electronics

Editorial

Big Data Analytics Using Artificial Intelligence

Amir H. Gandomi [1,2,*], Fang Chen [3] and Laith Abualigah [4,5,6,7,8,9,*]

1 Faculty of Engineering & Information Technology, University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007, Australia
2 University Research and Innovation Center (EKIK), Óbuda University, 1034 Budapest, Hungary
3 Data Science Institute, University of Technology Sydney, Ultimo, NSW 2007, Australia
4 Computer Science Department, Prince Hussein Bin Abdullah Faculty for Information Technology, Al Al-Bayt University, Mafraq 25113, Jordan
5 Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman 19328, Jordan
6 Center for Engineering Application & Technology Solutions, Ho Chi Minh City Open University, Ho Chi Minh 700000, Vietnam
7 Faculty of Information Technology, Middle East University, Amman 11831, Jordan
8 Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan
9 School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia
* Correspondence: gandomi@uts.edu.au (A.H.G.); aligah.2020@gmail.com (L.A.)

1. Introduction

Data analytics using artificial intelligence is the process of leveraging advanced AI techniques to extract insights and knowledge from large and complex datasets [1]. This involves utilizing machine learning algorithms, deep learning models, and natural language processing techniques to uncover patterns and relationships within big data that can inform decision making and drive innovation. The goal of big data analytics using AI is to automate data analysis and make the process faster, more accurate, and more scalable, enabling organizations to harness the full potential of their data and gain a competitive advantage.

2. The Present Issue

This Special Issue consists of fourteen articles covering different aspects of machine learning and artificial intelligence.

This study focuses on creating a machine learning model that can predict the likelihood of chronic kidney disease using publicly available data [2]. The data underwent several preprocessing steps, including the imputation of missing values, balancing through the SMOTE algorithm, and scaling of features. The chi-squared test was utilized to select the most relevant and highly correlated features. The machine learning model was built using a combination of supervised learning techniques, with support vector machine (SVM) and random forest (RF) achieving the lowest false-negative rate and highest test accuracy of 99.33% and 98.67%, respectively. SVM was found to perform better than RF upon validation through 10-fold cross-validation.

This study represents the first attempt to examine selected design research publications using a sophisticated method called "text mining" [3]. This method generates results based on the presence of specific research terms (i.e., keywords), which provides a more reliable outcome compared to other approaches that rely on contextual information or authors' perspectives. The primary objective of this research is to increase awareness and understanding of design research, and to identify potential future research directions by addressing gaps in the literature. Based on the literature review, it can be concluded that the field of design research still lacks a unifying theory. Text mining, with its features, enhances the validity and generalizability of the results compared to other methods in the literature. The text mining technique was applied to collect data from 3553 articles from 10 journals, utilizing 17,487 keywords. This research explores new topics in the field of

Citation: Gandomi, A.H.; Chen, F.; Abualigah, L. Big Data Analytics Using Artificial Intelligence. *Electronics* **2023**, *12*, 957. https://doi.org/10.3390/electronics12040957

Received: 6 February 2023
Accepted: 10 February 2023
Published: 15 February 2023

design concepts, drawing the attention of researchers, practitioners, and journal editorial boards. The key categories analyzed and presented in this paper provide insights into the growth and decline in various fields in the domain of design.

This paper presents a novel deep learning approach for detecting student emotions [4]. The main objective of the study is to explore the relationship between teaching practices and student learning, based on emotional impact. The system uses facial recognition algorithms to gather information from online platforms and image classification techniques to identify the emotions of students and teachers. Two deep learning models are compared for their performance, and the results show promising outcomes, as discussed in the Experimental Results section. The proposed system is validated using an online course with students, and the results indicate that the technique operates effectively. Various deep learning techniques are applied for emotional analysis, including transfer learning for a pre-trained deep neural network, which increases the accuracy of the emotion classification stage. The results of the experiment demonstrate that the proposed method is promising, as discussed in the Experimental Results section.

This paper proposes a deep learning solution for detecting masks worn in public to prevent the spread of coronavirus [5]. The system, designed for real-time use with a webcam, utilizes an ensemble method for high accuracy and improved detection speed. Transfer learning on pre-trained models and rigorous testing on objective data resulted in a dependable and cost-effective solution. The findings indicate the effectiveness of the solution in real-world settings, contributing to pandemic control. Compared to existing methods, the proposed solution achieves improved accuracy and performance metrics, such as specificity, precision, recall, and F measure, in three-class outputs. A careful balance is maintained between the number of parameters and processing time.

