



Article Info

Received: 02-04-2026

Revised: 23-04-2026

Accepted: 05-05-2026

¹Department of Computer Science,
Sokoto State University²Information and Communication
Technology Directorate, Sokoto State
University³Umar Ali Shinkafi Polytechnic
Sokoto⁴Kaduna State polytechnic⁵Hadhramout University
hadhramout yemen***Corresponding author's email:**sarkijabo24@gmail.com**Cite this:** *CaJoST*, 2026, vol,8 issue 2

Enhancing Optimization Stability in Deep Machine Learning-Based Intelligent Time Series Forecasting: A Systematic Review of Gradient Clipping Techniques

Shehu Malami SarkinTudu^{1*}, Anas Abdullahi², Anas Tukur Balarabe¹,
Mahmood Umar¹, Salisu Modi¹, Urwatu Lawal³, Naseer Sani⁴, Ajoge Sameera
Obaid Barraood⁵

Intelligent time series forecasting has become a vital tool across sectors such as finance, energy, and healthcare, where accurate predictions can enhance operational efficiency and strategic decisions. However, modern deep learning models for time series often encounter optimisation instability, including exploding and vanishing gradients, which can considerably impair forecasting performance. Gradient clipping has become a common technique to address these issues by limiting gradient magnitudes during training. Despite its significance, existing research on gradient clipping is scattered, with different methods, thresholds, and evaluation strategies applied across various architectures and fields. There is currently a lack of a comprehensive synthesis that systematically explores how gradient clipping affects optimisation stability, the effectiveness of different clipping strategies, and their practical implications for intelligent time series forecasting. To address this, this study presents a systematic literature review (SLR) of gradient clipping techniques, aiming to classify clipping strategies, assess their benefits and drawbacks, and identify open research challenges. The results offer a structured taxonomy of clipping methods, highlight areas needing methodological improvement, and suggest directions for future research to improve model stability and forecasting accuracy.

Keywords: Time series, optimization, clipping techniques, algorithms, systematic literature review.

1. Introduction

Sequences of observations or measurements gathered, recorded, or viewed at successive and evenly spaced time intervals constitute the conceptual definition of time series data [1]. On the other hand, the temporal ordering of observations is preserved because each data point in a time series corresponds to a particular timestamp or time period [2]. When analyzing the dynamic evolution of a phenomenon, variable, or system across time, this data structure is especially useful. According to [3], the essential characteristics of time series data are found in its temporal structure, where the dependency between successive data points and the order of observations are important. In order to better understand the behavior of the process being monitored, time series analysis looks for underlying

patterns, trends, seasonal impacts, and abnormalities [4].

Despite its importance, time series forecasting remains a challenging task due to issues such as non-stationarity, noise, missing values, and abrupt structural changes [5]. More than 70% of real-world industrial time series data exhibit nonlinear and irregular patterns [5], limiting the effectiveness of conventional statistical models such as ARIMA. As a result, deep learning and other intelligent forecasting models have gained widespread adoption. However, training such models introduces significant optimization stability challenges, including vanishing and exploding gradients, slow or unstable convergence, and sensitivity to hyperparameter configurations. These challenges become more

pronounced as data volume and model complexity increase, particularly in long-sequence forecasting tasks. The projected global data growth to 175 zettabytes by 2025 [7] further amplifies the need for robust optimization mechanisms capable of maintaining stability during large-scale training. Efficiency and accuracy are widely recognized as fundamental requirements in time series forecasting across domains such as finance [6], meteorology [7] healthcare, and agriculture [8].

Sw

While deep and recurrent architectures offer improved predictive performance, they often suffer from computational inefficiencies and unstable gradient propagation when scaling to high-dimensional or long-horizon data [9][10]. Studies have shown that achieving high accuracy frequently comes at the cost of increased computational overhead and that training large-scale deep learning models can consume significantly more energy than traditional approaches [15]. Within this context, gradient clipping has emerged as a critical optimization technique to stabilize training by constraining gradient magnitudes, preventing divergence, and enabling reliable learning in deep and recurrent models [20], [21], [22], [23], [24]. Despite its widespread use, the design choices, effectiveness, and limitations of gradient clipping strategies remain insufficiently synthesized, motivating a focused investigation into their role in achieving optimization stability for intelligent time series forecasting.

Although numerous studies and reviews have examined gradient clipping within optimization for deep learning and time-series forecasting, the existing literature remains fragmented and methodologically incomplete, creating a clear need for a more comprehensive synthesis [11][12][13]. Most prior studies [14], [15] focus on only one or two classes of clipping strategies, typically norm-based or value-based methods, without offering a unified taxonomy that systematically compares their principles, advantages, limitations, and application contexts. Moreover, existing reviews often emphasize empirical performance improvements while overlooking theoretical considerations such as convergence behavior, bias introduced by clipping, and interactions with optimization hyperparameters. As a result, there is limited guidance for practitioners on selecting appropriate clipping strategies across different intelligent time series models and application domains. This gap underscores the necessity for a systematic literature review that

integrates methodological rigor, theoretical insights, and application-driven analysis of gradient clipping techniques.

Furthermore, the rapid evolution of gradient clipping techniques between 2017 and 2025, driven by advances in deep learning architectures and large-scale time series forecasting, has not been systematically captured in existing SLRs. While gradient clipping has become a standard stabilization mechanism within first-order, gradient-based optimizers, prior reviews largely overlook how clipping strategies have evolved alongside adaptive optimizers, scalable training regimes, and automated hyperparameter tuning frameworks. Critical themes such as clipping behavior in large-scale and high-dimensional models, sensitivity to threshold selection, computational overhead, and robustness under non-stationary time-series data remain insufficiently examined. This gap creates limited guidance for researchers and practitioners seeking principled approaches to ensuring optimization stability in intelligent time series forecasting. Consequently, a new SLR covering the period 2017–2025 is necessary to provide a unified taxonomy of gradient clipping techniques, transparent review methodology, cross-strategy comparison, and application-specific insights that are absent from previous studies. The focus of this paper is exclusively on gradient clipping mechanisms used in conjunction with modern first-order optimizers (e.g., Adam, RMSprop, Adagrad, AdaDelta) to enhance stability and accuracy in time series prediction. Optimization paradigms that do not explicitly incorporate gradient clipping, such as evolutionary algorithms, swarm intelligence, reinforcement learning based optimization, derivative-free methods, second-order or convex optimization, heuristic search, multi-objective optimization, and AutoML search strategies, are beyond the scope of this review. The primary objective of this study is therefore to address the following research inquiries:

RQ1: How gradient clipping techniques evolved to stabilize optimization in time series forecasting?

RQ2: Which optimization algorithms benefit most from gradient clipping in forecasting?

RQ3: How can gradient clipping strategies be categorized into a taxonomy with applications?

RQ4: What are future research opportunities and open challenges in gradient clipping techniques?

The structure of the paper is as follows: Section 2 delves into related work; Section 3 outlines the research methods used; Section 4 details the findings from the SLR, by responding to research questions, including potential future research directions and Section 5 provides a conclusion to the paper.

