Data Mining Application with Deeplearning

*Note: Sub-titles are not captured in Xplore and should not be used

1st Syed Muhammad Abis Rizvi
ELctronics Engineering.
Hochschule Hamm Lippstadt
Soest, Germany
syed-muhammad-abis.rizvi@stud.hshl.de

***Abstract*—Due to the abundance of data on the Internet, mining is now a very active field of study. Algorithms are used in data mining to categorize data, find relevant trends, and generate forecasts. Data mining is still a difficult task despite several fields' triumphs. Data mining was one of the areas where findings in the past were frequently unsatisfactory. Data mining pulls pertinent information from files in order to identify associations, entities, and to classify data. New approaches were developed as mining methods improved.Deep learning techniques have enabled significant advancements in fields like data mining and natural language processing. A subset of machine learning called deep learning has been applied, among other things, to voice and facial recognition on smartphones. A form of artificial neural network known as a "deep learner" has numerous data processing layers and learns representations by raising the level of abstraction at each layer. These techniques have advanced the state-of-the-art in a variety of fields, including data mining, speech recognition, visual object recognition, natural language processing, genome mining, and therapeutic efficacy prediction. This paper discusses a few deep learning methods that have been applied recently in data mining research.**

***Index Terms*—Introduction, Datamining**

I. Introduction

Data mining (DM), which tries to extract intriguing and potentially helpful information from data, is the central stage of the knowledge discovery process (Goodfellow et al. 2016).[8]Even though large-scale data mining is the focus of this work, many strategies that work well for large-scale datasets can also be effectively used for smaller datasets. Artificial Intelligence and Machine Learning can both be built upon data mining. Several methods in this direction can be categorized under one of the following categories:(a)Artificial which is a technique that helps computer to carry out human actions. Machine learning, natural language processing (NLP), language synthesis, computer vision, robotics, sensor analysis, optimization, and simulation are all examples of artificial intelligence (AI) techniques.(b)A subset of AI techniques called machine learning (ML) enables computer systems to gain knowledge from the past (i.e., data observations) and adapt their behavior to a job at hand. Support Vector Machines (SVM), decision trees, Bayes learning, k-means clustering, association rule learning, regression, neural networks, and many other methods fall under the category of machine learning (ML).(c)Artificial neural networks (NNs), a subset of machine learning techniques, are loosely based on biological neural networks. They are typically defined as a layer-organized collection of interconnected units, or artificial neurons.(d)A subgroup of NNs called Deep Learning (DL) enables the computational multi-layer NN. Deep neural networks (DNNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative adversarial networks (GAN), and many more are examples of common deep learning architectures.

Identify applicable funding agency here. If none, delete this.

II. Data Mining

Data mining (DM) has grown in importance for both public and commercial enterprises as a result of the extraordinary amount of data being collected, stored, and made available online today. It is hardly unexpected that this subject has generated a lot of research. Only the advent of new technologies known as "Big Data" enabled the efficient processing and analysis of these massive data sets. Big Data and DM are frequently used interchangeably. Data mining (DM) is the analytical process of investigating data to find trends and connections between variables in data sets.DM is used, among other things, to evaluate crime patterns [1], anticipate consumer behavior [2], detect fraud [3], provide individualized medical care [4], study seismic activity to find new oil sources [5], and predict material defects [6], to mention a few. Information retrieval, visualization, machine learning, databases, statistics, and artificial intelligence are all included in the multidisciplinary topic of data mining [7].Practically every field has used DM, and fresh trends are always emerging. Access to a wide range of data sources and formats was made possible via the Internet. The interfaces of the various sources frequently are incompatible with one another. Data from these diverse data sources are analyzed using Distributed Data Mining (DDM) techniques that employ extremely complex algorithms. To calculate distances and topology for navigation and Geographic Information Systems, spatial and geographic DM analyzes geographic, environmental, and astronomical data (GIS). Mobile devices are used by Ubiquitous DM to examine how people behave and how people and machines interact. To study seasonal and cyclical trends and determine consumer behavior and purchasing habits, time series and sequential DM are used.Data mining process get completed with help of certaion process such as Data preprocessing which is also known as preparing the data for the primary operation is known as data preparation. In the real world, certain types of data are typically unprocessed, noisy, partial, and inconsistent, making them unsuitable for direct use in data mining.The next is data cleaning, in order to clean up data, data cleaning routines fill in missing values, reduce noise, find and remove outliers, and fix inconsistencies. Data integration procedures combine data from various sources (such as databases, data cubes, flat files, and data warehouses), store the combined data, and give users a unified view of the combined data. In reality, creating a data warehouse is similar to integrating data. Numerous researchers have produced such works. then comes dta transformation in this data can be changed via smooth aggregation, data generalization, and standardization into a specific form that is ideal for mining. For the goal of computational convenience, transformations typically involve replacing latent pattern distribution shapes or relationships. Then comes data reduction technique can be used to obtain a reduced representation of the dataset, which is much smaller but still close to preserving the integrity of the original data and producing the same or nearly the same results as the pre-reduction ones.

