Data transformation review in deep learning

Agata Kozina[1]

[1] *University of Economics and Business, Komandorska 118/120, 53-345 Wrocław, Poland*

Abstract
This paper delivers up-to-date bibliometric research dealing with data transformation used in deep learning. The study presents a theoretical approach that is developed by the bibliometric maps. The aim of this paper is the identification and exploration of the research gap. The study illustrates the relationships between different science areas, such as applied economics and finance, computer science, deep learning, and machine learning. Transformation in the literature is also defined as verification, integration, validation, cleaning, labeling of data. Creating a consistent picture of data and correcting erroneous data. Data transformation methods eliminate inconsistencies, add and aggregate data, and integrate data. In the literature on machine learning and deep learning, the term data transformation generally refers to the various operations performed on data before it is used to teach models. There is no breakdown of data transformation methods. Data transformation is a general term and there is no explanation of it in the context of deep learning. To fill this gap, this study uses VOSviewer software to explore queries' results in the Scopus database and to illustrate the knowledge about data transformation in deep learning. The results show that: (I) data transformation in deep learning is a process that is carried out in the first stage (data preparation) of building deep learning models; (II) decision support using deep learning is widely used in various business fields such as human resource management, market analysis, trading, insurance, risk analysis, logistics, industrial diagnostics, forecasting markets and economics, and business process automation; (III) there is a need to accurately distinguish between transformation and data augmentation methods to avoid potential problems associated with their misapplication in machine learning and deep learning; (IV) there is a need for further research on developing automated data transformation methods, especially through integration with feature selection or extraction methods, and testing their effectiveness in various business contexts. This scientific paper serves as a starting point for future studies dedicated to data transformation in deep learning.

Keywords
Data transformation, deep learning, machine learning, decision support, Artificial Intelligence

1. Introduction

Deep learning are machine learning methods based on deep neural networks, which are built from multiple layers of neurons that implement nonlinear transformations. These layers map successive levels of abstraction to form a hierarchical model [1]. Deep learning-based models, such as deep neural networks, are becoming an important tool in decision support due to their ability to learn from data and generate accurate predictions and risk assessments [2].

Currently, there are few articles in the literature on data transformation in deep learning in the area of management. They mainly concern the medical field (medical images [3], diabetic retinopathy [4], MRI, based on which it is possible to assess whether given patients will suffer from Alzheimer's disease in the future [5]) and handwriting recognition [6].

Existing publications mainly deal with automatic data transformation methods in the field of image analysis (techniques such as scaling, rotation, image rotation). Publications do not address issues related to automatic data transformation in the area of management with regard to decision-making [7], [8].

The main conclusions of the article are as follows: (I) data transformation in deep learning is a process that is carried out in the first stage (data preparation) of building deep learning models; (II) decision support using deep learning is widely used in various business fields such as human resource management, market analysis, trading, insurance, risk analysis, logistics, industrial

The First International Workshop of Young Scientists on Artificial Intelligence for Sustainable Development; May 10-11, 2024, Ternopil, Ukraine
agata.kozina@ue.wroc.pl (A. Kozina);
0000-0003-0447-4038 (A. Kozina);

diagnostics, forecasting markets and economics, and business process automation; (III) there is a need to accurately distinguish between transformation and data augmentation methods to avoid potential problems associated with their misapplication in machine learning and deep learning; (IV) there is a need for further research on developing automated data transformation methods, especially through integration with feature selection or extraction methods, and testing their effectiveness in various business contexts. The contributions of this work are presented in the form of bibliometric maps illustrating the researched research gaps identified in the queries. An additional feature of the network diagrams of these keywords is that they indicate the location of data transformation in deep learning.

2. Background

The beginning of neural networks dates back to the 1940s. At that time, a neuron model was developed capable of recognizing only two categories of objects based on the weights set by the operator. In the late 1950s, the first neural network called perceptron was built [9].

In the 1950s. Arthur Samuel made an important breakthrough by demonstrating that computers have the ability to transcend the simple execution of direct commands. He created a series of programs that operated on a game of checkers. They were capable of optimizing their movements. As a result, these programs over time reached a level of play higher than their creator himself [10].

A breakthrough in the field of artificial intelligence was the creation of the Dendral expert system in 1965 at Stanford University, which automated the analysis of chemical compounds [11]. The 1970s saw the emergence of Automated Mathematician, written in the LISP programming language, which aimed to automatically search for new mathematical laws through the use of heuristic algorithms [12]. After the initial phase of interest in the 1950s and 1960s came a period of criticism in the 1970s. Due to the technological limitations of the time, the development of neural networks was hampered. The network performed calculations over a very long period of time (weeks, months and even years of calculations) [13]. The period of criticism lasted until the development of the backward error propagation algorithm in 1986. This learning algorithm was crucial to the further development of neural networks and machine learning [14]. The artificial intelligence winter period was from 1987 to 1993, a term that refers to a period of low interest from consumers, the public and the private sector, which led to a reduction in research funding, which in turn resulted in few breakthroughs [10]. In the 1990s, Gerald Tesauro created the TD-Gammon program, capable of competing with world champions in the game of Backgammon. This program learned strategy by playing more than a million games, and its algorithms have found applications in neuroscience [15].