2. Related Works

Extant literature reviews have been widely published on optimization techniques [15], [16], [17]. We aim to highlight any existing gaps in research that these aforementioned SLRs may not have addressed. By doing so, we seek to understand the existing SLR publications on modern optimisation techniques research and determine areas that warrant further investigation. [12] conducted a review on deep learning algorithms for time series forecasting in geotechnics and geosciences. It highlights the dominance of RNN-based models, particularly LSTM, in handling complex and large-scale data. By comparing architectures such as CNNs, GANs, transformers, and hybrids, the review concludes that deep learning consistently outperforms traditional machine learning in predictive accuracy and data representation. [18] conducted systematic and bibliometric reviews, hybrid approaches combining optimization and machine learning for clustering and classification. It highlights the growing need for sophisticated models to handle complex data and the limitations of pure methods. Through a SWOT analysis of top algorithms and recent studies, it identifies key strengths, weaknesses, and opportunities for improving hybrid algorithm efficiency. However, the study does not explicitly examine gradient clipping or other optimization stability techniques, leaving training robustness and convergence control in deep and hybrid models insufficiently addressed.

[13] conducted comprehensive reviews of advances in computational modeling, simulation, and optimization across materials, mechanical, and energy systems. It emphasizes AI integration with traditional methods, highlighting improved accuracy and efficiency. By synthesizing state-of-the-art developments and proposing novel strategies, the study identifies key research trends and future directions, reinforcing the pivotal role of computational methods in modern engineering innovation. Despite its broad scope, the review lacks focused analysis of gradient clipping and related stabilization mechanisms, thereby overlooking their role in mitigating optimization instability in data-driven learning systems

Systematically reviews interval time series forecasting, which models data with upper and lower bounds to manage uncertainty and complexity. By analysing 125 selected papers, it identifies major research gaps such as hybrid modelling, multisource data use, improved evaluation, and broader applications. The review categorises existing work, highlights unresolved challenges, and outlines future research directions to advance interval forecasting methods. Nonetheless, the review does not investigate gradient clipping or optimisation control strategies, which are critical for ensuring stable training under uncertainty and bounded interval representations.

Provides a comprehensive review of deep learning algorithms for time series forecasting in geotechnics and geosciences. It evaluates models like RNNs, CNNs, GANs, transformers, and hybrids across applications such as seismology and construction. Highlighting RNNs, especially LSTMs, as the most effective, the study concludes that deep learning outperforms traditional methods in modelling complex patterns and large-scale geoscientific data. While the study demonstrates the superiority of deep learning models, it does not consider gradient clipping or other optimisation stabilisation techniques, limiting insights into how training instability and exploding gradients impact model reliability in complex geoscientific time series.

Table 1: *Comparison Between Our SLR and Prior Review Studies*

Author / Year	Review type	Search Covered	GB O	SEO	BPO	HEO
[12]	Narrative	Not stated	-	-	-	✓
[19]	Review	Not stated	✓	✓	-	✓
[13]	Review	Not stated	✓	✓	✓	✓
[15]	SLR	2019-2024.	✓	✓	✓	✓
[20]	SLR	Not stated	-	-	-	✓
[21]	Review	Not stated	-	✓	-	✓
[14]	Review	Not stated	-	✓	✓	-
Our SLR (2025)	SLR	2017–2025	✓	✓	✓	✓

Gradient-Based Optimization (GBO), SEO Swarm & Evolutionary, BP Bayesian/Probabilistic, HEO Hybrid/Ensemble Source: Authors

3.0 Methodology

Our systematic literature review on optimization techniques in computational approaches has been

carried out following the guidelines of [22]. In this section we will introduce how the examined papers have been selected, prepared for the analysis and an overview of the produced research (who are the main authors, how the papers are distributed over time and on the types of publication).

3.1 Selection of primary study

To search for primary papers, we have identified the following keywords: ("time series" AND (prediction OR forecasting)) AND ("optimization" OR optimizer OR "hyperparameter tuning") AND ("PSO" OR "GA" OR "GWO" OR "ACO" OR "ABC" OR "Bayesian optimization" OR SGD OR Adam OR "hybrid optimization" OR "ensemble optimization" AND ("machine learning" OR "deep learning" OR LSTM OR CNN OR Transformer) and set as a filter the years from 2017 to 2025. The platforms used for the search were Google Scholar, IEEE Xplore, Scopus, Digital Library, ACM Digital Library, and Science Direct. For each paper found, the title was analyzed, then the abstract, introduction, and conclusions to determine whether it pertained to the established inclusion criteria. Then we went on to analyze the rest of each paper to find significant contributions and open issues.

3.2 The Search process

In the Search step, the search strategy was developed. This strategy consists of the search queries and the respective databases. The search queries were focused on the following concepts: "intelligent time series forecasting: a systematic review of optimization techniques in computational approaches. The search was conducted in the title, abstract, and keyword search fields. To find all relevant studies, the above string was formed and used, considering possible synonyms.

3.3. Selection results

Considering the keywords, the following results were obtained, divided by platform: • Google Scholar: 42 papers, IEEE Xplore Digital Library: 12 papers, ACM Digital Library: 24 papers, Science Direct: 18 papers. Subsequently, duplicate studies were removed. Then, papers that did not meet the exclusion criteria were eliminated. Finally, 96 papers remain ready for the final screening based on the inclusion and exclusion criteria in the next section.

3.4 Insertion and elimination criteria

At this phase, papers published in scientific international journals and proceedings of international conferences were considered. All other kinds of work, like books, technical reports or master theses were not included. Papers dealing with topics not belonging to the corpus of knowledge of

intelligent time series forecasting have been excluded. In addition, papers not written in English and not indexed were not taken into account. The valuation step was performed according to the following predetermined insertion and elimination criteria:

Table 2: lists the insertion and elimination criteria for the papers in this review.

Actions	Criteria
Insertion Criteria	1. Must have publication dates ranging from 2017 to 2025. 2. Must present a theoretical framework for optimization techniques in time series data prediction. 3. Must evaluate optimization techniques in time series data prediction and paper must be written in English.
Elimination Criteria	1. Research lacking a theoretical framework despite examining optimization techniques 2. Research presenting a theoretical framework yet not related to optimization techniques 3. Research published in languages other than English.

Adopted Source: [23]

The publications were then arranged using the following variables: (i) Type, (ii) Author, (iii) Year, and (iv) optimization approaches. The Type variable indicated whether the paper focused on efficiency and accuracy, time series data prediction, or optimization strategies. The year and author variables were used to identify the paper. The optimization approaches variable illustrated the methodology employed to improve the study's accuracy and efficiency. An Excel spreadsheet containing all of the variables and associated data was prepared for analysis. All pertinent information was extracted from the data in the analysis process. The keywords utilized, the year of publication, the journal in which the paper was published, and the optimization methods employed by the study were all taken into consideration when analyzing the papers.

Further analysis was conducted on the category and its respective selection processes using the above-mentioned variables. 96 papers were identified from the literature search. All these papers were subjected

to a filtering and selection process before analysis (see Figure 1: papers selection process flow diagram). At the screening phases, 55 of these papers had to be eliminated based on the title, and 14 were eliminated based on the abstract. Another 8 papers from 50 papers had to be eliminated based on reading the full text. At the end of the filtering procedure, 42 papers were found to be suitable for qualitative analysis. Extraction and classification of data from the papers were conducted in the Synthesis step.

