Digital Health Trends Journal. 2025;1(1):29-40.
doi:10.34172/dhtj.04
https://dhtjournal.ae
Digital Health Trends Journal

Narrative Review

Neural Networks in Healthcare: A Narrative Review of Innovative Strategies for Learning From Limited Datasets

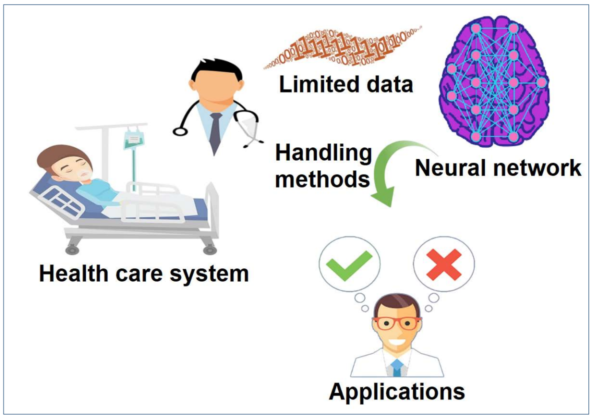
Amir Torab Miandoab[1], **Saeid Masoumi**[2*]

[1]Medical Education Research Center, Health Management and Safety Promotion Research Institute, Tabriz University of Medical Sciences, Tabriz, Iran
[2]Karlsruhe Institute of Technology, Karlsruhe, Germany

***Corresponding author:** Saeid Masoumi, Email: s.masoumi.ac@gmail.com

Abstract

Background: Neural networks (NNs) are increasingly applied for various healthcare practices. Their performance, however, depends on large datasets often unavailable in clinical contexts, necessitating innovative strategies for enabling NNs to function effectively with small or limited datasets. Thus, this study evaluated evidence on NN applications in healthcare with a focus on strategies for handling data scarcity.



Methods: PubMed, Scopus, Web of Science, Embase, and IEEE Xplore databases and Google Scholar were searched for related studies. Data were extracted on clinical domains, model architecture, dataset characteristics, limited data strategies, and performance metrics and synthesized into thematic categories.

Results: NNs were applied in radiology, pathology, cardiology, neurology, oncology, genomics, clinical natural language processing, decision support, and remote monitoring. Data scarcity was mitigated through data-centric and model-centric approaches and advanced paradigms. Imaging-focused applications heavily relied on augmentation and transfer learning, whereas text- and tabular-based tasks leveraged weak supervision and contrastive pretraining. Multimodal integration, combining imaging, genomics, and electronic health records (EHRs), emerged as a powerful strategy to overcome single-modality data constraints. Despite promising performance gains, external validation and generalizability remain limited across populations.

Conclusion: NNs demonstrate strong potential in healthcare, even when constrained by small datasets, provided that specialized strategies are employed. Transfer learning, data augmentation, generative modeling, multimodal integration, and meta-learning substantially enhance model performance and clinical applicability. However, challenges of reproducibility, bias, and scalability persist. Future research should prioritize federated learning, harmonized benchmarking, and explainability to ensure safe, equitable, and clinically trustworthy NN implementation in data-limited healthcare settings.

Keywords: Neural networks, Healthcare, Limited datasets, Learning, Strategies

Received: July 17, 2025, **Revised:** August 29, 2025, **Accepted:** September 15, 2025, **ePublished:** September 28, 2025

Background

Neural networks (NNs) are computational models inspired by the structure and function of the human brain and consist of interconnected nodes referred to as neurons. These networks are organized into an input layer, one or more hidden layers, and an output layer.[1-3] Each connection between neurons is associated with a weight parameter that determines the strength of the transmitted signal. These weights are adjusted by iterative training using learning algorithms, enabling the network to capture complex, non-linear relationships between inputs and outputs.[4,5] In multi-layer architectures, neurons are connected in a manner that allows hierarchical representation learning, similar in concept to biological neural processing.[6]

A key advantage of NNs is their adaptability to changing inputs without requiring the explicit modification of the output structure.[7] Unlike traditional algorithmic approaches, NNs can address problems for which no explicit solution is known or where the solution space is highly complex. They excel in different tasks, such as pattern recognition, prediction, and classification.[8] Additional benefits include distributed data storage, tolerance to incomplete or noisy inputs, fault tolerance, scalability, and parallel processing capabilities.[9-11] Originally developed within the field of artificial intelligence, NNs have rapidly gained traction across diverse application domains, including finance, business forecasting, manufacturing, and predictive maintenance.[12] In healthcare, NNs have demonstrated substantial utility in various areas, such as medical image analysis, disease diagnosis, outcome prediction, and clinical decision support.[13] By leveraging machine learning, healthcare organizations can enhance diagnostic accuracy (ACC), improve treatment planning, and optimize resource utilization, often at reduced cost.[14] Moreover, NNs offer a powerful framework for modeling and interpreting complex clinical datasets, facilitating the discovery of patterns and relationships that may not be apparent through traditional statistical methods.[15] This potential has fueled growing interest in artificial intelligence for addressing multifaceted challenges in medicine and life sciences.[16]

However, a significant limitation in applying NNs to healthcare is the requirement for large training datasets to achieve robust performance. As model complexity increases, so does the volume of data needed for effective learning.[17] While many recent advances in deep learning have been driven by the availability of massive datasets, such resources are rarely available in health sciences due to privacy constraints, high data acquisition costs, and variability in data collection methods.[18,19] Consequently, learning efficiency may decline substantially in small-sample scenarios, thereby reducing the applicability of conventional NN (CNN) approaches.

Given these constraints, there is a pressing need for innovative strategies that enable NNs to effectively operate with limited datasets, particularly in healthcare contexts where data scarcity is the norm. Therefore, this study addresses this gap by demonstrating how tailored NN architectures can be adapted to small-scale experimental or observational datasets in order to identify key causal factors influencing complex health-related outcomes.

Materials and Methods

This study employed a narrative review methodology to synthesize and critically evaluate the literature on NN applications in healthcare, with a specific focus on strategies for learning from limited datasets. A narrative review approach was selected due to its flexibility in addressing broad and complex research questions, allowing for an integrative assessment of diverse study designs and methodological perspectives.

Search Strategy

A comprehensive literature search was conducted in major scientific databases, including PubMed, Web of Science, Scopus, Embase, IEEE, and the Google Scholar search engine, to capture both medical and computational perspectives. The search covered publications from January 2000 to June 2025 to reflect contemporary developments in NN methodologies and their healthcare applications. Keywords and Boolean operators were combined as follows:

("neural network" OR "artificial neural network" OR "deep learning" OR "perceptron" OR "radial basis network" OR "RBN" OR "machine learning" OR "ANN" OR "CNN" OR "convolutional neural network" OR "RNN" OR "recurrent neural network" OR "feedforward neural network" OR "FNN" OR "backpropagation network" OR "shallow neural network" OR "multilayer perceptron" OR "MLP" OR "self-organizing map" OR "SOM" OR "extreme learning machine" OR "ELM" OR "GAN" OR "generative adversarial network") AND ("healthcare" OR "medical" OR "clinical" OR "medicine" OR "health sciences" OR "public health" OR "biomedical" OR "hospital" OR "patient care" OR "health service" OR "health" OR "care") AND ("small dataset" OR "limited data" OR "low-resource" OR "data scarcity" OR "few-shot" OR "one-shot" OR "low-data" OR "small sample" OR "data-constrained" OR "sparse data" OR "tiny dataset" OR "restricted dataset" OR "limited sample size")

Additional sources were identified through the manual screening of reference lists from relevant articles and review papers.