The latest breakthrough took place in 2006, along with the development of Big Data and computer technology. It continues to this day. Deep neural networks with effective methods for their learning are being implemented. No small part in this breakthrough was the introduction of graphics cards with thousands of cores, which significantly reduced the learning time of deep learning models [16].

The unquestionable success and rapid development of deep learning suggest that in the future it may become a tool to achieve Artificial General Intelligence (AGI) [17].

AGI refers to the development of software that has generalized cognitive abilities similar to humans. The task of an AGI system, when faced with unknown tasks, is to find their solutions[18].

In 2015, the nonprofit organization OpenAI was founded with a mission to conduct research and achieve general artificial intelligence, provided it would be safe and its goals would be in line with those of humanity [18].

OpenAI has released its latest model called "ChatGPT," which can communicate with humans through conversation. It is based on deep neural networks. The release took place on November 30, 2022 [19]. Interest in this chatbot was considerable. Less than a week after the launch, on

December 5, 2022, OpenAI CEO Sam Altman announced: "ChatGPT was launched on Wednesday. Today it surpassed one million users!" [18].

The process of building a deep neural network model consists of five steps [20]: (1) Data preparation; (2) Development of the deep learning model; (3) Model learning; (4) Model generalization; Testing, testing and running the model prototype in a real environment.

When developing deep learning models, the data preparation process is very important. It consists of the following steps:

1. Data acquisition and cleansing: determining data sources and structure. Acquired data is saved as a dataset and cleaned of data containing errors [21].
2. Data augmentation: increasing the baseline value of input data to the deep learning model by adding other data from internal and external sources in enterprises [22], [23].
3. Data transformation: transforming the data to reduce the complexity of the input data and increase the scalability of the modeling process while maintaining model performance [21], [24].

The result of this step is a data model (input vector), which is used for model development. A very important data preparation process is data transformation, which results in the transformation of the analyzed dataset into the input vector of the deep learning model.

Jason Brownlee emphasizes that it is the data preparation stage, especially data transformation, that is the most important when developing deep learning models. The reason is that each data set is different and highly specific to the problem being solved [1], [7]. As a result, it is not possible to directly use, for example, transformation methods related to the medical field in decision support problems in organizations, since the structure of the data can vary significantly.

The literature lacks automated data transformation methods for deep learning models used in decision support. Manual data transformation relies on procedures to unify data, remove inappropriate data, and often does not consider advanced transformations. Currently, data transformations require many manual database refinement procedures [25].

It should also be noted that the automation of data transformation in machine and deep learning models has reached a certain stage of advancement, however, there is still a need for manual analysis of variable types and dimensionality reduction process, especially in management decision-making. This task is not only time-consuming, but also subject to the risk of incorrect implementation (the risk of error by the expert making the selection and implementation of data transformation methods) [25], [26], [27].

Semi-automated methods that include the support of artificial intelligence, such as GPT chatbots, can be an attractive option because of their promises to reduce decision-making time. However, in practice, they come with some challenges. Using these tools requires the transfer of code and data information, which often leads to complex analysis and matching. This is a process that can consume a significant amount of time, as suggested transformations may prove to be inappropriate or inadequate for a particular problem. Analyzing the types of variables and tailoring suggestions to a given project problem are key steps that can be time-consuming and demanding. As a result, despite promising prospects, using semi-automatic methods can be a challenge, especially in terms of fast and efficient data processing in the process of building machine learning models [28].

3. Method and Results

The results of a study on the current knowledge of data transformation in deep learning are presented. A systematic literature review of the Scopus database was conducted (Table 1).

Table 1
Number of texts in particular years for selected keywords and their combinations

Keywords	First publication and quantity	2000	2006	2012	2019	2021	2023	Sum
„deep learning”	1980 (1)	20	58	193	33,677	66,508	104,599	398,030
“data transformation”	1960 (1)	61	154	214	390	416	516	6,632
„deep learning” AND „data transformation”	2011 (1)	-	-	1	17	30	57	216

Source: own study.

The earliest (1960) were publications on data transformation. The first publication on deep learning was published in 1980. Since 2012, the number of publications on deep learning has increased rapidly. The two keywords combined in 2011, but it is only since 2019 that publications on data transformation in deep learning have been increasing.

Deep learning is a relatively new field that has gained popularity in recent years due to tremendous advances in computing technology and the availability of big data [2].

Advances in technology and the availability of big data have made machine learning, and deep learning in particular, indispensable in many business fields. As researchers become more knowledgeable about these fields, the number of publications and studies on the subject is increasing [2].

