A Survey on Deep Learning in Big Data and its Applications

Zair Bouzidi
LIMPAF Laboratory, Computer Science Dept,
Science & Applied Science Faculty,
Bouira University, Bouira, Algeria
zair.bouzidi@gmail.com

Mourad Amad
LIMPAF Laboratory, Computer Science Dept,
Science & Applied Science Faculty,
Bouira University, Bouira, Algeria
zair.bouzidi@gmail.com

Abdelmalek Boudries
Laboratory LMA, Commercial Science Dept, Faculty of Economics, Business & Management, Bejaia University, Bejaia, Algeria
abdelmalekboudries@gmail.com

Abstract

Individuals can exchange real-time information thanks to the vast spread and reach of social networks. This active participation with the corporate data, as emails, documents, databases, business processor history, etc and content published on the Web, as age and contact details, reviews, comments, photos, images, videos, sounds, texts, famous cookies, or e-commerce transactions, exchanges on social networks, are very important. Data recovery from different sources can be a difficult task. A timely and correct assessment of an event currently under discussion is critical to the effectiveness of the used method. This information, collected in the Web can then be updated. Various ways are developed to automate this necessity, due to the extraction and analysis of correct social media content. Alleviation methods do not adequately incorporate these approaches. It may be necessary to reveal them in order to make further progress, particularly in the areas of energy efficiency and cleaner production.

Keywords: Big data, Clean energy, Cloud computing, Deep learning, Energy efficiency, Smart Data

I. INTRODUCTION

Social media has been increasingly pervasive in our daily lives in recent years. They have emerged as a viable resource for disseminating, detecting, tracking, and extracting information in order to better manage events [1,2]. Because of its ability to deliver a message to a potentially wide audience, social networking is used to acquire and transmit accurate and timely information. However, the massive amount of information provided during times of crisis event can make collecting relevant and actionable information even more challenging.

Recent events have used social media to allow the afflicted public to immediately submit a vast amount of event information, allowing managers to make speedly accurate and judicious decisions. The comparison of all Big Data models is shown in Table I. Using social media to convey timely information during catastrophes has become commonplace in recent years [2]. Major events can result in millions of messages being shared on social media. Researchers are paying more attention to the relation between social media, awareness and resilience [1,3,4]. Numerous studies [3,4] for the automated recognition of various types of social media material have been conducted in recent years.

With the rise of social media, a current event is being debated across all platforms. In order to acquire a complete picture of the event, it's critical to collect data from various sources. The information gathered from various sources has often qualitative variances. When it comes to recovering information published on social media, acquiring reasonable, great, and outstanding situational awareness, and good damage assessment, managers have an enormous problem. On the web, there are various sources of information where a current topic can be debated. Several automated tools [1] have been developed to assist event management in identifying and filtering useful material that has been posted online. Only a few studies have concentrated on warning [1,2], education [5], situational awareness, and assessment, with the great bulk of research focusing on leveraging social networking as a source of information for only a few phases of event management.

Databases, log files, online web applications, and social media networks are just some of the places where data can be found. It can be tough to work with a range of massive data sources and formats, as well as data stored locally. Data ingestion is the process of bringing raw data from several silo databases or files into a data lake on a data processing platform like Hadoop. A data lake is a storage repository for vast amounts of unstructured data in its original format, with data structure and standards defined only after the data is used. As a result, data lakes have a flat architecture and have schema-on-read, unlike data warehouses, which store data in a highly structured repository and employ a relational or dimensional data model. A data warehouse function called schema-on-write determines the data structure before it is stored. As a result, data lakes are more versatile, as data may be quickly updated and rebuilt to fit various research models. For importing various types of data into Hadoop, a number of data ingestion technologies are now available.

II. BIG DATA

TABLE I. COMPARATIVE TABLE OF ALL MODELS USING BIG DATA

Models	***Identification Usage of Big Data***
[6]	Making Sense of Big Data through Human Computing and Machine Learning
[7]	Decision-Making and emerging Big Data
[8]	Social Networking Data Quality in Big Data
[9]	Big Data privacy in public Social Media

IEECP'21, July 29-30, 2021, Silicon Valley, San Francisco, CA – USA

DOI : https://dx.doi.org/10.6084/m9.figshare.14737953

Big data refers to all the digital data produced by the use of new technologies for personal or professional purposes [10]. This includes corporate data (emails, documents, databases, business processor history, etc.) as well as data from sensors, content published on the Web (age and contact details, reviews, comments, photos, images, videos, sounds, texts, famous cookies), e-commerce transactions, exchanges on social networks, data transmitted by connected objects (electronic labels, smart meters, smartphones, etc.), geolocated data, etc [6,9,10]. Big data makes it possible to meet a huge technological challenge : to store a large amount of data from different channels on a huge hard drive, easily accessible from all corners of the planet. Data stored in a safe place and recoverable at any time in the event of any incident [10]. Massive data duplication is one of the keystones of the big data architecture. Cloud computing and distributed file systems (DFS) are among the main storage models currently available. These data provide very interesting clues about consumer behavior and market trends. Big data makes it possible to know your profile, but also your overall behavior: frequency of use of social networks and of your online purchases, channels used, hours of connection, etc. Table II shows the comparison of all tehnics and methods used in various OSN models.

