Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Enhancing Efficiency and Accuracy of Optimization Techniques in Time Series Data Prediction Using Machine Learning: A Systematic Literature Review

Tu Jun[1]✉, Azman B Yasin[2] and Nur Suhaili Binti Mansor[3]
[123]University Utara Malaysia, Kedah, Malaysia
Corresponding Author: Tu Jun, **E-mail**: tujun792324486@gmail.com

| ABSTRACT
Undoubtedly, time series data prediction stands as a primary focus of computational intelligence among researchers in both academia and industry, owing to its wide-ranging applications and significant impact. The efficacy of prediction models heavily relies on the optimization techniques employed to enhance both efficiency and accuracy. Within the realm of Machine Learning (ML), researchers have developed numerous optimization techniques and models, leading to a plethora of studies. Consequently, there exists a considerable body of literature comprising reviews of ML-based time series prediction. Recently, Deep Learning (DL) models have emerged in this domain, demonstrating performance levels that notably surpass those of traditional ML methods. Despite the burgeoning interest in advancing time series prediction models, there remains a noticeable absence of systematic review papers dedicated solely to enhancing the efficiency and accuracy of optimization techniques. Therefore, this paper is motivated by the need to present a systematic review of the efficiency and accuracy of optimizers studies concerning time series prediction implementations. We not only classify these studies based on their targeted optimization techniques, prediction applications—such as crop yield prediction, weather forecasting for farming, and pest detection and management—but also categorize them according to the types of optimization techniques and models employed, including Convolutional Neural Networks (CNNs), Deep Belief Networks (DBNs), and Long-Short Term Memory (LSTM) networks. Additionally, this study endeavor to provide insights into the future of the field by highlighting the challenges and potential future research opportunities, thereby offering guidance to interested researchers.

| KEYWORDS

Time series, optimization techniques, machine learning algorithms, systematic literature review

| ARTICLE INFORMATION

ACCEPTED: 10 November 2024 **PUBLISHED:** 26 November 2024 **DOI:** 10.32996/jcsts.2024.6.5.11

1. Introduction

Time series data conceptually defined as a sequence of observations or measurements collected, recorded, or observed over successive and evenly spaced intervals of time[1]. For example, Sharma et al. (2022) argued that in time series data, each data point is associated with a specific timestamp or time period, allowing for the temporal ordering of the observations[2]. The data format is particularly relevant for capturing and analyzing the evolution of a phenomenon, variable, or system over time. Adamides et al. the primary characteristics of time series data include its temporal structure, where the order of observations is critical, and the dependence between consecutive data points [3]. Time series analysis involves exploring patterns, trends, seasonality, and irregularities within the data to gain insights into the behavior of the underlying process [4].

However, efficiency and accuracy, are two fundamental pillars in the domain of time series prediction, wielding significant importance due to their implications across various fields such as finance meteorology, healthcare and agriculture [5][6][7] . Efficiency in time series prediction is crucial as it directly influences the responsiveness and scalability of predictive models[2]. As

Copyright: © 2024 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

highlighted by Mazambani & Mutambara, in financial markets, timely decision-making is paramount for maximizing returns and minimizing risks[8]. Therefore, predictive models must operate efficiently to process vast amounts of data swiftly, enabling traders and analysts to capitalize on fleeting market opportunities. Moreover, in real-time applications like traffic management systems or energy grid optimization, efficient prediction algorithms are indispensable for making rapid adjustments and optimizations based on incoming data streams[9]. Concurrently, accuracy holds paramount importance as erroneous predictions can lead to suboptimal decisions, resulting in significant financial losses or even jeopardizing lives in critical applications. In healthcare, inaccurate predictions in disease outbreak forecasting could lead to inadequate resource allocation and ineffective containment measures, potentially exacerbating public health crises[10]. Similarly, in climate modelling, inaccuracies in predicting extreme weather events may hinder disaster preparedness efforts and lead to inadequate mitigation strategies[11]. The synergy between efficiency and accuracy is evident in their complementary roles. As noted by Sze et al., achieving high accuracy often requires complex models that demand substantial computational resources[12]. However, these resource-intensive models may sacrifice efficiency, leading to longer prediction times and hindering real-time decision-making. Hence, there is a delicate balance to be struck between efficiency and accuracy, where optimizing one aspect may come at the expense of the other.

Despite the existence of a vast amount of review studies covering many fields of time series prediction using traditional soft computing techniques, to the best of our knowledge, no systematic reviews have been performed in the literature for accuracy and enhancing efficiency and accuracy. Therefore, this research opted for a systematic review of time series data prediction optimization methods. Our motivation is twofold: first, to offer an up-to-date overview of both academic and industrial viewpoints on the optimization techniques employed in time series data prediction and deep learning models, and second, to identify the key and unique features of each optimization model studied, thus aiding researchers and practitioners in making informed decisions during their system development phase. Additionally, we speculate on future industry trends. The primary objective of this paper was to address the following research inquiries:

1) Which optimization techniques are mostly used for time series prediction?
2) What are the current challenges of optimization techniques in time series studies?
3) What are the potential future directions for optimization techniques research for time series prediction?

Our focus was solely on modern optimization techniques implementations for enhancement of efficiency and accuracy in time series data prediction such as Adam, RMSprop, Adagrad and AdaDelta etc. For other traditional optimization techniques-based applications such as Genetic algorithms, Stochastic Optimization, Gradient Descent, Stochastic Gradient Descent (SGD), Swarm Intelligence etc., interested readers can check the recent survey paper [13]. The study conducted a systematic literature review focusing on optimization techniques for time series prediction, excluding ML or DL models that were not centered on optimization. Optimization algorithms were included if they had use cases, even if not directly intended for time series prediction. Algorithmic trading papers addressing efficiency and accuracy prediction were considered, while those lacking a time series component were excluded. Sources surveyed included journals, conferences, PhD theses, book chapters, and web searches, limited to English language articles. Despite the recent prevalence of the term "deep learning," older studies implementing deep models like Recurrent Neural Networks (RNN) were incorporated. This review closes a gap in the body of knowledge, offering insight into the efficiency and accuracy of optimization techniques for time series prediction. The aim is to aid researchers and developers in implementing optimization algorithms effectively. This study distinguishes itself from existing ML-focused reviews by emphasizing optimization techniques.

The rest of the paper is organized as follows: Following this brief introduction, in Section 2, the existing optimization techniques that are focused on ML and DL studies for any type of time series prediction are mentioned. In Section 3, we will cover the existing optimization algorithms that are used, such as CNN, LSTM, Deep Reinforcement Learning (DRL). In each subsection, the problem definition will be given, followed by the particular optimization implementations. In Section 5, the overall statistical results of our findings were presented including histograms about the yearly distribution of different subfields, models, publication types, etc. As a result, the state-of-the-art snapshot for time series prediction studies will be given through these statistics. At the same time, it will also show the areas that are already mature, compared against promising or new areas that still have room for improvement. Section 6 will provide discussions about what has been done through academic and industrial achievements and expectations of what might be needed in the future. The section will include highlights about the open areas that need further research. Finally, we will conclude in Section 7 by summarizing our findings.

