World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

Check for updates

Big data and machine learning and the impact on education

Panagiotis Leliopoulos [1], Sotiris Stavridis [2] and Athanasios Drigas [1,*]

[1] *Net Media Lab IIT NCSR Demokritos, Athens, Greece.*
[2] *Aristotelian University of Thessaloniki, Thessaloniki, Greece.*

World Journal of Advanced Research and Reviews, 2023, 18(03), 670–683

Publication history: Received on 24 April 2023; revised on 11 June 2023; accepted on 21 June 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.18.3.1054

Abstract

In this paper we mention deep learning. Deep learning can be used to improve computing service provider revenue in the mobile blockchain network or to use a machine to study standard mobile detection tasks (e.g., activity recognition). Using deep neural networks can compare results with learning techniques in more common use. Also, we see the deep learning for massive amounts of data and deep learning for a wide variety of data and high-velocity data. Beyond all applications Big Data and Machine learning can be used to improve education, and its practices and procedures.

Keywords: Big Data; Machine Learning; Mobiles; AI; Robots; Education

1. Introduction

Big Data will likely transform not only research but also schooling in the future. The use of data to influence instruction was one of the top five practices linked to substantial academic effects, according to a recent accurate significant similarity of various approaches used by 35 charter schools in NYC. The Cloud Computing is another form of collaborative technology that is based on Big Data. By providing access to low-cost content, online instructors, and communities of other learners, these technologies can enhance educational offerings for both young and adult learners. Big Data may also assist the traditional educational system by assisting teachers in analysing student knowledge and the most effective teaching methods for each student. Teachers can discover new techniques and approaches for their work in education in this way. As a result, technologies like data analytics and mining may give teachers and students quick feedback on their academic achievement. These techniques can offer a thorough investigation of some educational patterns and draw important conclusions from them. In order to prevent failure or dropout, big data collected collectively can identify which students will benefit most from additional support from the educational system. Finding pedagogical methods that seem to work best with particular students and learning styles is therefore necessary. The opposite is also true: "Big Data can easily find apply at online education." As we can see, the impact of online education on the education sector has grown significantly in recent years. Additionally, digital learning is a collection of data and analytics that can support teaching and learning. In this approach, a large number of students engage in online or mobile learning, where new data are generated. The students from various backgrounds are able to correlate one another and better comprehend the fundamental principles covered in the course thanks to these fresh facts, which is also made possible via social media. The conclusion is that the impact of big data will be extensive in every societal aspect and especially on education.

2. Machine Learning

Today's Big Data applications use potent machine learning methods like the SVM and deep learning machine to provide both predictable and comprehensive requirements [1]. Big Data has the potential to revolutionize how we work, live,

* Corresponding author: Athanasios Drigas

and think by streamlining processes, boosting information retrieval, and enhancing decision-making. The achievement of this enormous potential depends on the information that data analysis can glean from such massive amounts of data. The ability of machine learning to learn from data and deliver knowledge, decisions, and predictions based on data is at the foundation of what it does [2].

The development of multivariate statistics, pattern recognition, data mining, and advanced data analysis or forecasting is reflected in machine learning. It is especially important when deep and reliable learning is needed to extract hidden knowledge from huge, diverse, and dynamic data sets. In general, critical measures for assessing machine learning techniques are accuracy, scale, and speed. [1].

The collection of data that precisely fits the memory is one of several assumptions that previous approaches to machine learning are built on because they were created at a different time. Unfortunately, these assumptions no longer hold true in this new environment. Traditional methods are hindered by these occasional situations and Big Data properties. As a result, the issues with machine learning and Big Data are compiled, summarized, and organized in this article [2]. Sequence classification, which is applicable to traffic analysis and user behavior classification, is another crucial Machine Learning technique [1].

Machine learning's main goal is to establish a relationship between input data and output actions so that invisible data input patterns can be processed automatically. Sorting (identifying the type of input data) and regression analysis (adjusting the data), two common techniques in machine learning that can be applied to wireless communications, are among the many relevant techniques. They can be used for level prediction, sorting, or altering as well as mobile environment identification. Have the trajectory length, mobile location, and channel waiting times distributions as well (regression) [3].

2.1. Machine Learning Framework On Big Data

Figure 1 depicts the framework for Big Data ML (MLBiD). MLBiD focuses on the machine learning (ML) component, which interacts with the Big Data, user, sector, and system components. By contributing their expertise in the field, preferences, and usability feedback, users may engage with ML and use the learning outcomes to make better decisions. The system architecture affects the performance and efficiency of learning algorithms, and at the same time, addressing ML requirements may result in the joint design of system architecture [4].

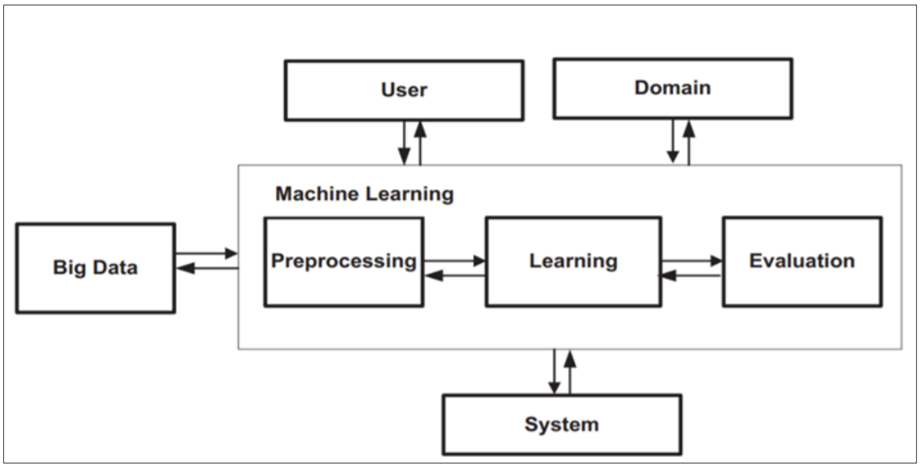

Figure 1 A framework of machine learning on Big Data (MLBiD) [4].

ML usually goes through pre-processing, learning, and data evaluation phases. Pre-processing data helps to prepare primary data in the "correct form" for the next learning steps. The pre-processing step converts such data into a format used as input to learning through data purification, export, conversion, and fusion. The learning phase chooses to learn algorithms and coordinates the parameters of the model to create the desired outputs using the pre-processed input data [4].

Some learning methods, especially representative learning, can also be used to pre-process data. The evaluation results can lead to the adjustment of the parameters of the selected learning algorithms and/or to the selection of different algorithms. ML can be characterized in many dimensions. The nature of learning comments, the purpose of learning tasks, and the availability of data [4].