TABLE II. COMPARATIVE TABLE OF ALL TECHNIQUES AND METHODS (SOCIAL NETWORKING) USED IN VARIOUS MODELS.

Ref	*Identification Methods*	*Used OSN*
[11]	Flood Disaster Game-based Learning	Twitter
[12]	Educational Purposes with Special Reference at the Faculty of Higher Education	Twitter
[13]	Summarization with social-temporal context	Twitter
[14]	Building a Tweet Summarization Dataset with a TREC Track	Twitter
[15]	Semi-automated artificial intelligence-based classifier for Disaster Response	Twitter
[16]	Summarizing situational tweets in crisis scenarios: An extractive-abstractive approach	
[17]	Based on Artificial NN (ANN)	Twitter & Facebook
[18]	Based on Artificial NN (ANN)	All the Web
[5]	Based on FeedForward NN (FFNN)	All the Web

Big data refers to all the digital data produced by the use of new technologies for personal or professional purposes [10,19]. Table III shows the advantages and limitations of Big Data.

TABLE III. BIG DATA ADVANTAGES AND LIMITATIONS

Advantages[1]	*Limitations*
Cost Savings	[20]
Time Reductions [21]	**Need for talented** [21]
Better sales insights	**Incompatible tools** [22]
Control online reputation	**Correlation Errors**
Understand the market conditions	**Costs**
Increased productivity	**Security and Privacy Concerns**
Fraud detection	
Clean energy with Construction of Global Energy Interconnection (GEI) [23]	

[1] https://www.vapulus.com/en/advantages-and-disadvantages-of-big-data/#:~:text=Big%20data%20is%20a%20term,a%20higher%20false%20discovery%20rate.

Contents were collected from all online channels tracked automatically by the Online Listening Tool [24,25] from websites to social media, such as Twitter, Facebook, LinkedIn, Instagram, Google+, Youtube and so on. Actually, many networking platforms allow access to their data via Application Programming Interface (API) [26].

Big data has opened up new possibilities in a variety of areas, including science, politics, communication, medicine, meteorology, ecology, economics, commerce and energy, to name a few. Researchers, companies and administrations can carry out trend or predictive analysis, draw up profiles, anticipate risks and monitor phenomena in real time, thanks to analytical tools and data modeling.

Big data makes it possible to meet a huge technological challenge, in storing, in a safe place and recoverable at any time in the event of any incident, a large amount of data from different channels on a huge hard drive, easily accessible from all corners of the planet. Big data is an essential tool for BtoB and BtoC companies. The data collected helps them to design personalized marketing campaigns adapted to the needs, preferences and behavior of consumers. This information helps improve the customer experience, attract prospects and retain existing customers. Improved targeting makes marketing campaigns more effective and reaches the desired segment, the one that is most likely to be interested in the company's products and / or services. Big data is also a competitive advantage for professionals who hold a multitude of data, because they can anticipate changes in behavior and better understand why consumers have turned to a particular service provider.

Big data is a valuable tool in vast private and public fields ranging from online sales to scientific research, including culture, politics, transport, insurance, the banking sector, industry and energy. Big data primarily designates masses of data that are too large, complex and heterogeneous or which change too quickly to be able to be analyzed and used correctly and quickly, with current methods of data processing. The smart data goes beyond this simple notion of large quantity data : it designates the useful and quality information obtained from masses of heterogeneous data, and for which we have not only taken into account the technical mastery of the mass of data but also of the quality, security and protection of data and its use. Knowledge is thus generated from the data. The use of this data is based on technologies that make complex processing possible and thus generate knowledge, a source of added value, and which will become the foundation of the new data economy. That's the reason for which big data must be transformed into smart data [27-30].

Big Data, peuvant être structurée, non structurée ou semi-structurée, permet de capturer, stocker, distribuer, gérer et analyser de vastes ensembles de données avec différentes structures à haute vitesse. Les données sont générées à partir de diverses sources et peuvent arriver dans le système à des rythmes différents. C'est grâce au parallélisme qui permet de traiter ces grandes quantités de données et grâce à Hadoop qui permet de structurer Big Data et le rendre utile à des fins d'analyse. Hadoop, étant un logiciel open source permettant le traitement distribué de grands ensembles de données, il est conçu pour passer d'un seul serveur à plusieurs machines, avec une grande tolérance aux pannes [31].