2. Literature Review

The application of machine learning methods in time series forecasting has widely and deeply impacted various fields. Traditional time series forecasting methods, such as ARIMA and exponential smoothing, while performing well in some cases, often fall short when dealing with complex nonlinear relationships and high-dimensional data. Machine learning methods, especially deep learning techniques such as Long Short-Term Memory Networks (LSTMs) and Convolutional Neural Networks (CNNs), excel in capturing complex patterns and long-term dependencies in data.

2.1 Optimization Techniques

Optimization techniques are algorithms and methods used to find the best solution from a set of possible solutions, typically in the context of mathematical or computational problems. In machine learning and deep learning, optimization techniques are crucial for training models and finding optimal parameter values.

Prior studies identified and investigated some common optimization methods used in various fields, including machine learning [14]–[16]. For example, Diab proposed an improvement of Stochastic Gradient Descent (SGD) algorithm performance in text classification. Utilizing a Grid-Search approach for hyperparameter tuning, the study explores various settings for feature representation, transformation, and weighting in terrorist attack incident summaries [17]. Testing SGD on SVM, Logistic Regression, and Perceptron classifiers via 10-K-fold cross-validation reveals that grid-search optimization enhances both pre-classification settings and classifier performance in accuracy and execution time. Liu et al. (2021) addresses the challenges in medical image processing using deep learning, focusing on the time-consuming nature of model training due to numerous parameters [18]. It introduces a novel algorithm, Particle Swarm Optimization-Stochastic Gradient Descent with Momentum (PSO-SGD), combining the advantages of both stochastic gradient descent and particle swarm optimization.

To overcome inadequacies associated with SGD, alternative methods, such as RMSProp, adapt learning rates individually based on historical gradient behavior, mitigating convergence challenges have been proposed [2], [19], [20]. For example, Uddin et al. argued that RMSProp uses the running average of squared gradients to normalize the gradient updates for each parameter. This normalization helps prevent the learning rates from becoming too large for parameters with large gradients which can lead to overshooting the optimal solution and too small for parameters with small gradients which can lead to slow convergence [21]. Li et al. introduce the Advanced Root Mean Square Propagation with Warm-Up (ARW) algorithm to enhance fiber coupling efficiency in applications like free-space optical communication [22]. The ARW algorithm adapts gain and perturbation rates, achieving higher effective range and steady-state coupling efficiency. It outperforms the SPGD algorithm, requiring fewer iterations to handle sudden disturbances and offering a 20% larger effective range in specific coupling platforms. Simulation and experimental results validate the ARW algorithm's efficacy. Most optimization techniques such as RMSPro, Adam, and Nadam are grounded in the same foundational principles as the original SGD [23]. Since the emergence of the ML and more recently DL, researchers on prediction have paid more attention to the building and training model with less emphasis on the optimization techniques ecosystem. As for the data science scholars, this requires a more vivid discussion [24]. Therefore, the main focus of this review is on the optimization techniques perspective of time series data prediction in the ML or DL models.

2.2 Machine and Deep Learning Model

ML frameworks are the foundation of numerous machine learning systems [24]. They are algorithms or mathematical representations that discover links and patterns in data, empowering them to make predictions, decisions, or classifications without being explicitly programmed [25]. These models form the core of various applications across industries, from recommendation systems in e-commerce to image recognition in healthcare [26], [27]. In an attempt to capture temporal dependencies, DL models, particularly recurrent neural networks (RNNs) and their variants like long short-term memory networks (LSTMs) and gated recurrent units (GRUs), are often preferred for time series data prediction over traditional machine learning models[28]–[30]. For example, Abiodun et al.,found Simple ANNs have been applied to pattern recognition and linear regression tasks[31]. Annotation networks (ANNs) have been developed in several studies to capture intricate relationships in data. The design and training of ANN models and model evaluation, tuning, testing, deployment, and continuous monitoring and maintenance are all covered by statistical model prediction techniques. Through the integration of several steps, this process guarantees that the models are theoretically valid but also dependable and effective in practical applications.

For example, Kadam et al. apply ANN and The Western Ghats, the Shivganga River Basin in India, uses multiple linear regression (MLR) techniques to predict whether the groundwater quality is suitable for drinking [32]. The outcomes verify that the performance is consistently acceptable in both seasons and that the ANN model's predictions are satisfactory. The suggested ANN model might be helpful in related research for predicting groundwater quality for human consumption. Xia et al. utilizes an ANN to analyze the impact of rolling parameters (e.g., thickness reduction, tension, speed, friction) on tandem cold rolling processes [33]. The ANN is trained with real data from practical rolling experiments. The best-performing ANN network topology has nine neurons in the hidden layer, achieving high correlation coefficients. The results also reveal the critical slip phenomena in cold rolling and demonstrate an accurate prediction of parameters, particularly slip, friction, and angular velocity. The Mean Squared Error for predicting slip is 4.2 × 10^(-4).In another study, Noori et al., created a hybrid model by combining the suggested process-based watershed model with ANN [34]. The two models work together to maximize the process of calibration and validation., and the hybrid model has good predictive power for nitrate, ammonium, and phosphate. In fact, it even outperforms the SWAT model calibrated at every location. Jagannath et al. present a practical and efficient Automatic Modulation Classification (AMC) system that leverages ANN [35]. The system's key strengths lie in its ability to provide robust performance across different SNRs, low computational complexity for easy implementation, and flexibility for future expansion of modulation format dictionaries, thereby showcasing its suitability for real-time commercial scenarios. Ibrahim et al., reported investigated the potential sources of pollution and the most important parameters leading to spatial variability, and used principal component analysis (PCA) and artificial neural networks (ANN) to develop optimal inputs for water quality modeling [36].

2.3 *Time Series Data Prediction*

Time series data conceptually defined as a sequence of observations or measurements collected, recorded, or observed over successive and evenly spaced intervals of time [1]. For example, Sharma et al. argued that in time series data, each data point is associated with a specific timestamp or time period, allowing for the temporal ordering of the observations [2]. The data format is particularly relevant for capturing and analyzing the evolution of a phenomenon, variable, or system over time. Adamides et al. the primary characteristics of time series data include its temporal structure, where the order of observations is critical, and the dependence between consecutive data points [3]. Time series analysis involves exploring patterns, trends, seasonality, and irregularities within the data to gain insights into the behavior of the underlying process [4]. A comprehensive literature review of time series data covers a wide range of topics, methodologies, and applications across various disciplines [37]–[40]. For instance, Büyükşahin & Ertekin introduce a new hybrid approach, joining Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN), operating in a more versatile framework. Experimental results highlight the significance of data decomposition and the hybridization process for improved forecasting accuracy, suggesting the method's effectiveness in enhancing traditional hybrid and individual forecasting methods [41]. The study discusses the importance of accurate time series forecasting across various domains and the limitations of existing methods. Ugurlu & Oksuz Umut explore intraday electricity markets, emphasizing the surge in research on price prediction. While many studies focus on market features, this work compares various models, highlighting the superiority of recurrent neural networks, especially the gated recurrent unit network architecture. Results reveal the day-ahead market price's pivotal role in intraday forecasting, suggesting it is an efficient method for prediction [42].

