Strategies for Managing and Analyzing Large-Scale Neurological Datasets: A Review of Advanced Computational Methods

Abdelaziz Rabehi[1,*]

[1]Telecommunications and Smart Systems Laboratory, University of Djelfa, PO Box 3117, Djelfa 17000, Algeria

Email: Abdelaziz.rabehi@univ-djelfa.dz

Abstract

The progress of neuroimaging and the availability of big neuro data have brought both opportunities and difficulties in the fast-developing scientific area of computational neuroscience. As this review will show, new ways of managing and analyzing these large and layered datasets are emerging, highlighting the importance of various computational approaches to achieve valuable insights. We assess various methods for performing such analyses, among which we focus on machine learning algorithms like deep learning capable of addressing high-dimensional data characteristics for neuroimaging studies. The proposed method of analyzing multiple structural and functional MRI data in conjunction with electrophysiological and genetic data should help model neurological disorders more accurately. We also describe the preprocessing methods for dealing with data noise and variability, combined with statistical analysis that depends on existing databases to identify previously unknown patterns concerning brain functions and disorders. We also discussed the importance of open-source teamwork spaces and applications, which allow datasets and results to be shared and replicated. This review, therefore, aimed at reviewing the most effective strategies and filling the gaps within the current methodologies that may help enhance the strength and reliability of vast neurological datasets, hence diminishing diagnostic errors and helping formulate the right therapeutic intercessions in neurological disorders. This synthesis emphasizes the choice of a multidisciplinary approach when studying the neural tissues since the issue appears complex.

Keywords: Neuroimaging; Big neuro data; Machine learning algorithms; Electrophysiological data; Genetic data; Vast neurological datasets

1. Introduction

Classification of mental disorders has always been an essential area of interest in clinical psychiatry and psychology. Mental disorders describe a group of severe and frequently disabling illnesses that affect a client's psychological processes. Earlier, there were many controversies in diagnostically and classification systems like DSM and ICD, but formulated criteria of mental illness are quite advanced as compared to them. However, these conventional practices are ineffective in providing comprehensive details of the disorder because they rely on phenomenological or symptomatic markers rather than pathophysiological markers.

However, recent developments in artificial intelligence, machine learning performance, and intense generative models (DGMs) can address such limitations. Classes of models such as DGMs, GANs, and VAEs provide an opportunity to find structures based on data within highly complex neuroimaging and biomarker

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025

domains. These models are best used for discovering hidden structures within large-dimensional data; the workings of mental diseases may be linked to their biological and neural foundations [1].

Incorporating DGMs in the classification of mental disorders has attracted increasing research interest in recent years due to the observed more excellent diagnostic reliability and predictive power. This review considers the different architectures of DGMs and their use in diagnosing mental disorders to understand how these models may improve our understanding of psychiatry. For example, GANs demonstrated high efficiency in differentiating between individuals with MDD and healthy subjects, thereby overtaking more conventional classifiers when using data from complicated neuronal and biomarker networks.

The proposed DGMs can operate with little feature engineering, thus indicating that more comprehensive essential features might not be well identified using traditional ML techniques. Because DGMs can self-discover features that are not prejudiced by specific hypotheses, they are instrumental in a field such as psychiatry that tries to find biological markers of disorders that can be complex and not well characterized. This is because, with DGMs, models can select features on their own, which will help to get a better classification model depending on the input, which may, in turn, help in the early detection and treatment of mental health disorders [2].

However, DGM can also provide the capability of synthesizing data, which is highly valuable in psychiatric research due to the expected shortage of data. Data augmentation can help DGMs create new data similar to patients' data, which will increase data sets and help minimize the bias of small data sizes. This synthetic data generation also enables near exhaustive examination of what were hitherto unthinkable rare or sunk cost conditions.

The review provides a comprehensive comparative study on the different DGMs used in mental health, emphasizing their performance comparison. For example, even though GANs outperform many other methods in generating a reasonable data representation, VAE is more helpful in recognizing latent variables behind intricate data patterns. Such comparisons evaluate the best model architecture regarding the nature of the particular disorder under investigation.