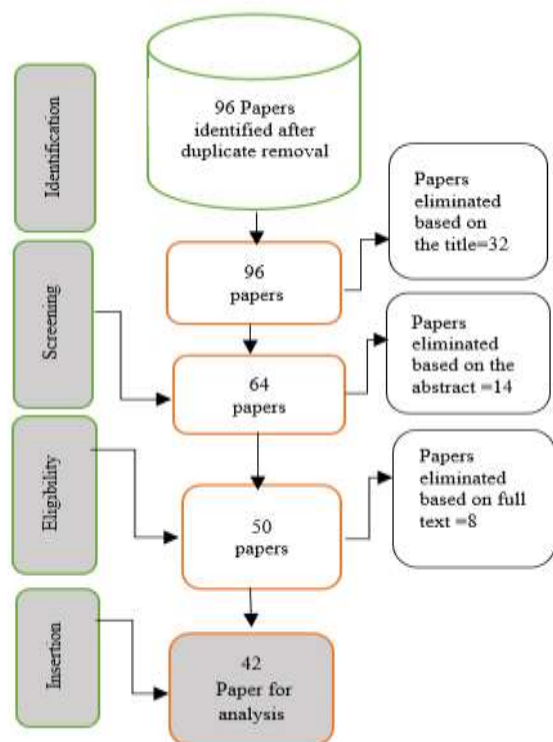


Figure 1: Flow diagram of papers selection process

According to the clipping techniques, the category they were dealing with: (i) clipping by norm (ii) clipping by value. The final papers for analysis dealing with (i.e., optimization techniques, time series data) are summarize in Table 3.

Table 3: Summary of the final papers for analysis

Clipping Techniques Category	No of Papers	Studies
Clipping by norm	31	[24],[25],[26],[27],[11][28],[29],[3],[30],[31],[32],[33],[34],[35],[36],[37],[27],[38],[39],[40],[41],[42],[24],[38],[43],[44],[45],[46],[47],[48],[49]
Clipping by value	11	[50],[51],[52],[53],[54],[55],

value	[56],[57],[58],[59],[50] [60]
-------	----------------------------------

Source: Authors

As shown in Table 4, norm-based gradient clipping has dominated the literature from 2017 to 2025, exhibiting steady growth until 2021 before stabilizing at a comparatively high level. This trajectory reflects the growing recognition of norm-based clipping as a reliable mechanism for maintaining optimization stability in deep and recurrent time series models. In contrast, value-based clipping remains marginal across all years, with only incremental increases and no comparable growth phase. The persistent disparity reported in Table 4 suggests that researchers such as (i.e. [24],[25],[26],[27]) increasingly favor norm-based clipping due to its ability to preserve gradient direction and ensure stable convergence, whereas value-based approaches are largely confined to niche or complementary applications rather than serving as a primary optimization strategy.

Table 4: Year-wise distribution of gradient clipping categories

Year	Norm-based Technique	Value-based Technique
2017	1	0
2018	2	1
2019	4	1
2020	6	2
2021	7	2
2022	5	2
2023	4	2
2025	2	1

Source: Authors

3.5 Quality assessment

The quality of the discovered research was then examined in addition to the insertion/elimination criteria. The quality assessment is one of the important elements that must be looked at, in addition to the insertion and elimination criteria [61]. To provide a way to evaluate the quality of the research papers that were kept for final analysis (N=42), a quality assessment checklist comprising nine criteria was created. Table 2 provides an

illustration of the quality assessment checklist. No scholar's work was meant to be criticized by the checklist [22]. The checklist was modified from those recommended by [22]. The three-point scale was used to rate each question: "Yes" was worth one point, "No" was worth zero, and "Partially" was worth half a point. As a result, each paper might receive a score between 0 and 8. The more points a study receives overall, the more thoroughly it answers the research questions. The quality assessment findings for each of the 42 papers are shown in Table 3. This makes it evident that every manuscript has passed the quality assessment, indicating that every study is eligible for additional analysis.

This paper used a variety of assessment criteria. The first step in evaluating the quality of the chosen papers for the review was to determine whether the study had defined objectives, detailed a suitable data analysis approach, offered an appropriate data gathering method, and presented clearly stated conclusions. For instance, the study on optimization approaches receives a score of 1 if its aims were explicit, a score of 0.5 if they were only partially clear, and a score of 0 if they were unclear. The criteria for evaluating quality were modified from [62]. [75]. The sum of the scores determines the overall score for each paper. Only six (6) of the 42 papers received a poor rating, while fifteen (15) received a very good rating. The findings of quality assessment 4 are displayed in Table 4.

Table 4: *Quality assessment 1 outcomes*

	Quality (scores)				Total
	Poor (1-1.99)	Fair (2-2.99)	Good (3-3.99)	V. Good (4)	
Number of papers	0	6	21	15	100
Percentage of papers (%)	0%	14.2%	50%	35.7%	

Source: [63]

Secondly, a set of nine criteria derived from [64][65] was utilized for quality assessment, shaping a method to evaluate the selected research papers (n = 42) for final analysis. The intent of this criterion wasn't to critique scholars' work but to gauge research quality, adapted from recommendations by [22]. Higher scores indicated a greater alignment with research objectives, determined through

scrutiny of nine (9) quality assessment criteria. The first and second authors independently assigned scores to each study, reconciling differences through discussion and further review of contentious papers. Table 5 outlines the quality assessment 2 criteria for all the 42 papers.

Table 5: *Quality Assessment criteria 2*

No	Checklist measurements
1	Are the paper's objectives clearly stated?
2	Was the aim of the paper accomplished?
3	Are the variables addressed in the study clearly outlined?
4	Is the paper's context clearly articulated?
5	Are the methods of data collection adequate?
6	Are the instruments reliable and valid?
7	Are the data analysis techniques adequate?
8	Do the findings of the paper contribute to existing literature?
9	Does the paper enhance knowledge?

Source : [64] [65]

In order to create a thorough approach for evaluating the 42 chosen research papers, an additional set of nine criteria, taken from [64], was used for quality assessment. This method, which was motivated by suggestions made by [66], sought to assess the caliber of research rather than criticize academics' work. A three-point rating system was used for each criterion, resulting in a score for each paper that ranged from 0 to 9. Examining the nine quality assessment criteria revealed that higher scores indicated better connection with study aims. Each study was given a score by the first and second authors separately, and disagreements were settled by discussion and additional analysis of controversial studies. An summary of the 42 publications' quality evaluation results is given in Table 6, demonstrating that every study satisfied the requirements and was judged appropriate for final analysis.