3. Machine Learning Categories

There are two main kinds of learning depending on the type of data available: supervised learning, where the system learns to map inputs to results when both the inputs and the predicted outputs (tags) are known, and unsupervised learning, where they are unknown [2]. A third sort of learning that exists is reward-based learning. By giving instances of input-output pairings, supervised learning aims to teach a learning system an operation that maps inputs to outputs. Instead than offering direct feedback or a desired outcome, unsupervised learning aims to find patterns in the data [4]. Despite the availability of numerous feature selection approaches, the volume of data and the dimensioning of prospective features within the sequences make it difficult to successfully categorize feature traits in a large data set [1].

As with reinforcement learning, an enhancement learning system is not presented with input-output pairs. Like supervised learning, reinforcement learning is when a machine or agent interacts with its environment, performs actions, and learns through a trial and error method [5]. In contrast to supervised learning, feedback on learning enhancement is a reward or punishment associated with actions instead of the desired performance or explicit correction of optimal actions [4].

Semi-supervised learning falls between supervised and unsupervised learning, where the system is presented with a small number of input-output pairs and a large number of unexplained inputs. With the exception of learning from both commented and uncommented data, the semi-supervised learning goal is comparable to supervising [4].

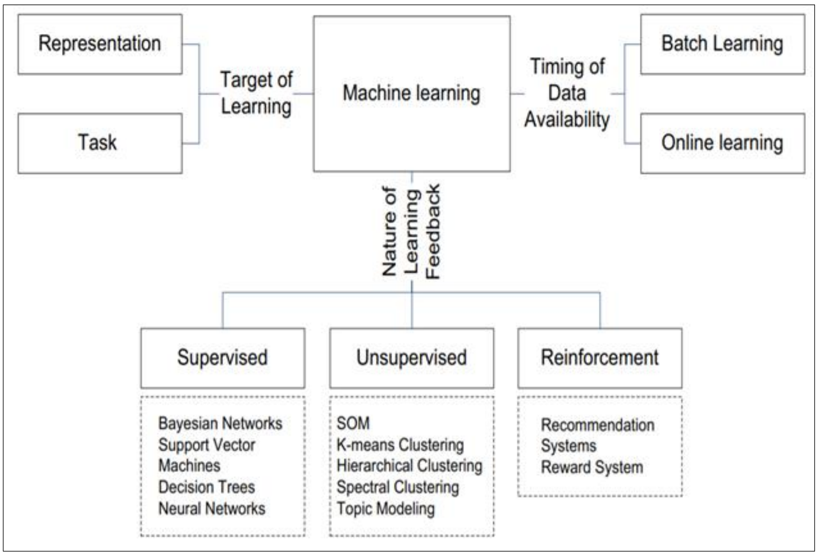

Figure 2 A multi-dimensional taxonomy of machine learning [4]

ML can be divided into representational learning and task learning depending on whether the learning target is specific tasks that use input features or the characteristics themselves. The goal of learning representation is to acquire new data structures that make it easier to extract relevant data for the development of predictions or classifiers. Estimating density and reducing dimensions frequently involve learning to represent. Setting a clear purpose or objective when learning to express oneself is frequently challenging. The categorization, regression, and grouping of learning to work, in contrast, typically produce the desired outcoIn classification, ML techniques produce a model that assigns invisible inputs to one or more predefined categories. Grouping creates data groups, and these groups are not known in advance, which is distinguished by the classification. Usually, classification and regression have been referred to as supervised learning and clustering as unsupervised learning (Figure 2) [4].

The ML can be divided into mass learning and online learning depending on the time available for the training data (for example, whether the training data are all available simultaneously or one at a time). While online learning informs models based on each new entry, mass learning informs models by studying the entirety of the educational data. The assumption made by a mass learning algorithm—which is typically not satisfied by actual data—is that the data are independent, dispersed simultaneously, or derived from the same probability distribution. Online education typically does not produce data-based statistics [4].

Although a generalization concept for online learning does not exist because the learning algorithm is supposed to accurately predict the labels of the samples it receives as input, a mass learning method is intended to be generalized. When training the full dataset is computationally impractical, the data is created over time, or both, and an educational system needs to be adjusted to new patterns in the data, online learning is used. There are various dimensions that can be assigned to each ML algorithm. Traditional decision trees, for instance, are supervised batch learning algorithms [4].

4. Machine Learning and Predictive Analytics

Regression and classification are examples of supervised learning. In classification, the results are given distinct values (class labels), but the outputs in regression are continuous. Support Vector Regression (SVR), linear regression, and polynomial regression are examples of regression techniques, whereas K-nearest neighbor, accounting regression, and Support Vector Machine (SVM) are examples of classification algorithms [2]. The SVM technique is useful in detecting characteristic sequences since it attempts to classify a given series into one category or another in a feature space and recognizes the underlying level of greatest margin to differentiate two types. Given the two sequences aand b, a similarity function s (a, b) can be used as a sorting kernel function [1]. Some methods, like as neural networks, can be applied to both regression and classification [2].

The placement and behavior of the kernel, such as the k-spectrum core, string core, polynomial type core, or the nuclei produced from potential models, must be understood. Using supervised and/or unobserved approaches for automatically learning hierarchical (or multi-level) representations in deep archives, deep learning emerges as an alternative to conventional learning techniques by examining architectural lessons with shallow structured learning [1].

Machine learning is the foundation of predictive analytics, which uses models created using historical data to forecast a variety of future outcomes. Algorithms like SVR, neural networks, and Nave Bayes can all be used for this. A typical ML assumption is that algorithms can learn more effectively and produce more accurate results with more data. Massive datasets, however, present a number of difficulties because traditional algorithms were not created to handle such demands. For instance, different ML algorithms were created for smaller datasets, assuming the dataset could fit entirely in memory. A significant issue of Big Data is to violate these assumptions, making traditional algorithms unusable or greatly hindering their performance [2].

5. Machine Learning Challenges On Big Data

Below we can see some descriptions according to the challenges of machine learning and how each challenge is a specific dimension of Big Data [2].

5.1. Processing Performance

The straightforward idea that scale or volume enhances computational complexity is one of the key obstacles faced in Big Data calculations. For instance, the SVM algorithm, where m is the number of training samples, has O (m3) training time complexity and O (m2) space complexity. Because of this, an increase in m size will significantly effect the time and memory needed to train the SVM algorithm, and in large data sets it may even be computationally unfeasible. [2].