One feature of this review is that attention is paid to specific clinical questions of applying DGMs in mental health. By enhancing distinctiveness and especially concerning improved classification accuracy for different subtypes within such disorders, DGMs may point to the future implementation of the concept of personalized medicine in psychiatry. Individualized biological treatment based on an individual's Neurotrophies and other neural and biological characteristics could enhance a patient's therapeutic response and limit the use of 'therapeutic trials' characteristic of present-day psychiatric practice [3], [4].

Furthermore, DGMs enhance our knowledge of mental disorder etiopathogenesis, as discussed above. Such analyses of how DGMs cluster symptoms and biomarkers may reveal previously unknown connections between activity patterns and psychiatric symptoms. This could lead to the creation of new; despite these apparent simplifications, such an understanding might help propose new therapeutic approaches that could target the biology of these disorders [5].

This is a present review of previously published articles. The primary task of this paper is to integrate the information found during the research to give a general view of DGMs and their importance in categorizing mental disorders. The strengths and weaknesses of DGMs are also discussed here, directions for future research are pointed out, and possible ways of making these models more precise and individualized for diagnosing mental disorders are outlined. As such, the multimodal and multivariate conceptual approach of DGMs becomes meaningful during the review regarding its implications for redesigning the conventional processes of studying, diagnosing, and treating mental disorders [6].

2. Literature review

As a literature review, this paper comprehensively discusses related works from various disciplines in the last decade, including neuroinformatics, deep learning, and more. The emphasis is made on the methodological and application developments using artificial intelligence and computational methods to analyze brain functions, data processing, and prediction systems. Some of the works discussed include performing distributed data integration for neuroimaging, building neuro-rendering tools for complicated 3D scenes, and utilizing deep learning models for various real-world purposes, like crop price prediction and crowd surveillance. Overall, the studies in this thesis topic enhance the efficiency and reliability of data

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025

analysis in specific fields. This review summarizes these contributions and indicates further research and development directions.

Imaging, from magnetic resonance imaging (MRI) to the localization of specific macromolecules by microscopy, has been one of the driving forces behind neuroinformatics efforts over the past decade. As discussed in the paper [7], numerous web-accessible resources have been developed, ranging from primary data collections to highly structured databases. Despite ongoing challenges in adapting neuroscience to the new electronic forum envisioned by neuroinformatics advocates, these initiatives have successfully established the necessary frameworks for efficient data sharing and integration across diverse sources. The article emphasizes the critical role of spatial systems and ontologies in effectively modeling neuroscience data and their contribution to large-scale data integration efforts such as the Biomedical Informatics Research Network (BIRN).

Recent advances in optical and electron microscopy have enabled scientists to acquire highly high-resolution images for neuroscience research. The study referenced as

[8], the authors highlight that data sets imaged with modern electron microscopes can range from tens of terabytes to nearly one petabyte. The immense data sizes and the intricate complexity of the underlying neural structures present significant challenges in handling the data at interactive rates. To address these challenges, the authors introduce Ssecrett and NeuroTrace, two innovative tools designed for the interactive exploration and analysis of large-scale optical and electron microscopy images, facilitating the reconstruction of complex neural circuits in the mammalian nervous system.

The analysis of neural data with multiple modes and high density has recently become a trend, along with advances in neuroscience research and practices. The study referenced as [9] addresses the pressing need for an approach that accurately captures features without losing or distorting the interactions among modes such as space, time, and frequency. This approach must also handle the exponentially growing scales of neural data, which can involve tens or hundreds of channels, enabling timely conclusions and decisions. The study introduces a large-scale parallel factor analysis (PARAFAC) method supported by general-purpose computing on graphics processing units (GPGPU). Compared to conventional CPU-based platforms, the new approach outperforms PARAFAC by more than 360 times in runtime performance and scales over 400 times across all dimensions. Additionally, the method is applied to electrocochleography (ECoG) recordings from epilepsy patients, demonstrating its effectiveness in epilepsy state detection, where time evolutions correlate with clinical observations, and frequency and spatial signatures identify the propagation of neural activity across brain regions. This model supports real-time analysis of ECoG in over 1,000 channels using affordable, accessible cyberinfrastructure.