This study proposes a deep learning method for the classification and analysis of scientific literature using convolutional neural networks (CNNs) [6]. The research is divided into three dimensions, publication features, author features, and content features, with explicit and implicit features forming a set of scientometric terms. The CNN model uses weighted scientometric term vectors to achieve dual-label classification of literature based on its content and methods. The study showcases the effectiveness of the proposed model through an application example from data science and analytics literature, with results showing improved precision, recognition, and F1 score compared to other machine learning classification methods. The proposed scientometric classification model also exhibits higher accuracy than deep learning classification using only explicit and dominant features. This study offers a guide for fine-grained classification of scientific literature and provides insight into its practical application.

This research aims to help science students identify butterfly species without causing harm to the insects during analysis [7]. The study employs transfer learning with neural network models to classify butterfly species based on images. The dataset consists of 10,035 images of 75 butterfly species and 15 unusual species were selected for the study, with various orientations, photography angles, lengths, and backgrounds. The imbalanced class distribution in the dataset resulted in overfitting, which was addressed with data augmentation. Transfer learning was applied using several convolutional neural network architectures, including VGG16, VGG19, MobileNet, Xception, ResNet50, and InceptionV3. The models were evaluated based on precision, recall, F measure, and accuracy. The results showed that the InceptionV3 architecture provided an accuracy of 94.66%, which was superior to all other architectures. This work proposes a new approach for identifying glaucoma from fundus images using a deep belief network (DBN), optimized by the elephant-herding optimization (EHO) algorithm [8]. The system is designed to be tested on various datasets, which can help to improve the accuracy of glaucoma diagnosis.

This paper examines 66 machine learning models using a two-stage evaluation process [9]. The evaluation was performed on a real-world dataset of European credit card frauds and used stratified K-fold cross-validation. Out of 330 evaluation metrics, the All K-Nearest Neighbors (AllKNN) undersampling technique with CatBoost (AllKNN–CatBoost) was found to be the best model, achieving an AUC of 97.94%, recall of 95.91%, and F1 score of 87.40%. The AllKNN–CatBoost model was compared to relevant studies and was found to outperform previous models.

This research presents a hybrid data analytics framework that combines convolutional neural networks and bidirectional long short-term memory (CNN-BiLSTM) to examine the effect of merging news events and sentiment analysis with financial data on stock trend prediction [10]. Two real-world case studies were conducted using data from the Dubai Financial Market between 1 January 2020 and 1 December 2021, in the real estate and communications sectors. The results demonstrate that incorporating news events and sentiment analysis with financial data improves the accuracy of stock trend prediction. The CNN–BiLSTM model achieved an improvement of 11.6% in the real estate sector and 25.6% in communications compared to benchmarked machine learning models.

This study introduces a four-layer model and proposes a hybrid imputation method (HIMP) for filling in multi-pattern missing data, including non-random, random, and completely random patterns [11]. HIMP starts by imputing non-random missing data patterns and then dividing the resulting dataset into two datasets with random and completely random missing data patterns. Next, different imputation methods are applied to each dataset based on the missing data pattern. The final dataset is created by merging the best imputed datasets from random and completely random patterns. The effectiveness of HIMP was evaluated using a real dataset named IRDia that had all three missing data patterns. HIMP was compared to other methods using accuracy, precision, recall, and F1 score with different classifiers, and the results showed that HIMP outperformed other methods in imputing multi-pattern missing values.

This paper presents a new Whale Optimization Algorithm (EWOA) to solve Optimal Power Flow (OPF) problems, with the aim of improving exploration capability and maintaining a balance between exploration and exploitation [12]. The movement strategy of whales in the EWOA is improved through the introduction of two new techniques: (1) encircling the target using Levy motion and (2) searching for the target using Brownian motion, which work in conjunction with the traditional bubble-net attacking method. To evaluate the performance of EWOA-OPF, it is compared with six well-known optimization algorithms in solving both single- and multi-objective OPF problems under system constraints. The comparison results show that the EWOA-OPF outperforms the other algorithms and provides better solutions for both single- and multi-objective OPF problems.