A. Smart Data

We produce information every day, in all situations of our life: whether at our workplace, at the doctor, behind the wheel of our car or at the baker, via our smartphone, our smartwatch or our computer. It is textual data, images or sensors. If we consider the fact that this data can be analyzed, processed and used, then we can really speak of it as raw material. Because these masses of data, this big data, contain immense potential. Streaming is a common Internet method of broadcasting and reading streaming content. It differs from a file download, which necessitates recovering all of a file's data. Streaming content, on the other hand, necessitates the use of an internet connection. Smart Data, a different concept of Big Data, is based primarily on real-time data analysis. This is a method of data analysis that analyzes data directly at the source rather than sending it to a centralized system for analysis. Big and smart data are transversal technologies, which will prevail in almost all areas of our life and will change them for the long term. Traditionally, data is first collected, converted, placed in a database, and processed in waves. However, with this approach, data is usually out of date when it is finally analyzed. Table IV shows Smart Data Areas.

TABLE IV. SMART DATA APPLICATION AREAS.

No	Smart Data application areas	Smart Data application application
1	Financial services	Fraud detection and prevention
2	Retail	Allowing brands to: - Analyzing the sentiment of their customers, - and Offering personalized and contextual promotions
3	Telecommunications industry	Possibility of - Better allocating bandwidth based on real-time needs, and - diagnosing the condition of antennas
4	Manufacturing	Preventive maintenance
5	Healthcare	- Monitoring patient vital signs and - Reducing readmission rates
6	Oil industry	- Proactive repairing infrastructures, and - Balancing the power delivered according to consumption
7	Public sector	- Detecting and preventing intrusion attempts on the network, and - Predicting the risk of epidemics
8	Transport sector	Possibility of detecting risky conduits

While the term Smart Data refers to an approach to data analysis that involves analyzing data directly at the source to enable immediate decisions to be made, without the need to send it to a centralized system. Processing time is reduced to just seconds. The term "*Smart*" refers to the intelligence of this concept, based on the fact that less data is processed, because statistical models are responsible for determining which are the relevant variables, namely the most correlated. For example, an autonomous car cannot wait for the data to be sent to the cloud and the results to be returned to the user. In this case, the data must be gathered directly by sensors so that the processors in the car can analyze it and the results are sent to the actuators that control the brakes and steering wheel of the vehicle. If the data is not analyzed immediately, the consequences can be tragic. Smart Data always combine the analysis of this data with human intelligence to provide you with comprehensive, synthesized and useful information for your business. The objective of Smart Data is first of all to remove uninteresting data for marketing actions, thus making a first cut in the mass of available data. The second step is to check the remaining information to distinguish which is still valid and which is no longer. This action is based on comparisons with data offered. It also involves the identification of data useful for a particular campaign. Finally, It can also be equipped with geographical criteria for sorting information, in order to highlight data that is only relevant in a specific catchment area. The presence and listening to social networks are also essential, because the E-reputation on social networks and the Web in the customer process are inseparable. Unlike Big Data, Smart Data makes it possible to control and qualify data. It is used for the establishment of a coherent marketing strategy. Smart Data can therefore generate leads and increase sales while building customer loyalty by providing them with optimal experience and satisfaction.

III. DEEP LEARNING

In recent years, Deep learning methodologies have achieved impressive results in computer vision [32], speech recognition, image processing, Disaster Management [19] and handwritten recognition of characters, while they is currently in its infancy in fault diagnosis [33]. Even if training appears to have been effective, Generative Adversarial Networks (GANs) with finite size discriminators and generators can struggle with distribution learning. Table V shows the Comparison of Machine Learning Technics.

TABLE V. COMPARING MACHINE LEARNING TECHNICS.

Algo-rithm	Lear-ning Type	Based-Class	Limitation	Advantage
K-Nearest Neigh-bor	Superv ised	Instance	Suufering from dimensiona lity curse	Distance-based problems
Naïve Baye-sian Classifi-cation	Superv ised	Probabi-listic	Imputs Independen cies	Positive Probability for all Classes
Hidden Markov	Superv ised	Marko-vian	With Markov Assump-tion	Time Series Data & Memory-less information
SVM	Superv ised	Deci-sion Boun-dary	Definite Distinction between Classifica-tions	Binary Classification
NN	Superv ised	Non Linear Func-tional Approxi-mation	Little Bias	Binary Inputs
Cluster-ing	Unsu-per-vised	Cluste-ring	/	Data Grouping with Distance: Euclidean, Manhatan
Ridge Regres-sion	Super-vised	Regres-sion	Low Limitation	Continus Variables
Filtering	Unsu-per-vised	Feature Transfor mation	/	Data/many variables to use Filtering

Restricted Boltzmann computer is a classifier, regression, subject modeling, collaborative filtering, and feature learning algorithm. RBM is concerned with the fundamental unit of composition, which has evolved into a variety of common architectures and is commonly used in a variety of large-scale industries. Neuroimaging, Sparse image restoration in mine planning, and Radar target recognition are all applications of the Restricted Boltzmann machine.