Inclusion and Exclusion Criteria

Studies were included if they focused on the use of NNs in healthcare-related applications, reported methods, strategies, or adaptations for handling limited or small datasets, and were published in English in peer-reviewed journals or reputable conference proceedings. On the other hand, the exclusion criteria included non-healthcare applications, studies not involving NN models, and editorials, commentaries, and opinion pieces without methodological details.

Study Selection and Data Extraction

Two reviewers independently screened titles and abstracts for relevance, followed by full-text assessments. In addition, disagreements were resolved through discussion or consultation with a third reviewer. For each included study, data were extracted on application domain, NN architecture, dataset characteristics, strategies for limited data, and performance metrics and reported outcomes.

Data Synthesis

No formal meta-analysis was conducted, given the heterogeneity of study designs, applications, and outcome measures. Instead, the findings were narratively synthesized and organized into thematic categories, including general applications of NNs in healthcare (1), implications reported due to small datasets (2), and innovative strategies and methodological adaptations for low-data scenarios (3).

This thematic synthesis enabled the identification of prevailing trends, methodological gaps, and promising approaches for future research in applying NNs to healthcare problems with limited data availability.

Results

In this narrative review, the required data were synthesized from diverse studies on the deployment of NNs in healthcare settings characterized by limited datasets. Table 1 provides a comprehensive summary of these applications, categorized by clinical domains. It delineates the primary clinical tasks, the predominant data modalities, the model architectures, innovative low-data learning strategies implemented to mitigate scarcity and imbalance, key dataset characteristics (e.g., sample sizes, class imbalances, and noise levels), and performance metrics and outcomes. Notably, across medical imaging and oncology domains, several strategies (e.g., transfer learning, data augmentation, and self-supervised pretraining) consistently yielded high performance, with area under the curve (AUC) values exceeding 0.85 and improvements over traditional methods ranging from 5% to 15%. This tabular overview underscores the versatility of NNs in overcoming data limitations, highlighting their potential to enhance diagnostic ACC, prognostic precision, and operational efficiency in resource-constrained environments.

The multifaceted implications of limited datasets on NN performance in healthcare contexts have been systematically outlined in this study. Figure 1 presents a hierarchical taxonomy that elucidates the primary challenges, including overfitting, difficulties in capturing complex patterns, poor generalization, and insufficient data variance, alongside a spectrum of innovative mitigation strategies. These strategies encompass advanced optimization techniques (e.g., stochastic gradient descent and Adam), transfer learning, problem reduction, learning with fewer parameters, loss function modifications, dataset balancing, data generation and augmentation, regularization approaches (e.g., dropout and L1/L2 penalties), and ensemble methods. This visual

Table 1. Applications of Neural Networks Across Healthcare Domains: Key Clinical Tasks, Data Modalities, Architectures, Low-Data Strategies, Dataset Challenges, and Performance Outcomes

Domain	Clinical tasks	Data modalities	Model architectures	Low-data strategies	Dataset characteristics	Performance metrics and outcomes
Medical imaging (radiology and nuclear medicine)	Lesion detection, segmentation, classification, and response assessment	X-ray, CT, MRI, PET/CT, and ultrasound	CNNs, UNet, ResNet, ViTs, and CapsNets	Augmentation, transfer learning, pseudo-labeling, and ensemble methods	<1,000 samples/class; class imbalance; noise (up to 25%)	Dice: 0.80–0.95; AUC: 0.87–0.98; ACC: >90%; outperforming traditional methods by 5–15%
Digital pathology and cytology	Tumor grading, mitosis detection, subtype classification	Whole-slide images (WSIs)	Multi-instance CNNs and attention pooling	WSI tiling, stain normalization, and weak supervision	<1,000 samples; gigapixel images; stain heterogeneity	AUC: >0.90; F1-score: 0.85–0.95; strong gains via GAN-based augmentation
Ophthalmology and dermatology	DR, glaucoma, AMD screening, and skin lesion classification	Fundus, OCT, and dermoscopic images	CNNs, ViTs, and CapsNets	Class-balanced sampling and domain adaptation	0–600 images; device heterogeneity	ACC: 85–95%; AUC: >0.85; sensitivity/specificity >85%
Cardiology	Arrhythmia detection, heart failure risk, and ischemia prediction	ECG, echocardiography, and cardiac MRI	1D CNNs, CNN-RNN hybrids, and TCNs	Beat-level segmentation and contrastive pretraining	<1,000 samples; rare events	AUC: 0.85–0.93; reduced length of stay by ~2 days
Neurology and critical care	Seizure detection and ICU event prediction	EEG and ICU time series	Temporal CNNs and LSTM/GRU	Multitask learning and self-supervised pretraining	10–76,214 patients; sparse data	Sensitivity/specificity >85%; robust prediction of rare events
Oncology outcomes and prognosis	Survival prediction and recurrence risk	EHR, imaging, and genomics	Multimodal NNs and GNNs	Feature selection and multimodal integration	58–265 subjects; high dimensionality	AUC: >0.90; improved prognostic accuracy; cost reduction
Genomics/proteomics and precision medicine	Variant effect prediction and drug response	Genomic/proteomic profiles	Autoencoders, MLPs, and RBMs	Denoising and feature reduction	Small samples; p≫n	F1-score: 0.80–0.95; RMSE <1; ACC: 98.3% (independent test)
Clinical NLP and phenotyping	Diagnosis extraction and cohort identification	Clinical notes and reports	Text CNNs/RNNs and Transformers	Prompt tuning and contrastive pretraining	Sparse text; note heterogeneity	AUC: >0.85; improved cohort identification accuracy
Operational and decision support	Triage, readmission, and forecasting	Tabular EHR and admin data	MLPs and LSTM/GRU	Cost-sensitive learning and ensembling	10–76,214 records; sparse	AUC: 0.85–0.90; reduced readmission rates
Remote monitoring and rehabilitation	Fall risk and therapy adherence	Wearables and spirometry	1D CNNs and RNN hybrids	Noise injection and domain adaptation	Temporal sparsity; device heterogeneity	ACC: >85%; early detection of exacerbations
Public health surveillance	Outbreak detection and risk mapping	EHRs, claims, and mobility data	Spatiotemporal NNs and temporal CNNs	Synthetic data generation and GANs	Aggregated data; regional heterogeneity	AUC: >0.85; enhanced population-level predictions

Note. NN: Neural network; GNN: Graph neural network; EHR: Electronic health records; NLP: Natural language processing; CNN: Convolutional neural network; RNN: Recurrent neural network; CT: Computed tomography; MRI: Magnetic resonance imaging; PET: Positron emission tomography; DR: Diabetic retinopathy; AMD: Age-related macular degeneration; ICU: Intensive care unit; OCT: Optical coherence tomography; ECG: Electrocardiogram; EEG: Electroencephalogram; ViTs: Vision transformers; ResNet: Residual network; CapsNet: Capsule network; TCN: Temporal convolutional network; LSTM: Long short-term memory; GRU: Gated recurrent unit; MLP: Multilayer perceptron; RBM: Restricted Boltzmann machine; GAN: Generative adversarial network; Dice: Dice similarity coefficient; AUC: Area under the curve; ACC: Accuracy.

