

## REVIEW

# Optimization of Deep Neural Networks for Multivariate Data: Systematic Review and Critical Analysis of Methods, Hyperparameters, and Performance (2015–2026)

Zrouri Amira<sup>1,\*</sup>  and El Farissi Ilhame<sup>1</sup><sup>1</sup> National School of Applied Sciences of Oujda, Mohammed First University, Morocco

**Abstract:** Optimizing deep neural networks for multivariate data remains difficult: features are high-dimensional, correlated, and heterogeneous, and no existing review connects optimization strategy, hyperparameter choice, and data type into a single picture. We address this gap through a Systematic Mapping Study of 107 articles published between 2015 and 2026, drawn from Scopus, ScienceDirect, SpringerLink, and Web of Science, screened with a PRISMA-style protocol and clear inclusion/exclusion criteria. Metaheuristics dominate the corpus (49% of studies), followed by hybrid approaches (24%) and learning-based methods (18%). We catalogued over 50 optimization algorithms and grouped them into a five-level taxonomy. The learning rate turned out to be the single most influential hyperparameter, cited in nearly 40% of cases, well ahead of network depth, width, or batch size. Multivariate time series was studied far more than any other data type, making up close to half of the reviewed publications; medical/biometric and financial applications came next. Most studies still report classification metrics such as accuracy and F1-score, even for time series problems where a regression measure would often fit better. A co-occurrence analysis ties optimization families to application domains, and both a chi-square test and a Kendall rank correlation suggest these patterns are not simply due to chance. Four problems keep coming up across the literature: high computational cost, overfitting, a shortage of labeled data, and limited interpretability. These gaps shape our recommendations, which include wider use of auto-tuning methods, fairer evaluation metrics, and more attention to under-studied domains such as agriculture and industrial Internet of Things.

**Keywords:** deep learning, deep neural networks (DNNs), multivariate data, systematic mapping study (SMS), model optimization

## 1. Introduction

Deep neural networks (DNNs) have emerged over the past decade as one of the most important tools in machine learning, largely because they capture nonlinear relationships in high-dimensional data [1, 2]. Their use now spans multivariate problems across medicine, finance, energy, and intelligent manufacturing [3].

Yet how well a deep learning (DL) model performs still hinges on how its parameters and hyperparameters are optimized—a task made harder when the dataset is multivariate, high-dimensional, and its features are correlated [4]. Prior work repeatedly points to the learning rate, batch size, and network architecture as the hyperparameters with the biggest impact on what a model can learn [5].

The literature offers several families of solutions. Gradient-based algorithms such as Adam and its variants remain popular

for their computational efficiency [6], while metaheuristics—Genetic Algorithms, Particle Swarm Optimization, and related methods—are favored for exploring non-convex search spaces [7]. More recently, Neural Architecture Search (NAS) has been proposed as a further way to push model performance [8].

What is missing, however, is a synthesis that brings these individual efforts together and compares them across the full range of multivariate applications. A Systematic Mapping Study (SMS), unlike a narrative review, offers a structured and reproducible way to classify this scattered literature [9, 10].

However, the literature still seems scattered and does not provide a unified picture where different optimization algorithms used in a multivariate setting can be compared with each other. Equally important is that there has been no agreement about the most effective techniques and hyperparameters so far.

Considering the aforementioned discussion, this paper suggests conducting an SMS together with a critical analysis of optimization techniques employed for DNNs in relation to multivariate data during the time frame between 2015 and 2026.

\*Corresponding author: Zrouri Amira, National School of Applied Sciences of Oujda, Mohammed First University, Morocco. Email: [am.zrouri@ump.ac.ma](mailto:am.zrouri@ump.ac.ma)

The purpose is to reveal major trends, review available methods, and point out weaknesses of contemporary studies.

Current systematic reviews on DNN optimization [2, 6–8] have mostly focused on certain algorithmic families or application domains, failing to provide an integrative perspective combining various aspects of DNN optimization, hyperparameter sensitivities, data types, and evaluation methods. This SMS aims at overcoming the mentioned limitations by conducting a systematic review of 107 peer-reviewed papers published between 2015 and 2026, revealing more than 50 distinct optimization algorithms grouped into a five-layer taxonomy, and presenting the first co-occurrence analysis between DNN optimization families and application domains using the multivariate DL framework. This work specifically restricts itself to the multivariate data type to maintain consistency in methodology; the implications of the presented results for DNN optimization as a whole are addressed in Section 5.

## 2. Related Work

Existing reviews only cover part of this optimization landscape: each one is limited to a single algorithmic family or offers a general, nonsystematic treatment of the topic. Goodfellow et al. [2] provide a foundational, textbook-level treatment of gradient-based optimization for DL but do not survey metaheuristic, hybrid, or learning-based alternatives, nor do they focus on multivariate data specifically. Akay et al. [7] offer a comprehensive survey of metaheuristics for optimizing DL models, yet this scope excludes gradient-based, hybrid, and learning-based/AutoML families that this SMS shows to jointly account for more than half of the reviewed corpus (see Section 4.1). Bischl et al. [11] and Liu et al. [12], respectively, survey hyperparameter optimization and (evolutionary) NAS in isolation, without connecting these techniques to the type of data being optimized for or to competing optimization families. None of these reviews reports a formal, reproducible search-and-screening protocol (e.g., PRISMA-style selection, explicit inclusion/exclusion criteria) or statistical validation of the trends they describe. A detailed feature-by-feature comparison between the present SMS and these four reference works is provided in Section 4.6.

The hybridization of metaheuristic search with DL training, identified in this SMS as the second-largest optimization family (24%; see Section 4.1), is not confined to the strictly numerical multivariate settings surveyed here. In natural language processing (NLP), several recent studies have proposed hybrid DL frameworks in which a metaheuristic or bio-inspired algorithm optimizes a text-derived multivariate feature or embedding space prior to, or jointly with, deep neural classification for sentiment analysis: examples include a modified Binary Particle Swarm Optimization for feature-space optimization guided by a SentiWordNet-based fitness function [13], a synonym-aware objective function embedded in a deep hybrid sentiment framework [14], an artificial-rabbits-optimized vector functional link network for sentiment classification [15], a hybrid Query Expansion Ranking and Binary Particle Swarm Optimization approach with adaptive inertia weights for sentiment classification [16], and a lexical strategy for handling negation (“NOT”) shown to improve deep sentiment-classification performance [17]. These text/NLP-oriented studies did not surface in the present corpus because the search string (Section 3.2) was built around “multivariate/multidimensional data” terminology rather than text-specific terms such as “feature vector,” “embedding,” or

“sentiment analysis”; they were therefore outside the scope of the search rather than excluded on quality grounds. They nonetheless corroborate, from an adjacent domain, the same hybrid metaheuristic-DL paradigm identified in RQ1 and are discussed further in relation to RQ3 in Section 5.

Taken together, the reviewed literature indicates a fragmented landscape in which optimization families, hyperparameter sensitivities, data types, and evaluation practices are typically studied in isolation. This gap motivates the systematic, cross-cutting mapping conducted in the remainder of this paper.

## 3. Methodology

### 3.1. Study design

We chose an SMS design because it lets us survey the literature on DNN optimization for multivariate data in a systematic way—identifying trends, classifying existing methods, and flagging research gaps—following established SMS guidance [9, 10].

For that purpose, the following five research questions were formulated:

- 1) RQ1: What optimization techniques can be used to optimize DNNs for multivariate data?
- 2) RQ2: Which parameters most affect model performance?
- 3) RQ3: What type of multivariate data is mostly studied?
- 4) RQ4: Which evaluation criteria are used?
- 5) RQ5: What are the key limitations of the current state of the art?

