen that the high citation numbers are in areas that are very important in terms of society and economy such as Supply Chain & Logistics and Power Systems & Electric Vehicles. These areas play an important role both in the search for solutions to current global problems (e.g. supply chain disruptions, sustainable energy) and in technological innovations. New and popular areas such as artificial intelligence, machine learning and telecommunications have gained great momentum in recent years and more and more researchers are working in these areas. Therefore, more citation numbers are reached in these subjects. When the citation distributions of publications are examined according to the micro criterion, the high citation numbers have an important place in the field of Particle Swarm Optimization (786 citations), machine learning and optimization. This situation reflects the influence of broader and multidisciplinary research. Some topics are much more specific and narrow in scope (e.g. Proton Conductivity and Doubly Fed Induction Generator). Such topics appeal to fewer researchers and therefore, it can be said that the citation numbers are lower. Developing and popular areas such as cloud computing, IoT, deep learning are seen to be gaining popularity with technological advances and the expansion of application areas. Therefore, it can be said that the studies conducted in these areas are increasing day by day.

When the publications are examined according to WoS indexes, it is revealed that scientific articles and conference proceedings are mostly concentrated in SCI-EXPANDED and CPCI-S indexes; social sciences and books are relatively less represented. This situation may be a typical distribution reflecting the interdisciplinary density in the literature.

Journal names according to publication titles show that the most popular journals are generally broad-scope engineering and scientific journals, but journals that deepen in certain areas also publish effectively. Such data is thought to be quite useful to see which journals publish more studies in which areas.

The study concluded that the majority of academic literature is in English and that other languages have a more limited representation in academic publications around the world.

In the distribution of publications by country, it was determined that developing countries such as China and India are places where scientific publications and research are rapidly increasing. This situation shows that education systems and research infrastructures are getting stronger. Developed countries (America, Europe) generally have more resources and more data, but the fact that fewer publications are seen in the data mentioned here may suggest that databases may be limited or focused on specific areas. Countries in regions such as the Middle East and North Africa (Egypt, Saudi Arabia, Iran) in particular indicate developing academic ecosystems.

In terms of publisher distribution, large publishers such as Elsevier, Springer Nature, IEEE have been found to have a great impact on scientific publishing and publish a large number of academic resources. It can be said that these publishers generally appeal to a wide range of audiences and have a great academic impact worldwide.

It was found that the highest number of citations in studies conducted with co-authors was obtained by Seyedali Mirjalili (58 publications, 15450 citations, 75 connection strength).

According to the number of citations of institutions, the highest value was obtained by Griffith University with 11,160.

According to the criterion of countries where co-authored publications work, China has the highest number of publications with 1,824 documents and the highest impact with 26,881 citations, and it was found to have the strongest collaborative network with a total connection strength of 650.

According to the keyword analysis, it was found that the themes of "Whale Optimization Algorithm" and "Optimization" are used intensively in academic studies and create strong collaborations.

The most cited document is (Yang et al., 2021) and it was found to have the highest impact with 723 citations.

According to the connection criterion between the number of citations in the published sources, IEEE Access was found to have the most documents (179). A term association map based on text data was created. According to this map, it was found that the words prediction model, prediction, optimization problem, dataset, classification, energy and feature came to the fore.

References

Agrawal, P., Abutarboush, H. F., Ganesh, T., & Mohamed, A. W. (2021). Metaheuristic Algorithms on Feature Selection: A Survey of One Decade of Research (2009-2019). *IEEE Access*, *9*, 26766-26791. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3056407

Al-Quraan, A., Al-Mhairat, B., Malkawi, A. M. A., Radaideh, A., & Al-Masri, H. M. K. (2023). Optimal Prediction of Wind Energy Resources Based on WOA—A Case Study in Jordan. *Sustainability*, *15*(5), Article 5. https://doi.org/10.3390/su15053927

Chandrasekaran, S. K., & Rajasekaran, V. A. (2024). Energy-efficient cluster head using modified fuzzy logic with WOA and path selection using enhanced CSO in IoT-enabled smart agriculture systems. *The Journal of Supercomputing*, *80*(8), 11149-11190. https://doi.org/10.1007/s11227-023-05780-5

Chen, X. (2024). Location strategy for logistics distribution centers utilizing improved whale optimization algorithm. *Journal of Intelligent Systems*, *33*(1). https://doi.org/10.1515/jisys-2023-0299

Cooper, I. D. (2015). Bibliometrics basics. *Journal of the Medical Library Association : JMLA*, *103*(4), 217-218. https://doi.org/10.3163/1536-5050.103.4.013

Dai, Y., Yu, J., Zhang, C., Zhan, B., & Zheng, X. (2023). A novel whale optimization algorithm of path planning strategy for mobile robots. *Applied Intelligence*, *53*(9), 10843-10857. https://doi.org/10.1007/s10489-022-04030-0

Deepa, R., & Venkataraman, R. (2021). Enhancing Whale Optimization Algorithm with Levy Flight for coverage optimization in wireless sensor networks. *Computers & Electrical Engineering*, *94*, 107359. https://doi.org/10.1016/j.compeleceng.2021.107359

Dewi, S. K., & Utama, D. M. (2021). A New Hybrid Whale Optimization Algorithm for Green Vehicle Routing Problem. *Systems Science & Control Engineering*, *9*(1), 61-72. https://doi.org/10.1080/21642583.2020.1863276

Hashim, F. A., Houssein, E. H., Hussain, K., Mabrouk, M. S., & Al-Atabany, W. (2022). Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems. *Mathematics and Computers in Simulation*, *192*, 84-110. https://doi.org/10.1016/j.matcom.2021.08.013

Hashim, F. A., & Hussien, A. G. (2022). Snake Optimizer: A novel meta-heuristic optimization algorithm. *Knowledge-Based Systems*, *242*, 108320. https://doi.org/10.1016/j.knosys.2022.108320

Houssein, E. H., Saad, M. R., Hashim, F. A., Shaban, H., & Hassaballah, M. (2020). Lévy flight distribution: A new metaheuristic algorithm for solving engineering optimization problems. *Engineering Applications of Artificial Intelligence*, *94*, 103731. https://doi.org/10.1016/j.engappai.2020.103731

Li, C., & Chen, L. (2024). Optimization for energy-aware design of task scheduling in heterogeneous distributed systems: A meta-heuristic based approach. *Computing*, *106*(6), 2007-2031. https://doi.org/10.1007/s00607-024-01282-1

Mirjalili, S., & Lewis, A. (2016). The Whale Optimization Algorithm. *Advances in Engineering Software*, *95*, 51-67. https://doi.org/10.1016/j.advengsoft.2016.01.008

Nadimi-Shahraki, M. H., Zamani, H., Asghari Varzaneh, Z., & Mirjalili, S. (2023). A Systematic Review of the Whale Optimization Algorithm: Theoretical Foundation, Improvements, and Hybridizations. *Archives of Computational Methods in Engineering*, *30*(7), 4113-4159. https://doi.org/10.1007/s11831-023-09928-7

Oliva, D., Abd El Aziz, M., & Ella Hassanien, A. (2017). Parameter estimation of photovoltaic cells using an improved chaotic whale optimization algorithm. *Applied Energy*, *200*, 141-154. https://doi.org/10.1016/j.apenergy.2017.05.029

Paul, C., Roy, P. K., & Mukherjee, V. (2023). Wind and Solar Based Multi-Objective Hydro-Thermal Scheduling Using Chaotic-Oppositional Whale Optimization Algorithm. *Electric Power Components and Systems*, *51*(6), 568-592. https://doi.org/10.1080/15325008.2023.2179130

Peddakapu, K., Mohamed, M. R., Srinivasarao, P., & Arya, Y. (2024). Assessment of energy storage and renewable energy sources-based two-area microgrid system using optimized fractional order controllers. *Journal of Energy Storage*, *86*, 111191. https://doi.org/10.1016/j.est.2024.111191

Pham, V. H. S., Nguyen, V. N., & Nguyen Dang, N. T. (2024). Hybrid whale optimization algorithm for enhanced routing of limited capacity vehicles in supply chain management. *Scientific Reports*, *14*(1), 793. https://doi.org/10.1038/s41598-024-51359-2

Yahya, W., Saied, K. M., Nassar, A., Qader, M. R., Al-Nehari, M., Zarabia, J., & Jian, Z. (2024). Optimization of a hybrid renewable energy system consisting of a of PV/wind turbine/battery/fuel cell integration and component design.

International Journal of Hydrogen Energy, *94*, 1406-1418. https://doi.org/10.1016/j.ijhydene.2024.11.187

Yang, Y., Chen, H., Heidari, A. A., & Gandomi, A. H. (2021). Hunger games search: Visions, conception, implementation, deep analysis, perspectives, and towards performance shifts. *Expert Systems with Applications*, *177*, 114864. https://doi.org/10.1016/j.eswa.2021.114864

Zhou, J., Zhu, S., Qiu, Y., Armaghani, D. J., Zhou, A., & Yong, W. (2022). Predicting tunnel squeezing using support vector machine optimized by whale optimization algorithm. *Acta Geotechnica*, *17*(4), 1343-1366. https://doi.org/10.1007/s11440-022-01450-7

CHAPTER II

Deep Learning Approaches For Accurate And Efficient Classification Of Eye Diseases

Kamil ORMAN[1]
Yıldız AYDIN[2]

1. Introduction

According to the World Health Organization, in the rapidly growing age of technology, eighty percent of the five senses are related to vision. Behind this increase in vision problems are various eye diseases such as cataracts, glaucoma, diabetic retinopathy, color blindness and refractive eye errors. Early diagnosis and treatment of these diseases play a critical role in preventing vision loss. In this direction, many companies worldwide focus on early diagnosis and classification applications by developing artificial intelligence-based models. Artificial intelligence determines eye diseases at an early stage and ensures that treatment processes are more effective.

[1] Department of Computer Engineering, Erzincan Binali Yildirim University, Erzincan 24000, Turkey.
[2] Department of Computer Engineering, Erzincan Binali Yildirim University, Erzincan 24000, Turkey.

Traditional methods include full-body radiological tests such as MRI, CT scans, live X-rays and 3D models, and these techniques require high-resolution, clear and well-focused images for doctors and technical staff to obtain error-free results. However, these methods are time-consuming and increase workload. The vast majority of vision disorders can be prevented or treated when diagnosed at an early stage. Diseases such as glaucoma, in particular, may not show symptoms until they cause irreversible vision loss. Therefore, regular eye examinations and optic nerve examinations performed by dilating the pupil are of great importance.

In recent years, high-resolution retinal imaging techniques such as fundus photography and optical coherence tomography (OCT) have been widely used in the diagnosis and treatment of ocular diseases. These techniques provide detailed information about the retinal structure and enable the detection of disease conditions such as ischemia, inflammation and neovascularization. However, these imaging methods can create a serious burden on doctors' interpretation skills due to the increase in the number of images produced (Tang, Yang, Wang, Wu, & Zhang, 2023) .

The rapid development of technology has made it possible to use deep learning techniques to overcome these difficulties. Deep learning analyzes large amounts of retinal images, provides fast and accurate diagnosis, reduces workload, increases the possibility of early intervention and eliminates the possibility of misdiagnosis (Elkholy & Marzouk, 2023). However, automated systems have been widely applied and provided significant benefits not only in the classification of eye diseases but also in different medical disciplines. Breast cancer diagnosis system (Dutta & Bandyopadhyay, 2020; Irmak, Tas, Turan, & Hasiloglu, n.d.), classification of brain tumors (Şahin, Özdemir, & Temurtaş, 2024), skin cancer diagnoses(Aydin, 2023) are just a few examples from these different disciplines. The main purpose of this study is to develop a fast, reliable and time-saving system for the classification and early diagnosis of the most common eye diseases worldwide using image processing and deep learning methods. The diseases covered in the study are as follows:

- Cataract
- Diabetic Retinopathy
- Glaucoma
- Normal (Healthy Eye)

Eye diseases are among the important health problems that need accurate and fast classification methods for early diagnosis. In the literature, various image processing and deep learning-based approaches have been developed to classify eye diseases. However, since these methods generally require large data sets and powerful computational resources, the difficulties of working with limited data continue. In addition, although some methods achieve high accuracy rates, they have difficulty coping with problems such as diversity and data imbalance encountered in the real world.

In this study, eye diseases were classified using InceptionV3, DenseNet121, MobileNetV2, ConvNeXtBase deep learning methods. In addition, data augmentation techniques were used to increase the generalization ability of the model. The model was terminated with softmax activation to recognize four different eye disease classes (cataract, diabetic retinopathy, glaucoma and normal).

Unlike many deep learning-based eye disease classification methods in the literature, the study aimed to achieve high accuracy with a limited data set. During the training process, the performance of the model was tested and the results obtained revealed that transfer learning methods provide an effective solution for such classification problems. This study shows that transfer learning offers great potential especially for medical image classification applications working with limited data. Future studies aim to further improve the performance of the model by testing it with larger data sets and different pre-trained models. This system enables rapid detection of diseases, allowing early intervention and aims to take the necessary steps to prevent disease progression. In this direction, performance comparisons were made using various deep learning methods and the aim was to determine the most effective model.

2. Method

In this study, a deep learning-based method was developed for the classification of eye diseases and transfer learning techniques were used. The main purpose of the study is to achieve high accuracy with a limited dataset. In this context, InceptionV3, DenseNet121, MobileNetV2, ConvNeXtBase deep learning methods were used as the basic structure in feature extraction and classification processes.

The dataset obtained from the Kaggle (Guna Venkat Doddi, n.d.) open access platform was divided into four classes: cataract, diabetic retinopathy, glaucoma and healthy eye. The same preprocessing steps were performed to ensure fair comparison to all deep learning methods. The images were resized to 256x256 to fit the model input dimensions and normalized in the range of 0-1. Data augmentation techniques were applied to increase the diversity of images in the dataset and improve the classification performance. These techniques include horizontal reflection, shift, zoom and brightness changes.

In the model design where the most successful results were obtained, the pre-trained weights of the DenseNet121 model (ImageNet) were used. The basic layers of DenseNet121 were kept constant and additional layers were added on top of these layers that provide deep feature classification. In particular, a Flatten layer, a fully connected layer with 1024 neurons ReLU activation, Batch Normalization and Dropout layers were added to the last part of the model. A softmax activation function that gives four class outputs was used in the last layer for the classification process. The flowchart of the suggested approach is given in the Figure 1.