There is an apparent increase in interest in the topics of data transformation and deep learning. There are several factors that may explain this trend:

1. The growing importance of data: in recent years, data has become an extremely important resource for organizations and businesses [29].
2. Advances in computing technologies: significant advances in computing technologies have been made in recent years, such as the development of better algorithms, the availability of powerful computing systems (e.g., GPUs) and the development of cloud infrastructure [1].
3. Growing understanding and awareness: as science and research in the areas of data science, machine learning and deep learning grows, so does the understanding and awareness of the importance of data transformation by business executives [28].
4. Market demand: with the development of technologies related to data processing and machine learning, market demand for specialists in these fields is also growing [30].

In the effect of the Scopus database queries and described procedure of the bibliometric review the results in the network diagram are represented in Figure 1.

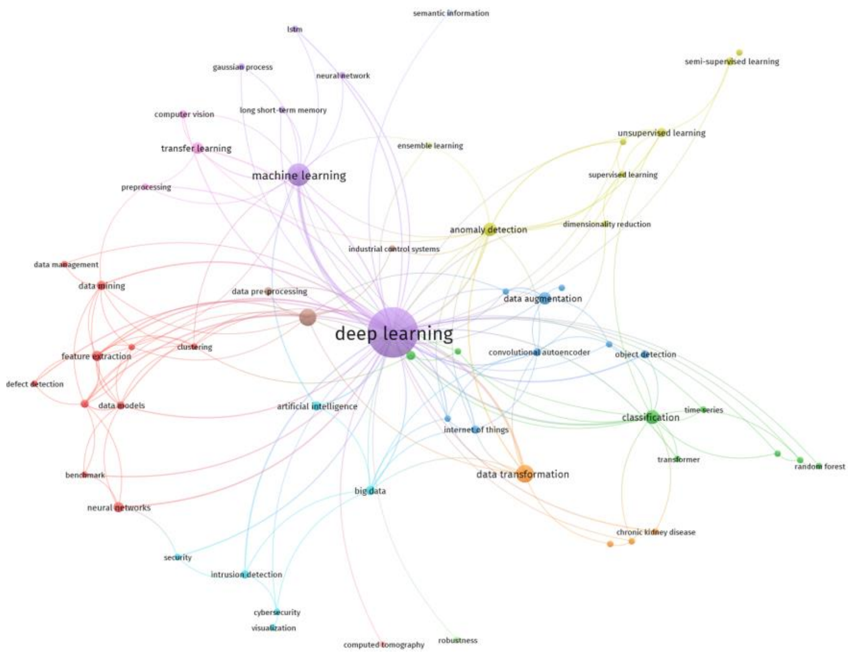

Figure 1: Vosviewer results
Source: own study.

The impact of the developed automatic data transformation method on decision support should be considered in parallel with the impact of deep learning on the process. Deep neural networks, based on multiple layers, can effectively map relationships and patterns in data that would be difficult or impossible to see by traditional analysis methods, and often even by traditional machine learning methods. Deep learning offers many advanced tools that support more rational, accurate and optimal decision-making. With its ability to analyze large data sets, make predictions, personalize, automate and solve complex problems, deep learning has wide applications in many fields, supporting decision-making and contributing to better performance and efficiency [2], [31], [32], [33], [34].

Multilayer deep neural network architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have gained recognition for their ability to automatically extract features from data. This makes it possible to recognize patterns and cause-and-effect relationships found in large and diverse data sets [2].

The use of deep learning in business data analysis makes it possible to identify hidden trends and customer preferences from purchasing data and consumer behavior. This allows companies to better adjust their products and marketing strategies, increasing their competitiveness in the market [35], [36].

With advanced deep learning methods, it is possible to analyze and process large amounts of data to make more accurate predictions, identify patterns and support key decision-making. In management, several major areas can be mentioned where deep learning has a significant impact on decision-making:

1. Human resource management: deep learning can help analyze large amounts of employee data, manage employee performance, understand training needs, and optimize the recruitment process [37].
2. Market analysis and forecasting: the use of deep neural networks in the analysis of economic, financial and market data allows the creation of more accurate forecasts, identification of market trends and assessment of investment risks [38].
3. Commerce and recommendations: deep learning is often used in e-commerce platforms and streaming services to generate personalized recommendations for products, movies or music, helping to improve user experience and increase sales. In the field of marketing and customer management, deep learning helps create personalized offers, product and service recommendations, tailored to individual customer preferences and behaviors. Deep learning-based models can analyze customer behavior, identify consumer characteristics, forecast sales, target marketing strategies, purchase histories and demographics to better understand customer needs and preferences [13], [39], [40].
4. Credit risk analysis: deep learning can help assess credit risk for potential borrowers by analyzing a variety of financial and social data to more accurately estimate customers' creditworthiness [41].
5. Insurance: deep learning can be applied in insurance decision support and risk management in this area.
6. Sentiment analysis: deep learning is used to analyze large text datasets to understand customer opinions and sentiments, which is particularly important in the field of customer service and branding [42].
7. Logistics and supply optimization: the use of deep learning in logistics data analysis can help optimize delivery routes, manage warehouses and predict product demand [499].
8. Industrial diagnostics: deep learning is being used in industry to analyze sensor and equipment data to predict maintenance dates for key machines, schedule production interruptions, order necessary parts and schedule repair and maintenance work. These activities allow early detection of machine faults and failures, helping to minimize production downtime and repair costs [43].
9. Predicting trends and markets: in finance and economics, deep learning is used to predict changes in asset prices, market trends, stock market behavior and to analyze investor sentiment. Models based on deep neural networks can analyze historical data, observe

