

# Machine Learning Technologies for Big Data Analytics

Amir H. Gandomi <sup>1</sup>, Fang Chen <sup>2</sup> and Laith Abualigah <sup>3,4,\*</sup>

<sup>1</sup> Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia; gandomi@uts.edu.au

<sup>2</sup> Data Science Institute, University of Technology Sydney, Ultimo, NSW 2007, Australia; fang.chen@uts.edu.au

<sup>3</sup> Faculty of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan

<sup>4</sup> School of Computer Sciences, Universiti Sains Malaysia, Gelugor 11800, Malaysia

\* Correspondence: aligah.2020@gmail.com

## 1. Introduction

Big data analytics is one high focus of data science and there is no doubt that big data is now quickly growing in all science and engineering fields. Big data analytics is the process of examining and analyzing massive and varied data that can help organizations make more-informed business decisions, especially for uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful information. Big data has become essential as numerous organizations deal with massive amounts of specific information, which can contain useful information about problems such as national intelligence, cybersecurity, biology, fraud detection, marketing, astronomy, and medical informatics. Several promising machine learning techniques can be used for big data analytics including representation learning, deep learning, distributed and parallel learning, transfer learning, active learning, and kernel-based learning. In addition, big data analytics demands new and sophisticated algorithms based on machine learning techniques to treat data in real-time with high accuracy and productivity.

The papers published in this Special Issue (Machine Learning Technologies for Big Data Analytics) have covered various vital topics enriching the state of the art in artificial intelligence, machine learning, and big data analytics. Additionally, these search papers build upon the fundamental techniques and approaches accomplished earlier. The creativity in the established papers resides in the methods, reviews, and experimental techniques that present an outstanding value for beneficial applications. That presents one of the explanations for why this Special Issue has been named “Machine Learning Technologies for Big Data Analytics”. Nevertheless, there is another explanation: practical applications need researchers, scientists, and engineers to find solutions for the big data problems consistent with current technologies and react to the demands from the near future. That is why the searchers must utilize and develop artificial intelligence and machine learning methods for a specific need. The reader will see this Special Issue as valuable for that goal.

## 2. The Present Issue

This Special Issue contains a variety of proposed methods covering a wide range of issues related to machine learning in big data applications. The contents of these published papers are shortly presented as follows.

In this paper [1], a deep learning-based denoising approach Cross-Modality Guided Denoising Network (CMGDNet) for reducing Rician noise in T1-weighted (T1-w) magnetic resonance images (MRI) is suggested, motivated by deep learning performance in numerous medical imaging applications.

Handwritten scripts differ from person to person, which is easy for people to grasp but difficult for machines to recognize, especially when a single character has different forms. A suitable dataset for Pashto digits is required to overcome the difficulty of training



**Citation:** Gandomi, A.H.; Chen, F.; Abualigah, L. Machine Learning Technologies for Big Data Analytics. *Electronics* **2022**, *11*, 421. <https://doi.org/10.3390/electronics11030421>

Received: 24 January 2022

Accepted: 28 January 2022

Published: 30 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

a machine with Pashto digits. As a result, one of the primary reasons for this search [2] is the creation of a suitable dataset for Pashto digits.

Four data-driven predictive models based on deep neural networks (DNNs) with an attention mechanism are proposed in [3]. Data are prepared using a sliding time window approach to facilitate DNN feature extraction. The raw data gathered after normalization is input into the proposed network, necessitating no prior knowledge of prognostics or signal processing and greatly simplifying the use of the proposed technology.

In this paper [4], Twitter sentiment is examined to assess popular attitudes before, during, and after elections, and these opinions are compared to actual election results. Opinions are compared between the 2016 election, which Donald J. Trump won, and the 2020 election. The authors constructed a dataset using the Twitter API, pre-processed it, retrieved the relevant features using TF-IDF, and then used the Naive Bayes Classifier to gather public views.

This paper creates a machine learning classifier based on these Twitter accounts' profiles and bio information [5]. A feature selection strategy is employed to lower the number of features and increase the classifier's performance efficiently and effectively.

This paper proposed a host-based intrusion detection system (HIDS) based on lightweight approaches and leveraging fog computing devices that use a modified vector space representation (MVSR) N-gram and a multilayer perceptron (MLP) model for safeguarding the Internet of Things (IoT) [6].