A. *classification*

One of the most used techniques in data mining for figuring out how various conditions and the characteristics of distinct objects relate to one another is classification. On the basis of a training set of data comprising observations (or instances) whose category membership is known in advance, classification algorithms can be used to determine which specific component of a new observation belongs to categories (sub-populations). k-nearest neighbor (KNN), Naive Bayes (NB), decision trees (DT), support vector machines (SVM), neural networks (NN), and ensemble learning are a few common classification approaches that we illustrate in this section (EM). **K-nearest neighbor** The unclassified data point is found and assigned to a specific label according to the previously known k nearest neighbor (KNN) points as a type of lazy learning methods based on instances, and a vote mechanism is used to choose the target item belonging. Jen et al. used KNN, linear discriminate analysis (LDA), and sequential forward selection (SFS) as the main body of their classification to investigate the correlations between important variables affecting both healthy individuals and those suffering from chronic illnesses. The chronic class was then identified and classified using an early warning method.[9]**Naïve bayes** The nave Bayesian (NB) is a probabilistic statistical classifier that is built on the Bayes' theorem. The so-called nave or simple assumption is that characteristics or features are conditionally independent, hence reducing computer complexity throughout the numerous operations of probability. David J. Spiegelhalter, et al. used Bayesian probabilistic and statistical concepts to develop expert systems that make probabilistic inferences about specific circumstances, possibly by means of a generic propagation technique.[10]**Decision tree** In a decision support flowchart, classification learning of observations is represented by a decision tree (DT), which connects features (denoted by nodes) to outcomes of targeted class values (represented by leaves) through logical conjunctions of those attributes (calculated by branches). The typical top-down divide-and-conquer strategy using the information entropy of several features is the key methodology underlying a decision tree. In order to alter magnetic resonance imaging (MRI), store large amounts of data on patients, and then utilize fuzzy decision tree (FDT) classifiers to extract features and combine them for decision-making in classification tasks, Estella, Francisco, et al. built and implemented a system.[11]**Support vector machine** The hyper-plane, which is created by implicitly mapping the original input space to a higher-dimensional space in order to maximize the distance between two separated classes, is the central component of the support vector machine (SVM). Typically, user-specified kernel functions such the radial basis function (RBF), sigmoid, Gaussian, polynomial, etc. are used during the mapping step. To fit the complexity of the problem and automatically modify the effective number of parameters, bounds on the generalization performance based on the leave-one-out approach and the VC-dimension are provided.[12]**Neural Network** Neural networks are designed to simulate neurological function system with multiple layers of grouped and interconnected processing nodes known as neurons with adjustable weighted links, neural network (NN) is inspired by the physical neural network in biological field. Its goal is to address the specific classification problem under the overall unity of aggregated neurons. In terms of the information flowing both internally and externally via network during studying steps, the training and learning process is reflected in the adjustment of connection weights and changes in network structure as an adaptive manner[13].**Ensemble** A number of classification models can cooperate and produce better classification results than one alone, which is the rational and illogical premise underlying ensemble (EM), which can be seen as a mix of different methodologies. As a basic example of ensemble learning, random forest (RF) is an ensemble classifier built up of numerous decision trees to be a forest with outputs of major classes from all individual trees. Enriched random forest (ERF) was suggested by Amaratunga, et al. as an extension of improved RF.[14]