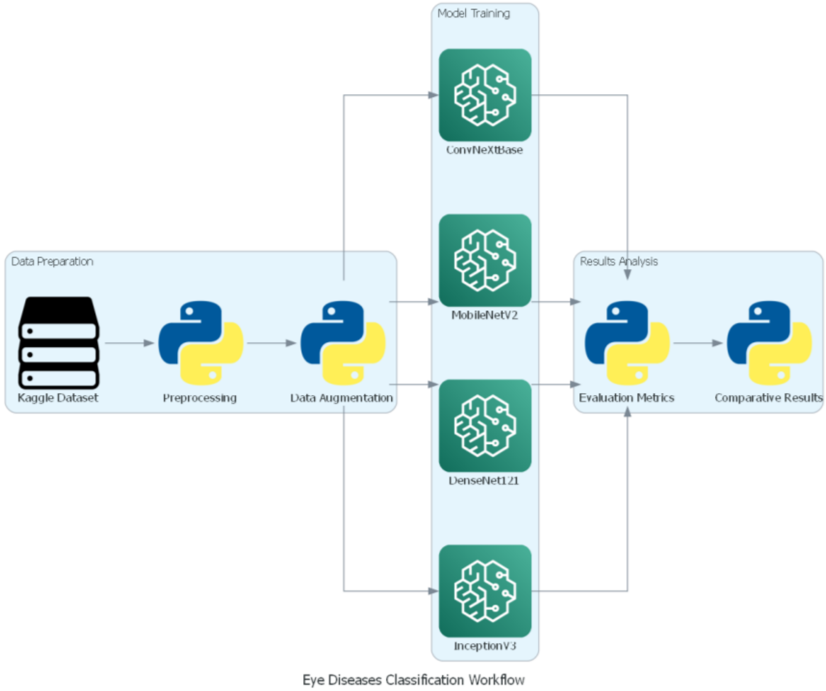


Figure 1. Flowchart of the applications

In the training process of the model, the Adam optimization algorithm was used and training was performed with categorical cross-entropy loss. In order to prevent over-learning during training and to increase the generalization ability of the model, early stopping and saving the best weights mechanisms (ModelCheckpoint) were applied. The model was trained with 70% of the training data and validated with a 30% test dataset.

3. Experimental Results

In this study, we conducted a comprehensive performance comparison of deep learning-based methods for the classification of eye diseases. The dataset used was sourced from the Kaggle open-access platform and contained images categorized into four classes: cataract, diabetic retinopathy, glaucoma, and normal (healthy eye). The class distribution was as follows: cataract (1038 images), diabetic retinopathy (1098 images), glaucoma (1007 images), and normal (1074 images).

All experiments were implemented on the Google Colab platform, leveraging its computational resources. The deep learning architectures utilized include DenseNet121, MobileNetV2, ConvNeXtBase, and Xception. These models were chosen based on their popularity and proven performance in image classification tasks. The highest accuracy, 89%, was achieved using DenseNet121, underscoring its robustness in handling the dataset. Table 1 summarizes the performance metrics—precision, recall, F1-score, and accuracy—for each model.

Table 1. Comparative Results of Deep Learning Methods

Technique Name	Recall	Precision	F1-score	Accuracy
InceptionV3	0.85	0.85	0.85	0.85
DenseNet121	**0.89**	**0.89**	**0.89**	**0.89**
MobileNetV2	0.88	0.88	0.88	0.88
ConvNeXtBase	0.64	0.62	0,61	0.62

To further analyze model performance, confusion matrices were generated for each technique (Figure 2). These matrices provide insights into the classification performance across different classes, highlighting areas where misclassifications occurred.

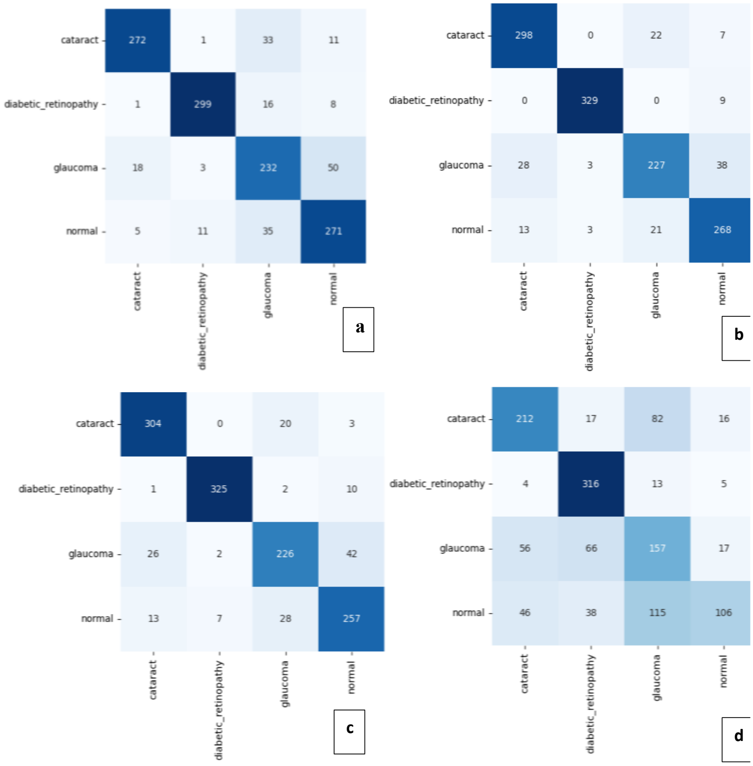


__Figure 2.__ Confusion Matrices for Deep Learning Methods. (a) DenseNet121, (b) MobileNetV2, (c) ConvNeXtBase, (d) InceptionV3.

The results demonstrated that DenseNet121 outperformed other models in all metrics, achieving the highest precision, recall, F1-score, and accuracy. The superior performance of DenseNet121 can be attributed to its densely connected architecture, which facilitates better feature propagation and reuse.

In summary, the experimental findings highlight the potential of DenseNet121 as an effective solution for eye disease classification tasks, particularly in scenarios with limited datasets.

Future work will aim to improve these results by expanding the dataset and incorporating additional pre-trained models to further enhance performance.

4. Conclusions

In this study, a comparative analysis of deep learning-based methods for the classification of eye diseases using public dataset is presented. Among the tested models, DenseNet121 showed the best performance, achieving the highest accuracy and balanced metrics. These results indicate that DenseNet121 can be effective in medical image classification problems.

Future research aims to improve the performance on different datasets by using different pre-trained architectures and advanced optimization techniques to improve the classification performance. The findings of this study are preliminary for future research in the field of automatic eye disease detection and highlight the potential of utilizing deep learning models to improve diagnostic accuracy in ophthalmology.

References

Aydin, Y. (2023). A Comparative Analysis of Skin Cancer Detection Applications Using Histogram-Based Local Descriptors. *Diagnostics*, *13*(19). https://doi.org/10.3390/diagnostics13193142

Dutta, S., & Bandyopadhyay, S. K. (2020). Early Breast Cancer Prediction using Artificial Intelligence Methods. *Journal of Engineering Research and Reports*, *13*(2), 48–54. https://doi.org/10.9734/jerr/2020/v13i217105

Elkholy, M., & Marzouk, M. A. (2023). Deep learning-based classification of eye diseases using Convolutional Neural Network for OCT images. *Frontiers in Computer Science*, *5*. https://doi.org/10.3389/fcomp.2023.1252295

Guna Venkat Doddi. (n.d.). https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification. Retrieved from https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification

Irmak, M. C., Tas, M. B. H., Turan, S., & Hasiloglu, A. (n.d.). Comparative breast cancer detection with artificial neural networks and mdiagachine learning methods. *SIU 2021 - 29th IEEE Conference on Signal Processing and Communications Applications, Proceedings*, 7–10. İstanbul, Turkey, 9–11 June 2021. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85111415921&doi=10.1109%2FSIU53274.2021.9477991&partnerID=40&md5=d5e0b611834622953ce912196ff280dd

Şahin, E., Özdemir, D., & Temurtaş, H. (2024). Multi-objective optimization of ViT architecture for efficient brain tumor classification. *Biomedical Signal Processing and Control*, *91*(November 2023). https://doi.org/10.1016/j.bspc.2023.105938

Tang, Q. Q., Yang, X. G., Wang, H. Q., Wu, D. W., & Zhang, M. X. (2023). Applications of deep learning for detecting ophthalmic diseases with ultrawide-field fundus images. *International Journal of Ophthalmology*, *17*(1), 188–200. https://doi.org/10.18240/ijo.2024.01.24

CHAPTER III

Artificial Intelligence Methods in Aquaculture: Review, Application and Challenges

Özge ZENCİR TANIR[1]
Volkan KAYA[2]

1. Introduction

Aquaculture, also known as aquaculture, is a sector that covers the production of animal (fish, shellfish, arthropods and mollusks) and plant (algae) aquatic organisms that people need for a healthy and balanced diet in controlled or semi-controlled environments. These activities, carried out for food supply, stock renewal, ornamental fish production, sports and scientific purposes, contribute to employment and provide significant support to the economy by creating high added value (Başçınar, 2004).

[1] Assoc. Prof. Dr., Erzincan Binali Yıldırım University, Faculty of Arts and Science, Department of Biology, Erzincan/Türkiye, Orcid: 0000-0002-2111-7019, zencir06@hotmail.com
[2] Assoc. Prof. Dr., Erzincan Binali Yıldırım University, Faculty of Engineering and Architecture, Department of Computer Engineering, Erzincan/Türkiye, Orcid: 0000-0001-6940-3260, vkaya@erzincan.edu.tr

Aquaculture in Turkey began in the late 1960s and early 1970s with the production of rainbow trout and mirror carp. In the 1980s, marine fish were targeted, and significant progress was made in the sector with the cultivation of sea bream and European sea bass. In the 1990s, there was a significant increase in the production of trout, sea bream, and sea bass, and studies were initiated on the cultivation of alternative species. Over time, this sector has not only grown in volume but also in technological sophistication, incorporating more modern farming practices, advanced feed systems, and aquaculture techniques. As Turkey continues to develop its aquaculture capabilities, the focus is increasingly shifting toward sustainability and the introduction of new, more resilient species that can thrive in diverse environmental conditions. This evolution has positioned Turkey as a key player in the global aquaculture industry, offering significant potential for export and meeting the growing demand for seafood. In the 2000s, there were significant developments in tuna cultivation in the Aegean and Mediterranean regions (Yıldırım & Okumuş, 2004; Demir, 2023). The share of aquaculture in total aquaculture production, including marine fish, has also increased. Today, almost half of the aquaculture products consumed come from the aquaculture sector.

Surrounded by sea on three sides, Turkey has an ideal potential for aquaculture with its 8,333 kilometers of coastline. The country has a total of 1,193 inland water resources, including 200 natural lakes, 243 dam lakes and 750 irrigation ponds. It also has a river network of 177,714 kilometers in length. In our seas, 247 fish species have been identified in the Black Sea, 200 in the Sea of Marmara, 300 in the Aegean Sea, and nearly 500 in the

Mediterranean. Around 100 of these species hold significant economic value (Anonim, 2012).

The total annual production in aquaculture across seas, lakes, dam lakes, and rivers ranges between 800,000 and 850,000 tons. According to the Ministry of Agriculture and Forestry's 2022 Aquaculture Report, Turkey's aquaculture production grew by 6% compared to the previous year, reaching 849,808 tons. In 2022, hunting production increased by 2% and aquaculture production increased by 9%. The majority of aquaculture production in our country consists of sea bass, sea bream and trout (TUİK, 2022)

The total surface area of our seas and inland waters is 25 million hectares, which is almost equivalent to our agricultural lands. This situation reveals the importance of using aquaculture resources effectively and sustainably. In order to obtain maximum efficiency from these resources in the future, sector-specific problems must be solved and necessary precautions must be taken. In recent years, the use of technology has come to the fore in issues such as the labor and costs spent in increasing the amount of production, disease control, protection of water resources due to global warming, etc. (Mustapha et al. 2021). Innovative technologies have led to the development of various ideas in the field of aquaculture, one of which is the application of artificial intelligence techniques in the sector.

2. Overview of Artificial Intelligence Methods

Emerging in the 1950s, Artificial Intelligence is a technology that enables machines to perform tasks with human-like abilities (Yılmaz, 2019). While weak AIs are limited to performing tasks as programmed, strong AIs are systems capable of performing algorithmic calculations to enhance the program’s performance and learn from mistakes. Since the 2010s, deep learning has become an important tool for solving complex problems using multilayered artificial neural networks. Furthermore, technological advancements like big data analytics and cloud computing are supporting and shaping AI research.

Driven by the rapid advancement of technology, artificial intelligence techniques are now widely used. In its simplest form, artificial intelligence refers to systems designed to imitate certain aspects of human intelligence. More broadly, it can be defined as computer software aimed at replicating human cognitive functions. Artificial intelligence is a collection of software and hardware systems that can imitate human intelligence, exhibit human-like behavior, reason, improve themselves with the experience they gain, and turn these developments into action (Russell et.al., 2016; Jiang et al., 2017).

Artificial intelligence applications are focused on areas such as pattern recognition, system modeling and classification rather than a conscious computer system. Artificial intelligence is effectively used in many sectors such as medicine, construction, urban planning and energy due to its advantages such as low cost, fast processing, solving complex problems, working with small samples and being efficient compared to classical methods.

2.1. Artificial Intelligence Methods

Artificial intelligence methods consist of many different algorithms to solve complex problems and are used in various fields. Among the artificial intelligence methods that include different techniques and methods, the most commonly used are Machine Learning and Deep Learning methods.

2.1.1. Machine Learning

Machine learning, a subfield of artificial intelligence, focuses on teaching computers how to perform tasks through algorithms, rather than pre-programming each step to solve a problem. This approach means developing algorithms that find solutions to problems that may arise dynamically, using various previously defined or learned content, rather than programs that follow a predetermined path (Gürsakal, 2017; Shinde et al., 2018; Alpaydin, 2021; Sharma et al., 2021). Machine learning, which gained popularity in the 1980s with the advent of data mining, can be categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning is a core technique in machine learning, where models are trained with labeled data to understand the relationship between inputs and their expected outputs. Through this process, the model identifies patterns and learns to map input features to correct results. This enables the model to make reliable predictions on new, unseen data, improving its capacity to generalize to real-world situations. Supervised learning is essential for applications in classification and regression, where the goal is to predict specific outcomes, making it highly valuable for tasks like image recognition, financial forecasting, and medical diagnosis (Sharma et al., 2021).