Toxicity has become associated with online hate speech, trolling, and, at times, outrage culture. Using the bidirectional encoder representations from transformers, this paper developed an effective model for detecting and classifying toxicity in social media from user-generated material (BERT) [7].

A vast dataset is proposed in [8], which consists of 10,742 carefully identified comments in Albanian. Furthermore, attempts are made in this research to design and create a sentiment analyzer based on deep learning. Consequently, the authors provided the experimental results derived from our proposed sentiment analyzer, trained and validated on our gathered and curated dataset using several classifier models with static and contextualized word embeddings, namely fastText and BERT.

This paper coupled a linear weighted regression with the energy-aware greedy scheduling (LWR-EGS) technique to manage large amounts of data [9]. The LWR-EGS approach first chooses tasks for an assignment. It then chooses the most excellent available computer to find the best answer. The problem was initially modeled as an integer linear weighted regression program to choose tasks for the assignment to achieve this goal. The best available machines were then chosen to discover the optimum solution.

This study provides a complete overview of meta-heuristic optimization techniques for text clustering applications, highlighting their major approaches [10]. Because of their adequate capacity to address machine learning challenges, particularly text clustering difficulties, these Artificial Intelligence (AI) algorithms are considered promising swarm intelligence technologies. This work examines the whole body of research on meta-heuristic-based text clustering applications, including several versions such as primary, modified, hybridized, and multi-objective techniques.

This study aimed to examine various intervention and preventative strategies for this condition in teenagers [11]. The requirements for admission were satisfied by 14 programs in total. The study of the programs enables the formulation of successful intervention strategies for prevention and the treatment of present issues resulting from teenage users' use of the Internet and technical gadgets.

This study summarizes the essential communication techniques (ground, aerial, and underwater vehicles) [12]. It sheds light on trajectory planning, optimization, and numerous challenges. This level of in-depth study is uncommon in the literature. Hence, an attempt has been made to fill the gap for readers interested in path planning.

This review study provides a thorough summary of the approaches presented in the existing research for evaluating ML explanations [13]. This paper established explain

ability qualities based on a survey of explain ability definitions. The established explain ability features are utilized as targets for assessment measures. According to the survey, quantitative metrics for model-based and example-based explanations are primarily used to assess the parsimony/simplicity of interpretability. In contrast, quantitative metrics for attribution-based explanations are primarily used to assess the soundness of fidelity of explain ability.

This research aims to fill that gap by developing a model for understanding and estimating the prevalence of cyberstalking victims [14]. This paper's model is based on habitual behaviors and lifestyle exposure theories, and it comprises eight assumptions. This paper's data were gathered from 757 Jordanian university students. This review study employs a quantitative method and structural equation modeling for data analysis. The findings demonstrated a small prevalence range that is more reliant on cyberstalking.

### 3. Future

Some important future directions are given here according to the conclusions of the published papers, which can help future researchers easily find starting points in their research in the domain of machine learning for big data analytics problems. It would be intriguing to expand the research to other organs such as the liver, lungs, and other multi-modal medical imaging modalities in the future. A new deep CNN model can be utilized for different handwriting styles. Future work can address the limitations such as identifying new features, developing classifiers with other machine learning techniques. Moreover, other works might be performed in the future to improve the classification toxic comments model's suitability for dealing with specific social media data, studying more colloquial textual data on social sites such as Twitter and Instagram and proposing deep learning models that can be enhanced with semantically rich representations to successfully extract people's ideas and attitudes. As a result, this study should be regarded as a beginning point for future research that integrates programs that have been implemented and verified with students at different academic levels.

**Funding:** This research received no external funding.

**Acknowledgments:** First and foremost, we would like to thank all scholars who contributed papers to this Special Issue for their outstanding work. I am also thankful to all of the reviewers who assisted in examining the articles and provided constructive ideas to improve the quality of the contributions. We would like to thank the *Electronics* editorial board for inviting me to guest edit this Special Issue. I am particularly thankful to the *Electronics* Editorial Office team, who worked tirelessly to keep the rigorous peer-review schedule and prompt publishing on track.