B. *Clustering*

Clustering is the process of dividing sample data into clusters and organizing a collection of items so that members of the same group—referred to as a cluster—are more similar (in some way) to one another than to members of other groups [15]. We illustrate some common clustering methods in this section, including hierarchical clustering (HC), partitioning relocation clustering (PRC), density-based clustering (DBC), grid-based clustering (GBC), and model-based clustering (MBC).

Hierarchical clustering

In a hierarchical fashion, hierarchical clustering (HC) groups data objects into small clusters, which are then divided into larger clusters from the button to the up layer, and so on. In order to cluster data with both continuous and categorical qualities, Zhang et al. devised the BIRCH distance measurement, which is particularly well suited for very large datasets [16]. Neighbor-Net, a hierarchical distance-based method that Bryant, David, and Vincent Moulton presented, offers a view of the data for building phylogenetic networks and directing more in-depth investigation. The Neighbor-Joining (NJ) algorithm forms the foundation of the Neighbor-Net.[17]

Partitioning relocation clustering

A number of data segmentations are organized and built using partitioning relocation clustering (PRC), each of which is a subdivision of the original collection and represents a distinct cluster. It attempts to alter the appropriate amount of partitions at a time using an iterative relocation strategy. Chiu, Tom, et al. proposed a clustering algorithm based on the BIRCH framework that performs a pre-clustering step by scanning the entire dataset and storing the dense regions of data records in terms of summary statistics, allowing the algorithm to automatically determine the right number of clusters and a novel approach to assigning cluster membership to noisy data.[18]

Density-based clustering Density-based clustering uses the density of data point distribution within the specified radius of neighbors as the criteria for cluster partitioning rather than distance measurements to create arbitrary shaped clusters (DBC). The clusters were identified by Jiang, Daxin, Jian Pei, and Aidong Zhang using a density-based, hierarchical clustering (DHC) method, which addressed the issue of effectively clustering and enhanced the clustering quality and durability.[19]

Grid-based clustering The space of objects is divided into separately quantized cells by multi-resolution grid structures, and all clustering calculations are carried out using grid-based clustering (GBC). Natalia Maltsev, et al. demonstrated a grid-based high-throughput PUMA2 interactive system that enables users to submit batches of sequence data for automated functional analysis and the creation of metabolic models to contrast and develop analyses of genomic data and metabolic networks in the context of taxonomic and phenotypic information.[20]

Model-based clustering

Model-based clustering's fundamental premise is the replacement of the best fitting model with the presumptive model for each cluster first. The statistical learning method and the neural network learning methodology are used as the two technical underpinnings of MBC algorithms. The weighted representation of belonged memberships is iteratively refined by expectation-maximization (EM) analysis, which is based on statistical modeling, and an object is assigned to it in accordance with cluster mean values. Si et al. used EM to estimate the clustering model parameters on RNA sequential data.[21]

III. Data Mining And Deep learning

Data mining, as the term implies, consists on extracting hidden information from vast amounts of data. According to its definition, the object being mined here is a sizable collection of imperfect, noisy, fuzzily distributed, and random practical application data. The information alludes to implicit, common, prior unknown knowledge that may be valuable and ultimately comprehendable knowledge. Data mining is more application-oriented in the real world of business. Machine learning uses computers, probability theory, statistics, and other information to complete data mining technologies. It is possible to create artificial intelligence by feeding data into computer programs and allowing them pick up new information, but this method of learning will not enable machines to develop awareness. Through training data, machine learning seeks to identify the goal function. The accuracy of machine learning can be impacted by data quality, hence data pretreatment is crucial. Deep learning, a brand-new category and area of machine learning, is primarily driven by the establishment, simulation, and analysis of the neural network of learning. It can be referred to as a particular type of brain that mostly mimics the way the human brain interprets data like pictures, sounds, messages, and signals.