Other ML methods also exhibit long-term complexity. For instance, the primary component analysis has a time complexity of O (mn2 + n 3), accounting regression has a time complexity of O (mn2 + n 3), and locally weighted linear regression has a time complexity of O (mn2). Because of this, all of these methods' calculation times would get exponentially longer as the size of the data increases, potentially rendering them worthless for very large datasets. Additionally, algorithm performance is more affected by the architecture used to store and transmit data as data sizes increase. Parallel data structures, data distribution, and data placement and reuse become increasingly important as data size increases. Resistant distributed datasets (RDDs) are a new emerging technique for memory computations in large groups. As a result, the amount of the data not only affects performance but also necessitates reconsidering the standard architecture for developing and implementing algorithms [2].

5.2. Curse of Modularity

Many learning algorithms are built to assume that the data being processed can be stored entirely in memory or a single file on a disk. However, when data size leads to the lack of success of this assumption, the whole categories of algorithms are affected. One of the approaches to this issue is Map-Reduce, an extensible programming example for processing large data sets through parallel execution in a large number of nodes [2].

5.3. Category Report

As data sets expand, the presumption that data is dispersed equally across all classes is frequently broken. This presents the issue of the class disparity. When datasets include class data with varying probabilities of occurrence, a machine learning algorithm's performance may suffer. Big Data is not the only source of order imbalance, which has been the focus of research for more than a decade. There is a strong possibility that the classes will be adequately represented by a large enough sample size, according to researchers [2].

The difficulty of Big Data's task, on the other hand, is anticipated to be considerable, which could have serious consequences for class inequality. Given that the scale of the data has expanded, there is a high likelihood that this problem will become more prevalent, serious, and difficult as a result of big data. Therefore, the Big Data framework's unaltered execution without taking into account class imbalance can produce subpar outcomes. As a result, the likelihood of class imbalance in the setting of big data is considerable due to the magnitude of the data. Therefore, the potential implications of class imbalance on machine learning are substantial because of the complex difficulties buried in such data [2].

5.4. Dimensions Procedure

Another issue brought on by the sheer amount of Big Data is what is known as the "curse of dimensionality," which describes the difficulties encountered when working in a high-dimensional domain. The quantity of features or qualities included in the dataset is particularly referred to by the dimension. Unfortunately, because there are more likely prospects, the likelihood of a high dimension rises as more data are employed to characterize an event. Therefore, as Big Data volume increases, the possibility of high dimensions increases as well. Furthermore, the extent has an effect on processing effectiveness. Therefore, it is evident that the time and space complexity of machine learning algorithms is directly impacted by the dimensionality of the data [2].

5.5. Feature Engineering

The high dimension is intimately related to machine characteristics, another volume difficulty. It gets harder to develop new, highly relevant features as the dataset expands both vertically and horizontally. Therefore, similar to dimension, as the dataset size increases, are the difficulties associated with the machine properties. The technical characteristics are related to the selection of features. While the machine characteristics create new features to improve learning outcomes, selecting features (reducing dimensions) aims to select the most relevant features. However, the selection of features minimizes the dimension and perhaps shortens the ML processing time. Due to spurious correlations and subsequent endo-genesis (correlation of an explanatory variable with the error term), it is difficult in high dimensions. Overall, the selection of features and the infrastructure are still crucial when it comes to Big Data, but they are also getting more complicated [2].

5.6. Non Linearity

The vastness of the data makes it difficult to apply established approaches for assessing the dataset's properties and the algorithm's performance. Despite the fact that this issue is not specific to big data, non-linearity is more pronounced in huge datasets. Big Data's lack of linearity is partly a result of the challenges in measuring linearity. However, in the case of Big Data, numerous points can combine to form a significant cloud, making it challenging to identify relationships and judge linearity. Due to the difficulty of evaluating both linearity and non-linearity, machine learning techniques create challenges when applied to Big Data [2].

5.7. Bonferonni's Principle

According to Bonferonni's principle, there is a high likelihood of discovering a particular sort of event if it is searched for within a given data set. Therefore, the majority of the time, these instances are phony, meaning they have no justification and don't fit any dataset. The Bonferonni adjustment theorem in statistics offers a way to steer clear of these dataset false-positive searches. But as data sizes grow, these events become more common. As a result, it is anticipated that there will be an increase in the likelihood of discovering an occurrence of interest, whether or not it is lawful. Researchers have recently talked about the incidence and effects of phony associations on Big Data. Therefore, the inclusion of a means of preventing these false positives is critical to consider in machine learning with Big Data [2].

5.8. Variance And Bias

Generalization is the foundation of machine learning. Representations can be generalized to enable analysis and prediction through observations and data manipulation. Variation and bias are the two parts of the generalization error. Prejudice refers to an algorithm's capacity to learn by error, whereas variation refers to the outcome of an algorithm's capacity to forecast random events (Figure 3). The learner algorithm may get very narrowly biased with the entire set as the volume of data rises and may not generalize well enough for new data. We must therefore exercise caution when working with big data because it can introduce prejudice and put generalization at risk. The term "normalization" describes methods for enhancing generalization and lowering overload. Techniques for normalization include Ridge, Lasso, and premature cessation, for instance. Despite the fact that these strategies increase generalization, they also include new parameters that must be coordinated in order to produce a usable application for data that does not show up. However, especially when dealing with huge datasets, these take more time to process. Although legalization methods are well-established in machine learning, more study is required to determine how well they work with Big Data [2].

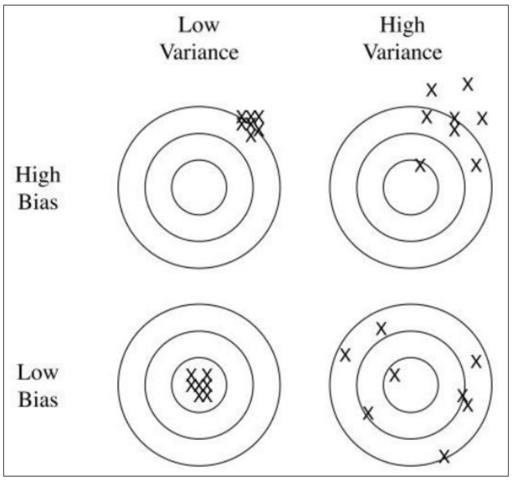

Figure 3 Variance and bias [2].

5.9. Data Locality

The location of the data is a problem with many Big Data. The complete dataset is once more presumed to be in memory or a disk file by machine learning algorithms. When transferring enormous datasets, there could be significant processing delays and increased network traffic. As a result, a method for computing data rather than transferring it for analysis evolved. An example is also used in Map-Reduce. Each map work processes its local data at the nodes where the data is situated, where map work is carried out. The systems employ the Map-Reduce example to compute the data after storing the data across a number of nodes as distributed storage solutions. Physical location is unimportant with tiny datasets. However, with Big Data, the data location is a primary challenge that must be addressed in any successful Big Data system [2].