Gurrib et al. introduce a method for predicting Bitcoin price direction, combining linear discriminant analysis (LDA) and sentiment analysis [43]. The model, utilizing current Bitcoin prices and Twitter headline news, achieves high accuracy, surpassing benchmark targets with a forecast accuracy of 0.828. Heintz et al. explore Linear Discriminant Analysis (LDA) for supervised classification, demonstrating that with minimal prior information, it is possible to compute the exact LDA projection vector from unlabeled data. Three pieces of information are identified as sufficient, and the proposed Minimally Informed LDA (MILDA) model, validated through experiments, matches the performance of supervised LDA, offering adaptability to non-stationary data [44]. Karimuzzaman et al. demonstrated the application of linear data analysis in predicting stock prices based on historical data, highlighting the practical relevance of linear models in financial decision-making. The stock price prediction using linear classification models like Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Partial Least-Square Discriminant Analysis (PLS-DA) [25]. Despite the dataset having characteristics favoring other models, LR performs the best, suggesting its suitability in scenarios with high correlation, multicollinearity, multivariate normality, and high dimensionality. The related work dealing with the time series data are given in a summary in Table1, with the application domain.

Table 1 Summary of some selected time series data research

Author	Dataset	Number	Domain Application
Büyükşahin & Ertekin	Wolf's sunspot data, Canadian lynx data, and the British pound/US dollar exchange rate data	731 data points	Finance
Ugurlu & Oksuz Umut	Electricity price of Turkey intraday market	26,280 data point	Finance
Heintz et al.	Image	100	N/A
Gurrib et al.	Daily bitcoin prices	990 data point	Finance

3. Methodology

For the analysis of indicators and indicator selection processes, the systematic literature review (SLR) methodology was followed. SLR aims to capture and evaluate the current research findings in a specific area of knowledge by answering research questions with data that satisfies pre-defined inclusion criteria [45]. SLR was chosen to guarantee a systematic and unbiased approach that produces reliable results [46]. The PSALSAR framework suggested by Mengist et al.[47] and search and selection procedure by Kitchenham, B & Charters [48] respectively was followed for the review. PSALSAR stands for the six steps in the framework: Protocol, Search, Appraisal, Synthesis, Analysis and Report (See Figure 1). After the research scope was determined, the research questions were formed in the Protocol step. The research scope of this study.

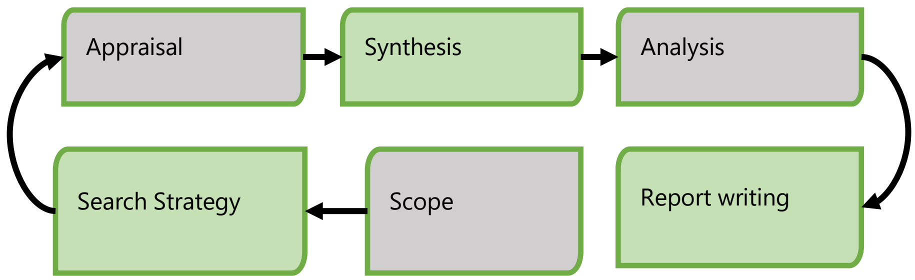

Figure 1 The PSALSAR framework

3.1 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 was conducted in the title, abstract, and keyword search fields. To find all relevant studies, the following string was formed and used considering possible synonyms: "Time series data*" AND "time series", "smart series data*" AND "optimization algorithm", "efficiency" AND "accuracy", "prediction*" AND "forecasting", "linear data*" AND "non-linear data", "Enhancing Efficiency" AND "Accuracy of Optimization Techniques".

3.2 The Inclusion and exclusion criteria

At this stage, the papers were selected based on these three criteria. The Scopus, ScienceDirect, and Google Scholar databases were searched as they are internationally recognized, prominent and widely used abstract and indexing databases for peer-reviewed scientific literature. For testing the effectiveness of the search, queries were tried out in the databases determined for the study. The search has resulted in the retrieval of papers that were previously identified as relevant. Therefore, the search queries were considered valid. The Assessment step consisted of the evaluation of papers according to the following predetermined inclusion and exclusion criteria:

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit the use of hard returns to only one return at the end of a paragraph. Do not add any pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

Table 2 lists the inclusion and exclusion criteria for the papers that were critically evaluated in this review

Actions	Criteria
Inclusion Criteria	1) Must have publication dates ranging from 2019 to 2024. 2) Must present a theoretical framework for optimization techniques in time series data prediction. 3) Must evaluate optimization techniques in time series data prediction must be written in English.

Exclusion 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.

114 papers were selected from the literature using the above inclusion criteria. All these papers were used for the optimization techniques for time series data prediction. The papers were subjected to a filtering procedure before analysis. 65 of these papers had to be excluded based on the title, and abstract text did not include one of the keywords such as optimization techniques or time series data prediction. Another 11 of the papers from 49 papers had to be excluded based on reading the full text. At the end of the filtering procedure, 38 papers were found to be suitable for qualitative analysis. Extraction and classification of data from the papers were conducted in the Synthesis step. Firstly, the papers were classified into two sub-categories according to the concepts they were dealing with: (i) optimization techniques (ii) time series data (See Table 3).

Table 3 Categorization of the reviewed articles according to the concepts they were dealing with (i.e., optimization techniques, time series data).

Category	Number of Papers	Studies
Optimization Techniques	33	Diab, Gorbunov et al., Liu et al.,Zou & Sugihara, Li et al., Prilianti et al., Uddin et al., Bock et al., Reddi et al., Prabowo et al., Attrapadung et al., Sharma et al., Zhang et al., Xie et al. , Fatima, Kuppusamy et al., et al.
Time series data	5	Ugurlu & Oksuz Umut, Büyükşahin & Ertekin, Heintz et al., Gurrib et al., Karimuzzaman et al

After that, variables were identified to organize the papers: (i) Type, (ii) Author, (iii) Year, (iv) optimization techniques. The Type variable showed if the paper was on optimization techniques, time series data prediction, or efficiency and accuracy. The Author and Year variables were used for the identification of the paper. The optimization techniques variable demonstrated the methodology used in enhancing the efficiency and accuracy of the study was implemented. All the variables and the related data were extracted to an Excel spreadsheet and made ready for analysis. In the Analysis step, all relevant information was retrieved from the data. The articles were analyzed according to the keywords used, year of publication, journal the paper was published in, and the optimization techniques used by the study. Further analysis was conducted on the category and their respective selection processes using the above-mentioned variables.