patterns and detect signals that help better understand market volatility and make more accurate investment decisions. Deep learning has also found application in assessing risk prediction of financial investments [44].

10. Risk and fraud assessment: in risk management and anti-fraud, deep learning can help identify suspicious transactions, detect money laundering, verify customer identities, and predict potential financial problems for companies or institutions. Advanced algorithms can analyze large amounts of transaction data, detect anomalies and issue warnings about potential risks [45].
11. Forecasting economic performance: in the field of economics, deep learning can support macroeconomic forecasts, such as GDP, inflation, unemployment, etc. Models based on deep neural networks can analyze a variety of economic indicators and economic data, allowing for more accurate forecasts and responses to changing economic conditions [46], [47].
12. Energy: in this area, deep learning can forecast electricity demand, monitor and optimize energy consumption, plan an appropriate level of supply that guarantees the minimum cost of purchasing energy resources, and minimize the cost of energy production.
13. Automation and optimization of business processes: in management and economics, deep learning can support automation of business processes, optimization of logistics operations or supply chain management. Machine learning models can analyze operational data, detect patterns and suggest optimal operating strategies [48].

The division of data transformation methods in machine learning and deep learning depends mainly on the area of the decision-making problem:

1. Management, where examples of data transformation can be: logarithmization of customer opinion data about products (sentiment analysis) [49], standardization of employee data [50] (transforming employees' skills to a standardized scale, making it easier to compare their skills against the requirements of different projects or tasks), normalization of production management data [51] (a deep learning model can analyze machine performance data in the production process, predicting potential failures, where, for a balance between different machine parameters, data normalization can be applied to scale the data to a certain range), or logistics [52] (transforming delivery route data into graphs or matrices and then using deep neural networks, such as convolutional neural networks (CNNs), to optimize delivery routes according to traffic conditions, time and other variables, where a data normalization method can be applied and normalize data on location distances or transportation costs, to a set range). Coding of categorical variables, such as age, gender, education level, so that they can be used as features in a predictive model dealing with social behavior is also an example [53].
2. Finance, where examples of data transformation can be: logarithmizing stock prices to transform non-linear price changes into more linear trends [54], applying Label Powerset (LP) transformation in investment risk classification. For example, for three types of risk (market, credit, liquidity), where each can appear as “low,” “medium” and “high,” LP will transform the problem into a multi-class problem, where each combination of risks (e.g., “low market risk,” “medium credit risk,” “high liquidity risk”) will be treated as a separate class [55].
3. Economics, where examples of data transformation may include: discretizing macroeconomic indicators by, for example, defining ranges to determine unemployment levels (low, medium, high) and assigning unemployment values to the corresponding ranges based on specified thresholds [56]. Demand forecasting is also exemplified by transforming temporal data, such as daily order quantities, into seasonal variables to account for seasonal demand patterns [57].
4. Industry, where examples of data transformation can be: standardization of manufacturing process parameters such as temperature, pressure, humidity, to compare and analyze process stability [58], [59].
5. Other areas of decision-making problems such as:
    - Medicine, where examples of data transformation include: normalizing patients' laboratory results to account for differences in units of measurement and reference

ranges [60], thresholding of radiological images to isolate areas of interest, such as tumors or other pathologies [61].

- Natural language processing, where examples of data transformation include removing unnecessary Stopwords (frequently used words such as “a,” “the,” “is”) from text to improve the quality of text analysis [62].
- Geographic data analysis, where examples of data transformation could be: normalizing geographic coordinates, such as longitude and latitude, to improve the comparability of spatial data [63].

Methods for dealing with missing data are called data augmentation. Data augmentation are methods of generating new data by altering the original data. This is often used to deal with missing data or to increase the number of samples in a learning dataset [64]. Data transformation involves changing or transforming the original data without generating new data. In the literature, sometimes the terms "augmentation" and "transformation" are used interchangeably, which can lead to confusion. However, many authors point out the need to divide these methods [64], [65], [66], [67], [68].