5.10. Data Heterogeneity

Integrating various data from many sources is a common step in big data analytics. The type, format, data model, and semantics of this information can all be added. The term "editorial heterogeneity" describes the variety of data kinds, file formats, data encoding, data models, and similar elements. To prepare the data for machine learning and make it match a particular model, these steps are frequently necessary. However, when using data from several sources, this data may take on a different form. Similar to syntax, when numerous datasets created from various sources are combined, semantic heterogeneity also rises in big data. As we can see, machine learning techniques have not been developed to handle data with varying semantic nuances, therefore heterogeneity needs to be handled before such techniques can be used [2].

In statistics, the term "heterogeneity" also refers to the statistical differences that exist within a whole dataset. Due to the fact that datasets frequently contain features from several sources, this problem applies to both small and large datasets. Because of this statistical heterogeneity, machine learning's joint hypothesis that all of the data in a set exhibits the same statistical features is invalidated. Long-running research areas include statistical heterogeneity, syntactic and semantic heterogeneity, and they have all recently gained more interest due to the rise of big data. The correlation of several datasets is typically what gives data analysis its economic worth, and it is essential for machine learning to be conducted on such datasets that it be completed [2].

5.11. Noisy Data

The data has its own set of different features that can be used for characterization:

- The situation determines the readiness of the data for analysis.
- The location refers to the place of the data.
- The population describes the entities and sets of standard features that together form the dataset [2].

Big Data is frequently referred to as a terrible condition because it takes time and money to prepare them for analysis. These data are referred to as noisy data by researchers. One of the three basic difficulties in analyzing big data is noisy data. The researchers propose that the processes of data collection, integration, and analysis be followed by a step to separate the signal from the noise. In a similar vein, when using Big Data for machine learning, noise must be considered. Big Data is not the only source of dirty and noisy data, however handling it may not be as simple for huge datasets [2].

Additionally, the data may be wrong. For instance, tags, information, or readings based on situations relevant to the data may be inaccurate. This is different from incorrect data for machine learning. Numerous sources increase uncertainty, usually when utilized for participative detection, but they can also provide noisy data since labels are assigned to the data using the human crisis. In addition to affecting the correctness of the data, the presence of inaccurate or noisy labels may also negatively impact machine learning performance by potentially supplying incorrectly labeled data [2].

5.12. Data Availability

As we can see, many approaches to machine learning depend on data availability, which means that before learning, it was considered that there was a whole set of data. However, in the data flow, where new data is continuously arriving, such a requirement cannot be met. In machine learning, a model typically absorbs knowledge from the entire dataset before applying the knowledge to fresh data to execute the learnt job, such as classification or prediction. In this illustration, the model executes the already updated task in new data rather than automatically learning from newly arriving data. Therefore, algorithms must support gradual learning in order to adapt to new inputs. Sequential learning is the process by which an algorithm can adapt its learning in response to the addition of new data without needing to retrain on an incomplete dataset. Instead of waiting for the entire training set to be ready before learning can begin, this approach evaluates new data as it comes in. recovery [2]Although elementary learning is a relatively old idea, it is still a topic of ongoing research since it is challenging to adapt particular algorithms to continuous data retrieval [2].

5.13. Real-Time Processing And Streaming

Real-time processing is required since traditional machine learning methods were not designed to handle fixed data flows. In light of this, it is slightly different from data availability. While data availability refers to the need to update the ML model as new data is received, real-time processing refers to the demand to process real-time or almost real-time quick data access. However, these algorithms' complexity and the lack of available online learning possibilities are problematic [2].

5.14. Concern Drift

Big Data is not stable. New data is continually arriving. Therefore, it is impossible to obtain the entire data set before processing, it implies that it is impossible to say if the distribution of the past, present, and future data is the same. This takes us to the intriguing issue of idea drift in machine learning using Big Data. While the concept of drift may alter in the conditional distribution of the desired output given the input, the distribution of the information itself may remain constant. Building machine learning models using ancient data that no longer accurately represents the distribution of new data is a problem as a result of this in particular. The energy consumption and demand forecasting models that can be developed using information from electricity meters cannot accurately reflect the changing energy characteristics, for example, if a building is reconstructed to boost its energy efficiency.

One solution to the problem of drift is a sliding window. In order to include only the most recent examples, the model is constructed using only samples from the training window. The cautious approach makes the possibly inaccurate assumption that the most recent data is more pertinent. The difficulties, however, typically lay in immediately spotting when the idea is changing and managing the transition model well during these changes. Concept drift is an old problem, like many of the concepts that have already been mentioned. The arrival and nature of Big Data, however, have increased its prevalence and rendered some older approaches worthless. A study on the impact of extensive Big Data on current displacement reduction strategies is one such example. Scientists conclude that the changes in the data significantly degraded the performance of the algorithm. Therefore, finding new ways to deal with the shift in Big Data is an essential task for the future of machine learning [2].

5.15. Independent And Identically Distributed Random Variables

The idea that random variables are independently and identically distributed (i.i.d.) is another prevalent one in machine learning. Big Data, on the other hand, might make it difficult to believe in i.i.d. Many data sets already have a non-random order, so it requires the data to be in a random order. To randomize the data before applying the algorithms is a stereotypical solution. This, however, is difficult and frequently impractical when it comes to Big Data [2].

Big Data is swift and continuous by nature. As a result, neither waiting for all the data to come in nor randomizing a set of data that is still insufficient are feasible options. Numerous common machine learning techniques, like support carriers and back-multiplication neural networks, have been found to rely on this notion and would substantially benefit from an accounting approach. This task should be taken on because of the Big Data scenario [2].

5.16. Data Provenance

Locating and documenting the source of data, as well as its transfers between locations, is known as data origin. As it identifies all the steps, transactions, and procedures presented by faulty data and provides data for machine learning, the recorded information, or source data, can be utilized to pinpoint the origin of the processing problem. The size of the data source itself also increases in the context of big data, so even while this data offers a great foundation for machine learning, its bulk presents its own set of difficulties [2]..

Therefore, the computational cost of conveying this charge becomes prohibitive in addition to the fact that this dataset is enormous. Although some techniques, such as Reduce and Map Provenance (RAMP), developed for Map-Reduce as an extension of Hadoop, have been created to preserve data provenance for specific data processing scenarios. Its already high complexity and computational cost often rise as a result of the extra source weight. Therefore, it is essential to strike a balance between the general processing costs and the truth value given that the source data offers a means of confirming the accuracy of Big Data [2].