To summarize, Table in figure 1 provides a list of the essential concepts and fundamental elements behind various data mining and deep learning algorithms so that readers can easily grasp them. From a technological standpoint, machine learning is the fundamental method of data mining, so it is only natural to compare machine learning and deep learning in the context of the following aspects:

Technique	The key ideas and basic structures
Data Mining	
Data Preprocessing	
Data Cleaning	Fill in missing values, smooth noise, identify or delete outliers, and resolve inconsistencies
Data Integration	Combine data from multiple sources, store them together and provide a unified view of data
Data Transformation	Replace pattern distribution or relationship with a specific form that is suitable for mining
Data Reduction	A reduced representation of the dataset which is much smaller but still close to preserving the integrity of the original data
Classification	
KNN	The unclassified data point is discovered and assigned to particular label according to the formerly known k nearest neighbor (KNN) points, and vote the determination of targeted object belonging
NB	A probabilistic statistical classifier where the naïve assumption is made that attributes or features are conditionally independent
DT	Split features to outcomes of targeted class values through logical conjunctions of those attributes with traditional top-down divide-and-conquer approach utilizing information entropy
SVM	Construct hyper-plane by implicitly kernel function mapping of the original input space to higher dimensional one in order to make the distance between two separated classes as maximal as possible
NN	Multiple layers of grouped and interconnected processing neurons with adjustable weighted links and nonlinear mapping
EM	A hybrid of various approaches with the hypothesis behind that many classification models are able to work collaboratively and get superior classification outcomes than single one only
Clustering	
HC	Gather data objects into tiny clusters via average-linkage, where those smaller clusters are categorized into larger clusters from button to up layer hierarchically
PRC	Organize and construct several segmentations of data, and each partition is a subgroup of original collection of data for every individual cluster measured by distance
DBC	The density of data points distribution within the given radius of neighbors is used for arbitrary shaped clustering and it is the criteria of cluster partitioning
GBC	Multiresolution grid structures of the space of objects are formed into independently quantized cells, wherein all calculations of clustering steps are implemented
MBC	The supposed model for each cluster is given firstly and replace the best fitting model for previous one
Deep Learning	
DNN	A hierarchy form consists of deep layers of input, multi hidden and output of a NN, processing input with feedforward and optimizing with backpropagation
CNN	Contain multiple convolution layers to capture feature maps from regional connectivity via the weighted filter, and polling layers to reduce data size, at last a fully connected NN is added as the classifier
SAE	A well-constructed model representation with the built encoder and decoder. The encoder transforms the input vector to a hidden layer and the decoder maps the hidden layer representation to the reconstructed input, which is regarded as the output result.
DBN	It is composed of multiple restricted Boltzmann machines (RBMs) as each single RBM contains a visual layer and a hidden layer, a two way direction link between two layers, where the visual layer simultaneously works as input and output multiplexing
RNN	Recurrent and cyclic connection unit is designed for handling sequential information therefore causing old data remembered implicitly in state vectors, hidden unit relies on the computation of current input and previous values stored in those state vectors, output is affected by both past and current inputs consequently

Fig. 1. Summary of the fundamentals of different data mining and deep learning algorithms[22]

The hardware environment is another notable distinction between the two learning methodologies. For typical machine learning methods, modest hardware setup on end computers is sufficient. On the other hand, because GPUs are capable of supporting deep learning quite well, it heavily depends on high performance end devices. Since deep learning tasks involve numerous and significant inherent mathematical calculations, such as matrix multiplication and algorithm optimization, the fundamental capabilities of GPUs can only just keep up with their demands.