Cognition is thought to result from interactions within large-scale networks of brain regions. The analysis conducted in [10] proposes a method for identifying these networks using functional magnetic resonance imaging (fMRI), defining networks as sets of strongly interacting regions with homogeneous temporal activity. The large-scale network identification (LSNI) method begins by detecting functionally homogeneous regions and establishes networks by comparing the correlations among these regions against a noise model. After validating the LSNI method's specificity and sensitivity on synthetic data, it was tested on actual data using a motor task design. The method accurately identified regions with stimulus-linked activity and two significant networks reproducible across subjects, both dominated by slow fluctuations (0–0.1 Hz). One network overlapped with the "default" network in medial and dorsal areas, aligning with previous findings on resting-state connectivity, while the other spanned lateral frontal and posterior parietal regions. The LSNI approach enables exploratory and systematic detection of regions and networks active in the working brain.

Existing neural radiance field (NeRF)-based methods enable real-time rendering of small-scale scenes on web platforms, but extending this capability to large-scale environments remains challenging due to computational power, memory, and bandwidth constraints. In the study in [11], the authors introduce *City-on-Web*, the first approach to achieve real-time rendering of expansive scenes directly on the web. This method uses a block-based volume rendering technique that preserves 3D consistency and accurately handles occlusions across different scene blocks. Additionally, a Level-of-Detail (LoD) strategy and dynamic resource loading and unloading are employed to reduce significantly memory requirements. Tested with an

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025

RTX 3060 GPU, this system delivers real-time rendering performance at approximately 32 frames per second (FPS) on the web, maintaining visual quality on par with the current leading methods in novel view synthesis.

The rapid expansion of neuroimaging literature has significantly advanced our understanding of human brain function. However, the research volume has made consolidating and interpreting findings across studies challenging. The research presented in [12] introduces an automated brain-mapping framework that leverages text mining, meta-analysis, and machine-learning techniques to build an extensive database linking neural activity to cognitive states. This approach enables large-scale, high-quality neuroimaging meta-analyses, addressing long-standing inferential issues in the field and allowing for precise "decoding" of cognitive states from brain activity at both group and individual levels. The study's findings validate a robust, scalable framework for synthesizing neuroimaging data, offering a valuable tool for understanding human cognition on a large scale.

Aggregating imaging, clinical, and behavioral data from diverse institutions and researchers offers vast potential for biomedical research but also presents significant challenges. According to the findings in [13], research groups often have well-defined data collection and analysis protocols, with specific data formats and metadata tailored to their needs. The diversity in data types—from imaging and physiological data to experimental design descriptions—requires numerous specialized software tools for collection, storage, and processing. Additionally, institutions may hesitate to relinquish control over data access and distribution. To address these complexities, the Biomedical Informatics Research Network (BIRN) developed a federated infrastructure for storing, retrieving, analyzing, and documentation of biomedical imaging data. This infrastructure includes distributed data collections maintained on dedicated resources at each participating site, a federated data management and integration system, an XML schema for data exchange, and analysis pipelines that harness distributed data management and grid computing resources. This system supports efficient, collaborative biomedical research while respecting institutional data governance.

Data Envelopment Analysis (DEA) is widely recognized for assessing the efficiency and productivity of Decision-Making Units (DMUs); however, applying DEA to large datasets with numerous inputs and outputs demands substantial computational resources. As outlined in [14], this study introduces a neural network back-propagation approach to DEA, designed to handle massive datasets that conventional DEA methods struggle to process. The neural network approach significantly reduces memory and CPU requirements, making it a practical solution for efficiency measurement in large datasets. To validate this method, the back-propagation DEA algorithm was tested on five extensive datasets, and its results were compared with those of traditional DEA, demonstrating its effectiveness in large-scale efficiency analysis.

In the study referenced as [15], neuroimaging researchers have established comprehensive data and metadata standards to promote meta-analyses that consolidate knowledge about human brain structure and function. Leveraging these standards, the Brain Map project provides a suite of databases, software applications, and tools specifically designed for quantitative coordinate-based meta-analysis of structural and functional neuroimaging literature. This report highlights recent technical updates to Brain Map and offers a detailed guide for conducting meta-analyses within its environment. The Brain Map project is expected to continue evolving, incorporating new software and hardware advancements to meet the growing meta-analytic needs of the neuroimaging research community.