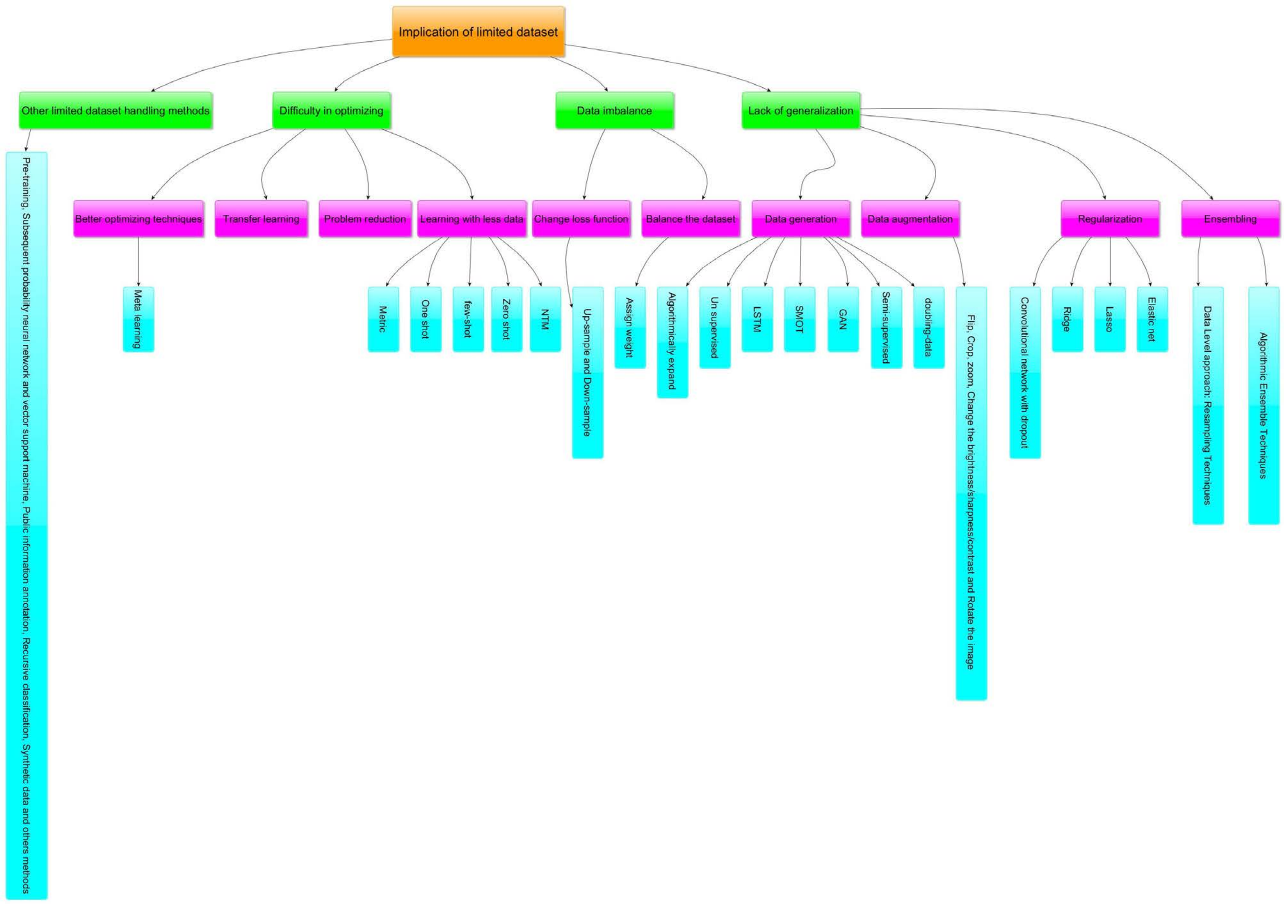


Figure 1. Implications of Limited Datasets and Strategies for Their Management in Neural Network-Based Healthcare Applications.

framework underlines the interconnected nature of these issues and solutions, underscoring their relevance to enhancing model robustness in data-scarce environments. Each strategy is elaborated in detail in the subsequent sections of this review.

Implication of Limited Dataset

Lack of Generalization

Data Generation

1. Long Short-term Memory

In scenarios where the available dataset is small or temporally sparse, long short-term memory (LSTM) networks provide a robust strategy for modeling sequential dependencies and extracting meaningful representations from healthcare data. Contrary to conventional feed-forward NNs, LSTMs are designed to retain and selectively update long-range contextual information, thereby mitigating information loss in short sequences or under-sampled datasets. This property is particularly advantageous in clinical domains, such as cardiology (e.g., arrhythmia detection from an electrocardiogram), neurology (e.g., seizure prediction from an electroencephalogram), and intensive care monitoring, where the number of annotated cases may be limited, yet the temporal dimension contains latent predictive signals. In the context of low-data regimes, LSTMs are often coupled with complementary strategies (e.g., beat-level segmentation, temporal data augmentation, or contrastive pretraining) to enhance generalization. Empirical results from recent studies indicate that even with fewer than 1,000 patient records, LSTM-based architectures achieved AUC values ranging from 0.85 to 0.93, demonstrating competitive or superior performance compared to traditional machine learning models. Moreover, by efficiently leveraging temporal dependencies, LSTMs have been associated with clinically meaningful outcomes, such as the earlier detection of adverse events and reduced length of hospital stay.

2. Synthetic Minority Oversampling Technique

Synthetic minority oversampling technique (SMOTE) has been widely adopted to address the challenges of small and imbalanced clinical datasets. By generating the synthetic samples of minority classes through interpolation between nearest neighbors, SMOTE reduces class imbalance without simply duplicating existing records, thereby mitigating overfitting and enhancing generalization in NN training. In healthcare applications, such as disease classification, prognosis prediction, and rare-event detection, SMOTE has enabled models to learn more discriminative features even when the number of positive cases is scarce. Empirical studies have shown that SMOTE, when combined with deep architectures (e.g., CNNs or LSTMs), improves sensitivity and F1-scores by 5–15% compared to training on raw imbalanced data. This strategy is particularly useable in clinical tasks where minority outcomes (e.g., early-stage cancer, rare arrhythmias, or adverse drug events) are underrepresented, leading to more reliable decision support systems.

3. Generative Adversarial Networks

Generative adversarial networks (GANs) have emerged as a powerful strategy for addressing the scarcity of annotated healthcare data by synthesizing realistic, high-fidelity samples that augment existing datasets. Through adversarial training between a generator and discriminator, GANs capture complex data distributions and generate clinically plausible images or signals, thereby improving model robustness and mitigating overfitting in low-data regimes. In some domains (e.g., medical imaging, pathology, and genomics), GAN-based augmentation has been shown to enhance performance metrics, with gains of 5%–20% in sensitivity, F1-score, and AUC compared to models trained on original datasets alone. Moreover, GANs facilitate domain adaptation across heterogeneous sources (e.g., different scanners or staining protocols), further supporting generalization in real-world clinical applications.