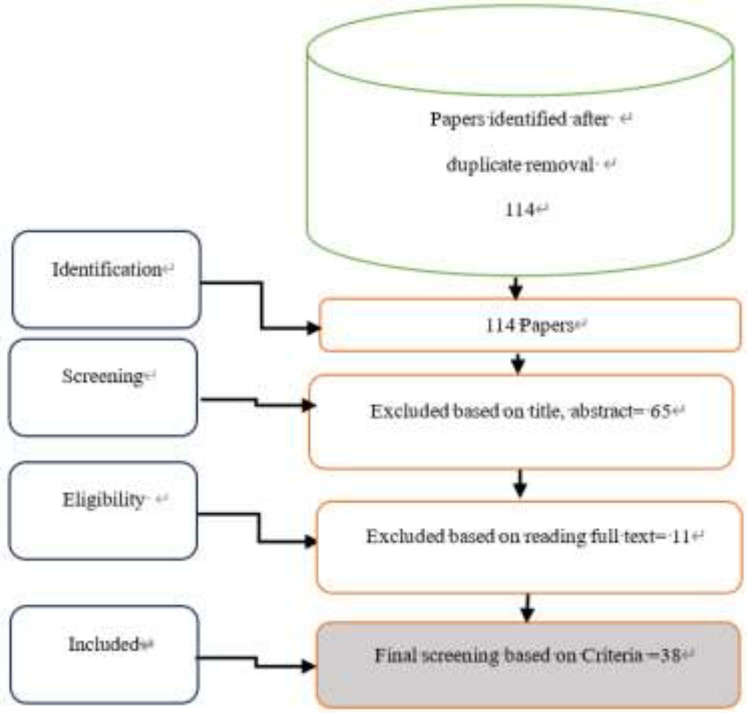

Figure 2 Search and selection procedure

3.3 Quality Assessment

In this paper multi, assessment criteria were deployed. First, the quality of the selected articles for the review was assessed based on whether the study presented clear objectives, provided appropriate data collection method, discussed appropriate data analysis method, and presented clearly stated findings. For example, if the objectives were clear, a score of 1 is given; if the objectives were partly clear, a score of 0.5 is given and if the objectives were not clear, a score of 0 is given. The quality assessment criteria were adapted from [49]. The total score for each article is calculated by adding the score. Table 4 shows the results of the quality assessment 1.

Table 4 Quality Assessment 1

Quality (scores)					Total
	Poor (1-1.99)	Fair (2-2.99)	Good (3-3.99)	V. Good (4)	100
Number of studies	2	5	10	21	
Percentage of papers (%)	5.2	13	26.3	55.2	

Out of 38 papers, 21 papers were rated very good and only two (2) papers were rated poor. Secondly, another set of nine criteria derived from Al-Emran et al. and Alqudah & Al-emran was utilized for quality assessment, shaping a method to evaluate the selected research papers (n = 38) for final analysis [50], [51]. The intent of this criteria wasn't to critique scholars' work but to gauge research quality, adapted from recommendations by Kitchenham & Charters [52]. Like the first criteria above each checklist question was rated on a three-point scale: if the objectives were clear, a score of 1 is given, if the objectives were partly clear, a score of 0.5 is given and if the objectives were not clear, a score of 0 is given, culminating in scores ranging from 0 to 9 for each study. Higher scores indicated a greater alignment with research objectives, determined through scrutiny of nine 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 4 outlines the quality assessment outcomes for all 38 papers, revealing that each study met the criteria and was deemed suitable for final analysis. As shown in Table 5:

Table 5 Quality Assessment

Numb	Checklist measurements
1	Are the research objectives clearly stated?
2	Was the aim of the study successfully accomplished?
3	Are the variables addressed the study clearly outlined?

4	Is the study's context clearly articulated?
5	Are the methods of data collection adequately?
6	Are the instrument reliable and valid?
7	Are the data analysis technique adequate?
8	Do the findings contribute to existing literature?
9	Does the study enhance knowledge?

Additionally, a supplementary set of nine criteria, drawn from Al-Emran et al. and Alqudah & Al-emran , was employed for quality evaluation, thus shaping a comprehensive method for assessing the 38 selected research papers, as delineated in Table 2 [50], [51]. This approach, inspired by recommendations by Kitchenham & Charters [52], aimed not to critique scholars' work but to evaluate research quality. Each criterion was rated on a three-point scale, contributing to a score ranging from 0 to 9 for each study. Higher scores indicated better alignment with research objectives, as determined through scrutiny of the nine quality assessment criteria. Both the first and second authors independently assigned scores to each study, resolving discrepancies through discussion and further examination of contentious papers. Table 6 provides an overview of the quality assessment results for all 38 papers, indicating that each study met the criteria and was deemed suitable for final analysis.

Table 6 outlines the quality assessment outcomes

Papers	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Sum	%
Paper1	0.5	1	0	1	1	0.5	1	1	1	7	77.7%
Paper2	1	1	1	0.5	1	1	0.5	1	1	8	88.8%
Paper3	0.5	1	1	1	1	1	1	1	0.5	8	88.8%
Paper4	1	1	1	0.5	1	1	1	1	1	8.5	94.4%
Paper5	0	1	0.5	1	1	0.5	1	1	1	7	77.7%
Paper6	1	1	1	1	1	1	0	0.5	1	7.5	83.3%
Paper7	0.5	1	1	0.5	1	1	1	0	1	7	77.7%
Paper8	1	1	1	1	1	1	0	0.5	1	7.5	83.3%
Paper9	0	0.5	1	1	1	1	1	1	0.5	7	77.7%
Paper10	1	1	1	1	1	0.5	1	1	1	8.5	94.4%
Paper11	1	1	1	1	1	1	0.5	0	1	7.5	83.3%
Paper12	0	0.5	1	1	1	1	1	1	0.5	7	77.7%
Paper13	1	0.5	1	1	0.5	1	1	0.5	1	7.5	83.3%
Paper14	1	1	1		1	1	1		1	7	77.7%
Paper15	1	1	0.5	1	1	1	0.5	0.5	0.5	7	77.7%
Paper16	1	0.5	1	1	0.5	1	1	1	1	8	88.8%
Paper17	1	1	1	0.5	1	1	1	0.5	0.5	7.5	83.3%
Paper18	0.5	1	1	1	0	1	0	0	1	5.5	61.1%
Paper19	1	1	1	0	1	1	1	0.5	1	7.5	83.3%
Paper20	1	1	1	0.5	1	1	1	1	0.5	8	88.8%
Paper21	1	0.5	1	1	0.5	1	1	1	1	8	88.8%
Paper22	0.5	1	1	1	1	1	1	1	1	8.5	94.4%
Paper23	1	0	1	0	0.5	1	1	1	1	6.5	72.2%
Paper24	0.5	1	1	1	1	0.5	1	1	0.5	7.5	83.3%
Paper25	1	0.5	1	1	1		0.5	1	1	7	77.7%
Paper26	1	0	1	1	1	1	1	1	1	8	88.8%
Paper27	0.5	1	1	1	1	0.5	1	1	1	8	88.8%
Paper28	1	0.5	1	1	1	1	0	0.5	1	7	77.7%
Paper29	0	1	0.5	1	1	0.5	1	1	1	7	77.7%
Paper30	1	1	1	1	1	1	0	0.5	1	7.5	83.3%
Paper31	0.5	1	1	0.5	1	1	1	0	1	7	77.7%
Paper32	1	1	1	1	1	1	0	0.5	1	7.5	83.3%
Paper33	0	0.5	1	1	1	1	1	1	0.5	7	77.7%
Paper34	1	1	1	1	1	0.5	1	1	1	8.5	94.4%
Paper35	1	1	1	1	1	1	0.5	0	1	7.5	83.3%
Paper36	0	0.5	1	1	1	1	1	1	0.5	7	77.7%

Paper37	1	0.5	1	1	0.5	1	1	0.5	1	7.5	83.3%
Paper38	1	0.5	1	1	0.5	1	1	0.5	1	7.5	83.3%

To address the research inquiries in this review, we've categorized the remaining 38 papers by various attributes such as optimization techniques and existing challenges in optimization techniques in time series data studies.