IV. Methodology

Many AI-related problems, such as visual object or pattern identification, speech perception, and language understanding, can be solved by creating intelligent systems that are capable of extracting high-level representations from high-dimensional data.[23]

Many learning algorithms employ shallow structures, including kernel logistic regression, support vector machines (SVM), neural networks with only one hidden layer, and many others. From high-dimensional input like image or video files, they frequently fail to extract meaningful patterns. The most important objects in an image have been selected through previous work in automatic image caption synthesis. For instance, removing Eigenfaces from images of a person is a common technique for human face recognition.[24] A one-dimensional rendering of a human face is called an Eigenface. The Eigenfaces are used for face recognition on fresh, undiscovered images and are kept in a database.

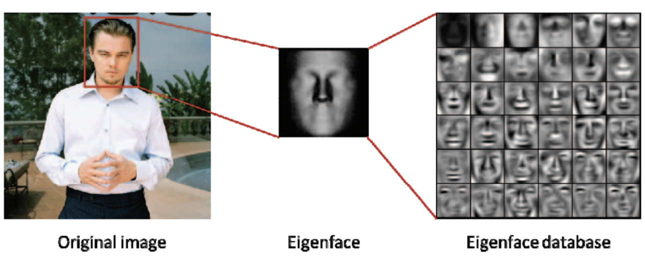

Fig. 2. Eigenface generation [25]

The disadvantage of this approach is that background information that may be used to create captions with more depth and description is lost. Recent years have seen the development of deep learning architectures that can extract high-level representations for a variety of domains, including speech perception, natural language processing, and visual object recognition. Deep Boltzmann machines, deep auto-encoders, deep belief networks, and sparse coding-based techniques are a few examples of deep learning schemes. This paper focuses on the artificial neural network subfield of Deep Learning (DL).

A. Shallow Artifical neural Network

Nature is an inspiration for aNNs. The astonishing simplicity with which humans can answer many categorization issues must be due to the brain's highly interconnected neurons, which enable the division of a problem into smaller, neuron-level subproblems [16]. The development of aNNs was motivated by this observation. Perceptrons, the neurons, and their weighted connections, the axons, make up a typical aNN. ANNs have a rich literature history. To compare shallow aNNs with deep aNNs, they are briefly described here. A perceptron's fundamental principle is to identify a linear function f: f(x)=x$^T X + b$

so that for one class, f(x) greater than 0 while for the other class, f(x) lesss than 0. During training, the bias b and the weights w = (w1, w2,..., wm) are changed until a loss function converges. The component parts of the feature vector are weighted together by the perceptron. The perceptron "fires," or classifies the input as falling within a particular category, if it is higher than a predetermined threshold. The result of this activation function, known as the step function, is either 0 or 1. A loss function, also known as an error function, is minimized throughout each training step until it conforms, or until the least is obtained.

A perceptron can only divide a space into two half-spaces that are separated from one another linearly. The term for this is binary classification. To represent nonlinear decision boundaries, perceptrons can be arranged into a hierarchical framework to create multilayer perceptrons, a sort of aNN. They are applicable to multiclass classification issues. An aNN uses a differentiable sigmoid function like the hyperbolic tangent function or radial basis function (RBF) in place of the perceptron's step function, which only outputs 0 or 1. However, there are additional activation mechanisms available. Due to the absence of cycles, multilayer perceptrons are a type of feed-forward network. Without any concurrence, the signal is transferred from the input layer to the output layer. For learning the weights of a neural network, optimization techniques like gradient descent and stochastic backpropagation are used. The weights of aNNs are modified during training using a reasonably straightforward method called backpropagation. Based on how each unit contributed to the outcome, backpropagation modifies the weights. It determines the gradient of the loss function in relation to each and every weight. This method is known as stochastic backpropagation since the overall error does not always go down with each repetition. The loss function also needs to be differentiable, just like the activation function does. Although the squared-error loss function is frequently employed, alternative loss functions, including negative log-likelihood, can also be used. Gradient descent is a mathematical optimization technique for minimizing the loss function. For learning the weights of a neural network, optimization techniques like gradient descent and stochastic backpropagation are used. The weights of aNNs are modified during training using a reasonably straightforward method called backpropagation. Based on how each unit contributed to the outcome, backpropagation modifies the weights. It determines the gradient of the loss function in relation to each and every weight. This method is known as stochastic backpropagation since the overall error does not always go down with each repetition. The loss function also needs to be differentiable, just like the activation function does. Although the squared-error loss function is frequently employed, alternative loss functions, including negative log-likelihood, can also be used. Gradient descent is a mathematical optimization technique for minimizing the loss function.