By using uncorrected label data and its reconstruction errors, RBM overcomes the issue of noisy labels. The unstructured data problem is solved by a feature extractor, which converts raw data into hidden units. The main drawback is that RBMs are difficult to train since the most common algorithm, Contrastive Divergence, necessitates precision sampling from a Monte Carlo Markov Chain. In addition, the energy of the model has a complicated partition function, making estimating the log likelihood difficult

In an autoencoder, on the other hand, cross entropy may at least be tracked. Figure 1 shows the Classification of Deep Learning.

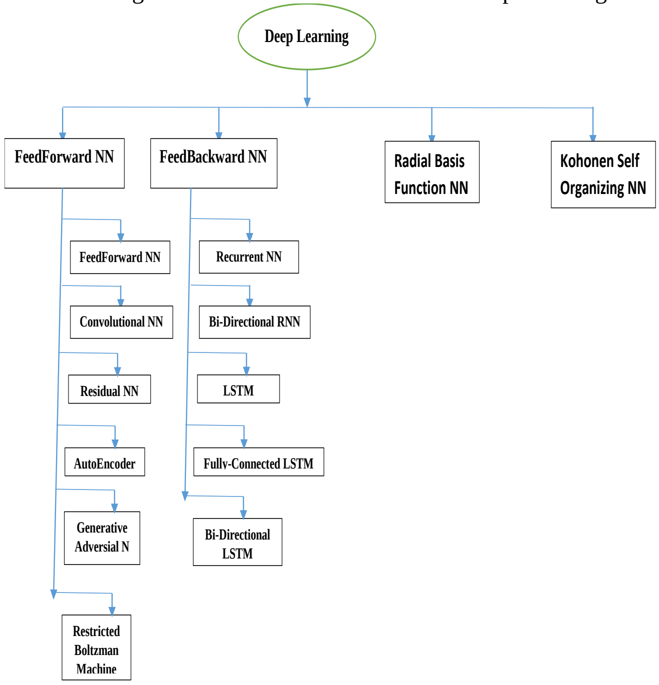

Fig. 1. Classification of Deep Learning

When a deeper feedforward NN is created, it can be giving the model the ability to capture more complex representations, as for Image recognition tasks (convolutional NNs), Natural language processing, Bio-informatics tasks.

If it is tried to create a deep NN to model more simpler phenomena, it is running the risk of over-fitting the data, losing the ability able to generalize to new examples and factoring the amount of resources takes to train a deep neural net.

Deep LSTM (DLSTM) and CLSTM are two LSTM variants that are important. The DLSTM model has a different number of layers than the LSTM model in general. The extraction of well-defined temporal data would be impossible with a single-layer LSTM. By piling more layers, the LSTM model, on the other hand, would be able to achieve better temporal features, making it better suited to capturing motion in the time dimension. To ensure that spatial features are preserved, CLSTM runs the data through convolutional layers first. The LSTM extracts temporal attributes from the Convolutionan NN (CNN) output, allowing the model to capture both spatial and temporal motion. These two variants can also be combined to create a convolutional deep LSTM, in which the outputs of a CNN are fed into a multilayer stacked LSTM, which guarantees a better result at the expense of increased computational complexity [34].

In comparison to RNN, the BRNN can be trained without the restriction of only using input data up to a predetermined future frame. As an example of an unsupervised task, gaps filling in high-dimensional time series with complex dynamics, where unidirectional RNNs have recently been trained successfully to model such time series, but inference in the negative time direction is non-trivial [35].

In a number of speech recognition tasks, both Convolutional NNs (CNNs) and Long Short-Term Memory (LSTM) have outperformed Deep NNs (DNNs). CNNs, LSTMs, and DNNs have complementary modeling capabilities, with CNNs excelling at reducing frequency variations, LSTMs excelling at temporal modeling, and DNNs excelling at mapping features to a more separable space. Combining CNNs, LSTMs, and DNNs into a single unified architecture [36] is an interesting opportunity to carry advantage of their complementarity.

CNNs present some limitations of temporal modeling [37]. The traditional models of Bi-directional Long Short-Term Memory (BLTSM) have encountered some limitations in presentation with multi-level features, but they can keep track of temporal information while allowing deep data representations [37]. Sentiment polarity classification (SPC) on social data is achieved using a Deep Bi-directional Long Short-Term Memory (DBLSTM) architecture with multi-level function presentation. By using DBLSTM, It can exploit more level features than BLTSM, while inheriting its temporal modeling (BLTSM) [37].