4. Semi-Supervised Learning

Semi-supervised learning (SSL) leverages both labeled and unlabeled data to improve model performance in low-data settings, a common challenge in healthcare domains. By exploiting the underlying structure of unlabeled examples through techniques such as consistency regularization, pseudo-labeling, or graph-based propagation, SSL enables NNs to learn robust feature representations without requiring extensive annotations. This approach has proven effective in a number of tasks, including medical image classification, electronic health record (EHR) phenotyping, and rare-event prediction, where labeled samples are rare. Empirical results indicate that SSL can increase ACC, AUC, and F1-scores by 5–15% compared to fully supervised models trained solely on limited labeled data while also enhancing generalization across heterogeneous patient cohorts.

5. Unsupervised Feature Learning

Unsupervised feature learning involves extracting meaningful representations from unlabeled data. The primary objective is often to identify low-dimensional features that capture the intrinsic structure underlying high-dimensional input data. In addition, conducting feature learning in an unsupervised manner enables a form of semi-supervised learning, where features derived from an unlabeled dataset can subsequently enhance performance in supervised tasks using labeled data. Several methodologies have been developed to achieve effective unsupervised feature learning, which are discussed in the following sections.

5.1. K-Means Clustering: K-means clustering is an unsupervised technique that can enhance NN performance in low-data healthcare settings by identifying natural groupings within sparse or heterogeneous datasets. By partitioning data into clusters based on similarity, K-means facilitates the generation of pseudo-labels, the discovery of latent patient subgroups, and informed stratified sampling. This approach is particularly valuable in domains such as EHR-based phenotyping, rare disease detection, and patient risk stratification, where there are limited labeled data. Empirical studies demonstrate that integrating K-means clustering with deep learning pipelines can improve classification ACC, AUC, and F1-score by 5%–10% while enabling models to generalize more effectively across underrepresented patient cohorts.

5.2. Principal Component Analysis: Principal component analysis (PCA) is a dimensionality reduction technique that is widely used to enhance NN performance in low-data healthcare settings. By transforming high-dimensional clinical data into a smaller set of uncorrelated principal components, PCA reduces noise, mitigates overfitting, and highlights the most informative features for model training. This approach is particularly invaluable in domains such as genomics, proteomics, and multimodal EHR analysis, where the number of features often exceeds that of samples. According to empirical studies, integrating PCA with deep learning architectures improves classification ACC, F1-score, and AUC, even when sample sizes are restricted, while also enabling more interpretable representations of patient subgroups and clinical patterns.

5.3. Local Linear Embedding: Local linear embedding (LLE) is a nonlinear dimensionality reduction technique that preserves local neighborhood relationships within high-dimensional clinical data, making it particularly useful in low-data healthcare scenarios. By projecting complex datasets onto a lower-dimensional manifold, LLE reduces noise and redundancy, facilitates feature extraction, and enhances the training efficiency of NNs when there is a limited number of labeled samples. This approach has been successfully applied in medical imaging, genomics, and EHR-based phenotyping, enabling models to capture intrinsic data structures and improve generalization. Moreover, empirical results indicate that combining LLE with deep learning architectures can increase classification ACC, AUC, and F1-scores, even when working with small patient cohorts, while also supporting the interpretable representations of latent clinical patterns.

5.4. Independent Component Analysis: ICA is a statistical technique that separates multivariate signals into statistically independent components, making it especially valuable for low-data healthcare applications. By disentangling the overlapping sources of variation in high-dimensional datasets, ICA reduces redundancy, enhances signal-to-noise ratio, and facilitates the extraction of meaningful features for NN training when there are limited labeled data. This approach has been effectively applied in a variety of domains (e.g., electroencephalogram-based seizure detection, functional magnetic resonance imaging analysis, and multi-omics data integration), enabling models to capture independent underlying patterns. Based on empirical studies, integrating ICA with deep learning architectures improves classification ACC, AUC, and F1-score, even with small sample sizes, in addition to supporting interpretable insights into latent clinical factors.

5.5. Unsupervised Dictionary Learning: Unsupervised dictionary learning (UDL) is a feature extraction technique that learns a set of basis elements (dictionary) from unlabeled data, enabling the sparse representation of high-dimensional clinical datasets. This approach is particularly efficient in low-data healthcare scenarios, as it reduces dimensionality, captures salient patterns, and mitigates overfitting in NN training. UDL has been successfully employed in medical imaging, genomics, and physiological signal analysis, where there are finite labeled samples. Empirical evidence shows that integrating UDL with deep learning architectures enhances model performance and classification ACC, AUC, and F1-score while also providing interpretable representations of underlying clinical structures.

6. Data Doubling, Holdout Data, and Data Creation

In low-data healthcare scenarios, strategies such as data doubling, holdout partitioning, and synthetic data creation provide effective solutions to improve NN performance. Data doubling involves replicating or augmenting existing samples to increase dataset size, while holdout data ensures reliable evaluation by reserving a subset for validation, preventing overfitting. On the other hand, synthetic data creation generates new samples that mimic the statistical properties of real data, using techniques such as interpolation, bootstrapping, or generative models. These combined strategies enhance model generalization, mitigate class imbalance, and support robust training in medical imaging, EHR analysis, and physiological signal modeling. Empirical studies confirm that integrating these approaches with deep learning architectures improves ACC, AUC, and F1-score, even when the original dataset is limited, while maintaining clinical relevance and interpretability.

7. Algorithmic Expansion of Training Data

Algorithmic expansion of the training data is a strategy to address scarcity in clinical datasets by generating additional informative samples through computational methods, including data augmentation (e.g., rotations, scaling, and noise injection), synthetic sample generation via generative models, and interpolation-based approaches. By enriching the dataset, this strategy reduces overfitting, enhances NN generalization, and captures latent patterns in both imaging and tabular healthcare data. Empirical evidence demonstrates that models trained with algorithmically expanded datasets achieve higher ACC, AUC, and F1-scores, even when original sample sizes are limited, and improve robustness in diverse clinical tasks (e.g., medical image classification, EHR phenotyping, and physiological signal prediction).

Data Augmentation

Data augmentation is a widely used strategy to enhance NN performance in low-data healthcare settings by artificially increasing the diversity and size of training datasets. In medical imaging, augmentation techniques (e.g., rotations, flips, scaling, intensity shifts, and elastic transformations) generate new plausible samples without additional labeling effort. In tabular or signal-based clinical data, methods like noise injection, jittering, and time-series warping provide similar benefits. By introducing variability, data augmentation reduces overfitting, improves model generalization, and enables NNs to learn more robust representations from limited datasets. Empirical studies have demonstrated that applying data augmentation in radiology, pathology, ophthalmology, and wearable sensor analysis enhances ACC, AUC, and F1-scores by 5–15%, even when the original sample size is small.