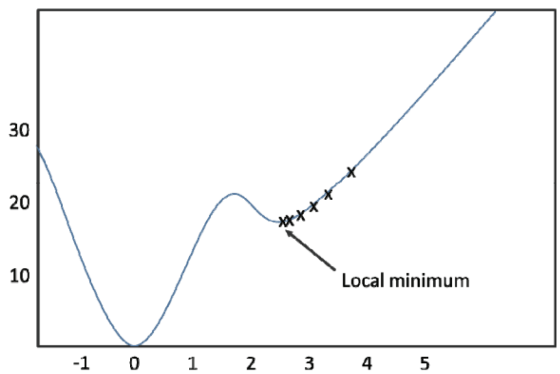

Fig. 3. Eigenface generation [25]

If there are multiple minima, gradient descent can only locate a local minimum, which is a significant drawback. Because they do not have this issue, SVM have historically been preferred over aNNs. When the objective function is differentiable, gradient descent can be used as a general-purpose optimization strategy [26]. A straightforward multi-layer perceptron with a hidden layer and sigmoid activation functions is shown in Fig. 3.

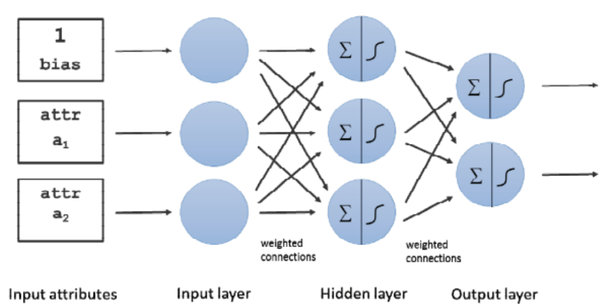

Fig. 4. Eigenface generation [25]

One of the most popular methods of training is supervised learning, in which the aNN is trained with labelled input data, such as pictures of a person's face from various angles and at various ages. Finding a function that correctly classifies a set of input data into its correct output—for example, correctly classifying a facial image as belonged to a certain person—is the aim. The trained network is then tested to see if it can accurately identify a human in brand-new, untrained images. Backpropagation is used to calculate the error and modify the weights if the image is incorrectly classified. When the aNN correctly recognizes a person in an image, the training process is repeated as many times as necessary.

Images must be turned into feature vectors because ML systems require them as input. This procedure is known as feature extraction. Given that images are composed of arrays of pixels, feature extraction entails locating pertinent parts in the training images, such as pertinent sections of a face. The pixel arrays of these locations would subsequently be contained in a feature vector. It is a binary classification problem when the aNN must recognize only one individual, and the output, for example, is 1 if the network has successfully positively recognized the individual, and 0 otherwise. It is a multiclass classification challenge if multiple people need to be identified. The probability that an input belongs to a specific class is output by the majority of ML algorithms.They can offer accurate estimates for extremely complex problems.