5.17. Deep Learning And Big Data

In recent years, the method of deep learning can automatically identify relevant features that have received particular attention. Deep learning, in principle, uses neural networks to encode any mapping from inputs to outputs. In particular, using a thoughtful downward slope, deep learning manages to find optimal solutions worldwide. For example, deep learning can be used to improve Edge Computing Service Provider (E.C.S.P.) revenue in the mobile blockchain network [6] or to use a machine to study standard mobile detection tasks (e.g., activity recognition) using Deep Neural Networks (D.N.N.s) and to compare results with learning techniques in more common use [7]. Also, in many cases where in-depth learning has been successfully related to conclusions that are important for mobile detection (e.g., emotion recognition, voice recognition) [7].

Typically, deep learning models use a non-supervised pre-training and a supervised optimization strategy to learn hierarchical features and descriptions of Big Data in deep architecture for classification and recognition tasks [8].

Massive attempts have been made to develop practical and scalable parallel methods for deep model training in light of the recent, unprecedented explosion of data in mobile networks. Even while providers frequently gather a large amount of unstructured data, using conventional models makes the task difficult [1]. Deep learning has exploded as a study topic in the machine learning community since it was first discussed in Science magazine. Numerous deep learning techniques have been developed recently [8].

The network of deep beliefs uses pre-training for uncontrolled learning and adapts supervised learning procedures, which in the end results in the development of a learning model [1]. Convolutional neural networks (CNN), stacked automatic encoders (SAE), and deep belief networks (DBN) are among the most prevalent deep learning models., and a repetitive neural network (RNN), which are also the most widely used models. Most of the other deep learning models can be variations of these four architectures in depth. In the following sections, we briefly review the four typical models of deep learning [8].

5.18. Deep Learning Networks

Conventional neural networks usually result in subpar implementations because they are prone to becoming stuck in local optima of a non-convex objective function. It includes supervised optimization and unsupervised pre-training methodologies for creating models. While supervised stages carry out a local search for optimization, uncontrolled stages are designed to learn data distribution without the use of tag information [9].

Figure 3 shows a typical DBN architecture, which consists of a stack of Boltzmann Limited Machines (RBM) and/or one or more additional levels for discrimination tasks. RBMs are probabilistic genetic models that teach a standard probability distribution of observed (training) data without using data labels. They can effectively use large amounts of unstamped data to exploit complex data structures. This is accomplished mostly through the uncontrolled learning of RBMs [9] A typical network of deep belief architecture is made up of a stack of RBMs, a genetic model that might be used to learn a common probability distribution of training data and numerous extra discriminations [1].

An RBM normally has two layers, and the nodes in one layer are completely connected to those in the other layer but not to those in the same layer [1]. As seen in Figure 3, this particular DBN includes three buried layers, each with three neurons. Both the input layer and the output layer each include five neurons. Any two neighboring layers can be combined to form an RBM that has been trained on unlabeled data. The outputs of the current RBM, such as h (1) i in the first RBM indicated in red and h (2) i in the second RBM highlighted in green, are the inputs of the subsequent RBM. With labeled data, the weights W can be modified following pre-training [9].

Each node in the layer is therefore independent of all other nodes, which makes it possible to train the genetic weights of each MMD using Gibbs sampling [1].

Before refining, RBMs are trained pre-layer by pre-layer. Following pre-training of each RBM, the outputs of one RBM are given as inputs to the next RBM. The method also has the advantage that the time complexity scales linearly with the size and number of RBMs. Recall that the number of accumulated RBMs is a parameter that has been predefined by users and that, for excellent generalization, pre-training just needs data without a label [9].

5.19. Convolutional Neural Networks (CNN)

A typical CNN has many levels of hierarchy, some levels for feature representations (or feature maps), and others that function for classification in a manner similar to traditional neural networks. Cohesive layers and sampling layers are two separate sorts of layers that are frequently present at the beginning.. Further sampling is done to minimize the dimensions following a nonlinear modification . The input level, which accepts 2D N N images as. Consists of multiple characteristic maps are constructed with assembled inputs with different filters (weight vectors). As we can see, the value of each unit on a feature map is the result that depends on a local receptive field at the previous level and the filter [9].

5.20. Deep Learning for Massive Amounts Of Data

While deep learning has demonstrated great results in a variety of applications, its training is still important for Big Data learning since the repetitive calculations that are a part of the majority of deep learning algorithms are frequently difficult to parallelize. Due to the massive rise of commercial and research datasets over the past several years, there has been an increase in interest in practical and scalable parallel training strategies for deep models. Contrary to shallow architectures, where few parameters are chosen to avoid problems, deep learning approaches succeed with numerous hidden neurons, frequently resulting to millions of free parameters. Consequently, massive deep learning usually makes use of both enormous models and vast volumes of data. Scientists have proposed the Deep Stacking Network (DSN), a parallelizable deep architecture modification. Tensor Deep Stacking Network (T-DSN), a novel deep architecture built on DSN, has been created to scale parallel computations [9].

5.21. Deep Learning For High Variety Data

Big Data's diversity is its second feature. Today, data is available in a wide variety of formats and likely comes from a variety of sources. The capacity of in-depth learning to learn representations makes it special. To develop usable representations of categorization traits, deep learning can be utilized either by supervised or unsupervised approaches, or by combining both. The challenge of data integration can therefore be solved naturally by first learning data representations from each data source using in-depth learning techniques, and then integrating the gained features at multiple levels. When employing deep learning, it has been found to be quite successful to integrate data from many sources. To learn representations using audio and video data, for instance, researchers have created a new application for deep learning algorithms. Deep learning has been proven to be usually effective in learning to represent distinctive ways through a variety of ways with unknown data, as well as in learning common representations that can capture links between various ways [9].

In order to build a unified representation, researchers created a multimodal Deep Boltzmann Machine (DBM) that integrates real-world dense image data with text data with sparse frequency words. In the multimodal entrance space, it learns a standard distribution that enables learning even with faulty techniques. Many problems still need to be answered, despite recent research showing that deep learning can employ heterogeneous inputs to significantly improve system performance. Deep learning's capacity to learn abstract representations and the diversity of the underlying data make it suited for integrating heterogeneous data in a variety of ways [9].

5.22. Deep Learning For High Velocity Data

High speed has led to new obstacles for Big Data learning. Online learning strategies are one way to learn from such fast data. Every time someone uses online learning, a presence is learned. Additionally, the actual label of each presentation will soon be made available and can be utilized to enhance the model. Given that existing machines are unable to store the whole amount of data in memory, this sequential learning approach works extremely well for big data. While traditional neural networks have been investigated for online learning, deep online learning has only advanced recently. It's interesting to note that deep learning is frequently taught using a downward tilt method, where each time the model's parameters are updated, an example of training with a well-known label is utilized [9].