Table 6: *Quality assessment outcomes*

Papers	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	%
Paper1	0.5	1	0	1	1	0.5	1	1	1	77
Paper2	1	1	1	0.5	1	1	0.5	1	1	88.
Paper3	0.5	1	1	1	1	1	1	1	0.5	88.8
Paper4	1	1	1	0.5	1	1	1	1	1	94.4
Paper5	0	1	0.5	1	1	0.5	1	1	1	77.
Paper6	1	1	1	1	1	1	0	0.5	1	83.3
Paper7	0.5	1	1	0.5	1	1	1	0	1	77.7
Paper8	0.5	1	1	1	1	0.5	1	1	0.5	83.3
Paper9	1	0.5	1	1	1	1	.5	1	1	77.
Paper10	1	0	1	1	1	1	1	1	1	88.8
Paper11	0.5	1	1	1	1	0.5	1	1	1	88.8
Paper12	0	0.5	1	1	1	1	1	1	0.5	77.7
Paper13	1	0.5	1	1	0.5	1	1	0.5	1	83.3
Paper14	1	1	1		1	1	1		1	77.7

Paper15	1	1	0.5	1	1	1	0.5	0.5	0.5	77.7
Paper16	1	0.5	1	1	0.5	1	1	1	1	88.
Paper17	1	1	1	0.5	1	1	1	0.5	0.5	83.3
Paper18	0.5	1	1	1	0	1	0	0	1	61.1%
Paper19	1	1	1	0	1	1	1	0.5	1	83.3%
Paper20	1	1	1	0.5	1	1	1	1	0.5	88.8%
Paper21	1	0.5	1	1	0.5	1	1	1	1	88.8%
Paper22	0.5	1	1	1	1	1	1	1	1	94.4%
Paper23	1	0	1	0	0.5	1	1	1	1	72.2%
Paper24	0.5	1	1	1	0	1	0	0	1	61.1%
Paper25	1	1	1	0	1	1	1	0.5	1	83.3%
Paper26	1	1	1	0.5	1	1	1	1	0.5	88.8%
Paper27	1	0.5	1	1	0.5	1	1	1	1	88.8%
Paper28	0.5	1	1	1	1	1	1	1	1	94.4%
Paper29	0.5	1	1	1	0	1	0	0	1	61.1%
Paper30	1	0.5	1	1	1		0.5	1	1	77.7%
Paper31	1	0	1	1	1	1	1	1	1	88.8%
Paper32	0.5	1	1	1	1	0.5	1	1	1	88.8%
Paper33	0	0.5	1	1	1	1	1	1	0.5	77.7%
Paper34	1	0.5	1	1	0.5	1	1	0.5	1	83.3%
Paper35	1	1	1		1	1	1		1	77.7%
Paper36	1	1	0.5	1	1	1	0.5	0.5	0.5	77.7%
Paper37	0.5	1	1	1	0	1	0	0	1	61.1%
Paper38	1	1	1	0	1	1	1	0.5	1	83.3%
Paper39	1	1	1	0.5	1	1	1	1	0.5	88.8%
Paper40	1	0.5	1	1	0.5	1	1	1	1	88.8%
Paper41	1	0.5	1	1	1	0	0.5	1	1	77.7%
Paper42	1	0	1	1	1	1	1	1	1	88.8%

Source: Authors as suggested by [66]

3.6 Analysing and coding data

We have grouped the remaining 42 publications according to several criteria, such as optimization approaches and current issues in optimization strategies in time series data investigations, in order to address the research questions in this study.

3.7 Data Extraction

The reference information for each publication in this study was recorded using the Mendeley platform, a reference management service. The fields of study, methods used, and underlying theories are all included in the retrieved data used to answer the research questions.

4.0 Findings

With respect to the published 42 research papers about optimization techniques, time series data prediction from 2017 to 2025, the findings of this SLR are reported based on the four research questions.

RQ1: How gradient clipping techniques evolved to stabilize optimization in time series forecasting?

Early research positioned gradient clipping as a robustness mechanism to counter exploding gradients and instability caused by non-smooth objectives and heavy-tailed stochastic noise[60], which are prevalent in sequential and time-dependent learning tasks such as intelligent time

series forecasting [3],[64],[82]. Figure 2 illustrates the distribution and diversity of gradient clipping techniques reported in the literature, revealing a strong concentration around norm-based variants, including distributed clipping, gradient normalization, and differentially private (DP) clipping [24][58][34]. This dominance suggests that research has largely prioritized scalability, stability, and privacy preservation in large-scale and distributed training settings, particularly relevant for modern deep and time-series models. However, the long tail of less frequently explored techniques, such as bias-aware, velocity-based, quantization-aware, and reinforcement learning (RL)-based clipping, highlights a fragmented research landscape where many approaches remain under-investigated or lack rigorous comparative evaluation. Critically, while the proliferation of adaptive and DP-related methods reflects emerging concerns around robustness and privacy, their relatively low adoption compared to classical norm-based clipping indicates unresolved challenges related to computational overhead, hyperparameter sensitivity, and theoretical guarantees [3],[64],[82]. Overall, Figure 2 underscores both the maturity of conventional clipping strategies and the need for systematic, theory-driven assessments of newer variants to understand their true benefits for optimization stability in intelligent time series forecasting.

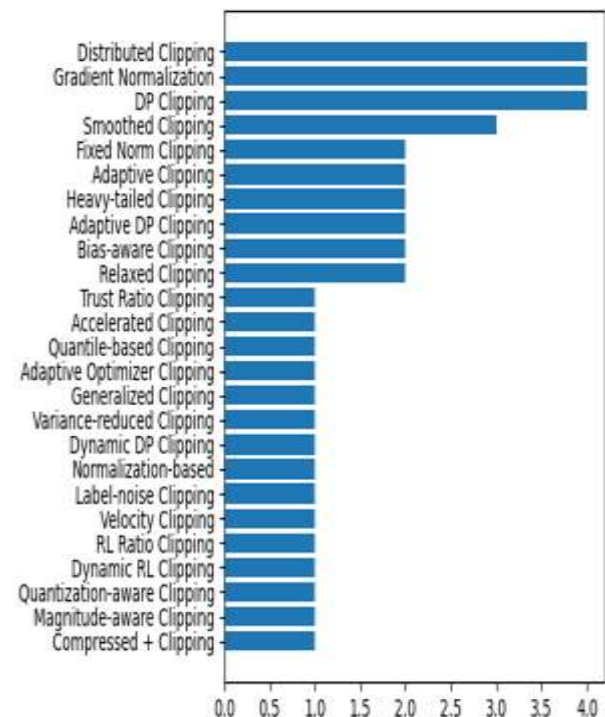


Figure 2: Distribution of gradient clipping techniques per number of studies

More recent studies, such as (e.g. [29],[58]) reflect an evolution toward adaptive, normalized, and decentralized gradient clipping, aligning with the growing complexity of intelligent time series systems deployed across distributed, privacy-sensitive and communication-constrained environments. Advanced methods incorporate normalization, dynamic thresholding, and error-feedback to mitigate over-clipping while preserving convergence optimality under heavy-tailed noise [67][32],[30]. In parallel, gradient clipping has been integrated with differential privacy, communication compression, and quantization to ensure optimization stability without sacrificing privacy guarantees or system efficiency [40] [68],[33],[4],[59]. Emerging policy-driven and dynamic clipping strategies further adapt thresholds based on gradient distributions and training dynamics, reflecting a shift from static safeguards toward context-aware stability control [69], [70][28][29]. Collectively, these developments indicate that gradient clipping has evolved into a central optimization stability mechanism underpinning reliable and scalable intelligent time series forecasting systems.