In this review, the authors examine the advancements and applications of the Harris Hawk Optimizer (HHO), a robust optimization technique that has gained popularity in recent years [13]. Through experiments conducted on the Congress on Evolutionary Computation (CEC2005) and CEC2017, HHO is compared to nine other state-of-the-art algorithms, showing its efficacy and effectiveness. The paper provides a comprehensive overview of HHO and delves into future directions and areas for further investigation of new variants of the algorithm and its widespread use.

This paper provides a comprehensive overview of effective communication techniques for space exploration of ground, aerial, and underwater vehicles [14]. The study not only summarizes the challenges faced in trajectory planning, space exploration, optimization, and other areas, but also highlights the future directions for research. Aiming to fill the gap in the literature for those interested in path planning, this paper includes optimization strategies for terrestrial, underwater, and airborne applications. The study covers numerical, bio-inspired, and hybrid methodologies for each dimension discussed. The goal of this paper is to establish a centralized platform for publishing research on autonomous vehicles on land and their trajectory optimizations, airborne vehicles, and underwater vehicles.

This review looks at the drawbacks of traditional TB diagnostic methods and provides a comprehensive overview of various machine learning algorithms and their use in TB diagnosis [15]. It also examines the integration of deep learning techniques with other systems, such as neuro-fuzzy logic, genetic algorithms, and artificial immune systems.

Finally, the review highlights several cutting-edge tools, such as CAD4TB, Lunit INSIGHT, qXR, and InferRead DR Chest, which are shaping the future of AI-assisted TB diagnosis.

3. Future Directions

The future of Big Data Analytics using Artificial Intelligence is expected to follow several key directions, including:

- Real-time analytics: the increasing demand for real-time insights and decision making will drive the development of AI-powered big data analytics platforms that can process large volumes of data in near real time.
- Edge analytics: with the proliferation of IoT devices, there will be a growing need for edge analytics, where data are analyzed and processed at the source, reducing the need for data to be transferred to centralized data centers.
- Explainable AI: as AI-powered analytics become more widespread, there will be a growing need for explainable AI, where the reasoning behind AI-generated insights and predictions is made transparent and understandable.
- Integration with other technologies: the integration of AI-powered big data analytics with other technologies, such as cloud computing, blockchain, and quantum computing, will enable organizations to take full advantage of the potential of big data.
- Personalized analytics: the development of AI algorithms that can tailor insights and predictions to specific individuals and organizations will drive the growth of personalized analytics, making big data analytics even more accessible and relevant.

These are some of the key directions that the future of big data analytics using AI is expected to take, enabling organizations to leverage the full potential of their data and drive innovation and growth.

Author Contributions: Conceptualization, A.H.G., F.C. and L.A.; formal analysis, A.H.G., F.C. and L.A.; writing—original draft preparation, A.H.G., F.C. and L.A.; writing—review and editing, A.H.G., F.C. and L.A.; supervision, A.H.G., F.C. and L.A.; project administration, A.H.G., F.C. and L.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not Applicable.