The benefits of radial basis function (RBF) networks include ease of design, good generalization, high input noise tolerance, and the ability to learn online.

It's possible that the popularity of Kohonen Self-Organizing NN stems from the fact that it has an easy-to-understand algorithm, is straightforward to use, and generates nice, intuitive results.

Abiodun et al. (2018) [38] recommend that future research can focus on combining various ANN models into a single network-wide application, as needed and depending on the characteristics of the various ANN models.

Neural learning is carried out by Feedforward or Feedback neural network. In Feedforward, we have *supervised learning* such as Feedforward NN itself for classification [5], convolutional NN [39-41] for image recognition/classification or Residual NN (ResNets) [42] for image recognition, *unsupervised learning* such as Autoencoder [36] for Dimensionality reduction and encoding, Generative Adversarial Network [42] network for generating realistic fake data, reconstruction of 3D models or image improvement and *supervised or unsupervised learning* such as Restricted Boltzmann machine [36] for dimensionality reduction, feature learning, topic modeling, classification, collaborative filtering or many body quantum mechanics (See Fig. 1).

In Feedbackward, we have only *supervised learning* such as Recurrent NN [19,34] for sequences recognition as precise timing, Bidirectional Recurrent NN [43] for natural language processing, Long Short-Term Memory [36] for temporal data such as stock market values over a period of time, video frames, Fully Connected-LSTM [44] for learning non-linear and complex processes in hydrological or meteorological modeling and Bi-Directional-LSTM [45] through time-natural language processing and language translation. Neural learning can be trained in either supervised/unsupervised ways by Radial Basic Function Network [46] for M-means clustering, Least square function, function approximation and time series prediction or unsupervised ways by Kohonen Self Organizing Netowork [46] for dimensionality reduction, optimization problems or clustering analysis.

IV. DISCUSSION ABOUT APPLICATIONS OF DEEP LEARNING AND BIG DATA

A. Big Data Application Areas

Smart Data has a wide range of applications (See Table IV). The judicious use of data may help almost every aspect of society and the economy. In banking and marketing, big data technologies are already widely used.

Edge computing and smart data devices are being tested by data scientists, marketers, and manufacturers to create more money, improve decision-making, and detect equipment flaws before they become costly.

When it comes to energy applications, Smart Data can be valuable in a variety of industries. Thanks to big data technologies in supporting Construction of Global Energy Interconnection (GEI) development, clean energy with GEI and its technology innovation direction uses Big Data for sharing, fusion, and comprehensive application of energy related data all over the world [23].

B. Deep Learning Application Areas

The potential use of industrial data via using Deep Learning algorithms for energy efficiency and greener production goals has been explored in the scientific literature [47].

As the benefits of Artificial Intelligence (AI) and Data Science become more apparent in the energy market, companies are developing an increasing number of apps and solutions that use AI to assist their clients in improving their energy efficiency. Deep Learning analyzes energy usage data to discover consumption trends, calculate the amount of energy utilized by a specific household appliance, and provide individualized recommendations on how to save energy and save money. Artificial intelligence detects equipment failure in the energy business, particularly on coal-fired power plants, potentially causing injuries or even death to workers, by using Deep Learning, sensors, and operational data to forecast when vital infrastructure will break, allowing for prompt maintenance and preventing accidents in power plants and power stations. Deep Learning processes data from smart meters, drones, and sensors installed on energy assets, as well as weather data, to anticipate energy demand, system load, outages, and the amount of renewable energy supplied by solar panels and wind turbines. It aids global power producing and distribution businesses in anticipating outages and boosting power flow through their enormous power networks.

AI-powered real-time management and analysis platforms connected with on-site equipment, as well as creating intelligent adaptive controllers linking with motor and other moving parts of lift systems to deliver both real-time control and maximize production, were developed to decrease expenses for sending people to remote locations for oil exploration and production. Energy businesses might use drone data to conduct virtual inspections of solar and wind installations (both offshore and onshore) and to develop AI-powered software that analyzes system drone imagery and generates actionable reports to track and manage solar asset health.

Deep learning models trained on an increasing number of datasets, satellite imaging, and climate physics can help on both fronts (precise weather and demand forecasts) and help the world achieve carbon neutrality by 2050 (climate neutrality with accurate supply-demand matching)[2].

Researchers employ a mixed architecture consisting of a convolutional neural network (CNN) integrated with an artificial neural network (ANN) to perform French energy demand predictions using weather data [48]. Our entire energy system's efficiency must increase. Blockchain, artificial intelligence, machine learning algorithms, and data software, all of which are critical, have a lot of potential in terms of minimizing the intermittency of some renewable energy sources[3].