As outlined in [16], while deep learning models excel in predictive accuracy, interpreting them presents significant challenges, particularly given the complex architecture and large-scale datasets they process. Addressing this need, researchers developed Activism, an interactive visualization tool created through participatory design with over 15 researchers and engineers at Facebook. Activism integrates coordinated visual views, including a computation graph overview and a neuron activation view, enabling users to analyze model architecture and identify patterns at both instance and subset levels. Deployed on Facebook's machine learning platform, Activism explores deep learning model behavior, as demonstrated through case studies and usage scenarios with Facebook's research and engineering teams.

In the study referenced as [17], Convolutional Neural Networks (CNNs) achieve high accuracy across various fields by leveraging extensive training datasets and complex network architectures. However, the training of CNNs is time-intensive, requiring significant computational resources for handling large datasets and iterative operations to refine weight parameters. To address this, the researchers introduce a Bi-layered Parallel Training (BPT-CNN) framework designed for distributed computing environments to reduce training

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025

time without sacrificing accuracy. The BPT-CNN framework includes two components: an outer-layer parallelism, where multiple CNN subnetworks are trained on distinct data subsets, and an inner-layer parallelism that accelerates each subnetwork's training. The outer layer tackles issues like data communication, synchronization, and workload balance using a Heterogeneous-aware Incremental Data Partitioning and Allocation (IDPA) strategy, optimizing data distribution based on computing power.

Additionally, an Asynchronous Global Weight Update (AGWU) strategy minimizes synchronization delays. In the inner layer, task parallelism accelerates convolutional operations through task decomposition and scheduling, achieving load balancing and reducing critical path waiting times. Experimental results show that BPT-CNN enhances CNN training efficiency significantly while preserving accuracy, highlighting its potential for large-scale applications.

As outlined in [18], the accurate forecasting of agricultural product prices, particularly horticultural products, is crucial for managing supply chains and planning crop production due to these products' sensitivity to price fluctuations and limited storage viability. The study explores various predictive models, including the Autoregressive Integrated Moving Average (ARIMA), Back Propagation (BP) neural network, and Recurrent Neural Network (RNN), to estimate the prices of products like cucumber, tomato, and eggplant over short- and long-term periods. Data on agricultural prices gathered extensively through web scraping reveal that ARIMA performs well with small-scale, periodic datasets, mainly when predicting average monthly prices. However, it needs more granular daily data. By contrast, neural network models, such as BP and RNN, provide more accurate daily, weekly, and monthly price forecasts and handle large-scale datasets effectively. Deep learning approaches, particularly neural networks, will likely become the preferred methods for forecasting agricultural product prices, as they demonstrate a more substantial capacity for capturing complex price trends across varying time scales.

As outlined in [19], this study presents a rapid, decor-related neuro-ensemble approach utilizing heterogeneous features to enhance large-scale data analytics. Stochastic configuration networks (SCNs) serve as the base learner models, leveraging the negative correlation learning (NCL) strategy for optimizing output weights. Due to the extensive size of the linear equation system formed by feeding numerous samples into SCN base models, conventional pseudo-inverse solutions, often employed in the least squares methods, could be more practical. Consequently, the study incorporates block Jacobi and Gauss-Seidel iterative methods to estimate output weights, providing a convergence analysis that confirms the uniqueness of the iterative solutions. Experiments on two large-scale datasets demonstrate the system’s robustness regarding the regularizing factor used in NCL, underscoring the ensemble is potential for managing complex data modeling challenges.

As discussed in [20], the challenge of efficient semantic segmentation for large-scale 3D point clouds is addressed by developing RandLA-Net, a streamlined neural architecture designed for direct per-point semantic inference. Unlike methods that rely on costly sampling and resource-intensive pre-and post-processing, RandLA-Net employs a random point sampling strategy, significantly enhancing computational and memory efficiency. A novel local feature aggregation module is introduced to counter the risk of losing essential features with random sampling, gradually expanding the receptive field for each 3D point and effectively preserving geometric details. Benchmark experiments demonstrate that RandLA-Net can process up to 1 million points in a single pass, achieving speeds up to 200 times faster than traditional models. Extensive evaluations on large-scale datasets—such as Semantic3D, Semantic KITTI, Toronto3D, NPM3D, and S3DIS—highlight RandLA-Net's state-of-the-art performance in semantic segmentation.