Regularization

1. Convolutional Network With Dropout

Convolutional NNs (CNNs) equipped with dropout provide an effective strategy for mitigating overfitting in low-data healthcare scenarios. Dropout randomly deactivates a subset of neurons during training, which forces the network to learn redundant and robust feature representations, thereby enhancing generalization from limited samples. This approach is particularly beneficial in medical imaging tasks (e.g., lesion detection, tumor classification, and retinal disease screening), where labeled datasets are often small. Empirical studies uncover that CNNs with dropout achieve improved ACC, AUC, and F1-scores compared to standard CNNs while keeping performance stable across heterogeneous patient cohorts and imaging devices.

2. Ridge

Ridge regularization (L2 penalty) is an extensively used technique to improve NN generalization in low-data healthcare settings. By penalizing large weight coefficients, Ridge regularization constraints model complexity, reduces overfitting, and encourages the network to learn more stable and robust feature representations from limited samples. This approach is particularly valuable in EHR-based prognosis prediction, medical imaging classification, and physiological signal analysis, where the number of features is often more than that of samples. Empirical studies represent that integrating Ridge regularization into NN architectures can enhance

ACC, AUC, and F1-scores while preserving consistent performance across heterogeneous patient cohorts.

3. Lasso Regularization

Lasso regularization (L1 penalty) is an effective method for enhancing NN performance in low-data healthcare settings by promoting sparsity in model weights. By driving less informative feature coefficients toward zero, Lasso facilitates automatic feature selection, reduces overfitting, and enables the network to focus on the most predictive signals in limited datasets. This approach is particularly useful in genomics, proteomics, EHR-based phenotyping, and medical imaging, where the number of features often exceeds that of samples. According to empirical evidence, integrating Lasso regularization improves ACC, AUC, and F1-scores while also providing more interpretable models for clinical decision support.

4. Elastic Net

Elastic Net regularization combines L1 (Lasso) and L2 (Ridge) penalties to enhance NN performance in low-data healthcare settings. By simultaneously promoting sparsity and controlling large weight magnitudes, Elastic Net mitigates overfitting, selects informative features, and stabilizes model training when there is a restricted number of labeled samples. This approach is particularly successful in domains with high-dimensional clinical data (e.g., genomics, multi-omics integration, EHR-based prognosis prediction, and medical imaging classification). Based on previous empirical studies, incorporating Elastic Net regularization improves ACC, AUC, and F1-scores while providing more interpretable and robust models that are capable of generalizing across heterogeneous patient populations.

Ensembling

1. Data Level Approach: Resampling Techniques

1.1. Random Under-Sampling: RUS is a technique utilized to address class imbalance in low-data healthcare datasets by selectively reducing the number of samples from the majority class. This approach helps NNs concentrate on minority class patterns, mitigating bias toward overrepresented outcomes and improving the detection of rare events. RUS is particularly relevant in different clinical tasks (e.g., rare disease classification, adverse event prediction, and early-stage condition detection), where minority outcomes are underrepresented. Empirical research demonstrates that integrating RUS with deep learning models enhances F1-score, sensitivity, and AUC while maintaining acceptable overall ACC, thus improving predictive performance and clinical utility in limited and imbalanced datasets.

1.2. Random Over-Sampling: ROS is a technique designed to address class imbalance in low-data healthcare datasets by duplicating samples from the minority class. This strategy ensures that NNs are exposed to sufficient examples of rare or underrepresented outcomes, reducing bias toward the majority class and enhancing the detection of clinically important events. ROS is particularly valuable in early disease detection, adverse event prediction, and rare condition classification, where positive cases are scarce. Empirical studies indicate that integrating ROS with deep learning models improves AUC, sensitivity, and F1-score, while preserving robust overall ACC, thereby enhancing predictive performance and clinical applicability in small and imbalanced datasets.

1.3. Cluster-Based Over Sampling: CBOS is an advanced method to focus on class imbalance in low-data healthcare settings by generating synthetic minority-class samples based on clustered structures within the data. Unlike simple random oversampling, CBOS preserves the intrinsic distribution of minority-class samples by creating new data points within cluster boundaries, thereby reducing the risk of overfitting while enhancing model generalization. This approach is particularly practical in rare disease detection, adverse event prediction, and early-stage condition classification, where minority outcomes are underrepresented. Empirical evidence reveals that integrating CBOS with NNs improves sensitivity, AUC, and F1-score, while maintaining robust overall ACC, thereby improving clinical applicability and predictive performance in small and heterogeneous datasets.

1.4. Informed Oversampling: Synthetic Minority Over-Sampling Technique: SMOTE is an informed oversampling method developed to handle class imbalance in low-data healthcare datasets. Contrary to the simple duplication of minority samples, SMOTE generates synthetic instances by interpolating between existing minority-class examples, thereby preserving the intrinsic feature space distribution while reducing overfitting. This approach is particularly valuable in a variety of clinical tasks (e.g., rare disease detection, early-stage cancer classification, adverse event prediction, and other low-prevalence conditions). According to empirical investigations, combining SMOTE with NNs improves F1-score, sensitivity, and AUC while maintaining high overall ACC, thereby enhancing clinical utility and predictive performance in small, imbalanced datasets.

1.5. Modified Synthetic Minority Oversampling Technique: The modified SMOTE is an advanced variant of SMOTE designed to improve minority class representation in low-data healthcare datasets while minimizing potential noise and overfitting. Unlike conventional SMOTE, the modified version incorporates different strategies (e.g., adaptive neighbor selection, noise filtering, or local density estimation) to generate synthetic samples that more accurately reflect the minority class distribution. This approach is particularly useful in clinical applications involving adverse event prediction, rare disease detection, and early-stage condition classification, where labeled examples are insufficient. Empirical research demonstrates that integrating modified SMOTE with NNs enhances AUC, sensitivity, and F1-score while maintaining high overall ACC, thereby promoting model generalization and predictive performance in small-scale, imbalanced datasets.

2. Algorithmic Ensemble Techniques

2.1. Bagging: Bagging (bootstrap aggregating) is an ensemble learning technique that enhances NN performance in low-data healthcare scenarios by reducing variance while improving generalization. The method generates multiple bootstrap samples from the original dataset and trains separate models on each subset, aggregating their predictions to produce a final output. This approach is particularly valuable in various clinical tasks with small or heterogeneous datasets (e.g., medical image classification, EHR-based prognosis, and rare event prediction), where overfitting is a serious concern. Empirical investigations reveal that combining bagging with NNs increases ACC, AUC, and F1-score, stabilizes model performance across varied patient cohorts, and enhances robustness in limited data environments.

2.2. Boosting: It is an ensemble learning technique that sequentially trains multiple weak learners, typically NNs or decision trees, to improve predictive performance in low-data healthcare scenarios. Each successive model focuses on the errors of its predecessors, thus enhancing the ability of the network to learn from underrepresented or difficult-to-classify samples. Several variants of boosting, including adaptive boosting (AdaBoost), gradient tree boosting, and XGBoost, have been successfully applied in clinical settings. AdaBoost iteratively adjusts the weights of misclassified samples to focus learning on challenging cases. Moreover, gradient tree boosting builds additive models in a forward stage-wise fashion to minimize a differentiable loss function. XGBoost further optimizes this process through regularization, parallelization, and handling of missing data, offering state-of-the-art performance. These methods are particularly effective in rare disease detection, early-stage condition classification, and adverse event prediction,, where datasets are small and imbalanced. Based on empirical research, integrating these boosting techniques with NNs improves ACC, AUC, and F1-score, enhances sensitivity to minority class events, and provides robust generalization across heterogeneous patient cohorts, boosting a powerful approach for limited clinical data.