V. CONCLUSION AND FUTURE WORKS

The study included a wide range of topics, including disaster management improvements made possible by Big Data and, in particular, Deep Learning. Deep Learning enables the use of Big Data to a variety of fields. Humanitarian operations and crisis management are evolving radically as a result of Big Data and Deep Learning. This paper retains more energy area. This study examines Big Data, Cloud Computing and Deep Learning, finding all of their distinguishing features and classifying Deep Learning architectures. It concludes with a description of all the advancements in the energy sector made possible by Big Data, Cloud Computing and Deep Learning.

Based on the examined literature, certain areas of improvement for additional research and future research development in Big Data, particularly in Deep Learning, can be given to professionals, researchers, and novice researchers.

VI. ACKNOWLEDGMENTS

We acknowledge support of "Direction Générale de la Recherche Scientifique et du Développement Technologique (DGRSDT')". MESRS, Algeria.

VII. REFERENCES

[1] Imran, M., Castillo, C., Lucas, J., Meier, P. & Vieweg, S., (2014), AIDR : Artificial intelligence for disaster response, *Proceedings of the 23rd International Conference on World Wide Web, (ICT-DM), 159-162.* https://doi.org/10.1145/2567948.2577034

[2] Olteanu, A., Vieweg, S., & Castillo, C. (2015), "What to Expect When the Unexpected Happens", Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15. doi:10.1145/2675133.2675242

[3] Toppel, M., Bartels, M., Nagel, C. and Hahne, M., (2016), "A Social Network to Identify Responsibilities and Expertises in Crisis Scenarios", 3rd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), 2016

[4] Vulic, I. and Moens, M.-F., (2015), "Monolingual and Cross-Lingual Information Retrieval Models Based on (Bilingual) Word Embeddings", SIGIR'15 Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Pages 363-372, Santiago, Chile — August 09 - 13, 2015

[5] Bouzidi, Z., Boudries, A. & Amad, M. (2020). Towards a Smart Interface-based Automated Learning Environment Through Social Media for Disaster Management and Smart Disaster Education. *Advances in Intelligent Systems and Computing. SAI 2020. Vol 1228. Springer, Cham, 443-468.* https://doi.org/10.1007/978-3-030-52249-0_31

[6] Ofli, F. and Meier, P. and Imran, M. and Castillo, C. and Tuia, D. and Rey, N. and Briant, J. and Millet, P. and Reinhard, F. and Parkan, M. and Joost, S., (2016), Combining Human Computing and Machine Learning to Make Sense of Big (Aerial) Data for Disaster Response, Big Data, vol. 4, No. 1

[7] Horita, Flavio E.A. and Albuquerque, Joao Porto (de) and Marchezini, Victor and Mendiondo, Eduardo M., (2017), Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil, Decision Support Systems, vol. 97, pp. 2-22, doi:10.1016/j.dss.2017.03.001

[8] Immonen, A. and Paakkonen, P. and Ovaska, E., (2015), Evaluating the Quality of Social Media Data in Big Data Architecture, IEEE Access, vol. 3, pp. 2028-2043, doi:10.1109/ACCESS.2015.2490723

[9] Smith, M., Henne, B., Szongott, C. and Voigt, G. von, (2012), Big data privacy issues in public social media, 6th IEEE International Conference on Digital Ecosystems Technologies, pp. 1-6, Doi:10.1109/DEST.2012.6227909

[10] Saranya, M. and Prema, A., (2017), Survey On Big Data Analytcis using Hadoop ETL, International Journal of Computer Trends and Technology (IJCTT), vol. 48, No. 1, doi:10.14445/22312803/IJCTT-V48P105

[11] Zaini, N.A., Noor, S.F.M. & Zailani, S.Z.M. (2020). Design and Development of Flood Disaster Game-based Learning based on Learning Domain. In *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(4), pp. 679-685, DOI:10.35940/ijeat.C6216.049420

[12] Vivakaran, M. V. & Neelamalar, M. (2018). Utilization of Social Media Platforms for Educational Purposes among the Faculty of Higher Education with Special Reference to Tamil Nadu. In *Higher Education for the Future*, 5(1), pp. 4-19, DOI:10.1177/2347631117738638

[13] He, R., Liu, Y., Yu, G., Tang, J., Hu, Q. & Dang, J. (2016). Twitter summarization with social-temporal context. In *World Wide Web*, 20(2), pp. 267-290, DOI:10.1007/s11280-016-0386-0

[14] Dussart, A., Pinel-Sauvagnat, K. & Hubert, G. (2020). Capitalizing on a TREC Track to Build a Tweet Summarization Dataset. In *Text REtrieval Conference*, (TREC'2020)