3.4 Data Extraction

In this study, Mendeley platform, a tool for managing references, was utilized to document the reference information of each article. The extracted data used to address the research inquiries encompasses the fields of study, methodologies employed, and the underlying theories.

4. Results and Discussion

The findings of this review were extracted according to the optimization techniques, time series data prediction and future research directions based on the comparative analysis of the current trend in optimization techniques and ML or DL model research field.

4.1 Optimization Techniques Mostly Used for Time Series Prediction

To answer research question one (1), Based on the results of extensive reading of the selected papers, the optimization algorithm methods used by the authors were gradient-based optimization and classified into their approaches to optimize model parameters. Which include first-order methods, second-order methods and adaptive learning rate methods respectively as listed below (Table 6). The findings reveal that only six (6) out of the 11 papers of Adaptive learning rate methods investigate Nadam optimizer a variant of Adam. For example, Sharma et al., examine LSTM networks for long-term solar PV power prediction, employing the Nadam optimizer [2]. Focused on the Indian solar context, it aims to enhance forecasting precision and efficiency. In addition, Zhang et al. introduce NAI-FGM, combining Nadam optimizer with a fast gradient method for adversarial attacks, enhancing efficacy especially in black-box scenarios[53]. Critically, it advances adversarial machine learning but lacks thorough empirical validation. Xie et al. explore LSTM with Nadam optimizer for solar PV forecasting in India, aiming to enhance accuracy[54]. Kuppusamy et al. compares Adam and Nadam in CNNs for facial emotion recognition, lacking detailed analysis[55]. Ezgi & Onan introduce two deep convolutional neural networks to automatically assess OA severity from X-ray images[56]. The first network detects knee joints accurately, while the second network, fine-tuned with an ordinal loss function, achieves state-of-the-art performance in classifying OA severity based on the Kellgren-Lawrence grading system. Results show high accuracy in knee joint detection and grading. As shown in Table 7:

Table 7 Optimization Techniques based Approaches to Optimizing Model

Gradient-Based Optimization Techniques Mostly Used for Time Series Prediction	Number of papers
First-order methods (i.e Gradient Descent, Stochastic Gradient Descent (SGD), and Mini-Batch Gradient Descent)	25
Second-order methods (i.e Newton's Methods and Quasi-Newton Methods)	2
Adaptive learning rate methods (i.e AdaGrad, RMSProp, Adam, Adadelta, AdamW and Nadam)	11

The evidence from this review indicates that data science scholars have yet to significantly contribute to reporting the scientific evidence of adaptive learning rate methods, specifically as depicted in Figure 3, the adaptive learning rate optimizer. This deficiency arises due to limitations in enhancing faster and smarter predictions for time series data. Therefore, it is imperative to explore this opportunity further.

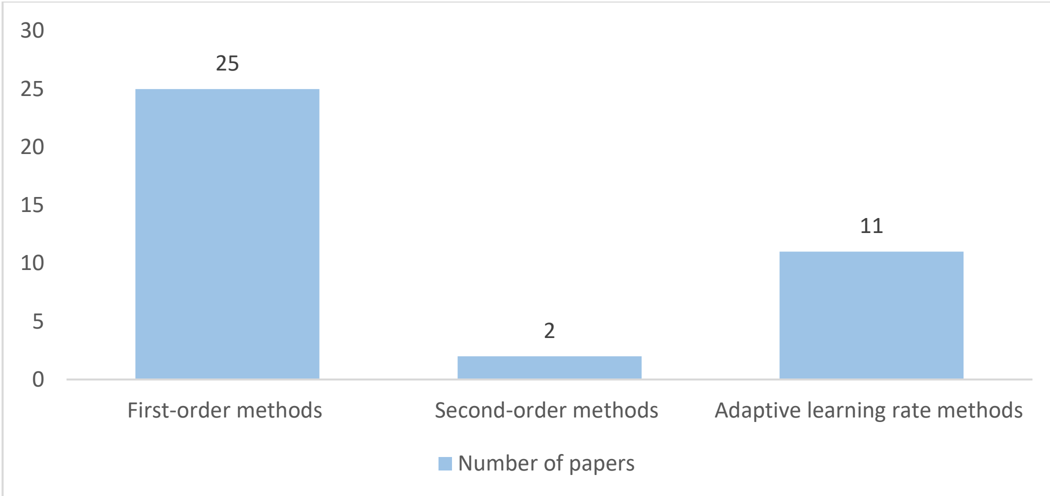

Figure 3 The Most Common Optimization Techniques Mostly Used for Time Series Prediction

4.2 What are the existing challenges in optimization techniques in time series data studies?

This study points out that although machine learning focuses on improving prediction accuracy and modeling complex patterns in time series data optimization, there is relatively little research on fast, intelligent prediction[57]. Time series data presents unique challenges such as trend, seasonality, and autocorrelation that require technical expertise. While the general approach is helpful, it is difficult to solve specific problems such as irregularity, missing values and non-stationarity. Especially in the context of infectious disease pandemics such as COVID-19, the importance of real-time or near-real-time predictive models is heightened[58]. Although the study by L. Zhou et al. demonstrated the effectiveness of the LSTM model in predicting COVID-19 trends, it failed to address the urgent need for rapid adaptive forecasting methods, and lacked a discussion of real-world deployment challenges and an assessment of the applicability of the updated technical framework.

Future research will focus on ammonia level prediction in specific fields such as aquaculture, using advanced models such as RNN and LSTM to capture complex time patterns, and combining domain knowledge to improve prediction accuracy and robustness. At the same time, real-time monitoring and adaptive forecasting methods will be the focus of research to respond to dynamic changes in aquaculture conditions and promote water quality maintenance and system health. Although studies have explored the potential of machine learning techniques, many lack a clear basis for research. Therefore, future research should focus on solving these key problems and provide stronger support for practical applications.