Misapplication of data augmentation and transformation methods in machine learning and deep learning can lead to serious problems, such as model overfitting, as well as degradation of prediction or classification quality. For example, if excessive data augmentation is applied, it can cause the model to learn overly detailed patterns that are specific only to the training set but do not reflect the overall structure of the problem, leading to overlearning. On the other hand, if an inappropriate data transformation is applied, it can result in loss of information or disruption of the data structure, which will negatively affect the model's ability to generalize and be effective against new data. Therefore, it is important to always properly select and apply data augmentation and transformation methods depending on the specifics of the problem, datasets and variables used [25], [69].

4. Discussion and Conclusions

This study presents an exploration of interdisciplinary connections between various scientific domains using VOSviewer software for bibliometric analysis. The aim is to systematically guide readers through the process of conducting bibliometric data analysis and generating bibliometric maps as networks. The analysis encompasses author collaboration and keyword-based queries in the Scopus scientific database. By following a step-by-step approach, the paper enables novice users to navigate VOSviewer effectively. Through this study, the authors facilitate data analysis and identification of research gaps within the scientific landscape. The novelty lies in demonstrating bibliometric data analysis pertaining to distinct scientific fields using VOSviewer.

The methodology involves conducting bibliometric analysis to visualize keyword maps based on predetermined topics. The analysis encompasses 216 documents published between 2011 and 2024. Data transformation in the context of deep learning is often neglected, when in fact it is a key element in achieving success and getting good model results. This issue is often overlooked or underestimated, yet proper data transformation can significantly impact the effectiveness of deep models. Through proper data preparation, such as cleaning, normalization or transformation, one can ensure that models have access to high-quality and consistent data, which in turn will improve their generalization ability and prediction accuracy. Therefore, it is important to pay attention to the data transformation process and understand its fundamental importance for the successful implementation of deep learning.

Based on the literature review, it is evident that there is a need for distinguishing between methods of data transformation and augmentation.

Further research could focus on, for instance, utilizing transformation methods from the feature selection or extraction group in constructing an automatic data transformation method, testing their effectiveness in various business contexts, and comparing them with existing manual methods. Additionally, future studies could concentrate on developing a unified terminology and precise definitions for data transformation and augmentation, which would enhance understanding of the role of these methods and their impact on the quality of decision-support deep learning models.

Acknowledgements

I would like to thank my supervisors Professor, Eng. Marcin Hernes and Dr. Artur Rot from the University of Economics and Business for their research assistance in my dissertation.

References

[1] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[2] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 6, p. 420, Nov. 2021, doi: 10.1007/s42979-021-00815-1.

[3] E. Gibson *et al.*, "NiftyNet: a deep-learning platform for medical imaging," *Comput. Methods Programs Biomed.*, vol. 158, pp. 113–122, May 2018, doi: 10.1016/j.cmpb.2018.01.025.

[4] N. Asiri, M. Hussain, F. Al Adel, and N. Alzaidi, "Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey," *Artif. Intell. Med.*, vol. 99, p. 101701, Aug. 2019, doi: 10.1016/j.artmed.2019.07.009.

[5] S. Spasov, L. Passamonti, A. Duggento, P. Liò, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *NeuroImage*, vol. 189, pp. 276–287, Apr. 2019, doi: 10.1016/j.neuroimage.2019.01.031.

[6] A. Baldominos, Y. Saez, and P. Isasi, "Evolutionary convolutional neural networks: An application to handwriting recognition," *Neurocomputing*, vol. 283, pp. 38–52, Mar. 2018, doi: 10.1016/j.neucom.2017.12.049.

[7] I. Ali, L. E. Cárdenas-Barrón, A. Ahmed, and A. A. Shaikh, Eds., *Optimal decision making in operations research and statistics: methodologies and applications*, First edition. in A Science Publishers book. Boca Raton London New York: CRC Press, Taylor & Francis Group, 2021.

[8] J. Wątróbski, "Studia w kierunku identyfikacji determinant zrównoważonego zarządzania," *Stud. Proc. Pol. Assoc. Knowl. Manag.*, vol. 89, pp. 85–90.

[9] K. Ciomcia, "Czym jest deep learning i sieci neuronowe," BulldogJob. Accessed: Mar. 29, 2024. [Online]. Available: https://bulldogjob.pl/readme/czym-jest-deep-learning-i-sieci-neuronowe

[10] "Historia sztucznej inteligencji," Sztuczna Inteligencja w Uniwersytecie Szczecińskim. Accessed: Mar. 29, 2024. [Online]. Available: https://ai.usz.edu.pl/historia/

[11] N. Brown, Ed., *Artificial Intelligence in Drug Discovery:* in Drug Discovery. Cambridge: Royal Society of Chemistry, 2020. doi: 10.1039/9781788016841.

[12] Y. Hakuk and Y. Reich, "Computational infrastructure for concepts discovery in science and technology," *Adv. Eng. Inform.*, vol. 56, p. 101938, Apr. 2023, doi: 10.1016/j.aei.2023.101938.

[13] M. L. Owoc and Ł. Piasny, "Adaptive Agent Systems in Contemporary e-Commerce Environment," *Probl. Zarzadzania*, vol. 13, no. 52, pp. 88–100, Jun. 2015, doi: 10.7172/1644-9584.52.7.