[2] https://techmonitor.ai/leadership/sustainability/ai-renewable-energy

[3] https://www.electrifying.world/machine-learning-wind-power/

[15] Lamsal, R. & Kumar, T. V. V. (2020). Classifying Emergency Tweets for Disaster Response. In *International Journal of Disaster Response and Emergency Management (IJDREM)*, 3(1), pp. 14-29, DOI:10.4018/IJDREM.2020010102

[16] Rudra, K., Goyal, P., Ganguly, N., Imran, M. & Mitra, P. (2019). Summarizing situational tweets in crisis scenarios : An extractive-abstractive approach. In *IEEE Transactions on Computational Social Systems*, 6(5), pp. 981-993, DOI:10.1109/tcss.2019.2937899

[17] Bouzidi Z., Boudries A. and Amad M., (2018), A New Efficient Alert Model for Disaster Management, Proceedings of Conference AIAP'2018 : Artificial Intelligence and Its Applications, El-Oued, Algeria,

[18] Bouzidi, Z., Amad, M. and Boudries, A., (2019), Intelligent and Real-time Alert Model for Disaster Management based on Information retrieval from Multiple Sources, International Journal of Advanced Media and Communication}, Vol. 7, No. 4, pp. 309-330, doi:10.1145/253260.253325

[19] Bouzidi, Z., Boudries, A. & Amad, M. (2021). Enhancing Crisis Management because of Deep Learning, Big Data and Parallel Computing Environment: Survey. Proceedings of the 3rd International Conference on Electrical, Communication and Computer Engineering (ICECCE), No. 443, 12-13 June 2021, Kuala Lumpur, Malaysia, Accepted

[20] Lefever, S., Dal, M. and Matthiasdottir, A., (2007), Online data collection in academic research : advantages and limitations, Journal British Journal of Educational Technology (BJET) of British Educational Research Association (BERA), vol. 38, No. 4, pp. 574-582, doi:10.1111/j.1467-8535.2006.00638.x

[21] Karma, S., Zorba, E., Pallis, G.C., Statheropoulos, G., Balta, I., Mikedi, K., Vamvakari, J., Pappa, A., Chalaris, M., Xanthopoulos, G. and Statheropoulos, M., (2015), Use of unmanned vehicles in search and rescue operations in forest fires: Advantages and limitations observed in a field trial, International Journal of Disaster Risk Reduction, vol. 13, pp. 307-312, doi:https://doi.org/10.1016/j.ijdrr.2015.07.009

[22] Bello, O. M. and Aina, Y. A., (2014), Satellite Remote Sensing as a Tool in Disaster Management and Sustainable Development : Towards a Synergistic Approach, Procedia - Social and Behavioral Sciences, vol. 120, pp. 365-373, 3rd International Geography Symposium, GEOMED2013, 10-13 June 2013, Antalya, Turkey, doi:10.1016/j.sbspro.2014.02.114

[23] Zhang, D. & Qiu, R. C., (2018), Research on big data applications in Global Energy Interconnection, Global Energy Interconnection, vol 1, No. 3, pp. 352-357, ISSN 2096-5117, doi:10.14171/j.2096-5117.gei.2018.03.006.

[24] Ruggiero, A. and Vos, M., (2014), Social Media Monitoring for Crisis Communication : Process, Methods and Trends in the Scientific Literature, In Online Journal of Communication and Media Technologies, Vol. 4, No. 1

[25] Young, S. D., Rivers, C. and Lewis, B., (2014), Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes, Preventive Medicine, vol. 63, pp. 112-115

[26] Imran, M., Ofli, F., Caragea, D. & Torralba, A., (2020), Using AI and Social Media Multimodal Content for Disaster Response and Management : Opportunities, Challenges, and Future Directions. *Information Processing & Management, 57(5), 1-9.* http://sci-hub.tw/10.1016/j.ipm.2020.102261

[27] Stillger, M., Lohman, G. M., Markl, V. and Kandil, M., (2001), LEO-DB2's learning optimizer, Very Large DataBases (VLDB), vol. 1, pp. 19-28

[28] Battre, D., Ewen, S., Hueske, F., Kao, O., Markl, V. and Warneke, D., (2010), Nephele/PACTs: a programming model and execution framework for web-scale analytical processing, Proceedings of the 1st ACM symposium on Cloud computing, pp. 119-130

[29] Alexandrov, A., Bergmann, R., Ewen, S., Freytag, J.-C., Hueske, F., Heise, A., Kao, O., Leich, M., Leser, U., Markl, V., Naumann, F., Peters, M., Rheinländer, A., Sax, M. J., Schelter, S., Hoger, M., Tzoumas, K. and Warneke, D., (2014), The stratosphere platform for big data analytics, The VLDB Journal, vol. 23, No. 6, pp. 939-964, Springer Berlin Heidelberg