Multilayer perceptrons frequently overfit and have a slow learning rate. If the aNN is overly complex and begins to detect noise rather than underlying relationships, overfitting may result. Minor fluctuations may be exaggerated by an overfitted model, which leads to subpar prediction accuracy. Other varieties of aNNs include radial basis function networks, self-organizing maps, and recurrent networks, to name a few. Every ML strategy needs a means to gauge the model's effectiveness. It takes a lot of test data to produce results that are statistically significant. This is frequently impractical because, for instance, there might not be enough images of a person taken from various angles or at various ages for face recognition. Shallow learners are effective for rough first- or second-order function approximations, but they are ineffective for modeling complicated higher-order features. Since object recognition utilizing pixel intensities of an image requires hierarchical feature extraction, they are therefore not well suited for it. To prevent overfitting, a vast volume of training data would be necessary in order to handle the complexity of the created models.The models' ability to generalize is highly unlikely if there is not enough data.

There are other non-linear learning strategies besides aNNs. Numerous nonlinear issues have been solved using kernelbase techniques like SVM. A kernel can be defined over any type of structure, including sets, strings, trees, and probability distributions [27]. A kernel is essentially a similarity function with specific mathematical qualities. Kernel approaches, on the other hand, limit the learner's ability to generalize outside of the training data samples. A local minimum of the loss function could persist throughout training, which is another risk. Automatic feature extraction is frequently difficult to accomplish. Manual feature extraction is a typical strategy [28]. The capacity to automatically extract useful characteristics using all-purpose learning techniques is a significant benefit of DL. ConvNets' primary advantage for many of these jobs is that the entire system is trained from initial pixels to final categories, eliminating the need to manually construct an appropriate feature extractor [29].

V. DECLARATION

I Name Syed Muhammad Abis Rizvi herewith declare that I have composed the present paper and work by myself and without use of any other than the cited sources and aids. Sentences or parts of sentences quoted literally are marked as such; other references with regard to the statement and scope are indicated by full details of the publications concerned. The paper and work in the same or similar form has not been submitted to any examination body and has not been published. This paper was not yet, even in part, used in another examination or as a course performance.

VI. References

[1] C. Xinyu, C. Youngwoon, and J. Suk young, ”Crime prediction using Twitter sentiment and weather.”, Systems and Information Engineering Design Symposium (SIEDS), pp. 63-68, 2015.

[2] T. Liu, X. Ding, Y. Chen, H. Chen, and M. Guo, “Predicting movie Boxoffice revenues by exploiting large-scale social media content,” Multimedia Tools and Applications, pp. 1-20, 2014/10/02, 2014.

[3] H. V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, and C. Shahabi, “Big data and its technical challenges,” Commun. ACM, vol. 57, no. 7, pp. 86-94, 2014.

[4] P. Wlodarczak, J. Soar, and M. Ally, ”Genome Mining Using Machine Learning Techniques,” Inclusive Smart Cities and e-Health, Lecture Notes in Computer Science A. Geissbühler, J. Demongeot, M. Mokhtari, B. Abdulrazak and H. Aloulou, eds., pp. 379-384: Springer International Publishing, 2015.

[5] J. O. Chan, “An Architecture for Big Data Analytics,” Communications of the IIMA, vol. 13, no. 2, pp. 1-13, 2013.

[6] A. Twinkle, and S. Paul, “Adressing Big Data with Hadoop,” International Journal of Computer Science and Mobile Computing, vol. 3, no. 2, pp. 459–462, 2014.

[7] P. Wlodarczak, M. Ally, and J. Soar, ”Data Process and Analysis Technologies of Big Data,” Networking for Big Data, Chapman Hall/CRC Big Data Series, pp. 103-119: Chapman and Hall/CRC, 2015.

[8] Nguyen, G., Dlugolinsky, S., Bobák, M. et al. Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. Artif Intell Rev 52, 77–124 (2019). https://doi.org/10.1007/s10462-018-09679-z

[9] Jen, C.-H., Wang, C.-C., Jiang, B. C., Chu, Y.-H., and Chen, M.- S., Application of classification techniques on development an early-warning system for chronic illnesses. Expert Syst. Appl. 39(10):8852–8858, 2012.