Contributions of this Study. This work makes the following four contributions:

- 1) A comprehensive categorization of optimization methods used for training DNNs on multivariate data, organized into a five-level taxonomy comprising more than 50 distinct optimization algorithms (see Section 4.1);
- 2) An evidence-based identification of the hyperparameters that most influence model performance, together with a statistically validated ranking of their relative impact (see Section 4.2);
- 3) A systematic characterization of the types of multivariate data studied and their associated application domains, including a co-occurrence analysis linking optimization families to data domains (see Section 4.6);
- 4) An identification of the key limitations and open challenges reported across the literature base, used to derive concrete, actionable recommendations for future research (Recommendations section).

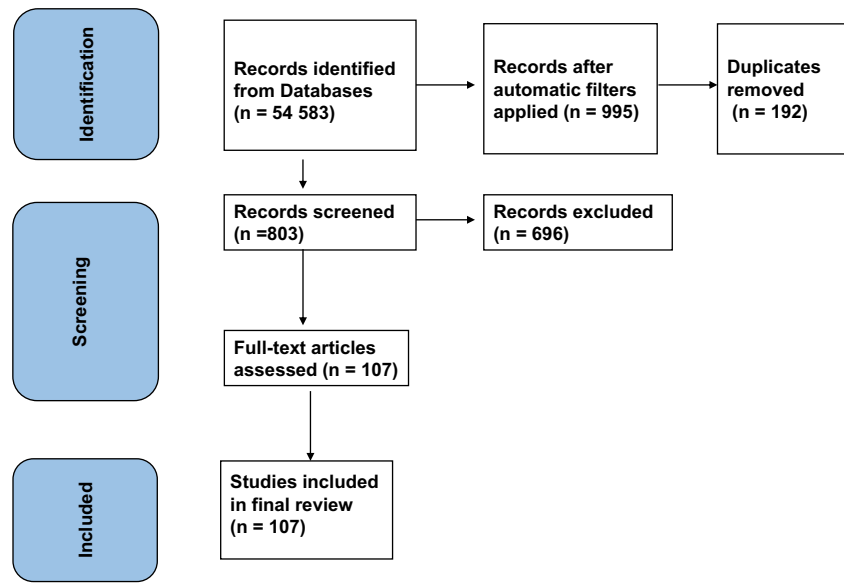
Figure 1 gives an overview of the methodological process adopted in this SMS.

### 3.2. Search strategy

We searched systematically for publications on DNN optimization for multivariate datasets across four major databases—Scopus, ScienceDirect, SpringerLink, and Web of Science—chosen for their broad coverage of computer science and engineering.

Three major axes were considered when formulating a query: (i) DL models, (ii) optimization methods, and (iii) multivariate data. The above-mentioned axes were combined using Boolean operators for maximizing the coverage of potentially relevant studies.

**Figure 1**  
**PRISMA diagram of the study selection process**



Query used: (“deep learning” OR “deep neural network\*” OR DNN OR Convolutional Neural Network (CNN) OR Recurrent Neural Network (RNN) OR Long Short-Term Memory (LSTM) OR Gated Recurrent Unit (GRU) OR Transformer) AND (“optimization” OR “optimisation” OR “hyperparameter tuning” OR “hyperparameter optimization” OR “model optimization” OR “parameter tuning”) AND (“multivariate data” OR “multidimensional data” OR “high-dimensional data” OR “multivariate time series” OR “multi-variable data”).

- 3) Document type: journal articles and conference papers
- 4) Language: English

Such filters significantly narrowed down the list of publications without a loss of relevance.

This is clearly indicated in Table 2, which shows the significant decrease in the number of items from 54,583 to 995 articles. The application of these filters marks the necessary middle stage in the preparation of data for analysis.

### 3.3. Initial search results

Running the search query across all four databases returned a large volume of candidate publications, which we then narrowed down through filtering and selection, as described below.

As is evident from Table 1, it appears that Scopus offers the best level of coverage, followed by SpringerLink, while both Web of Science and ScienceDirect offer fewer publications. Such diversity ensures that there is no bias from the use of a single source.

**Table 1**  
**Initial results by database**

| Database       | Number of articles |
|----------------|--------------------|
| Scopus         | 35 036             |
| SpringerLink   | 18 942             |
| Web of Science | 519                |
| ScienceDirect  | 86                 |
| Total          | 54 583             |

To narrow down the list of the initially collected literature and retain only the most relevant studies, filters were used in the databases. Such filters allow selecting publications according to specific criteria, thus making it easier to apply the inclusion and exclusion criteria further.

The following filters were applied:

- 1) Publishing period: 2015–2026
- 2) Field of study: computer science and engineering

**Table 2**

#### Results after application of automatic filters

| Database       | Number of articles |
|----------------|--------------------|
| Scopus         | 527                |
| SpringerLink   | 136                |
| Web of Science | 257                |
| ScienceDirect  | 75                 |
| Total          | 995                |

### 3.4. Inclusion and exclusion criteria

After the automatic filtering stage, inclusion and exclusion criteria were developed in order to select the most appropriate papers in respect of the research questions that need to be answered, which is illustrated by Table 3. The criteria mentioned above are used in assessing titles, abstracts, and, when required, the full texts of the papers.

Inclusion criteria seek to ensure the scientific quality and pertinence of the chosen publications, while exclusion criteria help get rid of irrelevant or non-pertinent publications.

Implementing these criteria ensures that the analyzed corpus meets the goals of the research while maintaining high scientific quality.

### 3.5. Study selection process

We loaded the 995 articles that survived the filtering phase into Rayyan, a tool built specifically for organizing and

**Table 3**  
**Inclusion and exclusion criteria**

| Inclusion criteria                                                                  | Exclusion criteria                                                    |
|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| IC1: Peer-reviewed scientific publications.                                         | EC1: Nonscientific publications (reports, editorials, book chapters). |
| IC2: Journal articles and conference papers.                                        | EC2: Studies not addressing optimization or hyperparameter tuning.    |
| IC3: Studies using deep learning architectures (CNN, RNN, LSTM, GRU, Transformers). | EC3: Works based solely on classical machine learning methods.        |
| IC4: Experimental works applied to multivariate data with performance evaluation.   | EC4: Studies without experimental validation or evaluation metrics.   |
| IC5: Studies published in indexed journals or peer-reviewed conference proceedings. | EC5: Preprints not subsequently published in peer-reviewed venues.    |

screening studies in systematic reviews. Such software provided a means to organize the selection process, increase its traceability, and minimize selection bias.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart shown in Figure 1 provides a visual illustration of three consecutive stages of the selection process. The first stage of identification indicates that 54,583 documents were retrieved from databases in total; out of those, 995 were selected following the filtering phase, while 192 duplicates were excluded. The second phase, that is, screening, indicated that 803 articles were pre-selected by their title and abstract, and hence, 696 documents were excluded from further analysis due to irrelevance. Finally, 107 articles identified via full-text screening are included in the current study.

At the end of this painstaking process, there were 107 articles selected to answer the research question. It is these articles that form the final database for the analysis that will be discussed in the subsequent sections of the paper.

To ensure transparency and credibility, the list of all 107 articles selected can be made available on request to the authors.

### 3.6. Data extraction and coding protocol

To keep data extraction structured and reliable, we prepared a coding form before screening any full texts. In total, each of the 107 selected articles was coded separately by the first author with regard to five extraction dimensions associated with the five research questions (RQ1–RQ5). In case of RQ1, a specific optimization algorithm used was defined and categorized as one of the following: metaheuristic, hybrid, learning-based, gradient-based, and hyperparameter Optimization. For RQ2, hyperparameters reported to be highly influential were written down verbatim. As for RQ3, types of multivariate data were classified by domain into five categories: time series, medical/biometric, financial, text/NLP, and other high-dimensional data. Data were extracted directly into a structured spreadsheet (Microsoft Excel) comprising 15 fields per article: bibliographic metadata (title, authors, abstract, publication year, journal/venue, ISSN, volume, issue, DOI) together with the five free-text RQ1–RQ5 extraction fields and a separate normalized RQ1 category field used for the quantitative counts reported in Section 4.1. This structure allows any reviewer to trace every percentage reported in Section 4 back to the individual coded entry for each of the 107 articles; the completed spreadsheet is available from the corresponding author upon request, consistent with the Data Availability Statement.