[30] Carbone, P., Katsifodimos, A., Ewen, S., Markl, V., Haridi, S. and Tzoumas, K., (2015), Apache flink : Stream and batch processing in a single engine, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, vol. 36, No. 4, IEEE Computer Society

[31] Harshawardhan, S. B. and Devendra, P. G, (2014), A REVIEW PAPER ON BIG DATA AND HADOOP, International Journal of Scientific and Research Publications, vol. 4, No. 10, pp. 756-764

[32] Krizhevsky, A. and Sutskever, I. and Hinton, G., (2012), ImageNet classification with deep convolutional neural networks, Proceedings of Advances in Neural Information Processing Systems, vol. 25, pp. 1090-1098

[33] Bouzidi, Z. Boudries, A. & Amad, M., Deep Learning and Social Media for Managing Disaster: Survey, IntelliSys 2021 Conference, Amsterdam, Accepted

[34] Roshan, S., Srivathsan, G., Deepak, K. and Chandrakala, S., (2020), Violence Detection in Automated Video Surveillance : Recent Trends and Comparative Studies, pp. 157-171, doi:10.1016/B978-0-12-816385-6.00011-8

[35] Berglund, M., Raiko, T., Honkala, M., Karkkainen, L., Vetek, A. and Karhunen, J., (2015), Bidirectional Recurrent Neural Networks as Generative Models, MIT Press, Cambridge, MA, USA

[36] Sainath, T., Vinyals, O., Senior, A. and Sak, H., (2015), Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks, pp. 4580-4584, doi:10.1109/ICASSP.2015.7178838

[37] Khuong, N., Cuong L. and Hong, P., (2016), Deep Bi-directional Long Short-Term Memory Neural Networks for Sentiment Analysis of Social Data, pp. 255-268, doi:10.1007/978-3-319-49046-5_22

[38] Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A. & Arshad, H. (2018). State-of-the-art in artificial neural network applications : A survey. In *Heliyon*, 4(11), DOI:10.1016/j.heliyon.2018.e00938

[39] Alam, Firoj and Imran, Muhammad and Ofli, Ferda, (2017), Image4Act : Online Social Media Image Processing for Disaster Response, Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, (ASONAM 17), pp. 601-604, doi:10.1145/3110025.3110164

[40] Nguyen, D. T., Al-Mannai, K., Joty, S. R., Sajjad, H., Imran, M. and Mitra, P., (2017b), Robust classification of crisis-related data on social networks using convolutional neural networks, ICWSM, pp. 632-635

[41] Kabir, Md Yasin and Madria, Sanjay Kumar, (2019), A Deep Learning Approach for Tweet Classification and Rescue Scheduling for Effective Disaster Management, Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, (SIGSPATIAL'19), pp. 269-278, doi:10.1145/3347146.3359097

[42] He, R., Liu, Y., Yu, G., Tang, J., Hu, Q. & Dang, J. (2016). Twitter summarization with social-temporal context. In *World Wide Web*, 20(2), pp. 267-290, DOI:10.1007/s11280-016-0386-0

[43] Canon, M. J., Satuito, A., Sy, C., (2018), Determining Disaster Risk Management Priorities through a Neural Network-Based Text Classifier, 2018 International Symposium on Computer, Consumer and Control (IS3C), Taichung, Taiwan, 2018, pp. 237-241, doi:10.1109/IS3C.2018.00067

[44] Zhao, J., Deng, F., Cai, Y. and Chen, J., (2018), Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction, Chemosphere, vol. 220, doi:10.1016/j.chemosphere.2018.12.128

[45] Kabir, Md Yasin and Madria, Sanjay Kumar, (2019), A Deep Learning Approach for Tweet Classification and Rescue Scheduling for Effective Disaster Management, Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, (SIGSPATIAL'19), pp. 269-278, doi:10.1145/3347146.3359097

[46] Pouyanfar, S. and Tao, Y. and Tian, H. and Chen, S.-C. and Shyu, M.-L., (2018), Multimodal deep learning based on multiple correspondence analysis for disaster management, World Wide Web, vol. 22, pp. 1893-1911, doi:10.1007/s11280-018-0636-4

[47] Narciso, D. A. C. & Martins, F.G., (2020), Application of machine learning tools for energy efficiency in industry: A review, Energy Reports, vol. 6, pp. 1181-1199, ISSN 2352-4847, doi:10.1016/j.egyr.2020.04.035.

[48] Real, A. J. d., Dorado F. & Durán, J., (2020), Energy Demand Forecasting Using Deep Learning: Applications for the French Grid, Energies, 13, 2242; doi:10.3390/en13092242 www.mdpi.com/journal/energies