The reviewed studies consistently find that gradient clipping substantially improves optimization stability by mitigating gradient explosion, variance inflation, and instability caused by heavy-tailed noise and non-smooth objectives, which are common in deep and sequential forecasting models [47][38]. Empirical and theoretical evidence show that, despite introducing bias, gradient clipping achieves faster and more reliable convergence through variance control, particularly in high-dimensional and non-convex settings [24],[58],[34]. These findings validate gradient clipping as a principled optimization mechanism rather than a heuristic safeguard. More advanced findings indicate that adaptive, normalized, and dynamic clipping strategies outperform static clipping, especially in decentralized, learning environments relevant to large-scale intelligent time series forecasting, as suggested by [71],[30],[72]. Studies further demonstrate that integrating gradient clipping with differential privacy, compression, and quantization preserves optimization stability while maintaining utility and convergence guarantees [40],[33],[50]. Overall, the findings reveal a clear evolution from static clipping toward context-aware and policy-driven clipping mechanisms, positioning

gradient clipping as a foundational component for robust and scalable intelligent time series forecasting optimization [28],[29].

RQ2: Which optimization algorithms benefit most from gradient clipping in forecasting?

Across gradient-based optimization (GBO) methods, gradient clipping is most extensively integrated into first-order and adaptive optimizers, where it plays a decisive role in stabilizing training and improving forecasting accuracy for deep time series models. Standard optimizers such as SGD, momentum-based SGD, and mini-batch SGD adopt norm-based or value-based clipping to control exploding gradients during the training of RNNs, LSTMs, GRUs, CNNs, and Transformer architectures, which are particularly sensitive to long temporal dependencies and non-stationary inputs [55]. Adaptive methods, including AdaGrad, RMSProp, and Adam, further incorporate clipping (e.g., AutoClip) to regulate gradient magnitudes alongside adaptive learning rates, leading to improved convergence consistency and reduced performance degradation in deep forecasting models [71]. Extensions to second-order or quasi-Newton methods with clipping variants have been shown to enhance optimization robustness in highly non-convex settings and long-horizon forecasting tasks, where uncontrolled curvature and gradient variance can otherwise impair predictive accuracy [55]. Collectively, these studies indicate that clipping-enhanced gradient-based optimizers consistently improve both training stability and forecast reliability across mainstream deep learning architectures. As shown in Table 7.

Table 7: Optimization classes

Main Class	Sub-Class	Algorithms	Models / Use Cases	Sources
Adaptive Gradient Methods	Adaptive Gradient Methods	AdaGrad, RMSProp, Adam, AutoClip,	Deep learning convergence ,	[54]
	Second-Order &	Newton, Free (with clipping variants)	Complex non-convex models, long-horizon forecasting	[38]
SEO	Evolutionary Algorithms	GA, DE	Hyperparameter tuning of deep and hybrid models	[43]

	Swarm Intelligence	PSO, GWO	LSTM/GRU parameter tuning	[44]	Clipping	fairness, and privacy		
BPO	Bayesian Optimization	GP-based BO	LSTM model architecture tuning	[25]	Task-Aware Optimization	Multi-Task Gradient Clipping	Gradient balance & clipping	Multi-task learning segmentation [52]
	Surrogate-Based Optimization	Random Forest	High-cost time-series model optimization	[47]	Source: Authors			
					Across gradient-based optimization (GBO) methods, gradient clipping is most extensively integrated into first-order and adaptive optimizers, where it plays a decisive role in stabilizing training and improving forecasting accuracy for deep time series models. Standard optimizers such as SGD and mini-batch SGD adopt norm-based or value-based clipping to control exploding gradients during the training of RNNs, LSTMs, GRUs, CNNs, and Transformer architectures, which are particularly sensitive to long temporal dependencies and non-stationary inputs [55]. Adaptive methods, including AdaGrad, RMSProp, and Adam, further incorporate clipping (e.g., AutoClip) to regulate gradient magnitudes alongside adaptive learning rates, leading to improved convergence consistency and reduced performance degradation in deep forecasting models [71]. Extensions to second-order or quasi-Newton methods with clipping variants have been shown to enhance optimization robustness in highly non-convex settings and long-horizon forecasting tasks, where uncontrolled curvature and gradient variance can otherwise impair predictive accuracy [55]. Collectively, these studies indicate that clipping-enhanced gradient-based optimizers consistently improve both training stability and forecast reliability across mainstream deep learning architectures.			
					RQ3: How can gradient clipping strategies be categorized into a taxonomy with applications?			
					Gradient clipping strategies can be categorized into a taxonomy reflecting their core principles, mathematical scope, and practical implications. Value-Based Clipping, which constrains individual gradient components, offers a low computational cost and simplicity, making it attractive for very early-stage models, simple RNNs, or low-resource environments [50], [51], [52],[53], [54],[55]. However, its element-wise nature may fail to address issues arising from accumulated gradient magnitudes in deeper or more complex networks, limiting its effectiveness and generalizability. Critically, while value-based clipping provides a straightforward safeguard against exploding gradients, it does not inherently consider the interdependence among			
HEO	Gradient Evolutionary	PSO with Adam (with clipping)	Joint training for complex models	[27]				
	Evolutionary	BO and GA	Efficient global search for time-series models	[73]				
	Ensemble Meta-Optimizers	Adaptive switches, voting-based optimizers	Robust optimization under varying data patterns					
Clipping Enhance Optimization	Stochastic Gradient Clipping	Norm clipping, clipping, U-Clip	Stable training under non-smooth or heavy-tailed noise	[26]				
	Accelerated Clipping	Clipped Momentum SGD, Nesterov with Clipping	Faster convergence in non-convex landscapes	[46]				
		Communication-efficient clipped SGD	RNN/LSTM Training	[43]				
DP-GBO	DP-SGD (Individual Clipping)	DP-SGD	Privacy-preserving deep learning	[52]				
	Batch & Group-wise Clipping	Batch Clipping, Group-wise, Layer-wise Clipping	Memory-accuracy trade-offs	[45]				
	Adaptive DP	DP-SGD-Adapt	Improved utility,	[25]				

gradient dimensions, which can compromise optimization stability in sophisticated architectures. Table 8: Taxonomy of gradient clipping techniques

Gradient Clipping

Value-Based Clipping

$$\tilde{g}_i = \begin{cases} -c, & \text{if } g_i < -c \\ g_i, & \text{if } -c \leq g_i \leq c \\ c, & \text{if } g_i > c \end{cases}$$

Where g_i is the i -th gradient component, c is the clipping threshold, and $\tilde{g}_i$ is the clipped gradient.

Table 8: Taxonomy of gradient clipping techniques

type	Clipping	Advantages	Limitation	Applications
Value-Based Clipping	Element-wise of gradients	Low computational cost	Less effective for deep models	Simple RNNs; low-resource settings
Norm-Based Clipping	Rescaling of gradient	widely adopted	Overhead and sensitivity	Sequence forecasting architectures such as LSTM

Source: Authors

In contrast, Norm-Based Clipping rescales the entire gradient vector to a fixed threshold, improving training stability and robustness in advanced models such as LSTMs, GRUs, Transformers, and time-series forecasting systems [24],[25],[26],[27],[11][28]. Despite its advantages, this approach introduces computational overhead due to norm calculations and is highly sensitive to threshold selection, which can lead to either under-clipping (insufficient stabilization) or over-clipping (slowed convergence).