Instead of using one example, modifications can be done in a mini-batch to expedite learning. Since data is frequently non-stop at high speeds, this presents another issue where data distribution evolves over time. Consequently, learning data as a flow is a key component of a deep learning algorithm for Big Data. Deep online learning is one topic that requires investigation. Online education frequently progresses smoothly, has a memory limit, is simple to parallel, and theoretically secure. Data is a vital element for learning Big Data. Deep learning can also take advantage of both the wide variety and the speed of Big Data through learning transfer or field adjustment, where training and testing data can be taken as a sample of different distributions [9].

Empirical results have shown that deep learning can produce a substantial and high-level representation shared in different areas. Scientists also applied in a depth learning of multi-level representations for learning transfer where the educational examples may not be well represented in test data.

5.23. Benefits of Big Data and Open Data in Education

New types of data, along with schooling becoming more executive and individualized, are helping researchers better comprehend learning. All of these data were acquired via online courses or other technologically-based learning platforms. In this circumstance, their analytics can assist students in learning more efficiently and point them in the direction of outcomes that are more beneficial than those of conventional education [10].

Big data in this case can give children and teenagers more opportunities to learn new things. As a result, students can interact with academic institutions to share information and develop their skills. Universities and other educational institutions may help and prepare their incoming freshmen [10]

Below we can see in summary some benefits of Big Data

Table 1 Benefits of Big Data and Open Data in Education

Improved instruction	Can enhance student performance and learning potential by making sessions more individualized. With the aid of analytics, teachers can modify the curriculum.
Matching students to programs	Parents and children can use open data to pick the best school or educational program.
Matching students to employment	Alternative and more effective tools might be found by businesses and prospective employees to leverage open data to qualify their talents against the necessary skills. Additionally, students can more effectively than before identify and apply for employment that fit their skills.

Transparent education financing	Students now have the opportunity to engage in educational activities that they were previously unable to accomplish. Additionally, you can select anything related to higher education and learn about it

6. Conclusion

Last but not least, we emphasize the significance of all digital technologies in the field of education and in ADHD training, which is very effective and productive and facilitates and improves the assessment, the intervention, and the educational procedures via mobile devices that bring educational activities anywhere [14-17], various ICTs applications that are the main supporters of education [18-24], and AI, STEM, and ROBOTICS that raise educational procedures to new performance levers [25-35] and friendly games [36-38]. Additionally, the improvement and blending of ICTs with theories and models of metacognition, mindfulness, meditation, and emotional intelligence cultivation [39-65], as well as with environmental factors and nutrition [11-13], accelerates and improves even more the educational practices and results, in every subdomain of education.

More specifically, as was demonstrated in the earlier chapters, big data can significantly enhance education. Can afford to create an innovative educational system that will benefit each student as much as possible. Teachers now have access to useful resources that they did not previously have, allowing them to choose from a wide range of innovative new teaching techniques and make more informed decisions.

As a result, Big Data are truly changing how several businesses, including education, operate. The classic problems will vanish in the new era of data while the effective strategies remain. New teaching techniques will improve the educational system and make it more effective and focused.

However, this new period is only in its infancy, and there are still many challenges to be overcome, such as the scarcity of experts in the fields of big data and data analytics. Finally, the students must embrace and use these new tools, and professors and academics must genuinely train them and involve them.

Compliance with ethical standards

Acknowledgments

The Authors would like to thank Net Media Lab Mind-Brain R&D Team for their support.

Disclosure of conflict of interest

The Authors proclaim no conflict of interest.

References

[1] Zheng K., Yang Z., Zhang K., Chatzimisios P., Yang K., and Xiang W., 'Big Data Driven Optimization for Mobile Networks towards 5G', Nov. 2015. doi: 10.13140/RG.2.1.2389.1923.

[2] L'heureux A., Grolinger K., Elyamany H. F., and M. A. Capretz, 'Machine learning with big data: Challenges and approaches', IEEE Access, vol. 5, pp. 7776–7797, 2017.

[3] Bi S., Zhang R., Ding Z., and Cui S., 'Wireless communications in the era of big data', IEEE Commun. Mag., vol. 53, no. 10, pp. 190–199, 2015

[4] Zhou L., Pan S., Wang J., and Vasilakos A. V., 'Machine learning on big data: Opportunities and challenges', Neurocomputing, vol. 237, pp. 350–361, 2017.

[5] 'Supervised vs Unsupervised vs Reinforcement Learning | Intellipaat', Intellipaat Blog, Dec. 26, 2019. https://intellipaat.com/blog/supervised-learning-vs-unsupervised-learning-vs-reinforcement-learning/ (accessed Sep. 06, 2020).

[6] Luong N. C., Xiong Z., Wang P., and D. Niyato, 'Optimal auction for edge computing resource management in mobile blockchain networks: A deep learning approach', in 2018 IEEE International Conference on Communications (ICC), IEEE, 2018, pp. 1–6.

[7] Lane N. D. and Georgiev P., 'Can deep learning revolutionize mobile sensing?', in Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications, 2015, pp. 117–122.

[8] Zhang Q., Yang L. T., Chen Z., and Li P., 'A survey on deep learning for big data', Inf. Fusion, vol. 42, pp. 146–157, 2018.

[9] Chen X.-W. and Lin X., 'Big Data Deep Learning: Challenges and Perspectives', Access IEEE, vol. 2, pp. 514–525, Jan. 2014, doi: 10.1109/ACCESS.2014.2325029.

[10] Podesta J., Pritzker P., Moniz E., Holdren J., and Zients J., "BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES," Executive Office of the President, The White House Washington, Study, May 2014.