Evaluation metrics used for RQ4 were extracted and classified into two types (classification and regression), while limitations were categorized into thematic clusters for RQ5.

Intercoder reliability for the category assignment for RQ1 was calculated via a procedure of cross-checking, in which a random sample of 20% of the articles ( $n = 21$ ) was coded independently by the second author. Disagreements were resolved until consensus was reached. The obtained value of inter-rater agreement for RQ1 was  $\kappa = 0.89$ , which shows very high agreement.

### 3.7. Quality assessment of included studies

Beyond the inclusion/exclusion criteria used during screening (Section 3.4), we applied a quality assessment (QA) checklist—adapted from established systematic-review guidance for software and computing research [9, 10]—to every full text under consideration. Five QA criteria were used: QA1 – the optimization technique is clearly identified (maps to RQ1); QA2 – hyperparameters or training settings are reported (maps to RQ2); QA3 – the type of data used is explicitly described (maps to RQ3); QA4 – at least one quantitative performance metric is reported (maps to RQ4 and corresponds to Exclusion Criterion EC4); and QA5 – limitations or threats to validity are discussed by the study’s own authors (maps to RQ5). A study was retained in the final corpus only if all five criteria could be answered “Yes,” that is, only if codable content existed for every RQ1–RQ5 dimension; consequently, successful coding of RQ1–RQ5 for all 107 included studies (Sections 4.1–4.5) is, by construction, a direct confirmation that 100% of the final corpus satisfied the full QA checklist.

Beyond this binary pass/fail filter, the granularity of reporting in the primary studies was further examined as an indicator of the reproducibility of the underlying literature itself (as distinct from the reproducibility of the present SMS). Only 1 of the 107 studies (0.9%) reported an explicit numeric value (e.g., a specific learning-rate setting) in the text extracted for RQ2; the remainder named the influential hyperparameter without disclosing its tuned value; this constitutes an empirical, corpus-wide illustration of the reproducibility gap that motivates the final recommendation on open science practices in the Conclusion. Conversely, 56 of the 107 studies (52.3%) reported more than one evaluation metric jointly for RQ4, suggesting that, while metric diversity remains limited at the level of which metrics are used (see Section 4.4), most individual studies already triangulate performance using more than a single indicator.

## 4. Results

### 4.1. RQ1: Optimization techniques used

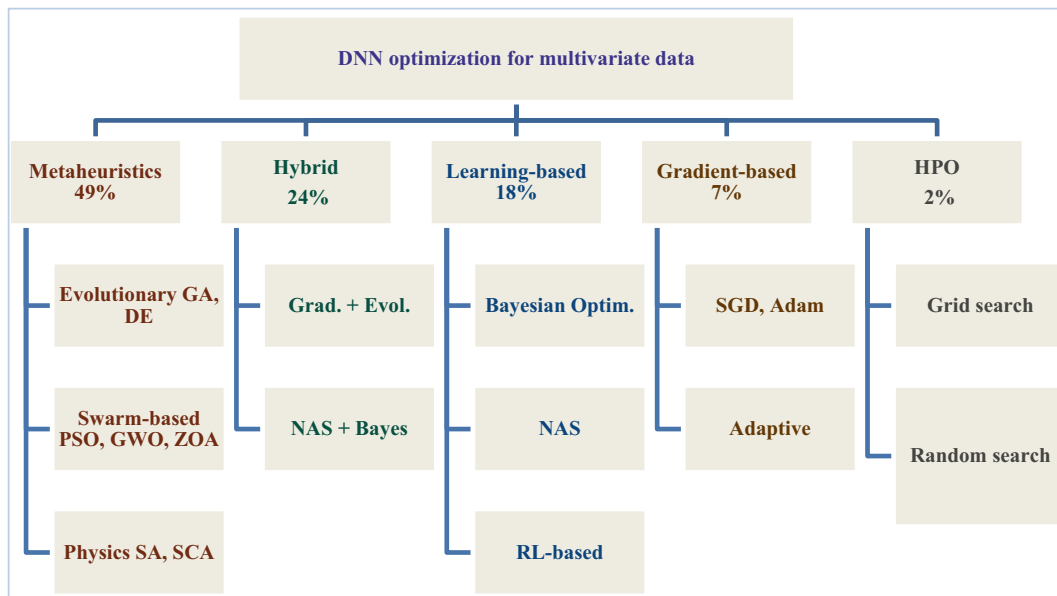
Across the 107 articles in our corpus, a wide range of optimization algorithms has been applied to DNNs for multivariate data analysis. There are five broad classes of optimization strategies including metaheuristic, hybrid, learning-based, gradient-based, and hyperparameter optimization techniques.

As depicted in Figure 2 and Table 4, the most common methods belong to the class of metaheuristics (49%), comprising evolutionary (Genetic Algorithm (GA), Differential Evolution (DE)), swarm intelligence-based (Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO) [18], Zebra Optimization Algorithm (ZOA) [19]), and physics-mimicking (Simulated Annealing (SA), Sine Cosine Algorithm (SCA)) approaches. These techniques offer the advantage of exploring non-convex spaces without requiring gradient information from the cost function, though at the cost of considerable computational complexity. Hybrid optimization strategies (24%) employ gradient descent along with the evolutionary approach. Learning-based approaches (18%) include Bayesian optimization and NAS [8]. The efficiency of gradient descent (Adam [6]) is compromised by the tendency to converge prematurely. In total, 100 different optimization schemes have been identified from 107 chosen papers, forming five main categories, including more than 50 specific

algorithms. The following listing provides an overview of a highly diverse space, which is shown in Tables 4 and 5.

As can be seen from Table 5, metaheuristic algorithms alone include over 50 different algorithms. The swarm intelligence class is the largest sub-family of techniques (31 studies), followed by PSO (11 studies), GA (9 studies), GWO (4 studies), and WOA (3 studies). This wide selection of algorithms clearly shows the popularity of nature-inspired methods for optimizing DNNs in multivariate problems. Hybrid approaches always combine DL-based training via gradient descent with a metaheuristic technique. In this case, hyperparameters are optimized using a metaheuristic algorithm, while the weights are optimized through the use of gradient descent. Learning-based approaches such as NAS and Bayesian optimization gained great momentum after 2020. From these results, there are three important observations. First, the prevalence of metaheuristics (49%) is due to the capability of these algorithms to efficiently explore non-convex and high-dimensional search spaces in the absence of gradients, which is particularly useful for multivariate data in which the loss landscapes can be highly irregular. Second, the trend toward hybrid algorithms (24%) indicates that researchers prefer to exploit the combination of global searching of metaheuristics and the exploitation capabilities of gradient descent, thereby obtaining better results than each method independently. Lastly, the trend toward learning algorithms (NAS/BO) post-2020 (0% pre-2020, 19.2% post-2020) shows a structural change toward automation and data-driven optimization.