Supervised learning follows a structured approach that ensures the development of a robust model. Initially, the model is trained with labeled data, which allows it to learn the connections between input features and their expected outputs. This training phase involves adjusting the model's internal parameters to reduce errors, enhancing its prediction accuracy. Afterward, the model undergoes testing on new data, allowing its ability to generalize to unseen situations to be evaluated. By comparing the model's predictions against actual outcomes, its performance is quantified, providing insights into its real-world applicability. The iterative nature of this process refines the model over time, strengthening its reliability and ability to make accurate predictions across various domains. This cycle of training, testing, and refinement is critical to creating models that can consistently deliver valuable insights and support decision-making (Muhammad and Yan, 2015; Sen et al., 2020; Verma et al., 2021; Tiwari, 2022).

Supervised learning consists of two main types: classification and regression. Classification categorizes input data into specific classes, using algorithms like logistic regression, SVM, naive Bayes, decision trees, and neural networks to make accurate predictions. Regression, on the other hand, predicts continuous numerical outcomes, utilizing algorithms such as linear regression, ridge regression, and lasso regression for forecasting or quantitative predictions. Both types are essential in machine learning, with classification used for tasks like image recognition and spam detection, and regression applied to price prediction and trend analysis (Sharma et al., 2020).

Supervised learning has many advantages and disadvantages. Since the data is labeled, the model makes more accurate predictions, resulting in higher performance, has a clear goal, is suitable for specific problems, and offers a wide range of algorithm options, since many different algorithms are used. The disadvantage of supervised learning is that it has difficulties such as labeling costs due to the time-consuming and costly nature of labeling large data sets, and generalization difficulties due to difficulty in generalizing to real-world situations when limited to only training data (Wuest et al., 2016; Saravanan & Sujatha, 2018). Supervised learning can be applied in almost any field where labeled data is available, such as finance and banking, healthcare, e-commerce, image and audio processing, natural language processing (NLP), and industry.

Unsupervised learning is a machine learning approach that deals with unlabeled data, aiming to discover hidden patterns or structures without predefined target labels. By analyzing input features, the model identifies relationships and groups similar data points, revealing insights based on natural similarities or differences. Unlike supervised learning, which relies on labeled data for predictions, unsupervised learning focuses solely on extracting valuable insights from the data itself, uncovering underlying patterns that may not be immediately apparent (Wuest et al., 2016; Sharma et al., 2021; Naeem et al., 2023; Valkenborg et al., 2023).

Unsupervised learning aims to uncover hidden patterns in data without relying on predefined labels. Rather than categorizing data, it presents results through techniques like clustering and dimensionality reduction. Clustering groups similar data points together, utilizing algorithms like K-Means and DBSCAN, while dimensionality reduction simplifies complex data to make it more manageable for analysis. Methods such as PCA, T-SNE, and Autoencoders are commonly used for reducing data complexity, allowing important features to be extracted while retaining key information in a more compact form (Sharma et al., 2020).

Unsupervised learning offers several advantages, including time and cost savings, the ability to reveal previously unnoticed data structures, and its applicability in situations where labeled data is scarce. However, it also has some drawbacks. Compared to supervised learning, interpreting outputs and utilizing results can be more difficult. Additionally, the accuracy of the results may be lower, as the patterns in the data are not always clearly defined. The effectiveness of unsupervised algorithms is closely tied to the choice of hyperparameters, meaning that selecting the right values is crucial but difficult. The sensitivity of these algorithms to hyperparameters implies that small changes can lead to significantly different outcomes, posing a challenge in optimizing them for the best performance. This requires careful experimentation and tuning to identify the most suitable configurations for a given dataset (Wuest et al., 2016; Wang & Biljecki, 2022). Unsupervised learning is used in data discovery, analysis and problem solving processes in many areas such as anomaly detection, genetic data analysis, visualization and exploratory analysis, and customer segmentation.

Reinforcement learning is a branch of machine learning where models learn by receiving feedback through rewards or penalties based on their actions. This approach helps the model improve its decision-making process over time, as it seeks to maximize rewards while minimizing penalties. In this approach, an artificial intelligence system, known as an agent, learns the best strategy to achieve a goal by continually interacting with its environment. Therefore, reinforcement learning is used as a strategy that determines which action an agent should choose by acting and making decisions in an environment by receiving feedback from many actions (Uprety and Rawat, 2020; Dong et al., 2020; Sharma et al., 2021).

Reinforcement learning involves an agent learning to make decisions by selecting actions based on its current situation and engaging with its environment. As the agent takes actions, it transitions to new states and receives feedback, either positive or negative. This feedback allows the agent to adapt and improve its decision-making process, gradually discovering the best strategy to accomplish its objectives (Matsuo et al., 2022; Shakya et al., 2023).

Reinforcement learning uses algorithms like Q-Learning, Policy Gradient Methods, and Deep Reinforcement Learning to improve how an agent makes decisions and learns optimal strategies by interacting with its environment. This method has several benefits, including the agent's ability to learn from its experiences, solve complex problems without needing a predefined mathematical model, and adjust to changing conditions in real-time. However, reinforcement learning has several disadvantages, including the high computational power required for complex environments and continuous action spaces, the challenge of balancing exploration of new situations with exploiting existing knowledge to maximize rewards, and the typically lengthy learning process (Wuest et al., 2016; Nian et al., 2020; Sivamayil et al., 2023). Reinforcement learning is frequently preferred in the fields of finance, games, autonomous vehicles and robotics, where it is effective in complex decision-making problems and provides better performance by learning from environmental feedback.

2.1.2. Deep Learning

Deep learning has become increasingly prevalent since 2010 and has gained significant popularity for analyzing large datasets (Sharma et al., 2021). Deep learning is a system that performs the calculations used in machine learning in one go, discovers the parameters that need to be defined, and can make evaluations with better parameters. Deep learning, a subset of machine learning, is capable of handling larger datasets and delivering more precise outcomes compared to conventional machine learning methods. It requires less human involvement in data labeling. Systems that autonomously perform feature extraction without human intervention or support are commonly referred to as "Deep Learning" (LeCun et al., 2015; Janiesch et al., 2021; Mathew et al., 2021; Hu et al., 2021; Roberts & Yaida, 2022).

Deep learning relies on methods like Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Generative Adversarial Networks (GAN), and Transfer Learning. Among these, ANNs, inspired by the human brain, are fundamental in establishing complex relationships between input and output data. Composed of multiple layers of neurons, ANNs process and pass data through the network, allowing them to solve problems like image recognition and natural language processing. As the layers deepen, ANNs learn increasingly abstract representations, enhancing their ability to handle diverse tasks (Choi et al., 2020; Yang & Wang, 2020; Abdolrasol et al., 2021; Tang & Yang, 2021). Artificial neural networks are made up of three primary components: the input layer, one or more hidden layers, and the output layer.

In an artificial neural network, data is processed through multiple layers to generate predictions. The input layer receives data, with each neuron representing a distinct feature. This data then passes through the hidden layers, where weights and biases are applied, and activation functions introduce non-linearity. As the network deepens with additional hidden layers, it becomes capable of capturing more complex patterns in the data. The output layer generates predictions, which may be class labels in classification tasks or continuous values in regression. Each neuron processes the data and forwards it through the network. Once predictions are made, they are compared to the actual values, and the error is calculated. This error is minimized through backpropagation, a process that adjusts the weights and biases to improve the network's accuracy over time. This iterative learning process enhances the model's ability to make reliable predictions and tackle complex tasks (Uzair & Jamil, 2020; Sharkawy, 2020; Dastres & Soori, 2021; Montesinos López et al., 2022).

Optimization algorithms like SGD, Adagrad, RMSProp, Adam, and AdaMax are used to improve network performance by adjusting weights to minimize error and enhance accuracy. Each algorithm updates the model's parameters differently, enabling more efficient learning and faster convergence (Zaheer & Shaziya, 2019; Shrestha & Mahmood; 2019 Soydaner, 2020).

Activation functions are crucial in artificial neural networks because they define the output of each neuron, enabling the network to learn and model complex patterns. These functions control the flow of information through the network, helping it detect and understand intricate relationships in the data. This allows the network to address more complex tasks. Common activation functions, including Sigmoid, Tanh, ReLU, Softmax, and Swish, each have distinct properties that impact how the network learns and processes information. For example, ReLU helps prevent vanishing gradients and speeds up learning, while Softmax is useful in classification tasks where probability distributions are required. Thus, selecting the appropriate activation function is crucial for optimizing the network's ability to process information and improve performance across different applications (Rasamoelina et al., 2020; Lederer, 2021; Dubey et al., 2022). Artificial neural networks excel at learning and representing non-linear relationships, allowing them to effectively handle a wide variety of data types, such as images, audio, and text. This capability enables them to perform exceptionally well on complex and large datasets, making them highly effective in a range of challenging tasks.

In addition, large data sets are needed to achieve successful results in artificial neural networks. This causes artificial neural networks to be more computationally intensive and the training time to be longer (Wu & Feng, 2018; Mijwel, 2021; Abdolrasol et al., 2021; Dragović, 2022). As technology continues to progress, artificial neural networks are becoming more widely adopted across various domains, including image processing, speech recognition, and natural language processing, showcasing their growing importance and versatility.

Convolutional Neural Networks (CNNs) are a type of deep learning model tailored for analyzing visual and spatial data. Unlike traditional neural networks, CNNs excel at extracting features from images by using layers such as convolutional, activation, pooling, and fully connected layers, which improve their efficiency in processing visual information (Wu, 2017; Gu et al., 2018; Li et al., 2021).

In the convolutional layer, key features are identified from the input, forming a feature map that highlights relevant patterns through an activation function. The pooling layer simplifies this map by reducing its size, retaining the crucial features and reducing the risk of overfitting. In the fully connected layer, the feature map is flattened, preparing it for the final classification task (Pinaya et al., 2020; Ketkar & Moolayil, 2021; Krichen, 2023).

Convolutional neural networks provide significant advantages in deep learning due to the processing of visual data, automatic extraction of features and prevention of over-learning. However, convolutional neural networks also have some disadvantages, including a high requirement for large datasets, limited interpretability, and significant computational power demands (Kamilaris & Prenafeta-Boldú, 2018; Sarıgül et al., 2019; Khan et al., 2020; Montesinos López et al., 2022). Convolutional neural networks (CNNs) are extensively applied across different domains like object detection, motion analysis, classification, and visual recognition. To tackle these tasks effectively, various CNN architectures, such as AlexNet, GoogLeNet, VGGNet, DenseNet, ResNet, Xception, MobileNet, and EfficientNet, are utilized, each offering unique advantages for specific challenges.

Recurrent Neural Networks (RNNs) are specialized artificial neural networks designed to process sequential data by remembering past information. Unlike traditional networks that treat each input separately, RNNs take into account previous inputs, enabling them to make predictions based on both current and historical data. RNNs are particularly effective for tasks like time series forecasting, speech recognition, and language modeling, as they excel at capturing time-dependent patterns and utilizing past information to make more accurate predictions (Medsker & Jain, 2001; Grossberg, 2013; Salehinejad et al., 2017; Caterini & Chang, 2018; Zhu et al., 2022; Mienye, et al, 2024). Recurrent Neural Networks (RNNs) are frequently applied in fields that require the analysis of sequential data, such as video processing, audio recognition, time series analysis, and natural language processing. Their ability to remember and use past information makes them ideal for tasks where the order of data points is crucial, enabling more accurate predictions and understanding of dynamic patterns over time. They provide several benefits, including the ability to predict future steps by analyzing previous ones, capturing the evolution of data over time, using fewer parameters effectively, and being applicable to a wide range of tasks. However, they have disadvantages such as longer training time compared to other networks, over-learning when working on complex data, an intensive preliminary examination process for the suitability of the input data and high computational cost (Bianchi et al., 2017; Khalilov et al., 2021; Tsantekidis et al., 2022; Ahmed et al., 2023).

Generative Adversarial Networks (GANs), developed by Ian Goodfellow in 2014, marked a significant breakthrough in deep learning. GANs consist of two networks: the Generator, which creates synthetic data, and the Discriminator, which differentiates between real and fake data. These networks engage in a competitive process where the Generator strives to generate more realistic data, while the Discriminator improves its ability to distinguish authentic from fabricated data. Through this iterative learning process, both networks improve over time, enabling GANs to produce high-quality data that closely resembles real-world examples (Wang et al., 2017; Creswell et al., 2018; Gonog & Zhou, 2019; Goodfellow et al., 2020; Aggarwal et al., 2021; Gui, et al., 2021).

GANs are increasingly utilized in various fields, including image creation, modification, and data enhancement. They provide notable benefits in generating realistic synthetic data, fostering creativity, and producing diverse and varied datasets, making them a powerful tool for innovation and expansion in data-driven tasks. However, GANs also present challenges, such as the complexity of the training process and reduced success rates when unbalanced training data leads to the generator producing examples of limited variety (Cao et al., 2018; Gonog and Zhou, 2019; Alqahtani et al., 2021; Atone & Bhalchandra, 2023).

Transfer learning is a technique in deep learning that accelerates the learning process for new tasks by applying knowledge acquired from a previously learned task. By utilizing the features and insights gained from a model trained on large, labeled datasets, transfer learning allows the model to effectively tackle new, often similar challenges with less data and effort. (Pan & Yang, 2009; Marcelino, 2018; Li et al., 2020; Hosna et al., 2022; Gupta et al., 2022). Therefore, transfer learning allows the new task process to be accelerated by using some or all of the weights used in the previously trained deep learning model (Weiss et al., 2016; Kaya, 2023; Wang & Chen, 2023).

Transfer learning involves two key stages: pre-training and adaptation. During pre-training, large neural networks are trained on a labeled dataset to learn general features. In the adaptation stage, the pre-trained model is applied to a new task, with some layers frozen and others retrained to adjust to the new data. This process can be carried out using two approaches: feature extraction and fine-tuning. Feature extraction reuses the earlier layers of the model, which capture general features, while only updating the final layers for the new task. Fine-tuning, on the other hand, involves retraining or modifying specific layers of the model to better fit the new task, allowing for a more in-depth adaptation while preserving knowledge from the original task (Ribani & Marengoni, 2019; Vrbančič and Podgorelec, 2020; Ayana et al., 2021; Pinto et al., 2022; Jain et al., 2023; Zhu et al., 2023).