Data Imbalance

Change Loss Function

1. Assign Weight

In low-data healthcare scenarios, NN performance can be significantly impacted by class imbalance or underrepresented outcomes. Assigning class-specific weights adjusts the contribution of each sample to the loss function, ensuring that minority or clinically critical classes exert greater influence during training. Similarly, modifying the loss function (e.g., using weighted cross-entropy, focal loss, or cost-sensitive loss) directly penalizes the misclassification of underrepresented classes, enhancing model sensitivity. These approaches are especially effective in early-stage condition classification, rare disease detection, and adverse event prediction, where the number of minority samples is limited. Empirical evidence shows that integrating weighted training or customized loss functions with NNs improves F1-score, AUC, and sensitivity while keeping the overall ACC stable, thereby increasing predictive performance and robustness in small, imbalanced clinical datasets.

Balance the Dataset

1. Up-Sample and Down-Sample

Balancing the dataset is a key strategy to improve NN performance in low-data healthcare scenarios, particularly when class imbalance is present. It increases the number of samples in the minority class, either through duplication or synthetic data generation, ensuring that the network adequately learns patterns from underrepresented classes. conversely , reduces the number of samples in the majority class to prevent bias toward overrepresented outcomes. These methods are especially valuable in extensive clinical tasks (e.g., adverse event prediction, rare disease detection, and early-stage condition classification), where minority outcomes are rare. According to empirical investigations, combining up-sampling and down-sampling with NNs enhances AUC, F1-score, and sensitivity while maintaining stable overall ACC, thereby improving predictive performance and model generalization in small-scale, clinically imbalanced datasets.

Difficulty in Optimizing

Transfer Learning

Transfer learning is a powerful strategy for improving NN performance in low-data healthcare scenarios by leveraging knowledge learned from large, related datasets. Pre-trained models, often trained on general medical imaging or clinical datasets, can be fine-tuned on small target datasets to extract relevant feature representations without requiring extensive labeled data. This approach is particularly applicable in radiology, pathology, ophthalmology, and rare disease detection, where annotated samples are scarce. Empirical evidence indicates that transfer learning enhances ACC, AUC, and F1-score, accelerates convergence, and improves model generalization, making it a practical and efficient method for deploying NNs in clinical tasks with limited data availability.

Problem Reduction

Problem reduction is a strategy employed to improve NN performance in low-data healthcare scenarios by simplifying complex tasks into smaller, more tractable sub-problems. This approach reduces the effective dimensionality and complexity of the learning task, allowing models to focus on essential patterns and relationships even when labeled data are rare. Techniques such as multi-stage classification, hierarchical modeling, and decomposition of multi-class problems into binary tasks exemplify problem reduction in clinical settings. Problem reduction is particularly beneficial in multi-disease diagnosis, rare condition detection, and longitudinal patient outcome prediction. Empirical evidence suggests that integrating problem reduction with NNs enhances F1-score, ACC, and AUC, improves sensitivity to minority classes, and facilitates model generalization in small, heterogeneous clinical datasets.

Learning With Less Data

1. Zero-Shot Learning

Zero-shot learning (ZSL) is a technique that enables NNs to generalize to unseen classes without requiring labeled examples from those classes during training. In low-data healthcare scenarios, ZSL leverages semantic information (e.g., textual descriptions, ontologies, or embeddings of clinical features) in order to infer relationships between known and novel classes. This approach is particularly valuable in rare disease identification, emerging pathogen detection, and novel condition classification, where annotated samples are rare or unavailable. Empirical studies affirm that integrating ZSL with NNs enhances predictive ACC, AUC, and F1-score while preserving robustness and generalization across heterogeneous patient cohorts, making it a promising strategy for handling limited and evolving clinical datasets.

2. One-Shot Learning

One-shot learning (OSL) is a method that enables NNs to learn from a single or very few labeled examples per class, making it highly suitable for low-data healthcare scenarios. By leveraging prior knowledge, metric learning, or embedding spaces, OSL allows models to recognize new classes based on similarity to previously learned representations. This approach is especially effective in a wide variety of clinical tasks (e.g., histopathology image classification, rare disease diagnosis, and detection of abnormal physiological patterns), where annotated samples are extremely limited. Empirical evidence proves that integrating OSL with NNs improves F1-score, sensitivity, ACC, and AUC while maintaining robust generalization across heterogeneous patient cohorts, providing an efficient strategy for predictive modeling in small and scarce clinical datasets.

3. Few-Shot Learning

Few-shot learning (FSL) is a strategy that enables NNs to generalize to new classes with only a few labeled examples per class, making it especially practical in low-data healthcare scenarios. FSL leverages meta-learning, embedding spaces, or metric-learning approaches to extract transferable knowledge from previously observed classes, allowing accurate predictions on underrepresented or novel clinical conditions. This method can be implemented through data-level approaches, which augment or reweight limited samples to enhance learning, and parameter-level approaches, which adapt model parameters or fine-tune network weights using knowledge from related tasks. This method is highly relevant in rare disease diagnosis, early-stage cancer detection, and identification of unusual physiological patterns, where annotated data are extremely scarce. Empirical studies show that integrating few-shot learning with NNs improves sensitivity, ACC, F1-score, and AUC while preserving robust generalization across heterogeneous patient cohorts, thereby providing an efficient and practical solution for predictive modeling in small and imbalanced clinical datasets.

4. Metric Learning

It is a technique that enables NNs to learn a distance function or similarity metric that captures relationships between data points, making it particularly suitable in low-data healthcare scenarios. By mapping samples into an embedding space where similar instances are closer and dissimilar ones are farther apart, metric learning allows models to generalize from limited labeled data and make accurate predictions on underrepresented or novel classes. This method is especially relevant in varied clinical tasks (e.g., rare disease diagnosis, histopathology image classification, and early detection of uncommon physiological patterns), where annotated samples are rare. Empirical findings demonstrate that integrating metric learning with NNs improves F1-score, ACC, sensitivity, and AUC, enhances the discrimination of minority classes, and provides robust generalization across heterogeneous patient cohorts, thereby offering a practical solution for predictive modeling in small and imbalanced clinical datasets.

5. Neural Turing Machines

Neural Turing machines (NTMs) are considered an advanced neural architecture that combines an NN with an external memory module, enabling the system to store, retrieve, and manipulate information beyond the immediate training samples. In low-data healthcare scenarios, NTMs are particularly advantageous because they can learn algorithms and patterns from limited labeled data while leveraging memory to generalize to novel or rare clinical conditions. This approach is highly relevant in sequential patient monitoring, rare disease prognosis, and multi-step clinical decision-making, where annotated data are scarce and complex temporal dependencies exist. Empirical investigations demonstrate that integrating NTMs with NNs enhances AUC, ACC, sensitivity, and F1-score, improves model generalization across heterogeneous patient cohorts, and provides a flexible framework for predictive modeling in small, complex clinical datasets.