Norm-Based Clipping

$$\tilde{\mathbf{g}} = \begin{cases} \mathbf{g}, & \text{if } \|\mathbf{g}\| \leq c \\ \mathbf{g} \cdot \frac{c}{\|\mathbf{g}\|}, & \text{if } \|\mathbf{g}\| > c \end{cases}$$

Where $\mathbf{g}$ is the gradient vector, $\|\mathbf{g}\|$ is the L2 norm, c is the clipping threshold, and $\tilde{\mathbf{g}}$ is the clipped gradient

Critically, norm-based clipping reflects a trade-off between control over gradient magnitudes and the potential for hyperparameter sensitivity, emphasizing the importance of careful tuning. Overall, the taxonomy highlights that the choice of gradient clipping strategy must be aligned with model

complexity, resource constraints, and optimization objectives. As Figure 2. Taxonomy of Gradient Clipping Techniques

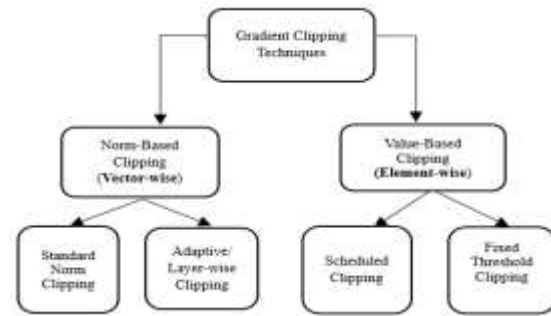


Figure 2. Taxonomy of Gradient Clipping Techniques

RQ4: What are future research opportunities and open challenges in gradient clipping techniques?

The studies presented so far provide insights that can be useful for both machine learning researchers and practitioners. Existing literature on gradient clipping primarily focuses on its application to improve optimization stability, but much of it remains introductory and empirical in nature. For example, studies by (eg. [37],[27],[38]) explore adaptive methods, theoretical dynamics, and interactions with other optimization strategies. However, many of these studies, such as (i.e [39],[40],[41],[42], [24]) lack systematic procedures for applying or evaluating clipping techniques across different model architectures and datasets. Although some papers present promising methods and theoretical insights, more is needed to comprehensively understand the mechanisms, limitations, and implications of gradient clipping in modern deep learning.

This review highlights several key research opportunities. First, there is a need for adaptive and context-aware clipping methods that can dynamically adjust thresholds during training to enhance stability without manual intervention. Second, the theoretical understanding of clipping dynamics particularly concerning convergence, bias, and gradient interactions in high-dimensional optimisation remains underexplored. Third, research should investigate how clipping interacts with other optimisation and regularisation techniques, ensuring stable training in complex models. Finally, applications in differential privacy and distributed learning offer promising yet challenging directions, as robust strategies and convergence guarantees in these areas are still limited. Collectively, these gaps present significant opportunities for future studies to

develop more rigorous, theoretically grounded, and application-aware gradient clipping techniques.

7. Conclusion

Despite extensive research on gradient clipping as a tool for stabilising optimisation in deep learning, significant opportunities still exist to enhance understanding and application. Current studies mainly concentrate on empirical implementations and specific model types, leaving gaps in systematic methodology and theoretical foundation. This review highlights the need for further exploration of the dynamics of gradient clipping, its interactions with other optimisation strategies, and its effects on model expressiveness and convergence. Although gradient clipping has been widely adopted, much of the literature does not fully incorporate existing theoretical frameworks or offer rigorous, reproducible methods. Additionally, challenges remain in applying clipping techniques to emerging areas such as large-scale, distributed learning and privacy-preserving training. Therefore, while gradient clipping is an established method, future research needs to address these limitations to improve both practical effectiveness and theoretical understanding. The insights from this review provide a foundation for developing adaptive, theory-informed, and application-aware gradient clipping approaches in modern deep learning.

Authors' Contributions

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

This research was conducted with institutional support from Sokoto State University, Faculty of Computing. Financial support from the Tertiary Education Trust Fund (TETFund) under the Institutional-Based Research Intervention scheme is gratefully acknowledged.

References

- [1] E. Yi, Y. Cho, S. Sohn, and K. Ahn, "After the splits: Information flow between bitcoin and bitcoin family," 2021. doi: 10.1016/j.chaos.2020.110464.
- [2] B. L. Tan, M. E. Norhaizan, and L. C. Chan, "Rice Bran : From Waste to Nutritious Food Ingredients," *mdpi*, vol. 15, no. 2503, pp. 1–25, 2023, doi: <https://doi.org/10.3390/nu15112503>.
- [3] E. Gorbunov, M. Danilova, and A. Gasnikov, "Stochastic optimization with heavy-tailed noise via accelerated gradient clipping," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, 2020.
- [4] Z. Liu, Z. Shen, S. Li, K. Helwegen, D. Huang, and K. T. Cheng, "How Do Adam and Training Strategies Help BNNs Optimization?," *Proc. Mach. Learn. Res.*, vol. 139, pp. 6936–6946, 2021.
- [5] K. R. Prilianti, T. H. P. Brotosudarmo, S. Anam, and A. Suryanto, "Performance comparison of the convolutional neural network optimizer for photosynthetic pigments prediction on plant digital image," in *AIP Conference Proceedings*, 2019, pp. 1–10. doi: 10.1063/1.5094284.
- [6] A. S. Prabowo, A. Sihabuddin, and A. SN, "Adaptive Moment Estimation On Deep Belief Network For Rupiah Currency Forecasting," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 1, p. 31, 2019, doi: 10.22146/ijccs.39071.
- [7] R. J. Wichink Kruit *et al.*, "Modelling trends in ammonia in the Netherlands over the period 1990–2014," *Atmos. Environ.*, vol. 154, pp. 20–30, 2017, doi: 10.1016/j.atmosenv.2017.01.031.
- [8] K. T. Lindner and S. Schwab, "Differentiation and individualisation in inclusive education: a systematic review and narrative synthesis," *Int. J. Incl. Educ.*, 2020, doi: 10.1080/13603116.2020.1813450.
- [9] M. Morin, I. Abi-Zeid, and C. G. Quimper, "Ant colony optimization for path planning in search and rescue