Data science scholars are working on breakthroughs in time series prediction using deep learning models. Although model contributions are emphasized, research into optimization techniques and model architectures is equally exciting. These studies have important implications for multiple industry roles such as data scientists, consultants, aquaculture, and machine learning developers. Challenges include balancing model complexity with interpretability, as well as missing values, outliers, and noise that are unique to time series data. The high dimensionality, nonlinearity and non-stationarity of time series data make it difficult for traditional optimization algorithms to cope with it. In addition, noise and outliers in the data, as well as complex patterns, trends and seasonality, add to the difficulty of optimization. Pretreatment technology is the key to solving these problems, but it is also challenging to choose the right pretreatment method and optimization model. The large number of models and parameters available makes the optimization process complex and time consuming. At the same time, optimization techniques can face compute-intensive and scalability issues, especially when dealing with large amounts of data or complex models. Therefore, future studies need to take these factors into account in order to develop more efficient and accurate time series prediction models [59] [60].

4.3 What are the potential future directions for optimization techniques research for time series prediction

While optimization techniques aim to improve the accuracy of time series forecasting models, ensuring the interpretability of these models remains important for decision-making. Complex optimization models may sacrifice interpretability, making it challenging for stakeholders to understand the underlying factors driving predictions. The studies presented so far provide an insight that can be useful to AI in general and more specifically ML/DL research communities and practitioners. Literature that focuses on examining the optimization techniques are merely simple integration of ML and optimization. For example, papers by Dogo et al.,Sharma et al., Q. Zhang et al., Z. Zhou et al., provide an insight into the leveraging the strengths of machine learning algorithms, such as deep learning and reinforcement learning, in conjunction with optimization techniques to develop hybrid models capable of handling complex temporal dependencies and improving prediction accuracy[2], [53], [61], [62]. It is also found that in many of these papers finding the optimal set of hyperparameters can be challenging and often requires significant computational resources and expertise. Although some scholars present their scientific findings using certain methodology, more needs to be done by the ML scholars in applying existing optimizer training ML/DL models by iteratively adjusting the model parameters to minimize a predefined loss function.

During training, gradients can grow exponentially, leading to unstable training and difficulty in convergence. Gradient clipping bounds the gradient values to a predefined threshold, preventing them from growing too large and causing numerical instability. This is particularly crucial in RNNs and their variants, where the vanishing or exploding gradient problem is common due to the nature of backpropagation through time. This study, so far, has provided significant evidence that might be useful for both the industry players and academic communities, especially the ML/DL scholars, to look into areas of cryptocurrencies platform ecosystem that are worth to be explored in the future. The findings also indicate that both the ML/DL or related focused on introducing and explaining the performance by emphasizing the performance of the model in terms of prediction rather than actual model training.

The choice of clipping threshold in gradient clipping is often based on empirical observations or heuristics. Further investigation could explore methods for automatically determining the optimal clipping threshold based on characteristics of the model architecture, dataset, or training dynamics. Adaptive or dynamic clipping strategies could also be explored to adjust the clipping threshold during training based on real-time feedback. While gradient clipping is known to stabilize training and prevent exploding gradients, its impact on overall model performance ability across different tasks and architectures may vary. Further investigation could systematically evaluate the effects of gradient clipping on various deep learning models and benchmark its performance against alternative optimization techniques.

While gradient clipping is a widely used technique in DL training, there is still room for further investigation to better understand its properties, optimize its parameters, and explore its interactions with other components of the training pipeline. Such investigations could lead to improved training methodologies, better-performing models, and deeper insights into the underlying optimization dynamics in DL especially using time series data.

Investigate how gradient clipping techniques can be adapted to better handle the temporal dynamics inherent in time series data. This could involve dynamic adjustment of clipping thresholds based on changes in the data distribution or temporal context, enabling more adaptive and effective gradient clipping strategies. In addition, exploring the interplay between gradient clipping and RNN architectures for time series analysis, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs). Investigating how different clipping strategies impact the training dynamics-based time series models could be another potential future research directions.

5. Conclusion

In spite of many research efforts by data science and other related fields to find accurate predictions through optimization techniques, there are still promising possibilities for enhancing faster and smarter predictions. In spite of many research efforts by data science, and AI fields in understanding the role of optimization techniques in iteratively adjusting the model parameters to minimize a predefined loss function, there are still promising possibilities of mitigating exploding gradients during the training DL models. This study has confirmed the need for more investigations on the need for gradient clipping is a valuable technique in deep learning training pipelines, providing stability, robustness, and faster convergence, ultimately leading to more reliable and effective models.

Even though the gradient clipping is a well-established technique in the field of DL, particularly for addressing issues related to exploding gradients during training, not much effort has been made as gradient clipping may ultimately introduce additional computational overhead, especially in distributed or parallel training settings. Further investigation could explore scalable and efficient implementations of gradient clipping algorithms that minimize computational overhead while maintaining effectiveness in stabilizing training. In addition, the use of time series data in the DL models may provide rigorous output in model training the emerging phenomenon is still lacking. Therefore, the conclusion cannot be recommended without acknowledging the limitations of this study. First and foremost, in conducting this study, the data gathered do not include the linear data that also investigate the issue of optimization techniques. Compared to other related SLR papers conducted previously, this study has moved forward by not just limiting the inclusion of traditional ML. Furthermore, the study focuses on the optimization techniques used in training ML/DL models, even though there are other emerging alternatives. The other optimizers will be considered as future work.

Funding: This research received no external funding
Conflicts of Interest: The authors declare no conflict of interest.
ORCID iD
https://orcid.org/0009-0007-2610-2530
https://orcid.org/0009-0009-8190-9227
https://orcid.org/0000-0002-7780-3691
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] R. Barzegar, M. T. Aalami, and J. Adamowski, "Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model," Stoch. Environ. Res. Risk Assess., vol. 34, no. 2, pp. 415–433, 2020, doi: 10.1007/s00477-020-01776-2.

[2] J. Sharma et al., "A novel long term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer: A case study of India," Energy Sci. Eng., vol. 10, no. 8, pp. 2909–2929, 2022, doi: 10.1002/ese3.1178.

[3] G. Adamides et al., "Smart farming techniques for climate change adaptation in Cyprus," Atmosphere (Basel)., vol. 11, no. 6, pp. 1–17, 2020, doi: 10.3390/ATMOS11060557.

[4] J. Qin, J. Liang, T. Chen, X. Lei, and A. Kang, "Simulating and predicting of hydrological time series based on tensorflow deep learning," Polish J. Environ. Stud., vol. 28, no. 2, pp. 795–802, 2019, doi: 10.15244/pjoes/81557.

[5] L. Li, P. Jiang, H. Xu, G. Lin, D. Guo, and H. Wu, "Water quality prediction based on recurrent neural network and improved evidence theory: a case study of {Qiantang} {River}, {China}," Environ. Sci. Pollut. Res., vol. 26, no. 19, pp. 19879–19896, Jul. 2019, doi: 10.1007/s11356-019-05116-y.

[6] P. G. Arepalli and K. J. Naik, "An IoT-based water contamination analysis for aquaculture using lightweight multi-headed GRU model," Environ. Monit. Assess., vol. 195, no. 12, pp. 1–21, 2023, doi: 10.1007/s10661-023-12126-4.