Transfer learning offers a more efficient training process by reducing computational costs and improving performance. However, its effectiveness can be limited if the pre-trained model's weights are either perfectly suited for the new task or not well aligned, potentially leading to suboptimal results (Sarkar, 2018; Castillo, 2021; Kim et al., 2022). Transfer learning has proven highly effective in domains like speech recognition, image recognition, and natural language processing, where it has led to significant success by leveraging pre-existing models for improved performance on related tasks.

3. Application Areas of Artificial Intelligence Methods in Aquaculture

Artificial intelligence technologies are used in many areas in the aquaculture industry.

3.1. Water Quality Parameters Monitoring

In aquaculture, ensuring ideal environmental conditions is crucial for the well-being and growth of aquatic species. Continuously monitoring key water quality parameters like temperature, oxygen levels, pH, salinity, and turbidity helps maintain an environment conducive to species survival and development, fostering their health and optimizing growth potential (Mustapha et.al., 2021). It is very difficult to keep these parameters under constant control, even on a daily basis, with human power. Data sets obtained from the measurement results of these parameters at regular intervals with artificial intelligence applications provide an important opportunity for aquaculture producers (Vo et al., 2021).

3.2. Fish Health and Management

The sustainability of aquaculture depends on fish health. Disease outbreaks can pose a significant risk to fish populations and lead to substantial economic losses, particularly if pathogens spread beyond the farm (Cascarano et.al.,2021). Artificial intelligence techniques are a technique that can help detect fish diseases and warn farmers in a timely manner. Early diagnosis of diseases in fish is important in terms of rapid intervention and preventing the spread of infectious diseases. One of the most important issues that must be kept under control in these facilities is that viruses and pathogens that cause fish diseases can be prevented in the face of any negative situation by applying computer-aided and artificial intelligence methods (Mandal & Ghosh, 2024).

3.3. Fish Mobility Measurement

In many fisheries, length and weight measurements are essential for assessing fish growth, mortality, reproduction, and population dynamics. However, obtaining these measurements can be time-consuming and may cause stress or injury to the fish, as they often need to be removed from the water. By combining artificial intelligence with image processing, it is possible to estimate the body length, weight, and mobility status of aquaculture species in real time without physical handling. To achieve this, fish detection is carried out using applications like Faster R-CNN and YOLO-v3. The length is calculated using the ordinary stereo matching algorithm, while the weight is determined with the linear regression method (Rahim et.al., 2010; Mustafa et.al., 2013; Chang et.al., 2021).

3.4. Measuring Fish Size

Estimating fish length and understanding catch composition are essential aspects of fisheries research, providing valuable insights into fish populations and ecosystem health. Length grading is a common practice during the rearing period, playing a vital role in aquaculture management. Automating the measurement and monitoring of aquatic species using machine learning can significantly reduce operational and management costs while enhancing profitability and production quality. Accurate measurement of fish for market purposes is also crucial for ensuring product consistency and meeting industry standards. Using artificial intelligence techniques, appropriate processing of stereo images obtained from fishing vessels and critical information on average fish size and catch composition is provided (Garcia et.al., 2020).

3.5. Intelligent Fishing

The advancement of smart technology in aquaculture has led to more efficient and sustainable farming practices globally. This transformation has created a more intelligent and resource-efficient system, improving the overall productivity and environmental sustainability of aquaculture operations (Zhao et.al., 2021). The integration of machine learning with advanced computing capabilities enables the extraction of complex, detailed information from data, offering groundbreaking solutions for modernizing aquaculture. This synergy is ushering in a new era for the fishing industry, enhancing its efficiency and intelligence (Samuel, 2000; Liakos et al., 2018). AI can predict optimal fishing locations, aiding in resource optimization and preventing overfishing. Furthermore, AI-driven smart sea cages can monitor fish behavior and regulate environmental conditions. Additionally, robots in the aquaculture industry can be programmed with AI to carry out specific tasks (Mustapha et.al., 2021; Wang et. al., 2021).

3.6. Feeding Control

AI can significantly optimize fish feed formulation in aquaculture by analyzing factors such as species-specific needs, nutritional demands, ingredient availability, and cost constraints.

This enables the development of customized, cost-efficient feeding strategies, contributing to more sustainable and productive aquaculture practices (Glencross at al., 2023). AI, through computer vision algorithms, can track fish movements and behaviors in real-time by analyzing video footage or images from tanks or ponds. This capability allows for precise monitoring of feed consumption, enabling the optimization of feeding strategies and improving resource management in aquaculture systems (Zhang et al., 2023). Feeding strategies in aquaculture can be improved by analyzing factors like fish behavior, growth rates, and environmental conditions. By integrating reinforcement learning algorithms, aquaculture systems can create adaptive feeding policies that dynamically adjust in real time according to the environment's changing conditions.

This dynamic approach not only optimizes feeding efficiency but also supports healthier fish growth by responding to changes in behavior and environmental factors. Furthermore, reinforcement learning allows for the fine-tuning of feeding strategies over time, ensuring that the system learns from past interactions to make more informed decisions, ultimately improving both sustainability and productivity in aquaculture practices (Xia et al., 2021). AI can help fish farmers optimize feeding strategies by analyzing the impact of environmental factors on fish feeding behavior, allowing for real-time adjustments that improve feeding efficiency and promote better fish health (Chiu et al., 2022). AI can be integrated into automated feeding systems to optimize feed distribution, adjusting quantities and schedules dynamically based on real-time data. By using machine learning, these systems can adapt to changes in factors like fish behavior, growth trends, and environmental conditions, allowing for more accurate and flexible feeding strategies. This not only enhances feed efficiency but also promotes fish health and growth by ensuring optimal nutrition at the right time. As the system learns from previous data, it improves its ability to predict future needs, contributing to the long-term sustainability and productivity of aquaculture operations (Hu et al., 2022).

Automated feeding systems are transforming aquaculture by improving feeding precision and reducing the need for manual intervention. These systems, already adopted in countries like Norway, Japan, and the United States, allow for accurate control over feed management. For instance, Norway's net cage system integrates real-time feeding adjustments with monitoring of water quality parameters, ensuring optimal conditions for fish growth. Similarly, Finland's robot feeding control system offers remote management, enhancing water quality and feeding accuracy through a web interface. This advancement in technology is helping to streamline aquaculture operations (Arvotec, 2021).

3.7. Inventory Management

AI can optimize resource management in aquaculture by predicting the ideal stocking densities, harvest times, and production schedules. By analyzing factors like growth rates, feeding behaviors, and market trends, AI helps to make more informed decisions, improving efficiency and profitability in aquaculture operations (Gladju et.al., 2022).

These are just a few of the potential applications of AI in the aquaculture industry. AI can be an important tool to make this industry more sustainable and use resources more effectively. One of the areas where the use of AI in aquaculture is needed the most is closed-circuit aquaculture facilities. In such facilities, where it is imperative to keep all dynamics under control, one of the most striking topics is the fight against diseases caused by viruses and pathogens. Computer-aided and artificial intelligence-monitored sensors, sensors and surveillance tools, as well as tools that have the ability to react to any negative situation detected by these tools, will be the right keys to preventing epidemics that may increase the risk of loss related to production.

4. Application of Artificial Intelligence Methods in Aquaculture

4.1. Classification and Identification

Traditional methods, such as classification and identification by sight, often struggle to provide accurate results, while genetic methods can be costly. As a result, machine learning and computer vision have become effective solutions for classifying and identifying fish quickly and with high accuracy. A large dataset of fish species that you can use to train a custom AI model may not be available on many platforms used for various AI applications. However, if you plan to build an AI model to classify fish species, you can use resources such as FishBase, Kaggle and other data science platforms, datasets from wildlife and ecosystem organizations, academic research and scientific articles, visual and audio data, and train a custom model by combining this data into a dataset.

By combining the data from the above sources, you can create a custom dataset that you can use to classify fish species. By separating this dataset into training and testing data, you can develop your AI model and optimize your model to increase accuracy.

4.1.1. Image and Preprocessing

To classify fish species using image processing and AI techniques, a dataset is created with labeled images of different fish species. Each image is tagged with the corresponding species, and the images are preprocessed by normalizing dimensions and adjusting color balance for better data suitability. The dataset is then split into training and testing subsets. For image recognition, a pre-trained Convolutional Neural Network (CNN) or a custom-built CNN can be employed. After training the selected CNN model on the dataset, it learns to classify fish species based on the images. Throughout the training process, the model's performance is evaluated using accuracy metrics. Once training is completed, the model is tested with new data and fine-tuned to enhance its classification accuracy.

Optimization techniques can include hyperparameter adjustments, data augmentation, learning rate adjustments, and testing networks with different depths or structures. Using the optimized model, it can classify new and unknown fish images. The model can identify fish species and make predictions.

The model's performance is assessed by examining its accuracy and identifying areas for improvement, particularly by analyzing misclassifications. Feedback from this analysis is used to refine the model and boost its overall effectiveness.

4.2. Performance Evaluation Metrics

To assess AI models for fish species classification, key performance metrics like Accuracy, Precision, Recall, and the F1 Score are used, each providing valuable insights into the model's performance. Accuracy gives a general overview of correctness, while Precision and Recall specifically measure how well the model identifies true positive instances. Precision focuses on the proportion of correctly predicted positive instances, while Recall examines the number of actual positives that were correctly identified. The F1 Score, which combines both Precision and Recall, is especially useful when minimizing false positives and negatives is crucial, offering a more balanced evaluation. These metrics are essential not only for assessing the model's performance but also for ensuring its reliability in making accurate predictions, such as in fish species classification (Zhu et al., 2010).

4.3. Hardware and Software Used

The hardware and software used for AI applications in the aquaculture industry include specialized tools such as processing power, data analysis, and learning algorithms. Many AI hardware, software, and libraries are widely used in the aquaculture industry. Central Processing Units (CPUs), Graphics Processing Units (GPUs), Application-Specific Integrated Circuits (ASICs) and Edge Computing Devices are used as hardware. Central processing units are generally used to perform data processing and management tasks, while Graphics Processing Units (GPUs) are used as accelerators, especially in complex artificial intelligence applications such as deep learning and processing of large data sets. To perform faster and more energy-efficient calculations, Application-Specific Integrated Circuits (ASICs) and Edge Computing hardware devices that can run artificial intelligence models locally, perform data analysis, and perform real-time analysis under field conditions are used.

Commonly used software libraries in AI and machine learning include TensorFlow, PyTorch, Keras, Scikit-learn, Caffe, and H2O.ai. TensorFlow is an open-source library that supports a wide variety of models and works on both CPUs and GPUs. PyTorch, known for its dynamic computational graphs, is favored by researchers for its flexibility. Keras provides a high-level interface for building deep learning models, often utilizing lower-level frameworks like TensorFlow or Theano. Scikit-learn is a popular library for machine learning, offering efficient tools for implementing key algorithms. Caffe specializes in deep learning model design and training, particularly for tasks like image and video analysis. Lastly, H2O.ai is a machine learning platform designed to accelerate the development of AI models, enabling data scientists to build and deploy models more efficiently (Nguyen et al., 2019).

Thanks to the hardware, software and libraries used in artificial intelligence applications in aquaculture, faster and easier operations can be performed on data. Therefore, it can be applied in fish farming, water quality monitoring, species classification, and many other areas.

5. Challenges of Artificial Intelligence Applications in Aquaculture

AI applications in aquaculture face several challenges, including data collection and processing, model training and compliance, complex environmental variables, model reliability, and implementation costs. AI models in aquaculture face challenges in data collection and processing, as they often require large, diverse datasets for training. For example, estimating the size of juvenile fish involves capturing a wide range of species, body shapes, and postures at various growth stages, which increases the complexity of data collection. In artisanal fisheries, these datasets are often lacking or incomplete, making it difficult to train accurate models. Therefore, the development of comprehensive aquatic animal datasets and the ability to gather data more cost-effectively are key obstacles to overcome (Yang et. al., 2021).

Training AI models in aquaculture is challenging due to the difficulty in generating adequate and representative datasets. These datasets must also reflect the specific conditions of the aquaculture environment and local factors to ensure effective model performance. Additionally, maintaining optimal water quality parameters is a significant challenge, especially in large-scale aquaculture operations or areas with limited access to clean water, making consistent monitoring and management crucial (Jan et al., 2021).

In aquaculture, factors like water quality, temperature, oxygen levels, pH, plankton count, and weather conditions play a significant role in shaping the accuracy of AI models. The interaction between these variables can complicate model performance, affecting its reliability. Given that AI model predictions are crucial for decision-making in aquaculture, ensuring their dependability is essential. However, the complexity of AI models, often seen as "black boxes," can make it difficult to interpret how they arrive at conclusions, which may reduce trust in their outcomes. Additionally, AI applications can incur significant costs due to hardware, software, data collection, and model training. In aquaculture, it is vital that these expenses remain manageable and that the return on investment is properly assessed.

To successfully integrate AI into the aquaculture industry, it is crucial to address existing challenges. As technology progresses and new application areas emerge, overcoming these hurdles will become more achievable. Achieving successful implementation also depends on collaboration among researchers, industry players, and regulatory bodies, ensuring AI is used responsibly and effectively in aquaculture.

6. Conclusions

In the context of big data in aquaculture, the diverse range of aquatic species and the complexity of breeding environments in smart fish farms pose considerable challenges for efficient data collection. Although many studies are performed in controlled laboratory settings, applying these methods to real-world scenarios exposes the difficulties of accurately gathering data in natural environments. Traditional video image acquisition techniques, which are commonly employed to monitor fish diseases and abnormal behaviors, often struggle with low accuracy, hindering the collection of high-quality data required for big data analysis in aquaculture. This limitation hinders the full potential of technological advancements, as precise data collection is crucial for effective monitoring and decision-making. Furthermore, the integration of big data technology into the aquaculture industry often relies on applying generic intelligent methods that fail to consider the specific characteristics of the industry. These methods, while successful in other fields, do not always address the unique challenges posed by aquaculture, such as the diverse aquatic environments, the variability of species, and the dynamic nature of aquatic ecosystems.