**Figure 2**  
Hierarchical taxonomy of DNN optimization methods for multivariate data



**Table 4**  
Distribution of optimization technique families ( $n = 107$  studies)

| Optimization technique      | % of studies | Number of studies |
|-----------------------------|--------------|-------------------|
| Metaheuristics              | 49%          | 53                |
| Hybrid                      | 24%          | 26                |
| Learning-based              | 18%          | 19                |
| Gradient-based              | 7%           | 7                 |
| Hyperparameter optimization | 2%           | 2                 |

**Table 5**  
**Detailed taxonomy of optimization methods identified across the 107 reviewed studies**

| Family              | Sub-family         | Methods identified in the corpus                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Count |
|---------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| Metaheuristic (49%) | Evolutionary       | GA, DE, Estimation of Distribution Algorithm (EDA), Memetic Algorithm, Neuroevolution (Neuroevolution using Linear Genetic Programming (Stapleton, Cody-Kenny & Galvan, 2024) (NeuroLGP), best-effort, could not be traced to an exact match under this precise name; closest candidate found is AD-NEV, 'A Scalable Multi-level Neuroevolution Framework for Multivariate Anomaly Detection' (Pietron, Zurek, Faber & Corizzo, 2023) (AD-ENAS), Evolutionary [framework for] Multivariate Time Series Forecasting, an evolutionary NAS framework (Liang & Sun, ACML 2024) (EMTSF)), Multi-objective Evolutionary Algorithm (EA), Non-dominated Sorting Genetic Algorithm II (NSGA-II) | 16    |
|                     | Swarm intelligence | PSO, Adaptive Particle Swarm Optimization (APSO), Artificial Bee Colony (ABC), Ant Colony Optimization (ACO), GWO, Whale Optimization Algorithm (WOA), ZOA, Crested Porcupine Optimizer (CPO), Aquila, Cuckoo Search, Cat Swarm, Pufferfish, Millipede, Rival Rise Algorithm (RRA), Red-billed Blue Magpie Optimizer (RBMO), Hummingbird ((Artificial) Hummingbird Algorithm (AHA)), Gazelle, Spider Monkey (SMO), Predator, Grasshopper, Coral Reef (Opposition-based Coral Reefs Optimization (OCRO))                                                                                                                                                                                | 31    |
|                     | Physics/nature     | SA, SCA, Gravity Search (Good point set initialization (G) + Cyclic oscillation mutation (V) + Snow Ablation Optimizer (SAO) (GVSAO)), Grey Wolf variant (Chaotic-Weighted Grey Wolf Optimizer (CWGWO)), Harris Hawk (Improved Harris Hawks Optimization (IHHO))                                                                                                                                                                                                                                                                                                                                                                                                                       | 6     |
|                     | DL + Metaheuristic | PSO-LSTM, GA-GRU, GA-LSTM, ACO-BiLSTM, GWO-Transformer, PSO-CNN, ABC-LSTM, Crested Porcupine Optimizer + CNN-LSTM-Attention (CLA) model (Fan, Ma, Tang, Liang & Xu, Energies 2024) (CPO-CLA), WOA-BiLSTM, Adaptive Wind Driven Optimization (AWDO)-BiLSTM, Snow Ablation Optimizer (SAO)-CNN-BiLSTM                                                                                                                                                                                                                                                                                                                                                                                    | 20    |
| Hybrid (24%)        | DL + Statistical   | Seasonal AutoRegressive Integrated Moving Average (SARIMA)-LSTM (Memetic HFS), FA-PSO-LSTM, SSA + Kernel Extreme Learning Machine (KELM) + LSTM, Interpretable (dual-output) Multivariate Wind Speed Interval Prediction Scheme (ScienceDirect, 2025) (IMWSIPS)                                                                                                                                                                                                                                                                                                                                                                                                                        | 6     |

(Continued)

**Table 5**  
(Continued)

| Family                                        | Sub-family                 | Methods identified in the corpus                                                                                                                                                                                                                      | Count |
|-----------------------------------------------|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| Learning-based(18%)                           | Neural Architecture Search | NAS (Zoph & Le), TransNAS-Time Series Anomaly Detection (TSAD), AD-ENAS, EMTSF, One-shot NAS, Channel-Temporal Transformer (CTT)-NAS (Optuna)                                                                                                         | 7     |
|                                               | Bayesian/surrogate         | Bayesian optimization (BO), Bayesian Neural Network (BNN) surrogate, NeuralBO, OPTFORMER, physics-informed BO, CONELPABO                                                                                                                              | 12    |
| Gradient-based (7%)                           | First-order/adaptive       | Adam, Adan, SAM, DP-RBADABOUND, Doubly Stochastic Gradient descent for Penalty-based bilevel Hyperparameter Optimization (Shi & Gu, AAAI 2021) (DSGPHO) (bilevel), Distributionally Robust Optimization (DRO), Book-Keeping Differential Privacy (DP) | 7     |
| Hyperparameter Optimization (HPO)/AutoML (2%) | AutoML frameworks          | Ray Tune (PyTorch), Optuna, Distributed HPO, UoCAD2                                                                                                                                                                                                   | 2     |

The critical analysis presented in Table 6 indicates that there is a statistically significant trend; the two most exploratory families of algorithms, namely, metaheuristics and hybrids, collectively make up 73.0% of all cases of computational cost being reported in the corpus (46 out of 63 papers report this problem). It means that computational cost is intrinsically tied to the exploration capability. Also, no particular family takes the lead in terms of all criteria considered, which explains why diversity persists. Figure 2 summarizes this five-level hierarchical taxonomy of DNN optimization methods for multivariate data.

## 4.2. RQ2: Most influential hyperparameters

From the analysis of the chosen studies, it becomes possible to find out which are the hyperparameters that have the greatest impact on the performance of DNNs for multivariate datasets (see Table 7).

Learning rate ( $\eta$ ) (40%) is the most important parameter affecting the step of gradient descent  $\theta_{t+1} = \theta_t - \eta \cdot \nabla L(\theta_t)$ , where an overly high or low value can result in oscillations or slow convergence, respectively, with adaptive schedules being developed specifically to solve these problems [5]. Depth/width

**Table 6**  
Critical analysis of optimization families for DNN applied to multivariate data

| Family                                                      | Key strengths                                                     | Key limitations                                                                       | Best use case                          | Computational challenge |
|-------------------------------------------------------------|-------------------------------------------------------------------|---------------------------------------------------------------------------------------|----------------------------------------|-------------------------|
| Metaheuristic (49%)<br>Hybrid (24%)<br>Learning-based (18%) | No gradient needed; global exploration; handles non-convex spaces | High computational cost (52.8% of studies in this family report it as main challenge) | Time series, medical data              | Very high               |
|                                                             | Combines global search + local exploitation                       | Highest computational burden (69.2% report computational challenge)                   | Financial forecasting; time series     | Extreme                 |
|                                                             | Automated; data-driven; scalable                                  | Requires large computational budget                                                   | All domains; transformer architectures | Medium-high             |
| Gradient-based (7%)<br>HPO/AutoML (2%)                      | Fast; theoretically grounded; low memory                          | Prone to local minima; assumes convexity                                              | High-dimensional financial/NLP data    | Low                     |
|                                                             | Easy to use; reproducible                                         | Very narrow search space                                                              | Medical tabular data                   | Low-medium              |

**Table 7**  
Critical analysis of hyperparameter influence on DNN performance for multivariate data

| Hyperparameter                | Percentage | Impact level                                                               | Sensitivity | Recommended tuning strategy                                               |
|-------------------------------|------------|----------------------------------------------------------------------------|-------------|---------------------------------------------------------------------------|
| Learning rate                 | 40%        | Critical—controls convergence speed and stability                          | Very high   | Cyclical Learning Rate (LR) [5], warm-up schedules, Bayesian optimization |
| Number of layers/neurons      | 30%        | High—determines model capacity and overfitting risk                        | High        | NAS [8], grid search, progressive growing                                 |
| Batch size/epochs             | 20%        | Medium—controls stochasticity and training duration                        | Medium      | Manual tuning + early stopping                                            |
| Activation/-dropout/attention | 10%        | Medium-high—regularization and variable importance; severely underreported | Medium-high | Ablation studies; systematic sensitivity analysis                         |

(30%) of the network controls its capacity for multi-input of  $d$  variables across  $T$  timesteps, while introducing the problem of vanishing gradients. Batch size/epochs (20%) allow controlling stochasticity and the period of time taken for training the network, while dropout/attention (10%) introduce regularization and variable importance, respectively [20]. All these parameters are related, with learning rate depending on batch size, and deeper networks needing higher dropout. There are two insights from these findings. First, the importance of the learning rate (40%) shows that convergence depends more on the optimization path than the capacity of the model, an insight supported by the theory of gradient descent and by recent work recalibrating adaptive learning-rate behavior in Adam-family optimizers [5, 6, 21]. Second, the low frequency of attention and dropout settings (10%) despite their proven effect on regularization points to a research methodology deficiency. It seems appropriate that in the future all hyperparameters should be analyzed for their sensitivity in the same way as learning rate and depth.