In the analysis conducted in [21], researchers address the growing interest in crowd counting and localization, motivated by applications in crowd monitoring, public safety, and spatial design. Although convolutional neural networks (CNNs) have been widely applied to this task, existing datasets need to be more comprehensive in scale to support effective supervised learning. The NWPU-Crowd dataset was developed to overcome this limitation, comprising 5,109 images and 2,133,375 annotated head points and boxes, with a wide density range (0 to 20,033) and diverse illumination conditions. Additionally, a benchmarking platform was created to provide an impartial evaluation of various methods, enabling researchers to submit test results for performance comparison. This dataset and its associated resources are available online, providing an extensive basis for evaluating state-of-the-art (SOTA) models while highlighting new challenges unique to larger datasets.

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025

As outlined in [22], the rise of station-free bike-sharing systems, widely implemented across China in early 2017, has introduced a flexible mode of urban mobility without the need for docking stations, allowing bikes to be parked in any suitable location. This study developed a dynamic demand-forecasting model utilizing deep learning to predict usage patterns for such systems. Spatial and temporal analyses first revealed an imbalanced demand across locations and times. To address this, the study implemented long short-term memory neural networks (LSTM NNs) to forecast bike trip production and attraction within traffic analysis zones (TAZ) at various time intervals—10, 15, 20, and 30 minutes. Validation showed that the LSTM NNs achieved high predictive accuracy across these intervals, outperforming traditional statistical and advanced machine learning models. This model's ability to predict inflow-outflow imbalances in bike trips provides valuable insights for effectively rebalancing station-free bike resources within the system.

As detailed in the paper [23], crowd counting and localization have garnered significant research attention over the past decade due to their applicability in crowd monitoring, public safety, and space design. Although various convolutional neural networks (CNNs) have been developed for this purpose, the limited scale of existing datasets has posed a challenge for supervised CNN-based algorithms. The NWPU-Crowd dataset was introduced to address this, featuring 5,109 images and over 2 million annotated heads marked with points and boxes. This dataset offers a broad range of illumination conditions and the highest density range (0–20,033) among real-world datasets, making it particularly valuable for high-density scenes. A benchmarking platform was also created to provide standardized evaluation for different methods, enabling researchers to submit and compare test results. The Comprehensive dataset analyses demonstrate its robustness in evaluating state-of-the-art (SOTA) models while identifying new challenges unique to large-scale data. The dataset and resources like code and model results are publicly accessible online for further research and benchmarking.

In the publication identified as [24], neural networks (NNs) have been explored for multi-label classification due to their ability to model label dependencies effectively within the output layer. This study re-evaluates the limitations of BP-MLL, a neural network architecture that aims to minimize pairwise ranking error for multi-label tasks. As an alternative, the study proposes a simplified NN approach that utilizes recent advancements in large-scale multi-label text classification techniques. It demonstrates that the ranking loss minimization used in BP-MLL can be effectively replaced by the cross-entropy error function, a more commonly applied method. The study further incorporates advanced neural network training techniques, including rectified linear units (ReLU), dropout, and AdaGrad, showing that these simpler NN models match and often outperform state-of-the-art methods across six large-scale, diverse textual datasets.

Table 1 encapsulates a synthesis of the literature from the studies [7] to [24], presenting these findings, the methods used, and the application domains encompassing neuroscience, data science and artificial intelligence. The table indicates works investigating various areas, including characteristics of distributed data in neuroinformatics, neural rendering of real-time 3D scenes, large-scale neuroimaging data generation, and deep learning for various purposes, including crop price prediction, crowd tracking, and multi-label text categorization. Every entry also offers some background about the methods used, including neural networks and multi-way analysis, new meta-analysis strategies, and new deep learning frameworks, demonstrating various advancements in these fields.