As a result, there is a gap between the potential of big data analytics and the increasing demand for customized solutions in aquaculture. A promising way to close this gap is through the use of deep learning, which has become a significant advancement in smart fish farming. Compared to traditional machine learning techniques, deep learning excels at extracting meaningful features from agricultural images and structured data, providing enhanced accuracy and deeper insights. This ability is particularly valuable in aquaculture, where nuanced patterns and subtle changes in fish behavior, water quality, and environmental factors need to be captured. Furthermore, deep learning can be effectively incorporated into agricultural machinery, leading to the creation of smart aquaculture systems that enhance feeding, monitoring, and harvesting processes while improving overall efficiency and sustainability. As deep learning technology advances, it has the potential to bridge gaps in data collection and processing, revolutionizing aquaculture practices. By delivering real-time insights and predictive capabilities, deep learning enables fish farmers to make informed, data-driven decisions that improve fish health, optimize resource utilization, and foster more sustainable and profitable aquaculture operations.

To advance AI applications in the aquaculture industry, further innovations in core AI technologies are required, along with greater integration of deep learning, knowledge computing, and hybrid-assisted intelligence in aquaculture practices.

References

Abdolrasol, M. G., Hussain, S. S., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., ... & Milad, A. (2021). Artificial neural networks based optimization techniques: A review. *Electronics*, *10* (21), 2689.

Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, *1*(1), 100004.

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Ali, A. B. M. S., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, *56* (11), 13521-13617.

Alpaydin, E. (2021). *Machine learning*. London: MIT Press.

Alqahtani, H., Kavakli-Thorne, M., & Kumar, G. (2021). Applications of generative adversarial networks (gans): An updated review. *Archives of Computational Methods in Engineering*, *28*, 525-552.

Arslan, G. & Oguzhan Yildiz, P. (2021). Türkiye su ürünleri sektörüne genel bakış. *Menba Kastamonu Üniversitesi Su Ürünleri Fakültesi Dergisi*, *7*(1), 46-57.

Arvotec (2021) Fish feeding robot. (https://www.arvotec.fi/feeding-technology/feeding-robot. Accessed 17.12.2024)

Atone, S. A., & Bhalchandra, A. S. (2023). Generative Adversarial Networks in Computer Vision: A Review of Variants, Applications, Advantages, and Limitations. In *International Conference on Sustainable and Innovative Solutions for Current Challenges in Engineering & Technology* (pp. 589-608). Singapore: Springer Nature Singapore.

Ayana, G., Dese, K., & Choe, S. W. (2021). Transfer learning in breast cancer diagnoses via ultrasound imaging. *Cancers*, *13* (4), 738.

Başçınar, N. (2004). Dünyada su ürünleri yetiştiriciliği ve ülkemizin geleceğine bakış. *Yunus Araştırma Bülteni*, *4* (1), 6-8.

Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A., Jenssen, R. (2017). Recurrent neural network architectures. *Recurrent Neural Networks for Short-Term Load Forecasting: An Overview and Comparative Analysis*, 23-29.

Cao, Y. J., Jia, L. L., Chen, Y. X., Lin, N., Yang, C., Zhang, B., Liu, Z., Li, X. X., Dai, H. H. (2018). Recent advances of generative adversarial networks in computer vision. *IEEE Access*, *7*, 14985-15006.

Cascarano, M.C., Stavrakidis-Zachou, O., Mladineo, I., Thompson, K.D, Papandroulakis, N. & Katharios, P. (2021) Mediterranean aquaculture in a changing climate: temperature effects on pathogens and diseases of three farmed fish species. Pathogens, *10*, 1205.

Castillo, D. (2021). *Transfer Learning for Machine Learning*. (https://www.seldon.io/transfer-learning Accessed 21/10/2024).

Caterini, A. L., & Chang, D. E. (2018). *Deep neural networks in a mathematical framework* (pp. 59-79). Berlin/Heidelberg, Germany: Springer International Publishing.

Chang, C. C., Wang, J. H., Wu, J. L., Hsieh, Y. Z., Wu, T. D., Cheng, S. C., ... & Lin, C. Y. (2021). Applying artificial intelligence (AI) techniques to implement a practical smart cage aquaculture management system. *Journal of Medical and Biological Engineering*, *41*, 652-658.

Chiu, M. C., Yan, W. M., Bhat, S. A., & Huang, N. F. (2022). Development of smart aquaculture farm management system using IoT and AI-based surrogate models. *journal of Agriculture and Food Research*, *9*, 100357.

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science & Technology*, *9* (2), 1-14.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE signal processing magazine*, *35*(1), 53-65.

Dastres, R., & Soori, M. (2021). Artificial neural network systems. *International Journal of Imaging and Robotics (IJIR)*, *21*(2), 13-25.

Demir, M. (2023) Van İli'nde su ürünleri üretiminin mevcut durumu. *Çanakkale Onsekiz Mart Üniversitesi Deniz Bilimleri ve Balıkçılık Dergisi*, *6* (1), 15-23.

Dong, H., Dong, H., Ding, Z., Zhang, S., & Chang, T. (2020). *Deep Reinforcement Learning*. Singapore: Springer Singapore.

Dragović, S. (2022). Artificial neural network modeling in environmental radioactivity studies–A review. *Science of the Total Environment*, *847*, 157526.

Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, *503*, 92-108.

Garcia, R., Prados, R., Quintana, J., Tempelaar, A., Gracias, N., Rosen, S., Vågstøl, H., Løvall, K., Vagstol, H., &Lovall, K. (2020) Automatic segmentation of fish using deep learning with application to fish size measurement. *ICES Journal of Marine Science*, *77*, 1354–1366.

Gladju, J., Kamalam, B. S., & Kanagaraj, A. (2022). Applications of data mining and machine learning framework in aquaculture and fisheries: A review. *Smart Agricultural Technology*, *2*, 100061.

Glencross, B., Fracalossi, D.M, Hua, K., Izquierdo, M., Mai, K., Øverland, M. & Yakupityage, A. (2023). Leveraging the benefits of nutrition research to address global challenges in the 21st century. *Journal of the World Aquaculture Society*, *54* (2), 343-363.

Gonog, L., & Zhou, Y. (2019). A review: generative adversarial networks. In *2019 14th IEEE conference on industrial electronics and applications (ICIEA)* (pp. 505-510). IEEE.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63* (11), 139-144.

Grossberg, S. (2013). Recurrent neural networks. *Scholarpedia*, *8* (2), 1888.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Li, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, *77*, 354-377.

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2021). A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering*, *35*(4), 3313-3332.

Gupta, J., Pathak, S., & Kumar, G. (2022). Deep learning (CNN) and transfer learning: a review. In *Journal of Physics: Conference Series* (Vol. 2273, No. 1, p. 012029). IOP Publishing.

Gürsakal, N. (2017). *Makine öğrenmesi ve derin öğrenme*. Bursa: Dora Basım.

Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: a friendly introduction. *Journal of Big Data*, *9* (1), 102.

Hu, W. C., Chen, L. B., Huang, B. K., & Lin, H. M. (2022). A computer vision-based intelligent fish feeding system using deep learning techniques for aquaculture. *IEEE Sensors Journal*, *22* (7), 7185-7194.

Hu, X., Chu, L., Pei, J., Liu, W., & Bian, J. (2021). Model complexity of deep learning: A survey. *Knowledge and Information Systems*, *63*, 2585-2619.

Jain, S., Salman, H., Khaddaj, A., Wong, E., Park, S. M., & Mądry, A. (2023). A data-based perspective on transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3613-3622).

Jan, F., Min-Allah, N., & Düştegör, D. (2021). Iot based smart water quality monitoring: Recent techniques, trends and challenges for domestic applications. *Water*, *13* (13), 1729.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695.

Jiang, F., Jiang ,Y., Zhi, H., Dong, Y., Li, H., Ma, S., & et al. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, *2* (4), 230-243.

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, *156* (3), 312-322.

Kaya, V. (2023). A perspective on transfer learning in computer vision," *Advances in Engineering Sciences*, 1st ed., vol. 1, (pp. 332-359). Platanus Publishing.

Ketkar, N., & Moolayil, J. (2021). Convolutional neural networks. *Deep learning with Python: learn best practices of deep learning models with PyTorch*, 197-242.

Khalilov, D. A., Jumaboyeva, N. A. K., & Kurbonova, T. M. K. (2021). Advantages and Applications of Neural Networks. *Academic Research in Educational Sciences*, *2* (2), 1153-1159.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, 53, 5455-5516.

Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: a literature review. *BMC medical imaging*, *22* (1), 69.

Krichen, M. (2023). Convolutional neural networks: A survey. *Computers*, *12* (8), 151.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Lederer, J. (2021). Activation functions in artificial neural networks: A systematic Overview. *arXiv preprint arXiv:2101.09957*.

Li, X., Grandvalet, Y., Davoine, F., Cheng, J., Cui, Y., Zhang, H., Belongie, S., Tsai, Y. H., Yang, M. H. (2020). Transfer learning in computer vision tasks: Remember where you come from. *Image and Vision Computing*, *93*, 103853.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, *33*(12), 6999-7019.

Liakos, K., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: a review. *Sensors-Basel*, *18* (8), 2674.

Mandal, A., & Ghosh, A. R. (2024). Role of artificial intelligence (AI) in fish growth and health status monitoring: A review on sustainable aquaculture. *Aquaculture International*, *32* (3), 2791-2820.

Marcelino, P. (2018). Transfer learning from pre-trained models. *Towards data science*, *10* (330), 23.

Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep learning techniques: an overview. *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, 599-608.

Matsuo, Y., LeCun, Y., Sahani, M., Precup, D., Silver, D., Sugiyama, M., Uchibe, E., & Morimoto, J. (2022). Deep learning, reinforcement learning, and world models. *Neural Networks*, *152*, 267-275.

Medsker, L. R., & Jain, L. (2001). Recurrent neural networks. *Design and Applications*, *5* (64-67), 2.

Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, *15*(9), 517.

Mijwel, M. M. (2021). Artificial neural networks advantages and disadvantages. *Mesopotamian Journal of Big Data*, 29-31.

Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Fundamentals of artificial neural networks and deep learning. In *Multivariate statistical machine learning methods for genomic prediction* (pp. 379-425). Cham: Springer International Publishing.

Muhammad, I. & Yan, Z. (2015). Supervised machine learning approaches: a survey. *ICTACT Journal on Soft Computing*, *5*(3), 946-952.

Mustafa, M., Zaıdı, M. Z., Shafry, M. M. R., Ismaıl, M. A., & Norhaıda, A. (2013). FLUDI: Using digital images for measuring fish length. *Galaxea, Journal of Coral Reef Studies*, *15*(Supplement), 101-106.

Mustapha, U.F., Alhassan, A., Jiang, D., & Li, G. (2021) Sustainable aquaculture development: a review on the roles of cloud computing, internet of things and artificial intelligence (CIA). *Reviews in Aquaculture*, *13* (4), 2076-2091.

Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An unsupervised machine learning algorithms: Comprehensive review. *International Journal of Computing and Digital Systems*, *13* (1), 911-921.

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., García, Á. L., Heredia, I., Malík, P., Hluchý, L. (2019). Machine Learning and Deep Learning Frameworks and Libraries for Large-Scale Data Mining: A Survey. *Artificial Intelligence Review*, *52* (1), 77-124.

Nian, R., Liu, J., & Huang, B. (2020). A review on reinforcement learning: Introduction and applications in industrial process control. *Computers & Chemical Engineering*, *139*, 106886.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22* (10), 1345-1359.

Pinaya, W. H. L., Vieira, S., Garcia-Dias, R., & Mechelli, A. (2020). Convolutional neural networks. In *Machine learning* (pp. 173-191). Academic Press.

Pinto, G., Wang, Z., Roy, A., Hong, T., & Capozzoli, A. (2022). Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. *Advances in Applied Energy*, *5*, 100084.

Rahim, M., Abdullah, N., Amin, I., Zakaria, M., Man, M., & Othman, N. (2010). A new approach in measuring fish length using FiLeDI framework. *IAJIT First Online Publications*.

Rasamoelina, A. D., Adjailia, F., & Sinčák, P. (2020). A review of activation function for artificial neural network. In *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)* (pp. 281-286). IEEE.

Ribani, R., & Marengoni, M. (2019). A survey of transfer learning for convolutional neural networks. In *32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)* (pp. 47-57). IEEE.

Roberts, D. A., Yaida, S., & Hanin, B. (2022). *The principles of deep learning theory* (Vol. 46). Cambridge, MA, USA: Cambridge University Press.

Russell, S.J. & Norvig, P. (2016). Artificial intelligence: a modern approach. Pearson Education (3rd ed.). Inc: New Jersey.

Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2017). Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*.

Samuel, A.L. (2000). Some studies in machine learning using the game of checkers IBM *Journal of Research and Development*, 44, 206-226.

Saravanan, R., & Sujatha, P. (2018). A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. In 2nd *International Conference* *on Intelligent Computing and Control Systems (ICICCS 2018)*, 14-15 June 2018, Madurai, India, (pp. 945-949).

Sarıgül, M., Ozyildirim, B. M., & Avci, M. (2019). Differential convolutional neural network. *Neural Networks*, *116*, 279-287.

Sarkar, D. (2018). *A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning*. (https://towardsdatascience.com/acomprehensive-hands-on-guide-to-transfer-learning-with-real-worldapplications-in-deep-learning-212bf3b2f27a Accessed 21/10/2024).

Sen, P.C., Hajra, M., & Ghosh, M. (2020). Supervised classification algorithms in machine learning: A survey and review. In *Emerging Technology in Modelling and Graphics* (pp. 99–111). Springer: Berlin/Heidelberg, Germany.

Shakya, A. K., Pillai, G., & Chakrabarty, S. (2023). Reinforcement learning algorithms: A brief survey. *Expert Systems with Applications*, *231*, 120495.

Sharkawy, A. N. (2020). Principle of neural network and its main types. *Journal of Advances in Applied & Computational Mathematics*, *7*, 8-19.

Sharma, N., Sharma, R., & Jindal, N. (2021). Machine learning and deep learning applications-a vision. *Global Transitions Proceedings*, *2* (1), 24-28.

Sharma, R., Sharma, K., & Khanna, A. (2020). Study of supervised learning and unsupervised learning. *International Journal for Research in Applied Science and Engineering Technology*, *8* (6), 588-593.