- operations," *Eur. J. Oper. Res.*, vol. 305, no. 1, pp. 53–63, 2023, doi: 10.1016/j.ejor.2022.06.019.
- [10] I. Gurrib, F. Kamalov, and L. Smail, "Bitcoin price forecasting: Linear discriminant analysis with sentiment evaluation," in *ACM International Conference Proceeding Series*, 2022, pp. 1–5. doi: 10.1145/3485557.3485561.
- [11] X. Chen, Z. S. Wu, and M. Hong, "Understanding gradient clipping in private SGD: A geometric perspective," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, 2020.
- [12] F. F. Mojtahedi, N. Yousefpour, S. H. Chow, and M. Cassidy, "Deep Learning for Time Series Forecasting: Review and Applications in Geotechnics and Geosciences," *Arch. Comput. Methods Eng.*, vol. 32, no. 6, pp. 3415–3445, 2025, doi: 10.1007/s11831-025-10244-5.
- [13] J. Krzywanski, M. Sosnowski, K. Grabowska, A. Zylka, L. Lasek, and A. Kijo-Kleczkowska, "Advanced Computational Methods for Modeling, Prediction and Optimization—A Review," *Materials (Basel)*, vol. 17, no. 14, 2024, doi: 10.3390/ma17143521.
- [14] S. Hikmat Haji and A. Mohsin Abdulazeez, "COMPARISON OF OPTIMIZATION TECHNIQUES BASED ON GRADIENT DESCENT ALGORITHM: A REVIEW," *J. Archaeol. Egypt/Egyptology*, vol. 18, no. 4, pp. 2715–2743, 2021.
- [15] T. Jun and A. B. Y. and N. S. B. Mansor, "Enhancing Efficiency and Accuracy of Optimization Techniques in Time Series Data Prediction Using Machine Learning: A Systematic Literature Review," pp. 129–142, 2024, doi: 10.32996/jcsts.
- [16] E. H. Sumiea *et al.*, "Deep deterministic policy gradient algorithm: A systematic review," *Heliyon*, vol. 10, no. 9, p. e30697, 2024, doi: 10.1016/j.heliyon.2024.e30697.
- [17] S. M. Al-Selwi *et al.*, "RNN-LSTM: From applications to modeling techniques and beyond—Systematic review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 5, p. 102068, 2024, doi: 10.1016/j.jksuci.2024.102068.
- [18] B. F. Azevedo, A. M. A. C. Rocha, and A. I. Pereira, *Hybrid approaches to optimization and machine learning methods: a systematic literature review*, vol. 113, no. 7. Springer US, 2024. doi: 10.1007/s10994-023-06467-x.
- [19] E. Harmon-Jones and J. Mills, "An introduction to cognitive dissonance theory and an overview of current perspectives on the theory," *Cogn. dissonance Reexamining a pivotal theory Psychol. (2nd ed.)*, pp. 3–24, 2019, doi: 10.1037/0000135-001.
- [20] Sawssen Briki, Nesrine Khabou & I, and smael Bouassida Rodriguez, "A Comparative Analysis of Time Series Prediction Techniques a Systematic Literature Review (SLR)," in *Model and Data Engineering Conference paper*, 2023. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-49333-1_1
- [21] M. Nadhir, A. Wahab, S. Nefti-meziani, and A. Atyabi, "A Comprehensive Review of Swarm Optimization Algorithms," pp. 1–36, 2015, doi: 10.1371/journal.pone.0122827.
- [22] B. Kitchenham and C. Ebse, "Guidelines for performing systematic literature reviews in software engineering," 2007.
- [23] 2007 Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," *Tech. report*, Ver. 2.3

- EBSE Tech. Report. EBSE*, 2007.
- [24] A. Koloskova, H. Hendrikx, and S. U. Stich, "Revisiting Gradient Clipping: Stochastic bias and tight convergence guarantees," *Proc. Mach. Learn. Res.*, vol. 202, pp. 17343–17363, 2023.
- [25] S. Venkatesh, I. Hossen, L. Shan, and V. Srinivas, "Electronic Commerce Research and Applications DP-SGD-global-adapt-V2-S : Triad improvements of privacy , accuracy and fairness via step decay noise multiplier and step decay upper clipping threshold," *Electron. Commer. Res. Appl.*, vol. 70, no. July 2024, p. 101476, 2025, doi: 10.1016/j.elerap.2025.101476.
- [26] B. Elesedy and M. Hutter, "U-Clip: On-Average Unbiased Stochastic Gradient Clipping," *JCO*, vol. 3, no. 2, pp. 1–24, 2023, [Online]. Available: <http://arxiv.org/abs/2302.02971>
- [27] J. Zhang, T. He, S. Sra, and A. Jadbabaie, "Why Gradient Clipping Accelerates Training: a Theoretical Justification for Adaptivity," *8th Int. Conf. Learn. Represent. ICLR 2020*, no. 2016, pp. 1–21, 2019.
- [28] I. Merad and S. Gaïffas, "Robust Stochastic Optimization via Gradient Quantile Clip- ping," in *International Conference on Machine Learning, Vancouver, Canada.*, 2021.
- [29] S. Chezhegov et al., "Clipping Improves Adam-Norm and AdaGrad-Norm when the Noise Is Heavy-Tailed," in *International Conference on Machine Learning, Vancouver, Canada.*, 2025.
- [30] T. Pethick, K. Antonakopoulos, and A. Silveti-falls, "Generalized Gradient Norm Clipping," in *39th Conference on Neural Information Processing Systems (NeurIPS 2025).*, 2025.
- [31] X. Ye, Y. Wu, W. Zhang, X. Li, Y. Chen, and C. Jin, "Optimized Gradient Clipping for Noisy Label Learning," in *The Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25)*, 2025.
- [32] A. Reisizadeh, H. Li, S. Das, and A. Jadbabaie, "Variance-reduced Clipping for Non-convex Optimization," vol. 2, pp. 1–25, 2023.
- [33] C. Wei, W. Li, C. Gong, and W. Chen, "DC-SGD : Differentially Private SGD with Dynamic Clipping through Gradient Norm Distribution Estimation," 2025, pp. 1–18.
- [34] H. Florian, "From Gradient Clipping to Normalization for Heavy Tailed SGD," 2024.
- [35] A. K. Menon, A. S. Rawat, S. J. Reddi, and S. Kumar, "CAN GRADIENT CLIPPING MITIGATE LABEL NOISE?," no. Lemma 6, pp. 1–26, 2020.
- [36] R. K. Srivastava and N. E. Toklu, "ClipUp: A Simple and Powerful Optimizer for Distribution-based Policy Evolution," pp. 1–20, 2020.
- [37] H. Yong, J. Huang, X. Hua, and L. Zhang, "Gradient Centralization: A New Optimization Technique for Deep Neural Networks," 2020.
- [38] B. Zhang, J. Jin, C. Fang, and L. Wang, "Improved analysis of clipping algorithms for non-convex optimization," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, 2020.
- [39] A. Author and A. Address, "Gradient Clipping Helps in Non-Smooth Stochastic Optimization with Heavy-Tailed Noise," no. NeurIPS, 2021.
- [40] G. Andrew and H. B. McMahan, "Differentially Private Learning with Adaptive Clipping," vol. 1, no. NeurIPS, pp. 1–12, 2021.
- [41] X. Yang and W. Chen, "Normalized / Clipped SGD with Perturbation for Differentially Private Non-Convex Optimization," vol. 1, no. 1, pp. 1–25.
- [42] J. Wei, "DPIS : An Enhanced Mechanism for Differentially Private SGD with Importance Sampling," no. November 2022, 2025, doi: 10.1145/3548606.3560562.