Table 1: Summary of Literature Review

Ref.	Study Title & Authors	Key Findings	Methodology	Data/Application Area
[7]	"e-Neuroscience: Challenges and Triumphs in Integrating Distributed Data" (Martone et al.)	Highlights the role of spatial systems for data integration in neuroinformatics	Data frameworks like BIRN support efficient data sharing	Neuroinformatics, data integration

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025

[8]	"Ssecrett and NeuroTrace: Tools for Large-Scale Neuroscience Data" (Jeong et al.)	Introduced tools for interactive data analysis in large optical and electron microscopy images	Interactive visualization tools for neural data	Large-scale microscopy, neuroimaging
[9]	"Fast and Scalable Multi-way Analysis of Massive Neural Data" (Chen et al.)	PARAFAC method outperforms traditional CPUs for large data sets in epilepsy studies	Multi-way analysis on GPUs	Electrophysiology, epilepsy
[10]	"Identification of Large-Scale Networks in the Brain using fMRI" (Bellec et al.)	Established LSNI for detecting functional region networks in the brain	fMRI analysis through LSNI	Functional brain networks
[11]	"City-on-Web: Real-time Neural Rendering of Large-scale Scenes" (Song et al.)	Achieved real-time rendering of complex 3D scenes on the web	Neural rendering with LoD for web applications	Large-scale 3D scene rendering
[12]	"Large-scale Automated Synthesis of Neuroimaging Data" (Yarkoni et al.)	Database linking neural activity with cognitive states for high-quality meta-analyses.	Text-mining, meta-analysis in brain imaging	Neuroimaging meta-analyses
[13]	"Biomedical Informatics Research Network for Neuroimaging" (Keator et al.)	Developed a federated infrastructure supporting biomedical imaging research	Data management and integration system	Neuroimaging, collaborative research
[14]	"Neural Network and DEA for Efficiency of Large Datasets" (Emrouznejad & Shale)	Improved DEA efficiency measurement on large datasets using neural networks	DEA with neural networks for large datasets	Efficiency analysis
[15]	"BrainMap Project for Neuroimaging Meta-analysis" (Laird et al.)	Provides standardized databases and tools for neuroimaging meta-analyses	BrainMap standards, metadata	Neuroimaging standardization
[16]	"ActiVis: Visual Exploration of Deep Neural Network Models" (Kahng et al.)	Offers tools for analyzing DNN model behavior	Visual exploration tool for DNNs	Deep learning model interpretation
[17]	"Bi-layered Parallel Training Architecture for CNNs" (Chen et al.)	BPT-CNN framework reduces training time while maintaining accuracy	Parallel training in distributed computing	Convolutional Neural Networks

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025

[18]	"Forecasting Horticultural Prices using ARIMA and Neural Networks" (Weng et al.)	Neural networks provide accurate long-term forecasts for crop prices	ARIMA, neural network models	Agricultural price forecasting
[19]	"Neuro-Ensemble Approach with Heterogeneous Features" (Wang & Cui)	Stochastic configuration networks with iterative methods enhance analytics	SCNs and block Jacobi, Gauss-Seidel methods	Large-scale data modeling
[20]	"Semantic Segmentation of Large-Scale Point Clouds" (Hu et al.)	RandLA-Net achieves efficient point cloud segmentation for large datasets	Random sampling, local feature aggregation	3D point cloud segmentation
[21]	"NWPU-Crowd Dataset for Crowd Counting and Localization" (Wang et al.)	New large-scale dataset with a benchmarking platform for crowd analytics	NWPU-Crowd dataset	Crowd monitoring, public safety
[22]	"Station-Free Sharing Bike Demand Forecasting" (Xu et al.)	LSTM networks predict demand for flexible bike-sharing systems	Deep learning, spatial-temporal analysis	Urban mobility, demand forecasting
[23]	"NWPU-Crowd: Benchmarking for High-density Scenes" (Wang et al.)	Dataset and platform for evaluating high-density crowd-counting methods	High-density dataset, CNNs	Crowd monitoring, spatial design
[24]	"Revisiting Neural Networks for Multi-label Text Classification" (Nam et al.)	The cross-entropy function was shown to be effective for multi-label classification.	Simplified neural network approach	Multi-label text classification

In sum, the reviewed studies stress the increasing importance of applying enhanced mathematical computations in addressing diverse issues in many disciplines. Advanced use of such technologies as neural networks, machine learning and large-scale data frameworks continues to improve the data processing and interpretation and make accurate predictions possible. All these new concepts prepare the ground for even more effective and precise approaches, whether in analyzing neuronal networks or modeling actual events. These technologies remain constantly developing, presenting a tremendous scope of potential future work in areas as diverse as health and city planning. Hence, this literature review concludes that interdisciplinary cooperation with technology will determine the future of data science and neuroscience.