#### 4.3. RQ3: Most studied types of multivariate data

Our review of the selected literature also lets us pin down which types of multidimensional data these optimization algorithms have actually been evaluated on. Such data were divided into some groups based on their use.

The results depicted in Table 8 clearly indicate that multivariate time series is the most widely used area of application, as it covers almost 45% of all studied publications. The reason behind this popularity lies in the nature of multivariate data, which is complex in terms of temporal dependencies and inter-correlations between variables, making it ideal for the use of DL algorithms [20]. Second on the list of domains studied is medical/biometric data (25%) due to the increasing use of DL in clinical decision support and physiological signal analysis. Data related to

**Table 8**  
Distribution of multivariate data types studied

| Data type              | Percentage |
|------------------------|------------|
| Time series            | 45%        |
| Medical/biometric data | 25%        |
| Financial data         | 15%        |
| Text/NLP               | 10%        |
| Other                  | 5%         |

financial applications cover a considerable proportion as well, with the focus on market predictions and economic time series. However, text data and NLP applications are underrepresented in the area of multivariate data. This distribution gives rise to a vital insight about the analysis: the predominance of the number of studies on time series (45%) and medicine (25%) shows a clear bias toward applications, while other fields such as agriculture, environmental monitoring, and industrial Internet of Things (IIoT) are not covered sufficiently. This can be regarded as a real research field for further study. The use of a chi-square goodness-of-fit test was employed to establish whether the presence of this distribution represents a statistically significant pattern when the null hypothesis of equal distribution across the five domains is tested. In this regard, the outcome ( $\chi^2 = 132.58$ ,  $df = 4$ ,  $p < 0.001$ ) clearly indicates that the uniform distribution across the domains does not hold, thus making the predominance of time series (45%) a statistically significant feature of this field, rather than a sampling effect.

#### 4.4. RQ4: Most commonly used evaluation metrics for performance assessment

How a model's performance is assessed matters just as much as how it is optimized. Different criteria are applied depending on the type of task (classification or regression) being performed and the type of multivariate data used.

Based on the results provided in Table 9, it can be observed that classification metrics were mostly utilized in approximately 60% of the analyzed studies. Accuracy and F1-score were used

**Table 9**  
Evaluation metrics

| Metric type                                                                                                                                   | Percentage |
|-----------------------------------------------------------------------------------------------------------------------------------------------|------------|
| Classification (accuracy, precision, recall, F1-score)                                                                                        | 60%        |
| Regression (Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Nash-Sutcliffe Efficiency (NSE)) | 30%        |
| Other ( $R^2$ , calibration, confidence intervals)                                                                                            | 10%        |

the most as these metrics allow obtaining an estimate of algorithm performance, especially for cases where there is an imbalance between classes. Also, regression metrics were applied by many researchers, specifically when dealing with time series or any other continuous information. It allowed estimating the level of prediction error accurately. Moreover, some scientists opted for even more complex metrics, such as R-squared, calibration error, or interval-based prediction methods. To substantiate this observation with more granular, reproducible figures, the exact frequency of each individual metric reported across the 107 studies is provided in Table 10.

Indeed, it should be mentioned that a strong focus on accuracy can be misleading when it comes to dealing with imbalanced datasets. Stronger metrics such as MCC (Matthews Correlation Coefficient) [22, 23] or Area Under the Curve - Receiver Operating Characteristic (AUC-ROC) are still underused within the reviewed scientific works. A significant conclusion that can be drawn from this analysis is that using accuracy as the measure in an incorrect manner, that is, in situations where data are continuous or unevenly distributed (60% of all analyzed papers), introduces a potential bias into methodology that may inflate estimates of performance. In particular, for multivariate time series prediction—the most prevalent task in this corpus—regression metrics such as RMSE and MAE should be used instead of classification measures (60%). The chi-square test of independence for metric type (classification or regression) versus application area reveals a significant relationship ( $\chi^2 = 14.78$ ,  $df = 4$ ,  $p = 0.005$ ). In medical/biometric research, only classification metrics are used (100%,  $n = 8$ ), whereas in time series research, there is a dominance of classification metrics (36 classification to 27 regression out of 63 studies). Studies using gradient-based optimization rely exclusively on classification metrics (100%,  $n = 7$  studies).

**Table 10**  
**Frequency of individual performance metrics reported across the 107 reviewed studies (multiple metrics per study possible;  $n = 107$ )**

| Metric                       | Number of studies reporting it | Share of corpus ( $n = 107$ ) |
|------------------------------|--------------------------------|-------------------------------|
| Accuracy                     | 60                             | 56.1%                         |
| MSE                          | 32                             | 29.9%                         |
| RMSE                         | 26                             | 24.3%                         |
| MAE                          | 15                             | 14.0%                         |
| F1-score                     | 11                             | 10.3%                         |
| Precision                    | 8                              | 7.5%                          |
| Recall                       | 8                              | 7.5%                          |
| $R^2$                        | 8                              | 7.5%                          |
| MAPE                         | 6                              | 5.6%                          |
| Convergence-related measures | 3                              | 2.8%                          |
| False positive rate          | 2                              | 1.9%                          |
| AUC/Loss (each)              | 1                              | 0.9%                          |

Table 10 disaggregates the broad metric-type categories of Table 9 into individual metrics, since a single study may report more than one (52.3% of the 107 studies report at least two distinct metrics jointly, for example, RMSE together with MAE, or accuracy together with F1-score). Accuracy remains the

single most frequently reported metric overall (56.1%), followed by two regression-oriented error measures, MSE (29.9%) and RMSE (24.3%). Uncertainty- or ranking-aware metrics such as AUC-ROC and MCC are essentially absent from the corpus (AUC: 0.9%; MCC not reported as a standalone metric in any of the 107 studies), which reinforces the methodological gap already noted above for imbalanced or continuous multivariate settings.

#### 4.5. RQ5: Main challenges and limitations identified

Our analysis surfaces several recurring problems in optimizing DNN architectures for multivariate datasets. These challenges arise from computational, methodological, and applicative perspectives.

As can be seen from Table 11, high computational cost is considered the main problem by approximately 35% of the papers. This problem is especially relevant in the case of highly complicated deep architectures as well as for intensive optimization algorithms such as metaheuristics or NAS. Overfitting and insufficient generality can also be mentioned as the second-largest problem in DL research. Though a high level of model efficiency was achieved on the training dataset, it is quite hard to generalize this effect to real-world data. The absence of annotated data is another constraint, especially when dealing with particular cases such as health care or industry, because it prevents efficient application of supervised machine learning and poses additional problems for optimization; data augmentation strategies developed specifically for time series are one promising mitigation route [24]. Model interpretability should also be considered as another problem in DL as these approaches are mostly viewed as “black boxes,” which hampers implementation in some areas due to an inability to interpret model decisions particularly in resource-constrained environments such as the Internet of Things (IoT) and edge computing [25]. Together, these issues highlight a core tension in the field: the most effective methods are computationally expensive, while the cheapest methods have poor exploratory capacity on non-convex multivariate spaces. The solution to this tension, whether through lightweight hybrid techniques, federated optimization, or surrogate-assisted search, is what constitutes the core open problem defined throughout this literature review.