[7] I. Sansa, Z. Boussaada, M. Mazigh, and N. M. Bellaaj, "Solar radiation prediction for a winter day using ARMA model," 6th IEEE Int. Energy Conf. ENERGYCon 2020, vol. 2, no. 1, pp. 326–330, 2020, doi: 10.1109/ENERGYCon48941.2020.9236541.

[8] L. Mazambani and E. Mutambara, "Predicting FinTech innovation adoption in South Africa: the case of cryptocurrency," African J. Econ. Manag. Stud., vol. 11, no. 1, pp. 30–50, 2020, doi: 10.1108/AJEMS-04-2019-0152.

[9] S. Sun, H. Wu, and L. Xiang, "City-wide traffic flow forecasting using a deep convolutional neural network," Sensors (Switzerland), vol. 20, no. 2, 2020, doi: 10.3390/s20020421.

[10] C. Comito and C. Pizzuti, "Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review," Artif. Intell. Med., vol. 128, no. March, p. 102286, 2022, doi: 10.1016/j.artmed.2022.102286.

[11] M. Ehrendorfer, "Predicting the uncertainty of numerical weather forecasts: A review," Meteorol. Zeitschrift, vol. 6, no. 4, pp. 147–183, 1997, doi: 10.1127/metz/6/1997/147.

[12] V. Sze, Y. H. Chen, T. J. Yang, and J. S. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," Proc. IEEE, vol. 105, no. 12, pp. 2295–2329, 2017, doi: 10.1109/JPROC.2017.2761740.

[13] E. M. Dogo, O. J. Afolabi, N. I. Nwulu, B. Twala, and C. O. Aigbavboa, "A {Comparative} {Analysis} of {Gradient} {Descent}-{Based} {Optimization} {Algorithms} on {Convolutional} {Neural} {Networks}," in 2018 {International} {Conference} on {Computational} {Techniques}, {Electronics} and {Mechanical} {Systems} ({CTEMS}), Dec. 2018, pp. 92–99. doi: 10.1109/CTEMS.2018.8769211.

[14] M. M. Althobaiti, K. Pradeep Mohan Kumar, D. Gupta, S. Kumar, and R. F. Mansour, "An intelligent cognitive computing based intrusion detection for industrial cyber-physical systems," Meas. J. Int. Meas. Confed., vol. 186, no. September, p. 110145, 2021, doi: 10.1016/j.measurement.2021.110145.

[15] D. Li, X. Xu, Z. Li, T. Wang, and C. Wang, "Detection methods of ammonia nitrogen in water: A review," TrAC - Trends Anal. Chem., vol. 127, p. 115890, 2020, doi: 10.1016/j.trac.2020.115890.

[16] 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/ICCEM47450.2020.9219400.

[17] S. Diab, "Optimizing Stochastic Gradient Descent in Text Classification Based on Fine-Tuning Hyper-Parameters Approach.," vol. 16, no. 12, pp. 1–6, 2019.

[18] X. Liu, R. Feng, S. Zhou, and Y. Yang, "A Novel PSO-SGD with Momentum Algorithm for Medical Image Classification," Proc. - 2021 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2021, pp. 3408–3413, 2021, doi: 10.1109/BIBM52615.2021.9669876.

[19] 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," AIP Conf. Proc., vol. 2084, no. March, 2019, doi: 10.1063/1.5094284.

[20] M. Fakhry and A. F. A. Brery, "A Comparison Study on Training Optimization Algorithms in the biLSTM Neural Network for Classification of PCG Signals," 2022 2nd Int. Conf. Innov. Res. Appl. Sci. Eng. Technol. IRASET 2022, 2022, doi: 10.1109/IRASET52964.2022.9738309.

[21] M. J. Uddin, Y. Li, M. A. Sattar, Z. M. Nasrin, and C. Lu, "Effects of Learning Rates and Optimization Algorithms on Forecasting Accuracy of Hourly Typhoon Rainfall: Experiments With Convolutional Neural Network," Earth Sp. Sci., vol. 9, no. 3, 2022, doi: 10.1029/2021EA002168.

[22] Z. Li, Z. Pan, yuting li, X. Yang, C. Geng, and X. Li, "Advanced Root Mean Square Propagation with Warm-up algorithm for Fiber Coupling," Opt. Express, vol. 31, no. 15, pp. 23974–23989, 2023, doi: 10.1364/oe.494734.

[23] 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.

[24] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic MR image reconstruction," IEEE Trans. Med. Imaging, vol. 38, no. 1, pp. 280–290, 2019, doi: 10.1109/TMI.2018.2863670.