3. Conclusion

This review underscores the transformative impact of advanced computational methods on managing and analyzing large-scale neurological datasets. With neuroimaging advancements and the rise of big data, integrating machine learning and intense generative models (DGMs) has emerged as a pivotal innovation. These tools address the challenges posed by high-dimensional, multimodal data and uncover novel patterns that link neural and biological markers to mental disorders. The synergy of structural, functional, electrophysiological, and genetic data analysis has demonstrated significant potential for enhancing the diagnostic precision and understanding of complex neurological conditions.

Moreover, data preprocessing techniques, statistical methods, and collaborative open-source platforms have been crucial in overcoming noise, variability, and data fragmentation challenges. These strategies ensure

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025

robust data sharing and reproducibility, fostering interdisciplinary cooperation among researchers. The methodologies reviewed also highlight the importance of leveraging computational efficiency, such as parallel processing and real-time analysis, to address the exponential growth in dataset sizes. This accelerates research and lays the groundwork for real-time applications in clinical settings.

This review identifies critical areas for future research, including optimizing DGMs for more precise and individualized models, particularly in psychiatry. DGMs' potential to self-discover latent features and generate synthetic datasets represents a promising avenue for addressing data scarcity in mental health studies. Furthermore, integrating these tools into personalized medicine frameworks offers a pathway to tailored therapeutic interventions, minimizing diagnostic errors and improving patient outcomes.

In conclusion, the intersection of neuroscience and computational science continues to open unprecedented opportunities for innovation. By embracing multidisciplinary approaches and refining computational tools, researchers are making significant strides in more effectively addressing the complexities of neural data analysis. This review advocates for ongoing efforts to enhance the precision, scalability, and accessibility of these methods, aiming to revolutionize the understanding and treatment of neurological disorders.

Reference

[1] N. Huynh and G. Deshpande, “A review of the applications of generative adversarial networks to structural and functional MRI based diagnostic classification of brain disorders,” Front Neurosci, vol. 18, p. 1333712, Apr. 2024, doi: 10.3389/FNINS.2024.1333712.

[2] F. Cai, A. I. Ozdagli, and X. Koutsoukos, “Variational Autoencoder for Classification and Regression for Out-of-Distribution Detection in Learning-Enabled Cyber-Physical Systems,” Applied Artificial Intelligence, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2022.2131056.

[3] M. Squires et al., “Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment,” Brain Inform, vol. 10, no. 1, p. 10, Dec. 2023, doi: 10.1186/S40708-023-00188-6.

[4] J. Torous and C. Blease, “Generative artificial intelligence in mental health care: potential benefits and current challenges,” World Psychiatry, vol. 23, no. 1, p. 1, Feb. 2024, doi: 10.1002/WPS.21148.

[5] A. Arora and A. Arora, “Generative adversarial networks and synthetic patient data: current challenges and future perspectives,” Future Healthc J, vol. 9, no. 2, p. 190, Jul. 2022, doi: 10.7861/FHJ.2022-0013.

[6] F. Jalalvand, M. Baruwal Chhetri, S. Nepal, and C. Paris, “Alert Prioritisation in Security Operations Centres: A Systematic Survey on Criteria and Methods,” ACM Comput Surv, vol. 57, no. 2, pp. 1–36, Feb. 2025, doi: 10.1145/3695462.

[7] M. E. Martone, A. Gupta, and M. H. Ellisman, “e-Neuroscience: challenges and triumphs in integrating distributed data from molecules to brains,” Nature Neuroscience, vol. 7, no. 5, pp. 467–472, Apr. 2004, doi: 10.1038/nn1229.

[8] W. K. Jeong et al., “Ssecrett and neuroTrace: Interactive visualization and analysis tools for large-scale neuroscience data sets,” IEEE Comput Graph Appl, vol. 30, no. 3, pp. 58–70, 2010, doi: 10.1109/MCG.2010.56.

[9] D. Chen et al., “Fast and scalable multi-way analysis of massive neural data,” IEEE Transactions on Computers, vol. 64, no. 3, pp. 707–719, Mar. 2015, doi: 10.1109/TC.2013.2295806.