[11] Stavridou Th., Driga, A.M., Drigas, A.S., 2021. Blood Markers in Detection of Autism, International Journal of Recent Contributions from Engineering Science & IT (iJES) 9(2):79-86. https://doi.org/10.3991/ijes.v9i2.21283

[12] Zavitsanou, A., & Drigas, A. (2021). Nutrition in mental and physical health. Technium Soc. Sci. J., 23, 67. https://doi.org/10.47577/tssj.v23i1.4126

[13] Driga, A.M., Drigas, A.S. 2019 "Climate Change 101: How Everyday Activities Contribute to the Ever-Growing Issue", International Journal of Recent Contributions from Engineering, Science & IT, (IJES) vol. 7(1), pp. 22-31. https://doi.org/10.3991/ijes.v7i1.10031

[14] Stathopoulou, et all 2018, Mobile assessment procedures for mental health and literacy skills in education. International Journal of Interactive Mobile Technologies, (IJIM)12(3), 21-37, https://doi.org/10.3991/ijim.v12i3.8038

[15] Kokkalia G, AS Drigas, A Economou 2016 Mobile learning for preschool education. International Journal of Interactive Mobile Technologies (IJIM) 10 (4), 57-64 https://doi. org/10.3991/ijim.v10i4.6021

[16] Stathopoulou A, Karabatzaki Z, Tsiros D, Katsantoni S, Drigas A, 2019 Mobile apps the educational solution for autistic students in secondary education Journal of Interactive Mobile Technologies (IJIM) 13 (2), 89-101https://doi.org/10.3991/ijim.v13i02.9896

[17] Drigas A, DE Dede, S Dedes 2020 Mobile and other applications for mental imagery to improve learning disabilities and mental health International Journal of Computer Science Issues (IJCSI) 17 (4), 18-23 DOI:10.5281/zenodo.3987533

[18] Drigas, A. S., Koukianakis, L, Papagerasimou, Y. (2006) "An elearning environment for nontraditional students with sight disabilities.", Frontiers in Education Conference, 36th Annual. IEEE, p. 23-27. https://doi.org/10.1109/FIE.2006.322633

[19] Drigas A, Petrova A 2014 ICTs in speech and language therapy International Journal of Engineering Pedagogy (iJEP) 4 (1), 49-54 https://doi.org/10.3991/ijep.v4i1.3280

[20] Bravou V, Oikonomidou D, Drigas A, 2022 Applications of Virtual Reality for Autism Inclusion. A review Retos 45, 779-785 https://doi.org/10.47197/retos.v45i0.92078

[21] Chaidi I, Drigas A, 2022 "Parents' views Questionnaire for the education of emotions in Autism Spectrum Disorder" in a Greek context and the role of ICTs Technium Social Sciences Journal 33, 73-91 DOI:10.47577/tssj.v33i1.6878

[22] Bravou V, Drigas A, 2019 A contemporary view on online and web tools for students with sensory & learning disabilities iJOE 15(12) 97 https://doi.org/10.3991/ijoe.v15i12.10833

[23] Xanthopoulou M, Kokalia G, Drigas A, 2019, Applications for Children with Autism in Preschool and Primary Education. Int. J. Recent Contributions Eng. Sci. IT (IJES) 7 (2), 4-16 https://doi.org/10.3991/ijes.v7i2.10335

[24] Chaidi I, Drigas A, C Karagiannidis 2021 ICT in special education Technium Soc. Sci. J. 23, 187, https://doi.org/10.47577/tssj.v23i1.4277

[25] Chaidi E, Kefalis C, Papagerasimou Y, Drigas, 2021, Educational robotics in Primary Education. A case in Greece, Research, Society and Development 10 (9), e17110916371-e17110916371 https://doi.org/10.33448/rsd-v10i9.16371

[26] Lytra N, Drigas A 2021 STEAM education-metacognition–Specific Learning Disabilities Scientific Electronic Archives 14 (10) https://doi.org/10.36560/141020211442

[27] Mitsea E, Lytra N, A Akrivopoulou, A Drigas 2020 Metacognition, Mindfulness and Robots for Autism Inclusion. Int. J. Recent Contributions Eng. Sci. IT (IJES) 8 (2), 4-20. https://doi.org/10.3991/ijes.v8i2.14213

[28] Stavridis S, D Papageorgiou, Z Doulgeri 2017 Dynamical system based robotic motion generation with obstacle avoidance, IEEE Robotics and Automation Letters 2 (2), 712-718 DOI:10.1109/LRA.2017.2651172

[29] Kastritsi T, D Papageorgiou, I Sarantopoulos, S Stavridis, Z Doulgeri, 2019 Guaranteed active constraints enforcement on point cloud-approximated regions for surgical applications 2019 International Conference on Robotics and Automation (ICRA), 8346-8352 DOI:10.1109/ICRA.2019.8793953

[30] Stavridis S, Z Doulgeri 2018 Bimanual assembly of two parts with relative motion generation and task related optimization 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems DOI:10.1109/IROS.2018.8593928

[31] Stavridis S, P Falco, Z Doulgeri 2020 Pick-and-place in dynamic environments with a mobile dual-arm robot equipped with distributed distance sensors IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids) DOI: 10.1109/HUMANOIDS47582.2021.9555672

[32] Papageorgiou D, S Stavridis, C Papakonstantinou, Z Doulgeri 2021Task geometry aware assistance for kinesthetic teaching of redundant robots IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 2021, pp. 7285–7291. https://doi.org/10.1109/IROS51168.2021.9636209

[33] Kastritsi T, I Sarantopoulos, S Stavridis, D Papageorgiou, Z Doulgeri Manipulation of a Whole Surgical Tool Within Safe Regions Utilizing Barrier Artificial Potentials Mediterranean Conference on Medical and Biological Engineering and Computing DOI:10.1007/978-3-030-31635-8_193

[34] Stavridis S, D Papageorgiou, L Droukas, Z Doulgeri 2022 Bimanual crop manipulation for human-inspired robotic harvesting https://doi.org/10.48550/arXiv.2209.06074

[35] Stavridis S, Papageorgiou D, Zoe Doulgeri, 2022, Kinesthetic teaching of bi-manual tasks with known relative constraints, Conference: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2022) Kyoto, Japan

[36] Chaidi I, Drigas A 2022 Digital games & special education Technium Social Sciences Journal 34, 214-236 https://doi.org/10.47577/tssj.v34i1.7054

[37] Doulou A, Drigas A 2022 Electronic, VR & Augmented Reality Games for Intervention in ADHD Technium Social Sciences Journal, 28, 159. https://doi.org/10.47577/ tssj.v28i1.5728

[38] Kefalis C, Kontostavlou EZ, Drigas A, 2020 The Effects of Video Games in Memory and Attention. Int. J. Eng. Pedagog. (IJEP) 10 (1), 51-61https://doi.org/10.3991/ijep.v10i1.11290

[39] Drigas, A., & Mitsea, E. (2020). The 8 Pillars of Metacognition. International Journal of Emerging Technologies in Learning (iJET), 15(21), 162-178. https://doi.org/10.3991/ijet. v15i21.14907

[40] Drigas, A. S., and M. Pappas, 2017. "The Consciousness-Intelligence-Knowledge Pyramid: An 8x8 Layer Model," International Journal of Recent Contributions from Engineering, Science & IT (iJES), vol. 5, no.3, pp 14-25, https://doi.org/10.3991/ijes.v5i3.7680

[41] Drigas, A., & Mitsea, E. (2021). 8 Pillars X 8 Layers Model of Metacognition: Educational Strategies, Exercises &Trainings. International Journal of Online & Biomedical Engineering, (IJOE) 17(8). https://doi.org/10.3991/ijoe.v17i08.23563