[10] Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., Cowell, R. G., Bayesian analysis in expert systems. Statistical science: 219–247, 1993.

[11] Estella, F., Delgado-Marquez, B. L., Rojas, P., Valenzuela, O., San Roman, B., and Rojas, I., Advanced system for automously classify brain MRI in neurodegenerative disease. In: Multimedia Computing and Systems (ICMCS), 2012 International Conference on. IEEE, pp 250–255, 2012.

[12] Boser, B. E., Guyon, I. M., and Vapnik, V. N. A., training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory. ACM, pp 144–152, 1992. [13] Lan, K., Wang, Dt., Fong, S. et al. A Survey of Data Mining and Deep Learning in Bioinformatics. J Med Syst 42, 139 (2018). https://doi.org/10.1007/s10916-018-1003-9

[14] Amaratunga, D., Cabrera, J., and Lee, Y.-S., Enriched random forests. Bioinformatics 24(18):2010–2014, 2008.

[15] Jain, A. K., Murty, M. N., and Flynn, P. J., Data clustering: a review. ACM computing surveys (CSUR) 31(3):264–323, 1999.

[16] Zhang, T., Ramakrishnan, R., and Livny, M., BIRCH: an efficient data clustering method for very large databases. In: ACM Sigmod Record. vol 2. ACM, pp 103–114, 1996.

[17] Bryant, D., and Moulton, V., Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21(2):255–265, 2004.

[18] Chiu, T., Fang, D., Chen, J., Wang, Y., and Jeris, C., A robust and scalable clustering algorithm for mixed type attributes in large database environment. In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 263–268, 2001.

[19] Jiang, D., Pei, J., and Zhang, A., DHC: a density-based hierarchical clustering method for time series gene expression data. In: Bioinformatics and Bioengineering. Proceedings. Third IEEE Symposium on, 2003. IEEE, pp 393–400, 2003.

[20] Maltsev, N., Glass, E., Sulakhe, D., Rodriguez, A., Syed, M. H., Bompada, T., Zhang, Y., and D'souza, M., PUMA2—grid-based high-throughput analysis of genomes and metabolic pathways. Nucleic Acids Res. 34(suppl$_1$) : $D369$–$D372, 2006$.

[21] Si, Y., Liu, P., Li, P., and Brutnell, T. P., Model-based clustering for RNA-seq data. Bioinformatics 30(2):197–205, 2013.

[22]Lan, K., Wang, Dt., Fong, S. et al. A Survey of Data Mining and Deep Learning in Bioinformatics. J Med Syst 42, 139 (2018). https://doi.org/10.1007/s10916-018-1003-9

[23]R. Salakhutdinov, “Deep learning,” in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, New York, USA, 2014, pp. 1973-1973.

[24] M. Turk, and A. Pentland, “Eigenfaces for recognition,” J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[25] P. Wlodarczak, J. Soar and M. Ally, ”Multimedia data mining using deep learning,” 2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC), 2015, pp. 190-196, doi: 10.1109/ICDIPC.2015.7323027.

[26] H. Weilong, G. Xinbo, T. Dacheng, and L. Xuelong, “Blind Image Quality Assessment via Deep Learning,” Neural Networks and Learning Systems, IEEE Transactions on, vol. 26, no. 6, pp. 1275-1286, 2015.

[27] N. Lopes, and B. Ribeiro, “Towards adaptive learning with improved convergence of deep belief networks on graphics processing units,” Pattern Recognition, vol. 47, no. 1, pp. 114-127, 1//, 2014

[28] V. Mayer-Schonberger, and K. Cukier, Big Data: A Revolution That Will Transform How We Live, Work, and

Think, New York, USA: Houghton Mifflin Harcourt Publishing Company, 2013.

[29] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013.