- [43] M. Liu, Z. Zhuang, Y. Lei, and C. Liao, "A Communication-Efficient Distributed Gradient Clipping Algorithm for Training Deep Neural Networks," *Adv. Neural Inf. Process. Syst.*, vol. 35, no. NeurIPS, pp. 1–14, 2022.
- [44] M. van Dijk, P. H. Nguyen, T. N. Nguyen, and L. M. Nguyen, "Generalizing DP-SGD with Shuffling and Batching Clipping," *Comput. Math. Methods Med.*, vol. 2, no. 3, 2022, [Online]. Available: <http://arxiv.org/abs/2212.05796>
- [45] T. N. Nguyen, P. H. Nguyen, L. M. Nguyen, and M. Van Dijk, "Batch Clipping and Adaptive Layerwise Clipping for Differential Private Stochastic Gradient Descent," pp. 1–20, 2023, [Online]. Available: <http://arxiv.org/abs/2307.11939>
- [46] V. V. Mai and M. Johansson, "Stability and Convergence of Stochastic Gradient Clipping: Beyond Lipschitz Continuity and Smoothness," *Proc. Mach. Learn. Res.*, vol. 139, pp. 7325–7335, 2021.
- [47] E. Gorbunov, M. Danilova, I. Shibaev, P. Dvurechensky, and A. Gasnikov, "Near-Optimal High Probability Complexity Bounds for Non-Smooth Stochastic Optimization with Heavy-Tailed Noise," no. 2020, 2021, [Online]. Available: <http://arxiv.org/abs/2106.05958>
- [48] B. Li, "Convergence and Privacy of Decentralized Nonconvex Optimization with Gradient Clipping and Communication Compression," 2023.
- [49] S. Yu, "Decentralized Nonconvex Optimization under Heavy-Tailed Noise : Normalization and Optimal Convergence," 2025.
- [50] J. Fong, S. Chen, and K. Chen, "Improving Layer-wise Adaptive Rate Methods using Trust Ratio Clipping," *J. Comput. Commun.*, vol. 5, no. 1, 2020, [Online]. Available: <http://arxiv.org/abs/2011.13584>
- [51] Z. Bu, R. Liu, Y.-X. Wang, S. Zha, and G. Karypis, "On the accuracy and efficiency of group-wise clipping in differentially private optimization," pp. 1–18, 2023, [Online]. Available: <http://arxiv.org/abs/2310.19215>
- [52] Y. Guo and C. Wei, "Multi-Task Learning Using Gradient Balance and Clipping with an Application in Joint Disparity Estimation and Semantic Segmentation," *Electron.*, vol. 11, no. 8, 2022, doi: 10.3390/electronics11081217.
- [53] J. M. Ede and R. Beanland, "Adaptive learning rate clipping stabilizes learning," *Mach. Learn. Sci. Technol.*, vol. 1, no. 1, 2020, doi: 10.1088/2632-2153/ab81e2.
- [54] P. Seetharaman, G. Wichern, B. Pardo, and J. Le Roux, "Autoclip: Adaptive gradient clipping for source separation networks," *IEEE Int. Work. Mach. Learn. Signal Process. MLSP*, vol. 2020-Septe, no. 1, 2020, doi: 10.1109/MLSP49062.2020.9231926.
- [55] Y. Wang, H. He, and X. Tan, "Truly Proximal Policy Optimization," 2017.
- [56] S. Yang, C. Dou, P. Guo, and K. Lu, "DCPO: Dynamic Clipping Policy Optimization," pp. 1–21, 2025.
- [57] Y. Chen, H. Zhang, and C. Zhang, "ClipQ : Clipping Optimization for the Post-Training Quantization of Convolutional Neural Network," pp. 1–17, 2025.
- [58] N. Marshall, K. L. Xiao, A. Agarwala, and E. Paquette, "To Clip or not to Clip : the Dynamics of SGD with Gradient Clipping in High-Dimensions," 2024.
- [59] C. Sakr, S. Dai, R. Venkatesan, B. Zimmer, W. J. Dally, and B. Khailany, "Optimal Clipping and Magnitude-aware Differentiation for Improved Quantization-aware Training," 2022.
- [60] J. Tu, A. Yasin, and N. S. Mansor,

- "NadamClip : A Novel Optimization Algorithm for Improving Prediction Accuracy and Training Stability," pp. 1–33, 2025.
- [61] M. Al-Emran, I. Arpaci, and S. A. Salloum, "An empirical examination of continuous intention to use m-learning: An integrated model," *Educ. Inf. Technol.*, vol. 25, no. 4, pp. 2899–2918, 2020, doi: 10.1007/s10639-019-10094-2.
- [62] T. Dybå and T. Dingsøy, "Empirical studies of agile software development: A systematic review," *Inf. Softw. Technol.*, vol. 50, no. 9–10, pp. 833–859, 2008, doi: 10.1016/j.infsof.2008.01.006.
- [63] S. M. Sarkintudu, H. H. Ibrahim, and A. A. Wahab, "Cryptocurrency platform ecosystem : a systematic literature review from information systems perspective," vol. 9, no. 44, pp. 308–315, 2019.
- [64] M. Al-Emran, V. Mezhuyev, A. Kamaludin, and K. Shaalan, "The impact of knowledge management processes on information systems: A systematic review," *Int. J. Inf. Manage.*, vol. 43, no. October 2017, pp. 173–187, 2018, doi: 10.1016/j.ijinfomgt.2018.08.001.
- [65] A. A. Alqudah and M. Al-emran, "applied sciences Technology Acceptance in Healthcare : A Systematic Review," *Appl. Sci.*, no. 1, p. 140, 2021.
- [66] M. Alshamsi, M. Al-Emran, and K. Shaalan, "A Systematic Review on Blockchain Adoption," *Appl. Sci.*, vol. 12, no. 9, pp. 1–18, 2022, doi: 10.3390/app12094245.
- [67] S. Yu, "Smoothed Gradient Clipping and Error Feedback for Decentralized Optimization under Symmetric Heavy-Tailed Noise," pp. 1–24, 2024.
- [68] J. Peng, W. Li, and Q. Ling, "Byzantine-Robust Decentralized Stochastic Optimization with Stochastic Gradient Noise-Independent Learning Error arXiv : 2308 . 05292v1 [cs . LG] 10 Aug 2023," vol. 00, pp. 1–30, 2023.
- [69] L. Guo, M. Li, S. Xu, and F. Yang, "Application of stochastic gradient descent technique for method of moments," *Proc. 2020 IEEE Int. Conf. Comput. Electromagn. ICCEM 2020*, pp. 97–98, 2020, doi: 10.1109/ICCCEM47450.2020.9219400.
- [70] C. Fang, Z. Yang, and W. U. Bajwa, "BRIDGE: Byzantine-resilient Decentralized Gradient Descent," 2021.
- [71] A. Reisizadeh, S. Prakash, R. Pedarsani, and A. S. Avestimehr, "CodedReduce: A Fast and Robust Framework for Gradient Aggregation in Distributed Learning," *IEEE/ACM Trans. Netw.*, vol. 30, no. 1, pp. 148–161, 2022, doi: 10.1109/TNET.2021.3109097.
- [72] A. Reisizadeh, A. Mokhtari, H. Hassani, and R. Pedarsani, "UC Office of the President An Exact Quantized Decentralized Gradient Descent Algorithm," 2019.
- [73] D. Han, H. Jiang, L. Wang, X. Zhu, Y. Chen, and Q. Yu, "Collaborative Task Allocation and Optimization Solution for Unmanned Aerial Vehicles in Search and Rescue," *Drones*, vol. 8, no. 4, pp. 1–18, 2024, doi: 10.3390/drones8040138.