Shinde, P. P. & Shah, S. (2018). A review of machine learning and deep learning applications. *Fourth international conference on computing communication control and automation (ICCUBEA 2018)*, 16-18 August 2018, Pune, India, (pp. 1-6).

Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE access*, *7*, 53040-53065.

Sivamayil, K., Rajasekar, E., Aljafari, B., Nikolovski, S., Vairavasundaram, S., & Vairavasundaram, I. (2023). A systematic study on reinforcement learning based applications. *Energies*, *16* (3), 1512.

Soydaner, D. (2020). A comparison of optimization algorithms for deep learning. *International Journal of Pattern Recognition and Artificial Intelligence*, *34* (13), 2052013.

Tang, S., & Yang, Y. (2021). Why neural networks apply to scientific computing?. *Theoretical and Applied Mechanics Letters*, *11*(3), 100242.

Tiwari, A. (2022). Supervised learning: From theory to applications. In *Artificial intelligence and machine learning for EDGE computing* (pp. 23-32). Academic Press.

Tsantekidis, A., Passalis, N., & Tefas, A. (2022). Recurrent neural networks. In *Deep learning for robot perception and cognition* (pp. 101-115). Academic Press.

TÜIK (2022). *Su Ürünleri İstatistikleri 2022* (02/06/2024 tarihinde https://data.tuik.gov.tr/Bulten/Index?p=SuUrunleri-2021-45745 adresinden ulaşılmıştır).

Uprety, A., & Rawat, D. B. (2020). Reinforcement learning for iot security: A comprehensive survey. *IEEE Internet of Things Journal*, *8* (11), 8693-8706.

Uzair, M., & Jamil, N. (2020). Effects of hidden layers on the efficiency of neural networks. In *2020 IEEE 23rd international multitopic conference (INMIC)* (pp. 1-6). IEEE.

Valkenborg, D., Rousseau, A. J., Geubbelmans, M., & Burzykowski, T. (2023). Unsupervised learning. *American Journal of Orthodontics and Dentofacial Orthopedics*, *163*(6), 877-882.

Verma, R., Nagar, V., & Mahapatra, S. (2021). Introduction to supervised learning. Data Analytics in Bioinformatics. *A Machine Learning Perspective*, (pp. 1-34). USA:Wiley.

Vo, T. T. E., Ko, H., Huh, J. H., & Kim, Y. (2021). Overview of smart aquaculture system: Focusing on applications of machine learning and computer vision. *Electronics*, *10* (22), 2882.

Vrbančič, G., & Podgorelec, V. (2020). Transfer learning with adaptive fine-tuning. *IEEE Access*, *8*, 196197-196211.

Wang, C., Li, Z., Wang, T., Xu, X., Zhang, X., & Li, D. (2021). Intelligent fish farm—the future of aquaculture. *Aquaculture International*, *29*, 2681-2711.

Wang, J., & Biljecki, F. (2022). Unsupervised machine learning in urban studies: A systematic review of applications. *Cities*, *129*, 103925.

Wang, J., & Chen, Y. (2023). *Introduction to transfer learning: algorithms and practice*. Springer Nature Press.

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., & Wang, F. Y. (2017). Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, *4*(4), 588-598.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, *3*, 1-40.

Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, *5* (23), 495.

Wu, Y. C., & Feng, J. W. (2018). Development and application of artificial neural network. *Wireless Personal Communications*, *102*, 1645-1656.

Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research*, 4 (1), 23-45.

Xia, K., Sacco, C., Kirkpatrick, M., Saidy, C., Nguyen, L., Kircaliali, A., & Harik, R. (2021). A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence. *Journal of Manufacturing Systems*, *58*, 210-230.

Yang, G. R., & Wang, X. J. (2020). Artificial neural networks for neuroscientists: a primer. *Neuron*, *107* (6), 1048-1070.

Yang, X., Zhang, S., Liu, J., Gao, Q., Dong, S., & Zhou, C. (2021). Deep learning for smart fish farming: applications, opportunities and challenges. *Reviews in Aquaculture*, *13*(1), 66-90.

Yıldırım, Ö. & Okumuş, İ. (2004). Muğla İli'nde su ürünleri yetiştiriciliği ve Türkiye su ürünleri yetiştiriciliğindeki yeri. *Ege Üniversitesi Su Ürünleri Dergisi*, *21*(3), 361-364.

Yılmaz, A. (2019). *Yapay zeka*. İstanbul: Koblab Yayınevi.

Zaheer, R., & Shaziya, H. (2019). A study of the optimization algorithms in deep learning. In *2019 third international conference on inventive systems and control (ICISC)* (pp. 536-539). IEEE.

Zhang, L., Li, B., Sun, X., Hong, Q., & Duan, Q. (2023). Intelligent fish feeding based on machine vision: A review. *Biosystems Engineering*, *231*, 133-164.

Zhao, S.; Zhang, S.; Liu, J.; Wang, H.; Zhu, J.; Li, D.; Zhao, R. (2021). Application of machine learning in intelligent fish aquaculture: A review. *Aquaculture*, *540*, 1–19.

Zhu, J., Jiang, Q., Shen, Y., Qian, C., Xu, F., & Zhu, Q. (2022). Application of recurrent neural network to mechanical fault diagnosis: A review. *Journal of Mechanical Science and Technology*, *36* (2), 527-542.

Zhu, W., Zeng, N., Wang, N. (2010). Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations, *NESUG Proceedings: Health Care and Life Sciences*, Baltimore, Maryland.

Zhu, Z., Lin, K., Jain, A. K., & Zhou, J. (2023). Transfer learning in deep reinforcement learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. (p. 13344-13362).

CHAPTER IV

Top 1000 Books: Regressıon Analysıs And Recommendatıon System

Sena TINKIR[1]

Introduction

Recently, in the digital age, access to information has increased rapidly and the number of books that readers can choose has increased exponentially. With these developments, online readers have been given the opportunity to share their thoughts on bookstore sites and social reading platforms. Since online reviews do not have strict rules, they can be written by anyone. This situation has both advantages and disadvantages. While they can provide rich content for in-depth analysis, they can be highly subjective and contain internet-specific vocabulary. Goodreads.com, the world's largest book review sharing site, has a large user base and a large

[1] Lecturer, Cappadocia University, School of Vocational School,Department of Computer Programming, Nevşehir, Türkiye sena.tinkir@kapadokya.edu.tr , ORCID: 0009-0001-3637-8836

number of book reviews. For this reason, it provides a useful data source for evaluating books reviewed online (Wang et al., 2019).

This study aims to perform various analyses of the books in the dataset, such as the most or least voted, the most rated, the least rated, and the most randomly recommended books, in addition to regression analysis, in line with the available data, and to provide the user with book recommendations randomly. The study uses the "Good Reads Dataset (Top 1000 Books)", which is taken from the Goodreads platform official website and includes the best 1000 books on the Kaggle platform. The study conducted using this dataset aims to examine the data, perform regression analysis, and randomly recommend books to the reader from this dataset if desired. The study is based on a regression analysis that examines the relationships between the features of the books in the dataset. In addition, the user can access the score information about the book and the information in the dataset on the book he/she encounters, such as the number of ratings, by receiving recommendations from the books in the list. If the user wishes, he/she can determine this book as the next book to read, or if it is a book he/she has read before or does not want to read, he/she can run the function again.

In the rest of this paper, the literature review is first given, then the data set used is briefly explained in the material and method section, and the methods used and the codes of the developed application are explained in detail. The experimental results section includes the results obtained from the study. In the discussion and conclusion section, a summary of the study and the results obtained and future studies are included.

Literature Review

Xiaojuan and Yutong conducted a study on exploring Goodreads reviews for literacy impact assessment. The results of the study showed that the most cited arts and humanities books received the most criticism. It was found that books with high citation rates can also receive negative reviews. It was stated that while authors tend to give praise on Goodreads, librarians are more likely to give low reviews. Approximately 70% of book reviews contain emotional words in the first sentence. Reviewers usually discuss the content of the book, not its external features. In the study, Goodreads reviews of books indexed by the book citation index were collected and evaluated in three aspects. These aspects are the popularity of books with high citation rates on Goodreads, the influence of reviewer roles, and finally the emotions and opinions behind the reviewers' reviews. The study concluded that if online reviews are to be used as an indicator of book impact assessment, the basic aspects should include the subject discipline, the role of the reviewer, and the difference in emotion (Wang et al., 2019). Kousha et al. looked at Goodreads reviews to evaluate the impact of the books more broadly. In their study, they compared book metrics available on Goodreads with different book-based impact indicators for 15,928 academic books across large subjects. The interactions were so great that at least one of the books had 85% of the impact in the arts, 80% in the humanities, and 67% in the social sciences (Kousha et al., 2017).

Verma and Potnaik in their study have come up with an automatic university library book recommendation system. They have done this to increase the efficiency and user experience of the university library. Since the number of books in the library is very large and manual selection of books is very time consuming, an automatic recommendation system was deemed necessary. In their proposed model, Discriminant Analysis with Hidden Markov support and Weighted Fuzzy Sorting (HMCAHB_DA-WFR) and Chaotic Artificial Hummingbird models have been proposed to provide better recommendations to the students. Python has been used in the software and a real-time dataset has been used from Bhilai Institute of Technology, Drug (Verma & Patnaik, 2024).

Materials and Methods

The dataset used in the study is the "Good Reads Dataset (Top 1000 Books)" dataset on the Kaggle platform. The necessary regression analyses were performed in the study with the dataset used. The dataset size is 27 kb and there are 1017 books in total. The program codes were written using a Dell brand Insprion 15 7000 Gaming model computer and the Python language with the PyCharm 2023.3.2 idesi. In addition, the random library is used to recommend books to users from the best 1017 books and to recommend the next book to read. In this way, the user can specify the recommended book as the next book to read if desired. While performing the regression analysis, a class was created at the beginning and the "_ _init_ _" function was created to display the dataset. Then, the "book" function was created for the necessary regression analyses, values and visualization of the data, and the "randomBook" function was created for random book prediction. When the program is started, we are greeted by a menu consisting of 3 options. In this menu, 1- It is the Analysis heading, under this heading, regression analysis and data visualization stages can be accessed. 2- It is the book recommendation system and it randomly suggests 1 book from our data in the data set and presents us with the author, average score, number of reviews and Goodreads score data of the book. Finally, when the 3- Exit option is selected, the program ends. The program codes are given below in detail:

import pandas as pd (1)

import numpy as np (2)

import matplotlib.pyplot as plt (3)

from sklearn.model_selection import train_test_split (4)

from sklearn.linear_model import LinearRegression (5)

from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error (6)

import random (7)

While writing the codes, the pandas library (1) for data analysis, the numpy library (2) for numerical operations, the matplotlib library (3) for drawing graphs, the sklearn.model_selection library (4) for separating the dataset into training and test classes, the sklearn.linear_model library (5) for the linear regression model, the sklearn.metrics library (6) for evaluating the performance of the model, and the random library (7) for generating random numbers were loaded at the very beginning of our program.

class Book: (8)

def __init__(self): (9)

self.bookCsv = pd.read_csv("C:/Users/kullanıcıAdı/klasörAdı/dosyaAdı/good_reads.csv", encoding="utf-8") (10)

def book(self): (11)

print(self.bookCsv.head()) #veri setinden bir kaç satır yazdırır (12)

print(self.bookCsv.isnull().sum()) #boş değer var mı kontrolü (13)

print(self.bookCsv.columns) # sütün başlıklarını yazdırmak içiin (14)

Yazar isimlerini one-hot encoding ile sayısal hale getirme

df = pd.get_dummies(self.bookCsv, columns=['Author'], drop_first=True) (15)

Veri setini görüntüleme

print(df.head()) (16)

After the library uploads were completed, we created the class named "Book" (8). Then we created our function (9) to read our file and defined our file (10) in it. "book"

By creating our second function (11), 1. We start to write the analysis steps that will take place when the one is selected into our function.

X = df[['Average Rating']] # Bağımsız değişken: 'Average Rating' (17)

y = df['Number of Ratings'] # Bağımlı değişken: 'Number of Ratings' (18)

XTrain, XTest, yTrain, yTest = train_test_split(X, y, test_size=0.2, random_state=42) (19)

model = LinearRegression() (20)

```
model.fit(XTrain, yTrain)(21)
# Katsayılar
intercept = model.intercept_(22)
coefficients = model.coef_(23)
print(f"Intercept (β0): {intercept}") (24)
print(f"Coefficients: {coefficients}") (25)
# Tahminler
yPred = model.predict(XTest) (26)
# Performans ölçümleri
mse = mean_squared_error(yTest, yPred) (27)
r2 = r2_score(yTest, yPred) (28)
print(f"Mean Squared Error (MSE): {mse}") (29)
print(f"R-squared (R²): {r2}") (30)
print(df.describe())  #tanımlayıcı istatistikler (31)
```

Dependent variable (17) and independent variable (18) selections were made and the data were divided into training and test data (19). This separation was realized as 20% test and 80% training, and with the random_state parameter, it was ensured that the same training and test sets were obtained in each run. The model was then created (20) and trained (21). The coefficients were determined, then estimates and performance measurements were made and printed on the consol screen.

Representation of the relationship between the independent variables and the dependent variables in the training set with a scatter plot

plt.scatter(XTrain['Average Rating'], yTrain, color='blue', label='Training data') (32)

plt.scatter(XTest['Average Rating'], yTest, color='pink', label='Test data') (33)

plt.xlabel('Ortalama Puan') (34)

plt.ylabel('Oylama Sayısı') (35)

plt.title('Ortalama Puan ve Oylama Sayısı Arasındaki İlişki') (36)

plt.legend() (37)

plt.show()(38)

The relationship between the independent variables and the independent variables in the training set is shown with a scatter plot. The relationship between the x heading 'Average Score' (34) and the y heading 'Number of Votes' (35) (35) was shown in blue (32) and the test data were pink (33) (38). The graph of this relationship is given in the experimental results section.