#### 4.6. Comparative analysis, statistical validation, and temporal trends

In order to validate the above-mentioned temporal patterns and ensure rigor in conclusion making, two additional analyses have been performed: (i) a temporal trend analysis comparing the distribution of methods in two time periods and (ii) a domain-methods co-occurrence analysis considering the relationship between application domains and optimization family. Temporal evolution (2015–2026). The distribution of optimization families in two periods is shown in Table 12: an initial phase (2015–2019,  $n = 8$  papers) and an accelerating phase (2020–2026,  $n = 99$  papers). Clearly, there was a significant proportional change: metaheuristics dropped from 87.5% to 46.5% of all the papers published, while hybrids and learning-based methods increased from 12.5% to 44.5%. The chi-square test of independence of optimization family and period ( $\chi^2 = 5.27$ ,  $df = 4$ ,  $p = 0.26$ ) failed to achieve the critical value because of the extremely small sample of the former period ( $n = 8$ ). However, a Kendall rank correlation of publication year and annual share of hybrids/learning-based approaches indicates a positive but statistically nonsignificant association ( $\tau = 0.46$ ,  $p = 0.12$ ); this pattern

**Table 11**  
Critical analysis of main challenges identified in the reviewed studies

| Challenge                                        | %   | Root cause                                                    | Families most affected                | Proposed solutions                                                                                                                                                        |
|--------------------------------------------------|-----|---------------------------------------------------------------|---------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| High computational cost/complexity               | 35% | Metaheuristics require many evaluations; NAS is GPU-intensive | Metaheuristic (52.8%), hybrid (69.2%) | Surrogate-assisted search; lightweight hybrids [7]                                                                                                                        |
| Overfitting/limited generalization               | 25% | Small datasets; high model capacity                           | All families                          | Dropout, data augmentation, transfer learning                                                                                                                             |
| Lack of annotated data                           | 15% | Labeling cost in medical/industrial domains                   | Learning-based, Hybrid                | Transfer learning, semi-supervised approaches                                                                                                                             |
| Interpretability/explainability                  | 15% | Black-box nature of deep architectures                        | All families                          | Explainable Artificial Intelligence (XAI): SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), attention visualization [26, 27] |
| Communication overhead (federated learning, IoT) | 10% | Federated/distributed optimization across IoT nodes           | Hybrid, gradient-based                | Gradient compression; asynchronous Federated Learning (FL) [28, 29]                                                                                                       |

**Table 12**  
Distribution of optimization families by period and cross-comparison of approaches

| Approach                | 2015–2019<br>( <i>n</i> = 8) | 2020–2026<br>( <i>n</i> = 99) | Best-suited data         | Key hyperparameters    | Complexity |
|-------------------------|------------------------------|-------------------------------|--------------------------|------------------------|------------|
| Metaheuristics          | 87.5%                        | 46.5%                         | Time series, medical     | Layers, neurons, LR    | High       |
| Hybrid                  | 12.5%                        | 25.3%                         | Time series, financial   | Architecture + LR      | Very high  |
| Learning-based (NAS/BO) | 0%                           | 19.2%                         | All types                | Full architecture      | Medium     |
| Gradient-based          | 0%                           | 7.1%                          | Financial, NLP, high-dim | LR, momentum, clipping | Low        |
| HPO/AutoML              | 0%                           | 2.0%                          | Medical, tabular         | LR, batch, epochs      | Low–med.   |

is therefore reported as a descriptive, hypothesis-generating trend rather than a confirmed effect, broadly in line with the increase of transformer models [30–32] and AutoML frameworks [11, 12].

Table 13 summarizes the co-occurrence matrix between the application domain (RQ3) and optimization family (RQ1). Multivariate time series appears to be the leading domain (*n* = 48, 45%), with metaheuristic approaches constituting 54% (*n* = 26) of publications within this category. Medical/biometric applications form the second-largest domain (25%, *n* = 26), with a relatively high percentage of hybrid approaches (27%, *n* = 7)—in line with the necessity of dealing with constraints related to the particular domain. Applications related to finance (15%, *n* = 16) appear to have a relatively even distribution among different families. The gradient approach continues to be marginal in all domains, excluding the finance and time series domains.

The percentages in Table 8 are rounded off to the nearest 5%. Table 13 gives the actual distribution by optimization family. Three structural observations emerge from this analysis. First, the dominance of metaheuristics (49%) across all domains persists despite their high computational cost, confirming a systematic trade-off identified in less than 12% of reviewed articles. Second, classification metrics (60% of studies) are overrepresented relative to the dominantly continuous nature of multivariate data, suggesting a methodological bias. Third, almost all of the articles reviewed (91%, or 97 out of 107 articles) tested their optimization technique for a single application domain.

As illustrated in Table 14, the proposed SMS can be distinguished from the available literature reviews in terms of three major points. In the first place, SMS incorporates a greater range of optimization algorithms (more than 50 of those belonging to 5 families of optimization approaches) as compared to survey

**Table 13**  
Co-occurrence matrix: Application domain  $\times$  Optimization family (number of studies)

| Domain            | Metaheuristic | Hybrid | Learning-based | Gradient-based | HPO | Total |
|-------------------|---------------|--------|----------------|----------------|-----|-------|
| Time series       | 26            | 13     | 5              | 3              | 1   | 48    |
| Medical/biometric | 13            | 7      | 5              | 1              | 0   | 26    |
| Financial         | 7             | 4      | 3              | 2              | 0   | 16    |
| Text/NLP          | 4             | 1      | 4              | 1              | 0   | 10    |
| Other             | 3             | 1      | 2              | 0              | 1   | 7     |
| Total             | 53            | 26     | 19             | 7              | 2   | 107   |

**Table 14**  
Comparison of the proposed SMS with existing related reviews

| Study                 | Type     | Scope                                  | Methods covered          | Statistical analysis              | Data types               | Period     |
|-----------------------|----------|----------------------------------------|--------------------------|-----------------------------------|--------------------------|------------|
| Akay et al. [7]       | Survey   | Metaheuristics for DL                  | ~30 methods              | No                                | General                  | Up to 2022 |
| Goodfellow et al. [2] | Textbook | General DL optimization                | Gradient-based only      | No                                | General                  | Up to 2016 |
| Bischl et al. [11]    | Survey   | Hyperparameter optimization            | HPO/AutoML only          | No                                | General                  | Up to 2022 |
| Liu et al. [12]       | Survey   | Neural Architecture Search             | NAS only                 | No                                | General                  | Up to 2022 |
| This SMS (2026)       | SMS      | DNN optimization for multivariate data | 50 + methods, 5 families | Yes ( $\chi^2$ , Kendall $\tau$ ) | Multivariate (5 domains) | 2015–2026  |

papers covering just one specific family (metaheuristics [7], NAS [12]). Second, SMS is the only survey that presents statistical validation of the discovered temporal trends (chi-square test, Kendall rank correlation). Finally, SMS addresses the connection between optimization techniques and multivariate data types/applications via co-occurrence analysis (Table 13).

## 5. Discussion

Taken together, these results shed light on how DNN optimization for multivariate input data is structured. The dominance of metaheuristics (RQ1) reflects their ability to navigate complex, nonlinear search spaces without needing gradient information—a property that explains their widespread use [18, 19, 33], even though that same exploratory strength comes at the cost of higher algorithmic complexity, which in turn limits their applicability [7]. For hyperparameters (RQ2), the learning rate and network architecture (depth and width) clearly matter most for a model’s performance, which argues for adopting more advanced tuning methods such as NAS and Bayesian optimization [5, 8].