[14] W. Duch, "Notatki do wykładów wstępnych, według książki Fascynujący świat programów komputerowych," Uniwersytet Mikołaja Kopernika w Toruniu. Accessed: Mar. 29, 2024. [Online]. Available: https://web.archive.org/web/20070514051106/http://www.fizyka.umk.pl/~duch/Wyklady/komput/w12/ai.html

[15] S. Mannor, Y. Mansour, and A. Tamar, *Reinforcement Learning: Foundations*. -, 2022. [Online]. Available: https://sites.google.com/view/rlfoundations/home

[16] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.

[17] M. Świechowski, "Deep Learning and Artificial General Intelligence: Still a Long Way to Go," 2022, doi: 10.48550/ARXIV.2203.14963.

[18] V. Taecharungroj, "'What Can ChatGPT Do?' Analyzing Early Reactions to the Innovative AI Chatbot on Twitter," *Big Data Cogn. Comput.*, vol. 7, no. 1, p. 35, Feb. 2023, doi: 10.3390/bdcc7010035.

[19] K. Klincewicz, *Zarządzanie, organizacje i organizowanie – przegląd perspektyw teoretycznych*. Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego, 2016. doi: 10.7172/978-83-65402-29-5.2016.wwz.9.

[20] G. Yufeng, "The 7 Steps of Machine Learning." [Online]. Available: https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e

[21] K. G. Srinivasa, S. Anupindi, R. Sharath, and S. K. Chaitanya, "Analysis of Facial Expressiveness Captured in Reaction to Videos," in *2017 IEEE 7th International Advance Computing Conference (IACC)*, Hyderabad, India: IEEE, Jan. 2017, pp. 664–670. doi: 10.1109/IACC.2017.0140.

[22] A. J. Ratner, H. R. Ehrenberg, Z. Hussain, J. Dunnmon, and C. Ré, "Learning to Compose Domain-Specific Transformations for Data Augmentation," 2017, doi: 10.48550/ARXIV.1709.01643.

[23] J. Li, B. M. Chen, and G. H. Lee, "SO-Net: Self-Organizing Network for Point Cloud Analysis," 2018, doi: 10.48550/ARXIV.1803.04249.

[24] A. Sołtysik-Piorunkiewicz, K. Pyszny, and P. Majerczak, "Analiza metod zabezpieczeń w systemach elektronicznych płatności," *Uniwersytet Ekonomiczny w Katowicach*, no. 390, pp. 105–124, 2019.

[25] M. Mayo and M. Daoud, "Data Normalisation using Differential Evolution and Aggregated Logistic Functions," in *2019 IEEE Congress on Evolutionary Computation (CEC)*, Wellington, New Zealand: IEEE, Jun. 2019, pp. 920–927. doi: 10.1109/CEC.2019.8790251.

[26] I. Icke and A. Rosenberg, "Automated measures for interpretable dimensionality reduction for visual classification: A user study," in *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, Providence, RI, USA: IEEE, Oct. 2011, pp. 281–282. doi: 10.1109/VAST.2011.6102474.

[27] U. Brefeld, E. Fromont, A. Hotho, A. Knobbe, M. Maathuis, and C. Robardet, Eds., *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part II*, vol. 11907. in Lecture Notes in Computer Science, vol. 11907. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-46147-8.

[28] A. Tapinos, B. Constantinides, M. V. T. Phan, S. Kouchaki, M. Cotten, and D. L. Robertson, "The Utility of Data Transformation for Alignment, De Novo Assembly and Classification of Short Read Virus Sequences," *Viruses*, vol. 11, no. 5, p. 394, Apr. 2019, doi: 10.3390/v11050394.

[29] A. Juneja and N. N. Das, "Big Data Quality Framework: Pre-Processing Data in Weather Monitoring Application," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India: IEEE, Feb. 2019, pp. 559–563. doi: 10.1109/COMITCon.2019.8862267.

[30] J. Huang, J. Chai, and S. Cho, "Deep learning in finance and banking: A literature review and classification," *Front. Bus. Res. China*, vol. 14, no. 1, p. 13, Dec. 2020, doi: 10.1186/s11782-020-00082-6.

[31] S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi, and S. Jain, "Machine translation using deep learning: An overview," in *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, Jaipur, India: IEEE, Jul. 2017, pp. 162–167. doi: 10.1109/COMPTELIX.2017.8003957.

[32] S. Y. Lee and M. J. Park, "Design of decision support system for yield management in semiconductor industry: application to artificial intelligence," *Int. J. Bus. Inf. Syst.*, vol. 40, no. 1, p. 60, 2022, doi: 10.1504/IJBIS.2022.122891.

[33] D. Wu, X. Wang, J. Su, B. Tang, and S. Wu, "A Labeling Method for Financial Time Series Prediction Based on Trends," *Entropy*, vol. 22, no. 10, p. 1162, Oct. 2020, doi: 10.3390/e22101162.