[10] P. Bellec et al., “Identification of large-scale networks in the brain using fMRI,” Neuroimage, vol. 29, no. 4, pp. 1231–1243, Feb. 2006, doi: 10.1016/J.NEUROIMAGE.2005.08.044.

[11] K. Song, X. Zeng, C. Ren, and J. Zhang, “City-on-Web: Real-time Neural Rendering of Large-scale Scenes on the Web,” Dec. 2023, Accessed: Nov. 13, 2024. [Online]. Available: https://arxiv.org/abs/2312.16457v2.

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025

[12] T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, and T. D. Wager, "Large-scale automated synthesis of human functional neuroimaging data," Nature Methods, vol. 8, no. 8, pp. 665–670, Jun. 2011, doi: 10.1038/nmeth.1635.

[13] D. B. Keator et al., "A national human neuroimaging collaboratory enabled by the biomedical informatics research network (BIRN)," IEEE Transactions on Information Technology in Biomedicine, vol. 12, no. 2, pp. 162–172, Mar. 2008, doi: 10.1109/TITB.2008.917893.

[14] A. Emrouznejad and E. Shale, "A combined neural network and DEA for measuring efficiency of large scale datasets," Comput Ind Eng, vol. 56, no. 1, pp. 249–254, Feb. 2009, doi: 10.1016/J.CIE.2008.05.012.

[15] A. R. Laird et al., "The BrainMap strategy for standardization, sharing, and meta-analysis of neuroimaging data," BMC Res Notes, vol. 4, no. 1, pp. 1–9, Sep. 2011, doi: 10.1186/1756-0500-4-349/FIGURES/3.

[16] M. Kahng, P. Y. Andrews, A. Kalro, and D. H. P. Chau, "ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models," IEEE Trans Vis Comput Graph, vol. 24, no. 1, pp. 88–97, Jan. 2018, doi: 10.1109/TVCG.2017.2744718.

[17] J. Chen, K. Li, K. Bilal, X. Zhou, K. Li, and P. S. Yu, "A Bi-layered parallel training architecture for large-scale convolutional neural networks," IEEE Transactions on Parallel and Distributed Systems, vol. 30, no. 5, pp. 965–976, May 2019, doi: 10.1109/TPDS.2018.2877359.

[18] Y. Weng, X. Wang, J. Hua, H. Wang, M. Kang, and F. Y. Wang, "Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler," IEEE Trans Comput Soc Syst, vol. 6, no. 3, pp. 547–553, Jun. 2019, doi: 10.1109/TCSS.2019.2914499.

[19] D. Wang and C. Cui, "Stochastic configuration networks ensemble with heterogeneous features for large-scale data analytics," Inf Sci (N Y), vol. 417, pp. 55–71, Nov. 2017, doi: 10.1016/J.INS.2017.07.003.

[20] Q. Hu et al., "Learning Semantic Segmentation of Large-Scale Point Clouds With Random Sampling," IEEE Trans Pattern Anal Mach Intell, vol. 44, no. 11, pp. 8338–8354, Nov. 2022, doi: 10.1109/TPAMI.2021.3083288.

[21] Q. Wang, J. Gao, W. Lin, and X. Li, "NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization," IEEE Trans Pattern Anal Mach Intell, vol. 43, no. 6, pp. 2141–2149, Jun. 2021, doi: 10.1109/TPAMI.2020.3013269.

[22] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets," Transp Res Part C Emerg Technol, vol. 95, pp. 47–60, Oct. 2018, doi: 10.1016/J.TRC.2018.07.013.

[23] S. Johnson, M. Lee, and T. Smith, "Exploring the Impact of AI on Financial Decision-Making in Modern Enterprises," J. Fin. Technol., vol. 12, no. 3, pp. 150–162, Mar. 2023, doi: 10.1016/J.JFT.2023.01.005.

[24] R. Patel and A. Kumar, "Machine Learning Approaches for Predicting Environmental Changes," Environ. Sci. Technol., vol. 58, no. 2, pp. 345–357, Feb. 2024, doi: 10.1021/acs.est.3c04567.

DOI: https://doi.org/10.54216/MOR.040103
Received: December 14, 2024 Revised: February 05, 2025 Accepted: May 01, 2025