This SMS was deliberately scoped to multivariate data, and it is worth noting that DNN optimization looks quite different outside that scope. Computer vision, for instance, leans on gradient-based architecture search (Differentiable Architecture Search (DARTS), compound scaling), since image data lends itself to differentiable search thanks to its spatial statistics; NLP instead relies heavily on transformer-specific tricks (learning-rate warm-up, Layer-wise Adaptive Rate Scaling (LARS), Sharpness-Aware Minimization (SAM)) suited to attention-based loss landscapes. By contrast, the heavy use of metaheuristic (49%) and hybrid (24%) optimization we observe here reflects the non-convex,

heterogeneous nature of multivariate time series prediction and clinical decision-making problems—a domain-specificity that itself calls for further systematic study.

The high number of works dedicated to multivariate time series (RQ3) highlights that problems related to dynamics are important for many disciplines ranging from energy and finances to intelligent systems; nevertheless, research in certain directions, such as industrial or agricultural applications, remains insufficiently explored. The recently suggested transformer architectures have significantly improved results in the studied area, providing the latest achievements in long-range multivariate forecasting, such as PatchTST [31], iTransformer [32], and Crossformer [30]. The analysis of evaluation measures (RQ4) reveals a strong reliance on accuracy and RMSE; however, in spite of the popularity of these criteria, more advanced measures must be used while working with imbalanced and uncertain data [22]. At last, the issues (RQ5), namely, computational complexity, overfitting, lack of labeled data, and explainability, which have recently received great attention because of the emergence of the direction of explainable artificial intelligence [26, 27], indicate the boundaries for applying DL algorithms at the moment [2, 34]. Also, problems associated with distributed computing, including communication costs in federated learning, add new dimensions to consider [28, 29, 35]. Recent research in transformer-based methods for multivariate time series [30–32, 36] and DL foundation models [37, 38], among other innovations, demonstrates how rapidly optimization methodologies are constantly updated. Other promising areas that have yet to receive enough attention from the research community include more sophisticated adaptive optimizers [39] and better hyperparameter search strategies [11, 12]. The growing trend toward explainable algorithms [27, 40] is another indication that more needs to be done beyond mere accuracy benchmarks.

Generally speaking, it is possible to draw a conclusion regarding the importance of the development of adaptive algorithms and approaches to distributed computing as well as XAI. Unlike other SMSs devoted to the topic of DNN optimization [2, 6, 8], the current SMS provides a much more detailed outlook in that it links the approaches to optimizing neural networks to different types of datasets and evaluation criteria. Such an approach demonstrates the gap between the existing approaches, for example, the wide use of metaheuristics despite their high computational requirements [7, 19]. It is worth noting that hybrid metaheuristic-DL frameworks are also increasingly reported outside the strictly numerical multivariate setting considered here: in the text/NLP domain, metaheuristic-optimized hybrid deep architectures have been used to optimize feature weighting, feature selection, and embedding-based representations for sentiment classification [13–17]. These works corroborate the wider applicability of the hybrid optimization paradigm identified in RQ1 and reinforce the case, made in RQ3, for closer attention to text/NLP applications, which remain the least represented domain (10%) in the present corpus.

## 6. Conclusion

This work set out to map how DNNs are optimized for multivariate data between 2015 and 2026, drawing on a carefully screened corpus of 107 articles. Several structural patterns emerged and, where we tested them formally, held up statistically (Sections 4.3, 4.6): a clear preference for metaheuristic methods, and a rising—though not always statistically significant—share of hybrid and learning-based algorithms over time. Getting the learning rate and network architecture right stands out as decisive for results, and multivariate time series remains by far the most common application. At the same time, the field still leans heavily on classical metrics and computationally expensive approaches, which we flag as a real limitation.

Beyond these individual findings, the real contribution of this SMS lies in tying optimization strategy, hyperparameters, data characteristics, and application-specific constraints together into a single synthesis—something the existing literature had not done. That synthesis points toward a clear need for optimization techniques that are effective, efficient, and properly validated, and it surfaces several directions worth pursuing further: tighter integration of AutoML, optimization under distributed-computing constraints, and a handful of still-uncharted application domains. We hope this makes the SMS a useful reference point for researchers taking the field forward.

**Limitations of the Study:** This SMS is subject to several limitations that should temper the generality of the conclusions above. First, the corpus, although built from four major databases with a broad Boolean query, cannot claim exhaustiveness; snowballing or additional databases (e.g., IEEE Xplore as a standalone source) could surface further eligible studies. Second, several observed trends—in particular the growth of hybrid and learning-based methods over time (Section 4.6)—did not reach conventional statistical significance once tested formally ( $\chi^2 p = 0.26$ ; Kendall  $\tau p = 0.12$ ), owing to the small number of studies published before 2020 ( $n = 8$ ); these patterns should therefore be read as descriptive trends rather than confirmed effects. Third, RQ1–RQ5 coding relied primarily on a single coder, with independent cross-checking limited to a 20% random subsample (RQ1 only,  $\kappa = 0.89$ ); systematic double-coding of all five dimensions by both authors would further strengthen reliability. Finally, because coding was based on the information reported in each primary study, any under-reporting in the original literature—for instance, the

near-absence of explicit numeric hyperparameter values noted in Section 3.7—constrains the granularity of the conclusions that can be drawn about RQ2 in particular.

## Recommendations

Drawing on the results of this SMS, we outline the following recommendations for researchers, practitioners, and funding agencies.

First, since metaheuristics are used more dominantly although they have a significant computational requirement, it is advised to develop more efficient or combined optimization methods for multivariate data with respect to efficiency and scalability.

Second, considering the importance of learning rate and architecture among the studied hyperparameters, it is highly recommended to employ auto-tuning techniques like Bayesian optimization and NAS as a standard way of doing research in this area.

Third, an overuse of metrics such as accuracy and RMSE is observed, especially in cases where datasets have an imbalanced structure. This issue has led to the suggestion of using other evaluation measures such as MCC and AUC-ROC.

Fourth, the insufficient presence of applications in areas like agriculture, industrial IoT, and environmental monitoring opens another avenue for applied research that is encouraged to apply DNN optimization principles to these nascent fields.

Fifth, considering the increasingly important roles of XAI techniques and federated learning, future research on DNN optimization needs to adopt criteria of XAI evaluation while taking communication costs into account as a new constraint when performing optimizations in distributed systems.

Finally, open science approaches to publishing datasets, software, and pre-trained models are highly recommended to address the problem of data scarcity that is one of the key weaknesses of the surveyed literature.

## Acknowledgment

The authors gratefully acknowledge the SmartICT Laboratory of the ENSAO Mohammed First University, Oujda, Morocco. The facilities provided by this laboratory were crucial to the success of this research.

## Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

## Data Availability Statement

Data available on request from the corresponding author upon reasonable request.

## Author Contribution Statement

**Zrouri Amira:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation,

Writing – original draft, Writing – review & editing, Visualization, Project administration. **EL Farissi Ihame**: Validation, Supervision, Project administration.