[34] F. Doko, S. Kalajdziski, and I. Mishkovski, "Credit Risk Model Based on Central Bank Credit Registry Data," *J. Risk Financ. Manag.*, vol. 14, no. 3, p. 138, Mar. 2021, doi: 10.3390/jrfm14030138.

[35] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," *J. Bus. Res.*, vol. 94, pp. 290–301, Jan. 2019, doi: 10.1016/j.jbusres.2018.03.003.

[36] R. Rivera-Castro, I. Nazarov, Y. Xiang, I. Maksimov, A. Pletnev, and E. Burnaev, "An Industry Case of Large-Scale Demand Forecasting of Hierarchical Components," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, Boca Raton, FL, USA: IEEE, Dec. 2019, pp. 134–139. doi: 10.1109/ICMLA.2019.00029.

[37] J. Li and Z. Zhou, "Design of Human Resource Management System Based on Deep Learning," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, Jul. 2022, doi: 10.1155/2022/9122881.

[38] Y. Dexiang, M. Shengdong, Y. Liu, G. Jijian, and L. Chaolung, "An Improved Deep-Learning-Based Financial Market Forecasting Model in the Digital Economy," *Mathematics*, vol. 11, no. 6, p. 1466, Mar. 2023, doi: 10.3390/math11061466.

[39] X. Chen and Z. Long, "E-Commerce Enterprises Financial Risk Prediction Based on FA-PSO-LSTM Neural Network Deep Learning Model," *Sustainability*, vol. 15, no. 7, p. 5882, Mar. 2023, doi: 10.3390/su15075882.

[40] L. Zhou, "Product advertising recommendation in e-commerce based on deep learning and distributed expression," *Electron. Commer. Res.*, vol. 20, no. 2, pp. 321–342, Jun. 2020, doi: 10.1007/s10660-020-09411-6.

[41] S. Shi, R. Tse, W. Luo, S. D'Addona, and G. Pau, "Machine learning-driven credit risk: a systemic review," *Neural Comput. Appl.*, vol. 34, no. 17, pp. 14327–14339, Sep. 2022, doi: 10.1007/s00521-022-07472-2.

[42] A. Iqbal, R. Amin, J. Iqbal, R. Alroobaea, A. Binmahfoudh, and M. Hussain, "Sentiment Analysis of Consumer Reviews Using Deep Learning," *Sustainability*, vol. 14, no. 17, p. 10844, Aug. 2022, doi: 10.3390/su141710844.

[43] S. Qiu *et al.*, "Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review," *Sensors*, vol. 23, no. 3, p. 1305, Jan. 2023, doi: 10.3390/s23031305.

[44] Y. Sun and J. Li, "Deep Learning for Intelligent Assessment of Financial Investment Risk Prediction," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–11, Oct. 2022, doi: 10.1155/2022/3062566.

[45] M. P. Brito, M. Stevenson, and C. Bravo, "Subjective machines: Probabilistic risk assessment based on deep learning of soft information," *Risk Anal.*, vol. 43, no. 3, pp. 516–529, Mar. 2023, doi: 10.1111/risa.13930.

[46] K.-H. Chen and K. Khashanah, "The Reconstruction of Financial Signals Using Fast ICA for Systemic Risk," in *2015 IEEE Symposium Series on Computational Intelligence*, Cape Town, South Africa: IEEE, Dec. 2015, pp. 885–889. doi: 10.1109/SSCI.2015.130.

[47] K. He, Q. Yang, L. Ji, J. Pan, and Y. Zou, "Financial Time Series Forecasting with the Deep Learning Ensemble Model," *Mathematics*, vol. 11, no. 4, p. 1054, Feb. 2023, doi: 10.3390/math11041054.

[48] M. Schmitt, "Deep learning in business analytics: A clash of expectations and reality," *Int. J. Inf. Manag. Data Insights*, vol. 3, no. 1, p. 100146, Apr. 2023, doi: 10.1016/j.jjimei.2022.100146.

[49] A. Iqbal, R. Amin, J. Iqbal, R. Alroobaea, A. Binmahfoudh, and M. Hussain, "Sentiment Analysis of Consumer Reviews Using Deep Learning," *Sustainability*, vol. 14, no. 17, p. 10844, Aug. 2022, doi: 10.3390/su141710844.

[50] D. V. Hieu, "Incorporating Generative AI into Quality Management Systems Enhancing Process Optimization and Product Development," *Int. J. Appl. Mach. Learn. Comput. Intell.*, vol. 13, no. 11, pp. 1–8, Nov. 2023.

[51] M. Züfle, F. Moog, V. Lesch, C. Krupitzer, and S. Kounev, "A machine learning-based workflow for automatic detection of anomalies in machine tools," *ISA Trans.*, vol. 125, pp. 445–458, Jun. 2022, doi: 10.1016/j.isatra.2021.07.010.