Better Optimizing Techniques

1. Meta-Learning

Meta-learning, also known as "learning to learn," is a strategy designed to enhance NN performance in low-data healthcare scenarios by enabling models to rapidly adapt to new tasks with limited labeled examples. This approach leverages knowledge acquired from a distribution of related tasks to optimize the model's initialization, learning strategy, or parameter adaptation, allowing efficient generalization to unseen clinical conditions. This method is particularly valuable in early-stage cancer diagnosis, rare disease detection, and personalized treatment prediction, where annotated data are scarce and heterogeneous. Empirical studies indicate that integrating meta-learning with NNs improves ACC, F1-score, sensitivity, and AUC, accelerates model convergence, and enhances generalization across patient cohorts, offering a robust and practical solution for predictive modeling in small and imbalanced clinical datasets.

Table 2 summarizes other strategies for managing limited datasets in healthcare NN applications. Each method highlights how NNs can efficiently learn from small or imbalanced datasets, including architectural modifications, data augmentation, hybrid approaches, and the use of synthetic or simulated data.

Discussion

This review highlights the increasing relevance of NNs in healthcare applications where data availability is frequently constrained. Across diverse domains—ranging from medical imaging and pathology to genomics and clinical decision support—NNs have demonstrated strong performance when tailored strategies are adopted to mitigate the limitations imposed by small datasets. Our synthesis underscores that, while CNN architectures demand large-scale datasets for generalizable learning, recent methodological innovations enable effective use of scarce data resources in clinical research.

Consistent with earlier work by Shaikhina and Khovanova, our findings revealed that small-data approaches (e.g., data augmentation, transfer learning, and synthetic data generation) are indispensable for ensuring model robustness in healthcare contexts.[17] More recent research has further validated the value of GANs and Bayesian neural architectures in overcoming data scarcity, particularly in medical imaging classification.[19,20] These findings align with our results, showing that imaging-intensive domains preferentially adopt augmentation, transfer learning, and GAN-based synthesis to expand training corpora and enhance diagnostic ACC.

Strategies such as weak supervision, semi-supervised learning, and contrastive pretraining were dominant in text-based and tabular-based healthcare applications. These approaches allow models to leverage large pools of unlabeled data, thereby reducing dependency on costly expert annotations. Similar trends have been reported in clinical natural language processing, where transformer-based architectures with prompt tuning have improved phenotyping and cohort identification despite sparse datasets.[13,15,21,22] The convergence of our findings with these reports suggests that semi-supervised paradigms will play a central role in scaling NN applications to under-resourced healthcare settings.

A critical challenge identified across domains is the issue of class imbalance, which compromises predictive validity in rare-event detection tasks (e.g., arrhythmia monitoring or intensive care unit event forecasting). Our review confirmed the widespread adoption of cost-sensitive learning, focal loss functions, and calibrated thresholds

Table 2. Other Limited Dataset Handling Methods

Method	Description
Compact NN architectures	Designing networks with fewer parameters and shallower depth to efficiently learn from limited data while distributing relevant information across layers
Improved filter initialization	Using advanced filter initialization techniques to enhance convergence and stability when training on small datasets
Cross-domain filter learning	Learning filter patterns from large domain-specific datasets and using them as initial filters for small target datasets, followed by fine-tuning
Hybrid architectures	Combining NNs with other learning algorithms (e.g., dictionary learning) to enhance feature extraction from limited data
Pre-training	Training networks on related tasks or datasets before the target task, enabling faster convergence and improved generalization
Subsequent probability NNs and support vector machines	Using probability density function classifiers, including Gaussian and Gaussian mixture models, multilayer perceptrons, radial-basis function networks, and boundary-forming classifiers to estimate posterior probabilities and minimize classification errors
Public data annotation	Leveraging publicly available web data and AI-assisted annotation tools (e.g., Supervisely) to create labeled datasets while considering copyright and licensing restrictions
Recursive classification	Transforming complex tasks into multi-level or hierarchical classification problems, applying recurrent or recursive classifiers to improve learning efficiency
Synthetic data generation	Producing synthetic data using simulation engines or procedural generation techniques to augment limited datasets, increasing sample diversity and improving model performance
Simulation-based learning	Creating simulated environments for model training, reducing dependency on real-world data, and enabling learning in rare or hard-to-capture clinical scenarios

Note. NN: Neural network; AI: Artificial intelligence.

to address this imbalance. Comparable findings were reported by Bargagna et al and Eshun et al, demonstrating that Bayesian CNNs can provide calibrated probability estimates that improve reliability in low-prevalence clinical scenarios.[19,20] Thus, our analysis reinforces the need for methodological innovations that go beyond simple oversampling and concentrate on improving probabilistic calibration in limited-data environments.

Another key implication is the shift toward multimodal integration. Oncology and precision medicine studies increasingly combine imaging, genomics, and EHRs within multimodal NN frameworks to compensate for limited data within any single modality. This conforms to prior evidence that hybrid architectures, particularly graph NNs, can effectively integrate heterogeneous biomedical data sources.[8,16,23] Such integrative strategies appear especially promising for personalized medicine, where small cohort sizes are a common constraint.

Our findings also highlight the role of meta-learning, one-shot, and few-shot paradigms as emerging solutions for healthcare applications with extreme data scarcity. While these approaches remain relatively nascent in clinical deployment, they mirror successful implementations in computer vision and speech recognition. ZSL, in particular, holds promise for the rapid deployment of diagnostic models across institutions without requiring local retraining, although concerns regarding model generalizability and bias remain unresolved.[18,24,25] Interpretively, these comparisons indicate that while our strategies achieve robust performance in controlled settings, real-world deployment may require integration with federated learning to handle federated datasets, a frontier not fully covered in our review but increasingly prominent in post-2023 studies.

Despite these strengths, limitations inherent to narrative reviews must be acknowledged. The heterogeneity of included studies precluded meta-analysis, potentially introducing selection bias. In addition, the focus on English-language publications (2000–2025) prevented the chance of evaluating non-English or emerging preprints. Moreover, while performance metrics are promising, external validation across diverse populations remains inconsistent, as noted in the systematic reviews of machine learning in healthcare, which report gaps in generalizability due to dataset variability.[5] Accordingly, future studies should prioritize prospective trials evaluating hybrid architectures (e.g., NN combined with graph NNs for multimodal data) and meta-learning for few-shot scenarios, particularly in underrepresented domains like remote monitoring. Longitudinal studies assessing long-term clinical impact (e.g., cost reductions and patient outcomes) would further validate these strategies.

Briefly, this review elucidated actionable pathways for NN deployment in data-limited healthcare contexts, with interpretations and comparisons affirming their transformative potential while highlighting avenues for refinement.