[42] Drigas, A. S., Karyotaki, M., & Skianis, C. (2018). An Integrated Approach to Neuro-development, Neuroplasticity and Cognitive Improvement. International Journal of Recent Contributions from Engineering, Science & IT (iJES), 6(3), 4-18. . https://doi.org/10.3991/ ijes.v6i3.9034

[43] Drigas A, Mitsea E 2020 A metacognition based 8 pillars mindfulness model and training strategies. International Journal of Recent Contributions from Engineering, Science & IT (IJES) 8(4), 4-17. https://doi.org/10.3991/ijes.v8i4.17419

[44] Papoutsi C, Drigas A, C Skianis 2021 Virtual and augmented reality for developing emotional intelligence skills Int. J. Recent Contrib. Eng. Sci. IT (IJES) 9 (3), 35-53. https://doi.org/10.3991/ijes.v9i3.23939

[45] Kapsi S, Katsantoni S, Drigas A 2020 The Role of Sleep and Impact on Brain and Learning. Int. J. Recent Contributions Eng. Sci. IT (IJES) 8 (3), 59-68. https://doi.org/10.3991/ijes.v8i3.17099

[46] Drigas A, Mitsea E, Skianis C 2021 The Role of Clinical Hypnosis & VR in Special Education International Journal of Recent Contributions from Engineering Science & IT (IJES) 9(4), 4-18. https://doi.org/10.3991/ijes.v9i4.26147

[47] V Galitskaya, A Drigas 2021 The importance of working memory in children with Dyscalculia and Ageometria Scientific Electronic Archives 14 (10) https://doi. org/10.36560/141020211449

[48] Chaidi I, Drigas A 2020 Parents' Involvement in the Education of their Children with Autism: Related Research and its Results International Journal Of Emerging Technologies In Learning (IJET) 15 (14), 194-203. https://doi.org/10.3991/ijet.v15i14.12509

[49] Drigas A, Mitsea E 2021 Neuro-Linguistic Programming & VR via the 8 Pillars of Metacognition X 8 Layers of Consciousness X 8 Intelligences Technium Soc. Sci. J. 26(1), 159–176. https://doi.org/10.47577/tssj.v26i1.5273

[50] Drigas A, Mitsea E 2022 Conscious Breathing: a Powerful Tool for Physical & Neuropsychological Regulation. The role of Mobile Apps Technium Social Sciences Journal 28, 135-158. https://doi.org/10.47577/tssj.v28i1.5922

[51] Drigas A, Mitsea E, C Skianis 2022 Clinical Hypnosis & VR, Subconscious Restructuring-Brain Rewiring & the Entanglement with the 8 Pillars of Metacognition X 8 Layers of Consciousness X 8 Intelligences. International Journal of Online & Biomedical Engineering (IJOE) 18 (1), 78-95. https://doi.org/10.3991/ijoe.v18i01.26859

[52] Drigas A, Karyotaki M 2019 Attention and its Role: Theories and Models. International Journal of Emerging Technologies in Learning 14 (12), 169-182, https://doi.org/10.3991/ijet.v14i12.10185

[53] Drigas A, Karyotaki M 2019 Executive Functioning and Problem Solving: A Bidirectional Relation. International Journal of Engineering Pedagogy (iJEP) 9 (3) https://doi.org/10.3991/ijep.v9i3.10186

[54] Bamicha V, Drigas A 2022 ToM & ASD: The interconnection of Theory of Mind with the social-emotional, cognitive development of children with Autism Spectrum Disorder. The use of ICTs as an alternative form of intervention in ASD Technium Social Sciences Journal 33, 42-72, https://doi.org/10.47577/tssj.v33i1.6845

[55] Drigas A, Mitsea E, C Skianis 2022 Neuro-Linguistic Programming, Positive Psychology & VR in Special Education. Scientific Electronic Archives 15 (1) https://doi.org/10.36560/15120221497

[56] Drigas A, Mitsea E, Skianis C. 2022 Virtual Reality and Metacognition Training Techniques for Learning Disabilities SUSTAINABILITY 14(16), 10170, https://doi.org/10.3390/su141610170

[57] Drigas A,. Sideraki A. 2021 Emotional Intelligence in Autism Technium Soc. Sci. J. 26, 80, https://doi.org/10.47577/tssj.v26i1.5178

[58] Drigas A, Mitsea E, Skianis C.. 2022 Subliminal Training Techniques for Cognitive, Emotional and Behavioural Balance. The role of Emerging Technologies Technium Social Sciences Journal 33, 164-186, https://doi.org/10.47577/tssj.v33i1.6881

[59] Bakola L, Drigas A, 2020 Technological development process of emotional Intelligence as a therapeutic recovery implement in children with ADHD and ASD comorbidity. . International Journal of Online & Biomedical Engineering, 16(3), 75-85, https://doi.org/10.3991/ijoe.v16i03.12877

[60] Drigas A, Bakola L, 2021The 8x8 Layer Model Consciousness-Intelligence-Knowledge Pyramid, and the Platonic Perspectives International Journal of Recent Contributions from Engineering, Science & IT (iJES) 9(2) 57-72, https://doi.org/10.3991/ijes.v9i2.22497

[61] Bamicha V, Drigas A, 2022 The Evolutionary Course of Theory of Mind - Factors that facilitate or inhibit its operation & the role of ICTs Technium Social Sciences Journal 30, 138-158, DOI:10.47577/tssj.v30i1.6220

[62] Karyotaki M, Bakola L, Drigas A, Skianis C, 2022 Women's Leadership via Digital Technology and Entrepreneurship in business and society Technium Social Sciences Journal. 28(1), 246–252. https://doi.org/10.47577/tssj.v28i1.5907

[63] Mitsea E, Drigas A,, Skianis C, 2022 Breathing, Attention & Consciousness in Sync: The role of Breathing Training, Metacognition & Virtual Reality Technium Social Sciences Journal 29, 79-97 https://doi.org/10.47577/tssj.v29i1.6145

[64] Drigas A, Papoutsi C, 2021,Nine Layer Pyramid Model Questionnaire for Emotional Intelligence, International Journal of Online & Biomedical Engineering 17 (7), https://doi.org/10.3991/ijoe.v17i07.22765

[65] Drigas A, Papoutsi C, Skianis, 2021, Metacognitive and Metaemotional Training Strategies through the Nine-layer Pyramid Model of Emotional Intelligence, International Journal of Recent Contributions from Engineering, Science & IT (iJES) 9.4 58-76, https://doi.org/10.3991/ijes.v9i4.26189.