Acknowledgments: I would like to express my gratitude to all the researchers who submitted articles for this Special Issue and made outstanding contributions. My appreciation goes to the reviewers as well, who played a crucial role in evaluating the manuscripts and offering insightful suggestions to enhance the quality of the contributions. I recognize the editorial board of *Electronics* for granting me the opportunity to guest edit this Special Issue. Lastly, I am grateful to the *Electronics* Editorial Office staff for their dedication in maintaining a rigorous peer-review process and ensuring timely publication.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gandomi, A.H.; Chen, F.; Abualigah, L. Machine Learning Technologies for Big Data Analytics. *Electronics* **2022**, *11*, 421. [CrossRef]
2. Swain, D.; Mehta, U.; Bhatt, A.; Patel, H.; Patel, K.; Mehta, D.; Acharya, B.; Gerogiannis, V.C.; Kanavos, A.; Manika, S. A Robust Chronic Kidney Disease Classifier Using Machine Learning. *Electronics* **2023**, *12*, 212. [CrossRef]
3. Nusir, M.; Louati, A.; Louati, H.; Tariq, U.; Abu Zitar, R.; Abualigah, L.; Gandomi, A.H. Design Research Insights on Text Mining Analysis: Establishing the Most Used and Trends in Keywords of Design Research Journals. *Electronics* **2022**, *11*, 3930. [CrossRef]
4. AlZu'bi, S.; Abu Zitar, R.; Hawashin, B.; Abu Shanab, S.; Zraiqat, A.; Mughaid, A.; Almotairi, K.H.; Abualigah, L. A Novel Deep Learning Technique for Detecting Emotional Impact in Online Education. *Electronics* **2022**, *11*, 2964. [CrossRef]
5. Ai, M.A.S.; Shanmugam, A.; Muthusamy, S.; Viswanathan, C.; Panchal, H.; Krishnamoorthy, M.; Elminaam, D.S.A.; Orban, R. Real-Time Facemask Detection for Preventing COVID-19 Spread Using Transfer Learning Based Deep Neural Network. *Electronics* **2022**, *11*, 2250. [CrossRef]
6. Daradkeh, M.; Abualigah, L.; Atalla, S.; Mansoor, W. Scientometric Analysis and Classification of Research Using Convolutional Neural Networks: A Case Study in Data Science and Analytics. *Electronics* **2022**, *11*, 2066. [CrossRef]

7. Fathimathul Rajeena, P.P.; Orban, R.; Vadivel, K.S.; Subramanian, M.; Muthusamy, S.; Elminaam, D.S.A.; Nabil, A.; Abualigah, L.; Ahmadi, M.; Ali, M.A. A novel method for the classification of butterfly species using pre-trained CNN models. *Electronics* **2022**, *11*, 2016. [CrossRef]
8. Ali, M.A.S.; Balasubramanian, K.; Krishnamoorthy, G.D.; Muthusamy, S.; Pandiyan, S.; Panchal, H.; Mann, S.; Thangaraj, K.; El-Attar, N.E.; Abualigah, L.; et al. Classification of Glaucoma Based on Elephant-Herding Optimization Algorithm and Deep Belief Network. *Electronics* **2022**, *11*, 1763. [CrossRef]
9. Alfaiz, N.S.; Fati, S.M. Enhanced Credit Card Fraud Detection Model Using Machine Learning. *Electronics* **2022**, *11*, 662. [CrossRef]
10. Daradkeh, M.K. A Hybrid Data Analytics Framework with Sentiment Convergence and Multi-Feature Fusion for Stock Trend Prediction. *Electronics* **2022**, *11*, 250. [CrossRef]
11. Nadimi-Shahraki, M.H.; Mohammadi, S.; Zamani, H.; Gandomi, M.; Gandomi, A.H. A Hybrid Imputation Method for Multi-Pattern Missing Data: A Case Study on Type II Diabetes Diagnosis. *Electronics* **2021**, *10*, 3167. [CrossRef]
12. Nadimi-Shahraki, M.H.; Taghian, S.; Mirjalili, S.; Abualigah, L.; Elaziz, M.A.; Oliva, D. EWOA-OPF: Effective Whale Optimization Algorithm to Solve Optimal Power Flow Problem. *Electronics* **2021**, *10*, 2975. [CrossRef]
13. Hussien, A.G.; Abualigah, L.; Abu Zitar, R.; Hashim, F.A.; Amin, M.; Saber, A.; Almotairi, K.H.; Gandomi, A.H. Recent Advances in Harris Hawks Optimization: A Comparative Study and Applications. *Electronics* **2022**, *11*, 1919. [CrossRef]
14. Mir, I.; Gul, F.; Mir, S.; Khan, M.A.; Saeed, N.; Abualigah, L.; Abuhaija, B.; Gandomi, A.H. A Survey of Trajectory Planning Techniques for Autonomous Systems. *Electronics* **2022**, *11*, 2801. [CrossRef]
15. Singh, M.; Pujar, G.V.; Kumar, S.A.; Bhagyalalitha, M.; Akshatha, H.S.; Abuhaija, B.; Alsoud, A.R.; Abualigah, L.; Beeraka, N.M.; Gandomi, A.H. Evolution of Machine Learning in Tuberculosis Diagnosis: A Review of Deep Learning-Based Medical Applications. *Electronics* **2022**, *11*, 2634. [CrossRef]