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

## References

1. Naseem, R.; Cheikh, F.A.; Beghdadi, A.; Muhammad, K.; Sajjad, M. Cross-Modal Guidance Assisted Hierarchical Learning Based Siamese Network for MR Image Denoising. *Electronics* **2021**, *10*, 2855. [\[CrossRef\]](#)
2. Rehman, M.Z.; Nawari, N.M.; Arshad, M.; Khan, A. Recognition of Cursive Pashto Optical Digits and Characters with Trio Deep Learning Neural Network Models. *Electronics* **2021**, *10*, 2508. [\[CrossRef\]](#)
3. Muneer, A.; Taib, S.M.; Naseer, S.; Ali, R.F.; Aziz, I.A. Data-Driven Deep Learning-Based Attention Mechanism for Remaining Useful Life Prediction: Case Study Application to Turbofan Engine Analysis. *Electronics* **2021**, *10*, 2453. [\[CrossRef\]](#)
4. Chaudhry, H.N.; Javed, Y.; Kulsoom, F.; Mehmood, Z.; Khan, Z.I.; Shoaib, U.; Janjua, S.H. Sentiment Analysis of before and after Elections: Twitter Data of U.S. Election 2020. *Electronics* **2021**, *10*, 2082. [\[CrossRef\]](#)
5. Karami, A.; Lundy, M.; Webb, F.; Boyajieff, H.; Zhu, M.; Lee, D. Automatic Categorization of LGBT User Profiles on Twitter with Machine Learning. *Electronics* **2021**, *10*, 1822. [\[CrossRef\]](#)
6. Khater, B.; Wahab, A.A.; Idris, M.; Hussain, M.; Ibrahim, A.; Amin, M.; Shehadeh, H. Classifier Performance Evaluation for Lightweight IDS Using Fog Computing in IoT Security. *Electronics* **2021**, *10*, 1633. [\[CrossRef\]](#)
7. Fan, H.; Du, W.; Dahou, A.; Ewees, A.; Yousri, D.; Elaziz, M.; Elsheikh, A.; Abualigah, L.; Al-Qaness, M. Social Media Toxicity Classification Using Deep Learning: Real-World Application UK Brexit. *Electronics* **2021**, *10*, 1332. [\[CrossRef\]](#)
8. Kastrati, Z.; Ahmedi, L.; Kurti, A.; Kadriu, F.; Murtezaj, D.; Gashi, F. A Deep Learning Sentiment Analyser for Social Media Comments in Low-Resource Languages. *Electronics* **2021**, *10*, 1133. [\[CrossRef\]](#)

9. Kallam, S.; Patan, R.; Ramana, T.; Gandomi, A. Linear Weighted Regression and Energy-Aware Greedy Scheduling for Heterogeneous Big Data. *Electronics* **2021**, *10*, 554. [[CrossRef](#)]
10. Abualigah, L.; Gandomi, A.H.; Elaziz, M.A.; Al Hamad, H.; Omari, M.; Alshinwan, M.; Khasawneh, A.M. Advances in Meta-Heuristic Optimization Algorithms in Big Data Text Clustering. *Electronics* **2021**, *10*, 101. [[CrossRef](#)]
11. Cañas, E.; Estévez, E. Intervention Programs for the Problematic Use of the Internet and Technological Devices: A Systematic Review. *Electronics* **2021**, *10*, 2923. [[CrossRef](#)]
12. Gul, F.; Mir, I.; Abualigah, L.; Sumari, P.; Forestiero, A. A Consolidated Review of Path Planning and Optimization Techniques: Technical Perspectives and Future Directions. *Electronics* **2021**, *10*, 2250. [[CrossRef](#)]
13. Zhou, J.; Gandomi, A.; Chen, F.; Holzinger, A. Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics. *Electronics* **2021**, *10*, 593. [[CrossRef](#)]
14. Abu-Ulbeh, W.; Altalhi, M.; Abualigah, L.; Almazroi, A.; Sumari, P.; Gandomi, A. Cyberstalking Victimization Model Using Criminological Theory: A Systematic Literature Review, Taxonomies, Applications, Tools, and Validations. *Electronics* **2021**, *10*, 1670. [[CrossRef](#)]