[25] M. Karimuzzaman, N. Islam, S. Afroz, and M. M. Hossain, “Predicting Stock Market Price of Bangladesh: A Comparative Study of Linear Classification Models,” Ann. Data Sci., vol. 8, no. 1, pp. 21–38, 2021, doi: 10.1007/s40745-020-00318-5.
[26] N. Attrapadung et al., “Adam in Private: Secure and Fast Training of Deep Neural Networks with Adaptive Moment Estimation,” Proc. Priv. Enhancing Technol., vol. 2022, no. 4, pp. 746–767, 2022, doi: 10.56553/popets-2022-0131.
[27] D. Pang, G. Liu, J. He, W. Li, and R. Fu, “Automatic Remote Sensing Identification of Co-Seismic Landslides Using Deep Learning Methods,” Forests, vol. 13, no. 8, 2022, doi: 10.3390/f13081213.
[28] Y. Li, K. Li, C. Chen, X. Zhou, Z. Zeng, and K. Li, “Modeling Temporal Patterns with Dilated Convolutions for Time-Series Forecasting,” ACM Trans. Knowl. Discov. Data, vol. 16, no. 1, 2021, doi: 10.1145/3453724.
[29] P. G. Arepalli, M. Akula, R. S. Kalli, A. Kolli, V. P. Popuri, and S. Chalichama, “Water Quality Prediction for Salmon Fish Using Gated Recurrent Unit (GRU) Model,” 2022 2nd Int. Conf. Comput. Sci. Eng. Appl. ICCSEA 2022, pp. 1–5, 2022, doi: 10.1109/ICCSEA54677.2022.9936539.
[30] J. H. Lee et al., “Development of a short-term water quality prediction model for urban rivers using real-time water quality data,” Water Supply, vol. 22, no. 4, pp. 4082–4097, 2022, doi: 10.2166/ws.2022.038.
[31] O. I. Abiodun et al., “Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition,” IEEE Access, vol. 7, pp. 158820–158846, 2019, doi: 10.1109/ACCESS.2019.2945545.
[32] A. K. Kadam, V. M. Wagh, A. A. Muley, B. N. Umrikar, and R. N. Sankhua, “Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India,” Model. Earth Syst. Environ., vol. 5, no. 3, pp. 951–962, 2019, doi: 10.1007/s40808-019-00581-3.
[33] J. S. Xia et al., “Using feed-forward perceptron Artificial Neural Network (ANN) model to determine the rolling force, power and slip of the tandem cold rolling,” ISA Trans., vol. 132, pp. 353–363, 2023, doi: 10.1016/j.isatra.2022.06.009.
[34] N. Noori, L. Kalin, and S. Isik, “Water quality prediction using SWAT-ANN coupled approach,” J. Hydrol., vol. 590, 2020, doi: 10.1016/j.jhydrol.2020.125220.
[35] J. Jagannath et al., “Artificial neural network based automatic modulation classification over a software defined radio testbed,” IEEE Int. Conf. Commun., vol. 2018-May, 2018, doi: 10.1109/ICC.2018.8422346.
[36] A. Ibrahim et al., “Water quality modelling using principal component analysis and artificial neural network,” Mar. Pollut. Bull., vol. 187, no. September 2022, p. 114493, 2023, doi: 10.1016/j.marpolbul.2022.114493.
[37] R. Barzegar, J. Adamowski, and A. A. Moghaddam, “Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in {Aji}-{Chay} {River}, {Iran},” Stoch. Environ. Res. Risk Assess., vol. 30, no. 7, pp. 1797–1819, Oct. 2016, doi: 10.1007/s00477-016-1213-y.
[38] X. Chen, L. Tian, and E. S. Huebner, “Bidirectional Relations Between Subjective Well-Being in School and Prosocial Behavior Among Elementary School-Aged Children: A Longitudinal Study,” Child Youth Care Forum, vol. 49, no. 1, pp. 77–95, 2020, doi: 10.1007/s10566-019-09518-4.
[39] D. Li, T. Nanseki, Y. Chomei, and J. Kuang, “A review of smart agriculture and production practices in Japanese large-scale rice farming,” J. Sci. Food Agric., vol. 103, no. 4, pp. 1609–1620, 2023, doi: 10.1002/jsfa.12204.
[40] S. Song, L. Shen, and M. Valstar, “Human behaviour-based automatic depression analysis using hand-crafted statistics and deep learned spectral features,” Proc. - 13th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2018, pp. 158–165, 2018, doi: 10.1109/FG.2018.00032.
[41] Ü. Ç. Büyükşahin and Ş. Ertekin, “Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition,” Neurocomputing, vol. 361, pp. 151–163, 2019, doi: 10.1016/j.neucom.2019.05.099.
[42] I. and Ugurlu and Oksuz Umut, “Neural network based model comparison for intraday electricity price forecasting,” Energies, vol. 12, no. 23, pp. 1–14, 2019, doi: 10.3390/en12234557.
[43] 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.
[44] N. Heintz, T. Francart, and A. Bertrand, “Minimally Informed Linear Discriminant Analysis: training an LDA model with unlabelled data,” Comput. {Computing} {Technologies} {Agriculture} {V}, vol. 2, no. 1, pp. 1–13, 2023.
[45] Y. ten Hoeve, G. Jansen, and P. Roodbol, “The nursing profession: Public image, self-concept and professional identity. A discussion paper,” J. Adv. Nurs., vol. 70, no. 2, pp. 295–309, 2014, doi: 10.1111/jan.12177.
[46] I. Salisu, M. Bin Mohd Sappri, and M. F. Bin Omar, “The adoption of business intelligence systems in small and medium enterprises in the healthcare sector: A systematic literature review,” Cogent Bus. Manag., vol. 8, no. 1, 2021, doi: 10.1080/23311975.2021.1935663.
[47] W. Mengist, T. Soromessa, and G. Legese, “Method for conducting systematic literature review and meta-analysis for environmental science research,” MethodsX, vol. 7, p. 100777, 2020, doi: 10.1016/j.mex.2019.100777.
[48] S. Kitchenham, B & Charters, “Guidelines for performing systematic literature reviews in software engineering,” Tech. report, Ver. 2.3 EBSE Tech. Report. EBSE, vol. 1, no. January 2007, pp. 1–54, 2007.
[49] T. Dybå and T. Dingsøyr, “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.
[50] A. A. Alqudah and M. Al-emran, “applied sciences Technology Acceptance in Healthcare : A Systematic Review,” Appl. Sci., no. 1, p. 140, 2021.
[51] 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.
[52] B. Kitchenham and S. Charters, “Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3,” Engineering, vol. 45, no. 4ve, p. 1051, 2007, doi: 10.1145/1134285.1134500.
[53] Q. Zhang et al., “Boosting Adversarial Attacks with Nadam Optimizer,” Electron., vol. 12, no. 6, pp. 1–16, 2023, doi: 10.3390/electronics12061464.
[54] Z. Xie, Y. Wang, and S. Sun, “TCNNn Text Classification Model based on nAdam Optimizer,” ACM Int. Conf. Proceeding Ser., pp. 590–597, 2022, doi: 10.1145/3558819.3565149.

[55] P. Kuppusamy, P. Raga Siri, P. Harshitha, M. Dhanyasri, and C. Iwendi, "Customized CNN with Adam and Nadam Optimizers for Emotion Recognition using Facial Expressions," WiSPNET 2023 - Int. Conf. Wirel. Commun. Signal Process. Netw., no. June, pp. 1–5, 2023, doi: 10.1109/WiSPNET57748.2023.10134002.

[56] A. Ezgi and A. Onan, "5 th International Conference on Applied Engineering and Natural Sciences Automatic Knee Osteoarthritis Severity Grading using Deep Neural Networks: Comparative Analysis of Network Architectures and Optimization Functions," pp. 197–203, 2023.

[57] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., vol. 379, no. 2194, 2021, doi: 10.1098/rsta.2020.0209.

[58] L. Zhou, C. Zhao, N. Liu, X. Yao, and Z. Cheng, "Improved LSTM-based deep learning model for COVID-19 prediction using optimized approach," Eng. Appl. Artif. Intell., vol. 122, no. December 2022, p. 106157, 2023, doi: 10.1016/j.engappai.2023.106157.

[59] L. Guo, M. Li, S. Xu, and F. Yang, "Application of {Stochastic} {Gradient} {Descent} {Technique} for {Method} of {Moments}," in 2020 {IEEE} {International} {Conference} on {Computational} {Electromagnetics} ({ICCEM}), Aug. 2020, pp. 97–98. doi: 10.1109/ICCEM47450.2020.9219400.

[60] E. M. Dogo et al., "Optimization Algorithms on Convolutional Neural Networks," 2018 Int. Conf. Comput. Tech. Electron. Mech. Syst., pp. 92–99, 2018.

[61] E. M. Dogo, O. J. Afolabi, and B. Twala, "On the Relative Impact of Optimizers on Convolutional Neural Networks with Varying Depth and Width for Image Classification," Appl. Sci., vol. 12, no. 23, 2022, doi: 10.3390/app122311976.

[62] Z. Zhou, Q. Zhang, G. Lu, H. Wang, W. Zhang, and Y. Yu, "{AdaShift}: {Decorrelation} and {Convergence} of {Adaptive} {Learning} {Rate} {Methods}." arXiv, Jun. 2019. doi: 10.48550/arXiv.1810.00143.