[52] P. Dhawas, M. A. Ramteke, A. Thakur, P. V. Polshetwar, R. V. Salunkhe, and D. Bhagat, "Big Data Analysis Techniques: Data Preprocessing Techniques, Data Mining Techniques, Machine Learning Algorithm, Visualization," in *Advances in Business Information Systems and Analytics*, D. Darwish, Ed., IGI Global, 2024, pp. 183–208. doi: 10.4018/979-8-3693-0413-6.ch007.

[53] H. S. Sandhu and D. M. Vistro, "Data Management for Healthcare with a Focus on Privacy and Security for Cancer Patients," *Int. J. Curr. Res. Rev.*, vol. 12, no. 17, pp. 62–70, 2020, doi: 10.31782/IJCRR.2020.121712.

[54] D. Liu and C.-H. Lung, "P2P traffic identification and optimization using fuzzy c-means clustering," in *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, Taipei, Taiwan: IEEE, Jun. 2011, pp. 2245–2252. doi: 10.1109/FUZZY.2011.6007613.

[55] M. Hernes *et al.*, "Deep Learning for Repayment Prediction in Leasing Companies," *Eur. Res. Stud. J.*, vol. XXIV, no. Issue 2, pp. 1134–1148, May 2021, doi: 10.35808/ersj/2178.

[56] M. Curry, A. Trott, S. Phade, Y. Bai, and S. Zheng, "Analyzing Micro-Founded General Equilibrium Models with Many Agents using Deep Reinforcement Learning," 2022, doi: 10.48550/ARXIV.2201.01163.

[57] N. Phumchusri and P. Ungtrakul, "Hotel daily demand forecasting for high-frequency and complex seasonality data: a case study in Thailand," *J. Revenue Pricing Manag.*, vol. 19, no. 1, pp. 8–25, Feb. 2020, doi: 10.1057/s41272-019-00221-6.

[58] F. Liu and Y. Dai, “Product Processing Quality Classification Model for Small-Sample and Imbalanced Data Environment,” *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–16, Mar. 2022, doi: 10.1155/2022/9024165.

[59] A. Kopper, R. Karkare, R. C. Paffenroth, and D. Apelian, “Model Selection and Evaluation for Machine Learning: Deep Learning in Materials Processing,” *Integrating Mater. Manuf. Innov.*, vol. 9, no. 3, pp. 287–300, Sep. 2020, doi: 10.1007/s40192-020-00185-1.

[60] M. Greco, P. F. Caruso, and M. Cecconi, “Artificial Intelligence in the Intensive Care Unit,” *Semin. Respir. Crit. Care Med.*, vol. 42, no. 01, pp. 002–009, Feb. 2021, doi: 10.1055/s-0040-1719037.

[61] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, “Deep Learning in Medical Image Analysis,” in *Deep Learning in Medical Image Analysis*, vol. 1213, G. Lee and H. Fujita, Eds., in Advances in Experimental Medicine and Biology, vol. 1213. , Cham: Springer International Publishing, 2020, pp. 3–21. doi: 10.1007/978-3-030-33128-3_1.

[62] Y. Xie, L. Le, Y. Zhou, and V. V. Raghavan, “Deep Learning for Natural Language Processing,” in *Handbook of Statistics*, vol. 38, Elsevier, 2018, pp. 317–328. doi: 10.1016/bs.host.2018.05.001.

[63] A. Juneja and N. N. Das, “Big Data Quality Framework: Pre-Processing Data in Weather Monitoring Application,” in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India: IEEE, Feb. 2019, pp. 559–563. doi: 10.1109/COMITCon.2019.8862267.

[64] C. Dilmegani, “Top Data Augmentation Techniques: Ultimate Guide for 2024.” Accessed: Mar. 29, 2024. [Online]. Available: https://research.aimultiple.com/data-augmentation-techniques/

[65] A. Mumuni and F. Mumuni, “Data augmentation: A comprehensive survey of modern approaches,” *Array*, vol. 16, p. 100258, Dec. 2022, doi: 10.1016/j.array.2022.100258.

[66] A. Gandhi, “Data Augmentation | How to use Deep Learning when you have Limited Data—Part 2.” Accessed: Mar. 29, 2024. [Online]. Available: https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/

[67] M. Nakhwan and R. Duangsoithong, “Comparison Analysis of Data Augmentation using Bootstrap, GANs and Autoencoder,” in *2022 14th International Conference on Knowledge and Smart Technology (KST)*, Chon buri, Thailand: IEEE, Jan. 2022, pp. 18–23. doi: 10.1109/KST53302.2022.9729065.

[68] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Glob. Transit. Proc.*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.

[69] S. Zhu, B. An, and F. Huang, “Understanding the Generalization Benefit of Model Invariance from a Data Perspective,” *CoRR*, vol. abs/2111.05529, 2021, [Online]. Available: https://arxiv.org/abs/2111.05529