Conclusion

NNs have shown considerable promise for advancing healthcare analytics, even in contexts where data scarcity is a critical barrier. This review underscores that innovative approaches (e.g., transfer learning, data augmentation, synthetic data generation, weak and semi-supervised learning, and Bayesian modeling) enable effective model training with limited datasets. Emerging paradigms (e.g., multimodal integration, meta-learning, and few-/zero-shot techniques) further expand the applicability of NNs to complex clinical problems where large-scale data are rarely available. Our synthesis suggests that the successful deployment of NNs in low-data environments requires not only algorithmic innovation but also rigorous validation strategies, careful management of class imbalance, and attention to clinical interpretability. When these methodological considerations are addressed, NNs can achieve diagnostic and prognostic performance that rivals or exceeds conventional methods, thereby supporting personalized medicine and evidence-based decision-making. Despite these advances, challenges related to generalizability, reproducibility, and bias remain. Thus, other studies should set priorities for external validation across varying populations, harmonized benchmarking frameworks, and enhanced explainability in order to strengthen clinical trust. Addressing these gaps will be essential for ensuring that NNs trained on limited datasets can be safely and effectively translated into routine healthcare practice.

Acknowledgements

We would like to express our sincere gratitude to the researchers and contributors whose valuable work formed the foundation of this review. In addition, we acknowledge the support of academic institutions and database providers for granting access to critical literature.

Authors' Contribution

Conceptualization: Amir Torab-Miandoab.
Data curation: Amir Torab-Miandoab.
Formal analysis: Amir Torab-Miandoab.
Funding acquisition: Amir Torab-Miandoab, Saeid Masoumi.
Investigation: Amir Torab-Miandoab.
Methodology: Amir Torab-Miandoab.
Project administration: Amir Torab-Miandoab.
Resources: Amir Torab-Miandoab.
Software: Saeid Masoumi.
Supervision: Amir Torab-Miandoab.
Validation: Saeid Masoumi.
Visualization: Saeid Masoumi.
Writing–original draft: Amir Torab-Miandoab, Saeid Masoumi.
Writing–review & editing: Amir Torab-Miandoab, Saeid Masoumi.

Competing Interests

The authors declare that they have no competing interests.

Consent for Publication

Not applicable.

Data Availability Statement

The dataset supporting the conclusions of this article is included within the article.

Ethical Approval

Not applicable.

Funding

This study was self-funded by the authors and received no external financial support from any funding organization.

Intelligence Use Disclosure

The authors used ChatGPT Plus software to paraphrase and natively translate the text of the article. The authors reviewed the text for accuracy and take full responsibility for the final content.

References

1. Yang GR, Wang XJ. Artificial neural networks for neuroscientists: a primer. Neuron. 2020;107(6):1048-70. doi: 10.1016/j.neuron.2020.09.005.
2. Miikkulainen R, Liang J, Meyerson E, Rawal A, Fink D, Francon O, et al. Evolving deep neural networks. In: Kozma R, Alippi C, Choe Y, Morabito FC, eds. Artificial Intelligence in the Age of Neural Networks and Brain Computing. 2nd ed. Academic Press; 2024. p. 269-87. doi: 10.1016/b978-0-323-96104-2.00002-6.
3. Mijwel MM. Artificial neural networks advantages and disadvantages. Mesopotamian Journal of Big Data. 2021;2021:29-31. doi: 10.58496/mjbd/2021/006.
4. Pulvermüller F, Tomasello R, Henningsen-Schomers MR, Wennekers T. Biological constraints on neural network models of cognitive function. Nat Rev Neurosci. 2021;22(8):488-502. doi: 10.1038/s41583-021-00473-5.
5. Sharkawy AN. Principle of neural network and its main types. J Adv Appl Computat Math. 2020;7:8-19. doi: 10.15377/2409-5761.2020.07.2.
6. Dastres R, Soori M. Artificial neural network systems. Int J Imaging Robot. 2021;21(2):13-25.
7. Khalilov DA, Jumaboyeva NA, Kurbonova TM. Advantages and applications of neural networks. Acad Res Educ Sci. 2021;2(2):1153-9.
8. Ahmed SF, Bin Alam MS, Hassan M, Rozbu MR, Ishtiak T, Rafa N, et al. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif Intell Rev. 2023;56(11):13521-617. doi: 10.1007/s10462-023-10466-8.
9. Werbos PJ. Why neural networks? In: Handbook of Neural Computation. 1st ed. CRC Press; 2020.
10. Wu YC, Feng JW. Development and application of artificial neural network. Wirel Pers Commun. 2018;102(2):1645-56. doi: 10.1007/s11277-017-5224-x.
11. Abdolrasol MG, Hussain SS, Ustun TS, Sarker MR, Hannan MA, Mohamed R, et al. Artificial neural networks-based optimization techniques: a review. Electronics. 2021;10(21):2689. doi: 10.3390/electronics10212689.
12. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: a survey. Heliyon. 2018;4(11):e00938. doi: 10.1016/j.heliyon.2018.e00938.
13. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review. PLoS One. 2019;14(2):e0212356. doi: 10.1371/journal.pone.0212356.
14. Zion I, Ozuomba S, Asuquo P. An overview of neural network architectures for healthcare. In: 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS). Ayobo, Nigeria: IEEE; 2020. p. 1-8. doi: 10.1109/icmcecs47690.2020.246980.
15. Pandit A, Garg A. Artificial neural networks in healthcare: a systematic review. In: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). Noida, India: IEEE; 2021. p. 1-6. doi: 10.1109/Confluence51648.2021.9377086.
16. Bhowmik BR, Varna SA, Kumar A, Kumar R. Deep neural networks in healthcare systems. In: Machine Learning and Deep Learning in Efficacy Improvement of Healthcare Systems. CRC Press; 2022. p. 195-226.
17. Shaikhina T, Khovanova NA. Handling limited datasets with neural networks in medical applications: a small-data approach. Artif Intell Med. 2017;75:51-63. doi: 10.1016/j.artmed.2016.12.003.
18. Pasini A. Artificial neural networks for small dataset analysis. J Thorac Dis. 2015;7(5):953-60. doi: 10.3978/j.issn.2072-1439.2015.04.61.
19. Bargagna F, De Santi LA, Martini N, Genovesi D, Favilli B, Vergaro G, et al. Bayesian convolutional neural networks in medical imaging classification: a promising solution for deep learning limits in data scarcity scenarios. J Digit Imaging. 2023;36(6):2567-77. doi: 10.1007/s10278-023-00897-8.
20. Eshun RB, Bikdash M, Islam AK. A deep convolutional neural network for the classification of imbalanced breast cancer dataset. Healthc Anal (N Y). 2024;5:100330. doi: 10.1016/j.health.2024.100330.
21. Mazurowski MA, Habas PA, Zurada JM, Lo JY, Baker JA, Tourassi GD. Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. Neural Netw. 2008;21(2-3):427-36. doi: 10.1016/j.neunet.2007.12.031.
22. Aminian K, Ameri S. Application of artificial neural networks for reservoir characterization with limited data. J Pet Sci Eng. 2005;49(3-4):212-22. doi: 10.1016/j.petrol.2005.05.007.
23. Goh AT, Kulhawy FH. Neural network approach to model the limit state surface for reliability analysis. Can Geotech J. 2003;40(6):1235-44. doi: 10.1139/t03-056.
24. Thorpe M, van Gennip Y. Deep limits of residual neural networks. ArXiv [Preprint]. October 28, 2018. Available from: https://arxiv.org/abs/1810.11741.
25. Zhang Z, Zohren S, Roberts S. DeepLOB: deep convolutional neural networks for limit order books. IEEE Trans Signal Process. 2019;67(11):3001-12. doi: 10.1109/tsp.2019.2907260.