Histogram of the number of votes

plt.hist(df['Number of Ratings'], bins=30, color='green', alpha=0.7) (39)

plt.xlabel('Oylama Sayısı') (40)

plt.ylabel('Frekans') (41)

plt.title('Oylama Sayısının Dağılımı') (42)

plt.show() (43)

The histogram of the number of votes in our training set is created in green (39). The distribution of the number of votes was shown with the titles 'Number of votes' (40) and 'Frequency' (41) of title X (43). This histogram chart is given in the experimental results section.

```
# Histogram of Average Score
plt.hist(df['Average Rating'], bins=30, color='red', alpha=0.7) (44)
plt.xlabel('Ortalama Puan') (45)
plt.ylabel('Frekans') (46)
plt.title('Ortalama Puanın Dağılımı') (47)
plt.show() (48)
```

The average score histogram in our training set was created in green (44). The mean score distribution was shown with the 'Average Score' (45) heading of the X heading and 'Frequency' (46) of the y heading (48). This histogram chart is given in the experimental results section.

```
# Goodreads Histogram of the score
plt.hist(df['Score on Goodreads'], bins=30, color='purple', alpha=0.7) (49)
plt.xlabel('Ortalama Puan') (50)
plt.ylabel('Frekans') (51)
plt.title('Ortalama Puanın Dağılımı') (52)
plt.show() (53)
```

The Goodreads score histogram in our training set is created in purple (49). The distribution of the number of votes was shown with the titles 'Average Score' (50) and 'Frequency' (51) of the x heading (53). This histogram chart is given in the experimental results section.

The book with the highest average

highAverageBook = self.bookCsv.loc[self.bookCsv['Average Rating'].idxmax()] (54)

print(f"En yüksek ortalamaya sahip kitap: {highAverageBook['Book Name']}") (55)

print(f"Yazarı: {highAverageBook['Author']}") (56)

print(f"Ortalama Puan: {highAverageBook['Average Rating']}") (57)

print(f"Değerlendirme Sayısı: {highAverageBook['Number of Ratings']}") (58)

print(f"Goodreads Puanı: {highAverageBook['Score on Goodreads']}") (59)

highRAbook = self.bookCsv[self.bookCsv['Author'] == highAverageBook['Author']] (60)

Sort data by highest average score

sortedBooksAR = self.bookCsv.sort_values(by='Average Rating', ascending=False) (61)

En yüksek ortalamalı ilk 5 kitap ve yazarları

print("\nEn yüksek ortalamaya sahip ilk 5 kitap ve yazarları:") (62)

print(sortedBooksAR[['Book Name', 'Author', 'Average Rating']].head()) (63)

In order to find the book with the highest average in our data set, we take the milk named 'Average Rating' in our file where our average scores are given and assign the information of the book with the highest average to the variable named 'highAverageBook' with idmax() (54). Then we print all the information of our book, which has the highest average, respectively. We then rank the authors with the highest average from highest to lowest score (61) and print the 5 books with the highest average with author and average rating information (63).

The book with the lowest average

lowAverageBook = self.bookCsv.loc[self.bookCsv['Average Rating'].idxmin()] (64)

print(f"En düşük ortalamaya sahip kitap: {lowAverageBook['Book Name']}") (65)

print(f"Yazarı: {lowAverageBook['Author']}") (66)

print(f"Ortalama Puan: {lowAverageBook['Average Rating']}") (67)

print(f"Değerlendirme Sayısı: {lowAverageBook['Number of Ratings']}") (68)

print(f"Goodreads Puanı: {lowAverageBook['Score on Goodreads']}") (69)

Print all the information of the author with the lowest average

lowRAbook = self.bookCsv[self.bookCsv['Author'] == lowAverageBook['Author']] (70)

Sort data by lowest average score

sortedBooksAR = self.bookCsv.sort_values(by='Average Rating', ascending=True) (71)

Top 5 books with the lowest average and their authors

print("\nEn düşük ortalamaya sahip ilk 5 kitap ve yazarları:") (72)

print(sortedBooksAR[['Book Name', 'Author', 'Average Rating']].head()) (73)

In order to find the book with the highest average in our data set, we take the milk named 'Average Rating' in our file where our average scores are given and assign the information of the book with the highest average to the variable named 'highAverageBook' with idmax() (54). Then we print all the information of our book, which has the highest average, respectively. We then rank the authors with the highest average from highest to lowest score (61) and print the 5 books with the highest average with author and average rating information (63). The same steps are repeated for the book with the lowest average score. While performing operations for the book with the lowest average, we print the information of our book with the lowest average with idmin() to our variable (64). By ranking from the lowest to the highest, we print our top 5 books with the lowest average with the author and average score information. Screenshots and explanations of the outputs of repeated codes, which have changed their naming for the number of evaluations and goodreads scores, are given separately for each in the experimental results section.

authorBookCounts = self.bookCsv['Author'].value_counts() (74)

print("\nHer yazarın kitap sayısı:") (75)

print(authorBookCounts) (76)

In order to find out which of our authors has how many books in line with the information in our data set, we count the number of times each author is mentioned with the value_counts() function (74). By printing our variable (76), in which we threw this information, we can also see how many books by which author are in our data set. Some of the information on the number of books belonging to the authors is given in the experimental results section.

def randomBook(self): (77)

randomIndex = random.randint(0, len(self.bookCsv) - 1) (78)

randombook = self.bookCsv.iloc[randomIndex] (79)

print(f"Rastgele Kitap Önerisi: {randombook['Book Name']}") (80)

print(f"Yazarı: {randombook['Author']}") (81)

print(f"Ortalama Puan: {randombook['Average Rating']}") (82)

print(f"Değerlendirme Sayısı: {randombook['Number of Ratings']}") (83)

print(f"Goodreads Puanı: {randombook['Score on Goodreads']}") (84)

A function called "randomBook" has been created to present a random book recommendation among the books in our data set (77).0 and self.bookCsv (78) then we print all the information of our randomly determined book by indexing (79) in the pandas data frame with the iloc method. The visual of the stage in which random book suggestions were made is shown in the experimental results section.

```
if __name__ == '__main__': (85)
        lib = Book() (86)
        while True: (87)
            print("\n--- MENU ---") (88)
            print("1. Analiz ") (89)
            print("2. Kİtap öneri sistemi ") (90)
            print("3. Çıkış") (91)
            choice = input("Lütfen bir seçeneği giriniz(1-3): ") (92)
            if choice == "1": (93)
                lib.book() (94)
            elif choice == "2": (95)
                lib.randomBook() (96)
            elif choice == "3": (97)
                print("Programdan çıkış yapıldı.") (98)
                break (99)
            else: (100)
                print("Lütfen geçerli bir seçenek seçiniz.") (101)
```

When the Python file is run, we wrote our condition for the following codes to be executed (85), then we create a lib object from the Book class (86) and loop it infinitely with While True (87). By printing the menu, we write the codes of which function will work when which option is selected with condition blocks. This field consists of the codes that appear when our program runs and perform the analysis in it by running the book function if 1 is selected according to the selected option, running the randomBook function if 2 is selected, making a random book recommendation with the book recommendation system, and terminating the program with the break command if 3 is selected. The menu view and the outputs of the operations performed by the functions in it are included in the experimental results section.

Experimental Results

The menu that appears when the program is run appears in figure 1. At this stage, the operation we want to do can be performed by selecting the operation we want to do.

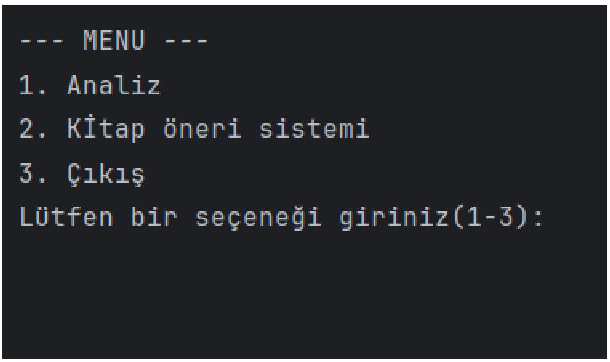


Figure 1 Menu view

In Figure 2, there is a part of the outputs we obtain when the 1st of the options in the menu, that is, the analysis step, is selected from the codes given in the Materials and Methods section. First, a few lines from our data set were printed, then the null value check was done respectively and it was seen that there was no null value in our data, the titles of the column in our data set were printed, and finally, our data was digitized with one-hot encoding.

```
Lütfen bir seçeneği giriniz(1-3): 1
                                           Book Name  ... Score on Goodreads
0                              To Kill a Mockingbird  ...            17358.0
1                                               1984  ...            15474.0
2                                Pride and Prejudice  ...            15135.0
3  Harry Potter and the Sorcerer's Stone (Harry P...  ...            12440.0
4                                   The Great Gatsby  ...            10828.0

[5 rows x 5 columns]
---------------------------------------------
Book Name             0
Author                0
Average Rating        0
Number of Ratings     0
Score on Goodreads    0
dtype: int64
---------------------------------------------
Index(['Book Name', 'Author', 'Average Rating', 'Number of Ratings',
       'Score on Goodreads'],
      dtype='object')
---------------------------------------------
                                           Book Name  ...  Author_Émile Zola
0                              To Kill a Mockingbird  ...              False
1                                               1984  ...              False
2                                Pride and Prejudice  ...              False
3  Harry Potter and the Sorcerer's Stone (Harry P...  ...              False
4                                   The Great Gatsby  ...              False

[5 rows x 688 columns]
```

Figure 2 Outputs of the first part of the codes

According to the data in Figure 3, the Intercept (β0) value refers to the value of the dependent variable, that is, the number of votes, when all of the independent variables, that is, the average score, are 0. The Coefficients value shows the effect of average point values on the number of votes. The Mse value, i.e. the mean square error, is the average of the squares of the differences between the predicted values and the actual values. The lower the value, the more accurate the model's predictions. The R2 value indicates how much of the variation in the dependent variable the model describes.

```
Intercept (β0): -1608207.1014717168
Coefficients: [459116.16226926]
Mean Squared Error (MSE): 580574309483.5581
R-squared (R²): 0.05419934127829651
-------
       Average Rating  Number of Ratings  Score on Goodreads
count     1017.000000       1.017000e+03         1017.000000
mean         3.867099       1.695201e+05          667.055064
std          0.254757       5.889344e+05         1639.846128
min          2.710000       1.500000e+01            1.000000
25%          3.730000       4.131000e+03           67.000000
50%          3.880000       1.584800e+04          136.000000
75%          4.040000       8.608300e+04          457.000000
max          4.600000       1.006313e+07        17358.000000
```

Figure 3 Performance evaluation outputs for regression

As a result of the analysis, the average score in the data set was determined as the independent variable and the number of votes was determined as the dependent variable, the rates were analyzed as 20% test and 80% training, and the result was shown in Figure 4 with the Scatter Plot, that is, the scattering diagram.

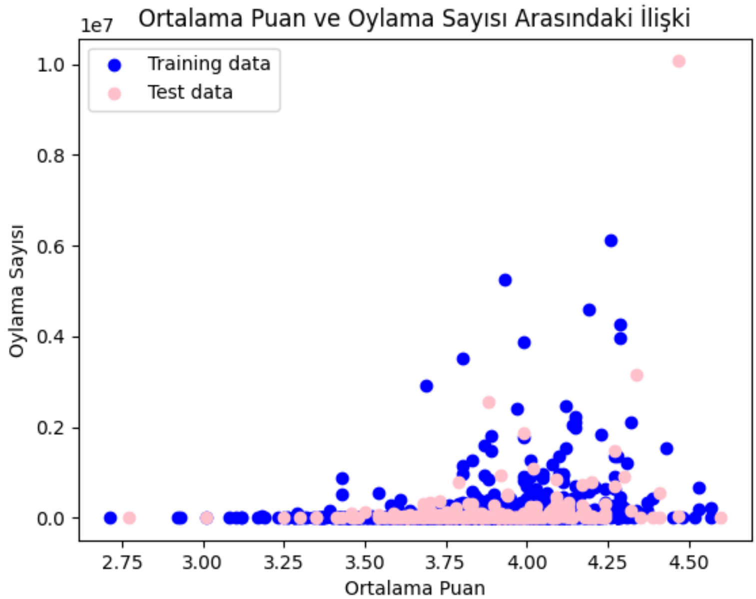


Figure 4 Scatter plot representation of the relationship between the average score and the number of votes

According to the data in the data set, histogram graphs of the number of votes, average score distribution and Goodreads score were created, respectively. These graphs are shown in Figure 5, Figure 6 and Figure 7, respectively.

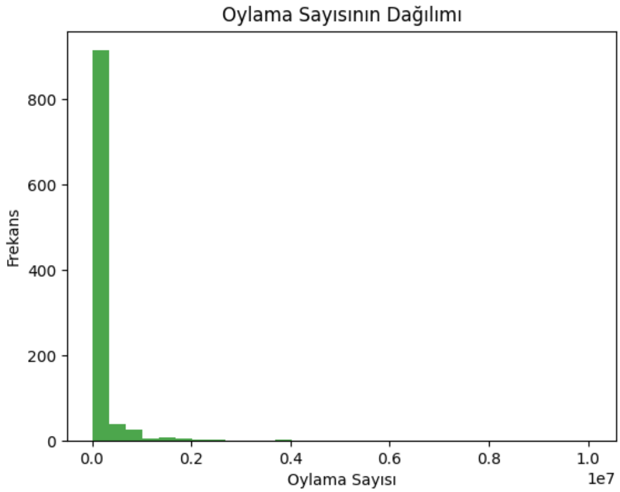


Figure 5 Histogram of the number of votes

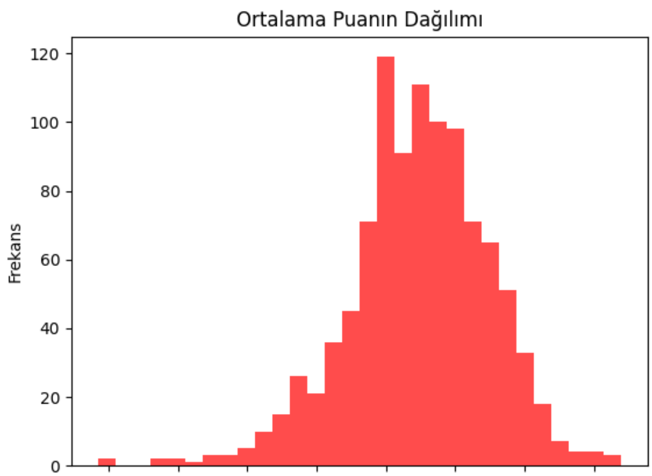


Figure 6 Histogram of the average score

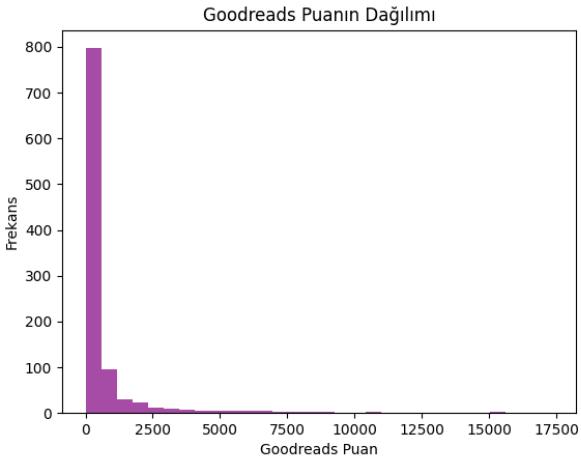


Figure 7 Goodreads Score Histogram

```
En yüksek ortalamaya sahip kitap: The Devil to Pay in the Backlands
Yazarı: João Guimarães Rosa
Ortalama Puan: 4.6
Değerlendirme Sayısı: 5382
Goodreads Puanı: 25.0

En yüksek ortalamaya sahip ilk 5 kitap ve yazarları:
                            Book Name                Author  Average Rating
944  The Devil to Pay in the Backlands  João Guimarães Rosa            4.60
858               Covenant With Death          John  Harris            4.57
154                  The Complete Maus        Art Spiegelman            4.57
245  Lonesome Dove (Lonesome Dove, #1)        Larry McMurtry            4.53
8                  The Lord of the Rings       J.R.R. Tolkien            4.53
```