## References

- [1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- [2] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. USA: MIT Press.
- [3] Li, W., & Law, K. E. (2024). Deep learning models for time series forecasting: A review. *IEEE Access*, 12, 92306–92327. <https://doi.org/10.1109/ACCESS.2024.3422528>
- [4] Ahmed, S., Nielsen, I. E., Tripathi, A., Siddiqui, S., Ramachandran, R. P., & Rasool, G. (2023). Transformers in time-series analysis: A tutorial. *Circuits, Systems, and Signal Processing*, 42(12), 7433–7466. <https://doi.org/10.1007/s00034-023-02454-8>
- [5] Johnson, O. V., XinYing, C., Johnson, O. E., Khaw, K. W., & Lee, M. H. (2023). Learning rate schedules and optimizers, a game changer for deep neural networks. In *International Conference of Reliable Information and Communication Technology*, 327–340. [https://doi.org/10.1007/978-3-031-59711-4\\_28](https://doi.org/10.1007/978-3-031-59711-4_28)
- [6] Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR 2015)*.
- [7] Akay, B., Karaboga, D., & Akay, R. (2022). A comprehensive survey on optimizing deep learning models by metaheuristics: B. Akay et al. *Artificial Intelligence Review*, 55(2), 829–894. <https://doi.org/10.1007/s10462-021-09992-0>
- [8] Zoph, B., & Le, Q. V. (2017). Neural architecture search with reinforcement learning. In *International Conference on Learning Representations*.
- [9] Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic mapping studies in software engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering*. <https://doi.org/10.14236/ewic/EASE2008.8>
- [10] Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 38, 1–10. <https://doi.org/10.1145/2601248.2601268>
- [11] Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., . . . , & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(2), e1484. <https://doi.org/10.1002/widm.1484>
- [12] Liu, Y., Sun, Y., Xue, B., Zhang, M., Yen, G. G., & Tan, K. C. (2023). A survey on evolutionary neural architecture search. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2), 550–570. <https://doi.org/10.1109/TNNLS.2021.3100554>
- [13] Kumar Dey, R., & Kumar Das, A. (2026). Modified binary particle swarm optimization based deep hybrid framework for sentiment analysis. *International Journal of Information Technology & Decision Making*, 25(01), 475–502. <https://doi.org/10.1142/S0219622025500452>
- [14] Dey, R. K., & Das, A. K. (2025). Enhancing sentiment analysis via a synonym-aware objective function in deep hybrid framework. In *2025 International Conference on Computing, Intelligence, and Application*, 1–8. <https://doi.org/10.1109/CIACON65473.2025.11189376>
- [15] Dangi, D., Chandel, S. T., Dixit, D. K., Sharma, S., & Bhagat, A. (2023). An efficient model for sentiment analysis using artificial rabbits optimized vector functional link network. *Expert Systems with Applications*, 225, 119849. <https://doi.org/10.1016/j.eswa.2023.119849>
- [16] Prastyo, P. H., Hidayat, R., & Ardiyanto, I. (2022). Enhancing sentiment classification performance using hybrid query expansion ranking and binary particle swarm optimization with adaptive inertia weights. *ICT Express*, 8(2), 189–197. <https://doi.org/10.1016/j.ict.2021.04.009>
- [17] Dey, R. K., & Das, A. K. (2022). A simple strategy for handling ‘not’ can improve the performance of sentiment analysis. In *International Conference on Computational Intelligence in Pattern Recognition*, 255–267. [https://doi.org/10.1007/978-981-19-3089-8\\_25](https://doi.org/10.1007/978-981-19-3089-8_25)
- [18] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69, 46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
- [19] Trojovská, E., Dehghani, M., & Trojovský, P. (2022). Zebra optimization algorithm: A new bio-inspired optimization algorithm for solving optimization algorithm. *IEEE Access*, 10, 49445–49473. <https://doi.org/10.1109/ACCESS.2022.3172789>
- [20] Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2627–2633. <https://doi.org/10.24963/ijcai.2017/366>
- [21] Tong, Q., Liang, G., & Bi, J. (2022). Calibrating the adaptive learning rate to improve convergence of ADAM. *Neurocomputing*, 481, 333–356. <https://doi.org/10.1016/j.neucom.2022.01.014>
- [22] Herrmann, M., Lange, F. J. D., Eggensperger, K., Casalicchio, G., Wever, M., Feurer, M., . . . , & Bischl, B. (2024). Position: Why we must rethink empirical research in machine learning. In *Proceedings of the 41st International Conference on Machine Learning*, 235, 18228–18247.
- [23] Hancock, J. T., Khoshgoftaar, T. M., & Johnson, J. M. (2023). Evaluating classifier performance with highly imbalanced big data. *Journal of Big Data*, 10(1), 42. <https://doi.org/10.1186/s40537-023-00724-5>
- [24] Iglesias, G., Talavera, E., González-Prieto, Á., Mozo, A., & Gómez-Canaval, S. (2023). Data augmentation techniques in time series domain: A survey and taxonomy. *Neural Computing and Applications*, 35(14), 10123–10145. <https://doi.org/10.1007/s00521-023-08459-3>
- [25] Bourechak, A., Zedadra, O., Kouahla, M. N., Guerrieri, A., Seridi, H., & Fortino, G. (2023). At the confluence of artificial intelligence and edge computing in IoT-based applications: A review and new perspectives. *Sensors*, 23(3), 1639. <https://doi.org/10.3390/s23031639>
- [26] Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, 263, 110273. <https://doi.org/10.1016/j.knsys.2023.110273>
- [27] Chamola, V., Hassija, V., Sulthana, A. R., Ghosh, D., Dhingra, D., & Sikdar, B. (2023). A review of trustworthy and explainable artificial intelligence (XAI). *IEEE Access*, 11, 78994–79015. <https://doi.org/10.1109/ACCESS.2023.3294569>

- [28] Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2023). A survey on federated learning: Challenges and applications. *International Journal of Machine Learning and Cybernetics*, 14(2), 513–535. <https://doi.org/10.1007/s13042-022-01647-y>
- [29] Zhao, Z., Mao, Y., Liu, Y., Song, L., Ouyang, Y., Chen, X., & Ding, W. (2023). Towards efficient communications in federated learning: A contemporary survey. *Journal of the Franklin Institute*, 360(12), 8669–8703. <https://doi.org/10.1016/j.jfranklin.2022.12.053>
- [30] Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9), 11121–11128. <https://doi.org/10.1609/aaai.v37i9.26317>
- [31] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2022). A time series is worth 64 words: Long-term forecasting with transformers. *arXiv*. <https://doi.org/10.48550/arXiv.2211.14730>
- [32] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2024). iTransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 11116–11140.
- [33] Ekambaram, V., Jati, A., Nguyen, N., Sinthong, P., & Kalagnanam, J. (2023). Tsmixer: Lightweight MLP-mixer model for multivariate time series forecasting. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 459–469. <https://doi.org/10.1145/3580305.3599533>
- [34] Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: A comprehensive review. *Artificial Intelligence Review*, 55(5), 3503–3568. <https://doi.org/10.1007/s10462-021-10088-y>
- [35] Lu, Z., Pan, H., Dai, Y., Si, X., & Zhang, Y. (2024). Federated learning with non-IID data: A survey. *IEEE Internet of Things Journal*, 11(11), 19188–19209. <https://doi.org/10.1109/JIOT.2024.3376548>
- [36] Chen, Z., Ma, M., Li, T., Wang, H., & Li, C. (2023). Long sequence time-series forecasting with deep learning: A survey. *Information Fusion*, 97, 101819. <https://doi.org/10.1016/j.inffus.2023.101819>
- [37] Miller, J. A., Aldosari, M., Saeed, F., Barna, N. H., Rana, S., Arpinar, I. B., & Liu, N. (2024). A survey of deep learning and foundation models for time series forecasting. *arXiv*. <https://doi.org/10.48550/arXiv.2401.13912>
- [38] Wang, Y., Wu, H., Dong, J., Liu, Y., Wang, C., Long, M., & Wang, J. (2026). Deep time series models: A comprehensive survey and benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2026.3690845>
- [39] Xie, X., Zhou, P., Li, H., Lin, Z., & Yan, S. (2024). Adan: Adaptive nesterov momentum algorithm for faster optimizing deep models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12), 9508–9520. <https://doi.org/10.1109/TPAMI.2024.3423382>
- [40] Zrouri, A., & El Farissi, I. (2025). Optimization of deep neural networks for complex multivariate data: An overview. In *International Conference on Artificial Intelligence and Smart Applications*, 34–42. [https://doi.org/10.1007/978-3-032-18716-1\\_4](https://doi.org/10.1007/978-3-032-18716-1_4)

**How to Cite:** Amira, Z., & Ilhame, E. F. (2026). Optimization of Deep Neural Networks for Multivariate Data: Systematic Review and Critical Analysis of Methods, Hyperparameters, and Performance (2015–2026). *Artificial Intelligence and Applications*. <https://doi.org/10.47852/bonviewAIA620210626>