Figure 8 The book with the highest average

Among the books, The Devil to Pay in the Backlands by João Guimarães Rosa has the highest average score of 4.6 in the data set. In total, the book has been reviewed by 5382 people and has a goodreads rating of 25.0. The top 5 books with the highest average are displayed in Figure 8. The book with the lowest average is Moon Over Africa by Ida Pollock. The book has 21 reviews, a rating of 2.71, and a GoodReads score of 97.0. The top 5 books with the lowest average and author information are displayed in Figure 9 along with their averages.

```
En düşük ortalamaya sahip kitap: Moon Over Africa
Yazarı: Ida Pollock
Ortalama Puan: 2.71
Değerlendirme Sayısı: 21
Goodreads Puanı: 97.0

En düşük ortalamaya sahip ilk 5 kitap ve yazarları:
                                             Book Name  ... Average Rating
600                                   Moon Over Africa  ...           2.71
453          Pamela: Virtue Rewarded (Forgotten Books)  ...           2.77
967  The History Of Pompey The Little: Or, The Life...  ...           2.92
738                               The Reluctant Orphan  ...           2.93
657                                            Caprice  ...           3.01
```

Figure 9 The book with the lowest average

```
En yüksek Değerlendirmeye sahip kitap: Harry Potter and the Sorcerer's Stone
Yazarı: J.K. Rowling
Ortalama Puan: 4.47
Değerlendirme Sayısı: 10063128
Goodreads Puanı: 12440.0

En yüksek değrlendirme sayısına sahip ilk 5 kitap ve yazarları:
                                           Book Name  ... Number of Ratings
3  Harry Potter and the Sorcerer's Stone (Harry P...  ...          10063128
0                              To Kill a Mockingbird  ...           6129090
4                                    The Great Gatsby  ...           5244056
1                                                1984  ...           4604557
2                                 Pride and Prejudice  ...           4273146
```

Figure 10 The book with the highest reviews

Among the books in our dataset, the book with the highest number of reviews is Harry Potter and the Sorcerer's Stone (Harry Potter, #1) by J.K. Rowling. The book has been reviewed for 10063128 times and has an average rating of 4.47 and a Goodreads rating of 12440.0. The top 5 books with the highest number of reviews are given in Figure 10 with the number of reviews. Figure

11 shows the book and author information for the lowest number of reviews. Here we can see the top 5 least reviewed books. La Comedie Humaine: Scenes from Political Life by Honore de Balzac has been reviewed 15 times and has an average rating of 4.13 and a Goodreads score of 92.0.

```
En düşük Değerlendirmeye sahip kitap: La Comedie Humaine: Scenes from Political
Yazarı: Honoré de Balzac
Ortalama Puan: 4.13
Değerlendirme Sayısı: 15
Goodreads Puanı: 92.0

En düşük değerlendirme sayısına sahip ilk 5 kitap ve yazarları:
                                              Book Name  ... Number of Ratings
632      La Comedie Humaine: Scenes from Political Life  ...                15
805                                     Office Politics  ...                16
833               Handley Cross: or, Mr.Jorrocks's Hunt  ...                20
600                                    Moon Over Africa  ...                21
589   A Modern Declaration of Independence: Crimes A...  ...                25
```

Figure 11 The book with the lowest reviews

```
En yüksek Goodreads Puanına sahip kitap: To Kill a Mockingbird
Yazarı: Harper Lee
Ortalama Puan: 4.26
Değerlendirme Sayısı: 6129090
Goodreads Puanı: 17358.0

En yüksek Goodreads Puanına  sahip ilk 5 kitap ve yazarları:
                                           Book Name  ... Score on Goodreads
0                              To Kill a Mockingbird  ...            17358.0
1                                               1984  ...            15474.0
2                                Pride and Prejudice  ...            15135.0
3  Harry Potter and the Sorcerer's Stone (Harry P...  ...            12440.0
4                                   The Great Gatsby  ...            10828.0
```

Figure 12 The book with the highest Goodreads rating

Harper Lee's To Kill a Mockingbird has the highest Goodreads score with a Goodreads score of 17358.0. The book has an average score of 4.26 with 6129090 number of reviews. The top

5 of the books with the highest Goodreads scores appear in Figure 12. Barbara Vine's book, A Fatal Inversion, has the lowest Goodreads score of 1.0. The book has an average score of 3.94 with 4491 reviews. The top 5 books with the lowest goodreads scores are included in Figure 13 along with their scores.

```
En düşük Goodreads Puanına sahip kitap: A Fatal Inversion
Yazarı: Barbara Vine
Ortalama Puan: 3.94
Değerlendirme Sayısı: 4491
Goodreads Puanı: 1.0

En düşük Goodreads Puanına sahip ilk 5 kitap ve yazarları:
                                         Book Name  ... Score on Goodreads
1016                             A Fatal Inversion  ...                1.0
1015                            A Dark-Adapted Eye  ...                2.0
1013  The Case of the Gilded Fly (Gervase Fen, #1)  ...                3.0
1014                      The Swimming-Pool Library  ...                3.0
1012                         The Old Men at the Zoo  ...                3.0
```

Figure 13 The book with the lowest Goodreads score

When we view the number of books belonging to the authors, it is seen that we have a total of 685 different authors in our data set. Among these authors, Charles Dickens has the most books in our data set with 9 books. It is followed by Graham Greene and Evelyn Waugh with 8 books. Figure 14 contains the number of books of the authors whose only part is given.

```
Her yazarın kitap sayısı:
Author
Charles Dickens          9
Graham Greene            8
Evelyn Waugh             8
P.G. Wodehouse           6
Jane Austen              6
                        ..
Geoff Ryman              1
Tim O'Brien              1
Vikram Seth              1
Matthew Gregory Lewis    1
Edmund Crispin           1
Name: count, Length: 685, dtype: int64
```

Figure 14 Author book count information

When the book recommendation system is run by selecting 2 while choosing from the menu, a random book in our data set is recommended to us with all its information. Figure 15 shows the information of the book The End of the World News by Anhony Burgess, which is recommended when the book recommendation system is run. Figure 16 shows George Orwell's 1984 book.

```
Lütfen bir seçeneği giriniz(1-3): 2
Rastgele Kitap Önerisi: The End of the World News
Yazarı: Anthony Burgess
Ortalama Puan: 3.81
Değerlendirme Sayısı: 766
Goodreads Puanı: 87.0
```

Figure 15 Book recommendation system

```
Rastgele Kitap Önerisi: 1984
Yazarı: George Orwell
Ortalama Puan: 4.19
Değerlendirme Sayısı: 4604557
Goodreads Puanı: 15474.0
```

Figure 16 Book recommendation system

When Exit, which is the last option of our menu, option number 3, is selected, the program is logged out as seen in Figure 17.

```
Lütfen bir seçeneği giriniz(1-3): 3
Programdan çıkış yapıldı.
```

Figure 17 3. Alternative

As can be seen in Figure 18, if a value other than the 3 options in the menu is entered, the program gives us the message please select a valid option and brings up the menu screen again.

```
--- MENU ---
1. Analiz
2. Kİtap öneri sistemi
3. Çıkış
Lütfen bir seçeneği giriniz(1-3): 5
Lütfen geçerli bir seçenek seçiniz.

--- MENU ---
1. Analiz
2. Kİtap öneri sistemi
3. Çıkış
Lütfen bir seçeneği giriniz(1-3):
```

Figure 18 Selecting different options in the menu

Discussion and Conclusion

In the study, regression analysis was performed using the Good Reads Dataset (Top 1000 Books) data set, the books with the highest and lowest values were found, and a system was developed that recommends a random book with a random library from the books in the data set. The book with the highest average in the dataset is The Devil to Pay in the Backlands by João Guimarães Rosa. The highest-reviewed book is Harry Potter and the Sorcerer's Stone (Harry Potter, #1) by J.K. Rpwling. The book with the highest Goodreads score is To Kill a Mockingbird by Harper Lee. The author with the most books in the dataset is Charles Dickens with 9 books.

In the study, the analyzes made for the books can also be carried out for many classified data sets such as movies and places to visit, and the suggestions made to the users can be adapted to many areas that can be diversified as movies, food, places to visit instead of books, and can be turned into a system that can make suggestions over different data sets. If it wants to be customized on the books, the recommendation system can be used on the books in our own library, on the children's books or on the books in a school library.

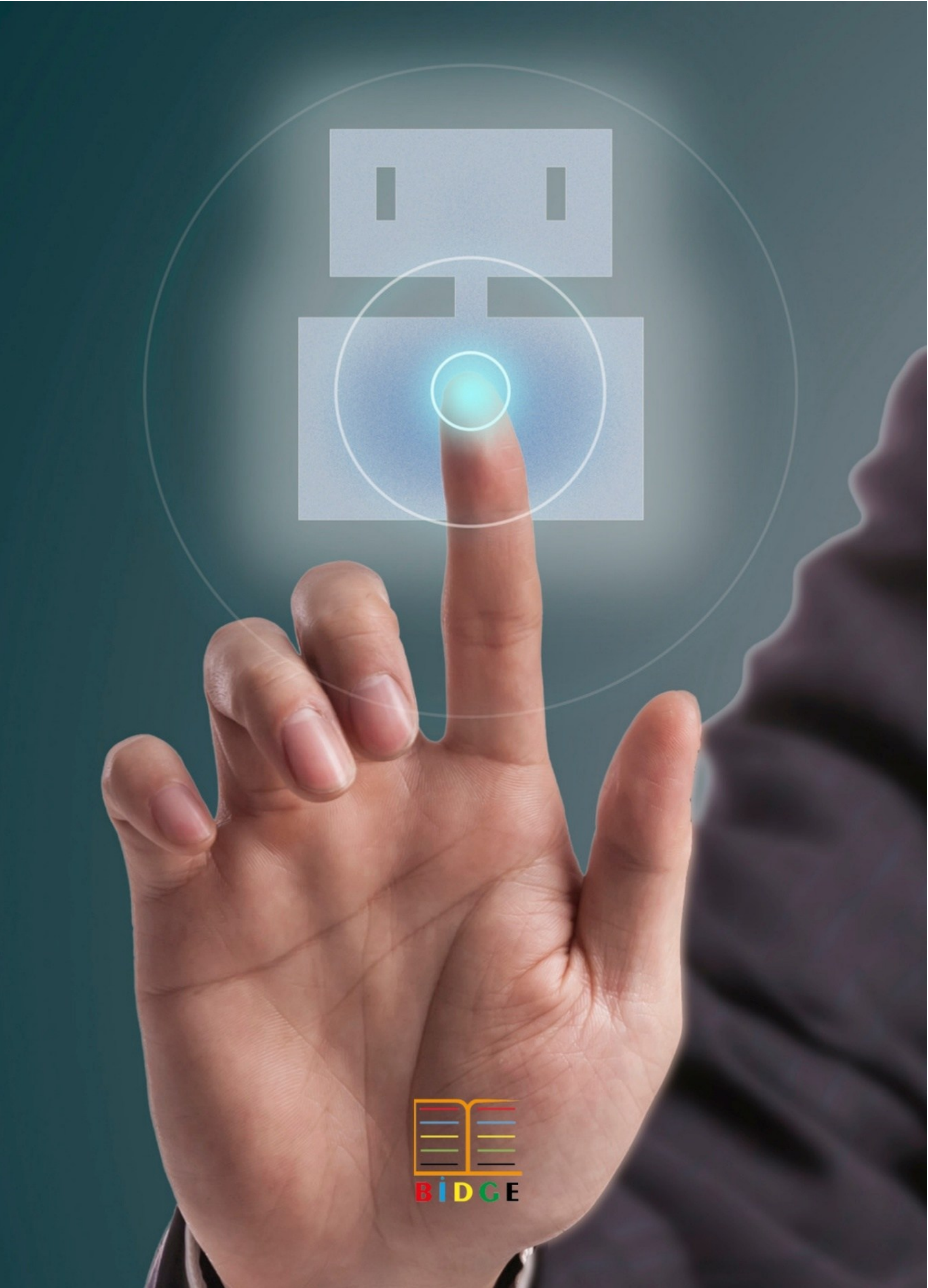
REFERANSLAR

Kousha, K., Thelwall, M., & Abdoli, M. (2017). Goodreads reviews to assess the wider impacts of books. Journal of the Association for Information Science and Technology, 68(8), 2004-2016.

Sawhney, P. (2020). Good Reads Top 1000 Books. [online]. https://www.kaggle.com/datasets/prishasawhney/good-reads-top-1000-books [Ziyaret Tarihi: 25.05.2024].

Verma, M., & Patnaik, P. K. (2024). An automatic college library book recommendation system using optimized Hidden Markov based weighted fuzzy ranking model. Engineering Applications of Artificial Intelligence, 130, 107664.

Wang, K., Liu, X., & Han, Y. (2019). Exploring Goodreads reviews for book impact assessment. Journal of Informetrics, 13(3), 874-886.


BİDGE