AI

MDPI

Review

Long Short-Term Memory Networks: A Comprehensive Survey

Moez Krichen [1,2,*] and Alaeddine Mihoub [3,*]

1 Department of Software Engineering, Faculty of Computing and Information, Albaha University, P.O. Box 7738, Al Baha 65779, Saudi Arabia
2 ReDCAD Research Laboratory, Sfax University, Sfax 3029, Tunisia
3 Department of Management Information Systems, College of Business and Economics, Qassim University, P.O. Box 6666, Buraydah 51452, Saudi Arabia
* Correspondence: m.krichen@redcad.org (M.K.); a.mihoub@qu.edu.sa (A.M.)

Abstract

Long Short-Term Memory (LSTM) networks have revolutionized the field of deep learning, particularly in applications that require the modeling of sequential data. Originally designed to overcome the limitations of traditional recurrent neural networks (RNNs), LSTMs effectively capture long-range dependencies in sequences, making them suitable for a wide array of tasks. This survey aims to provide a comprehensive overview of LSTM architectures, detailing their unique components, such as cell states and gating mechanisms, which facilitate the retention and modulation of information over time. We delve into the various applications of LSTMs across multiple domains, including the following: natural language processing (NLP), where they are employed for language modeling, machine translation, and sentiment analysis; time series analysis, where they play a critical role in forecasting tasks; and speech recognition, significantly enhancing the accuracy of automated systems. By examining these applications, we illustrate the versatility and robustness of LSTMs in handling complex data types. Additionally, we explore several notable variants and improvements of the standard LSTM architecture, such as Bidirectional LSTMs, which enhance context understanding, and Stacked LSTMs, which increase model capacity. We also discuss the integration of Attention Mechanisms with LSTMs, which have further advanced their performance in various tasks. Despite their strengths, LSTMs face several challenges, including high Computational Complexity, extensive Data Requirements, and difficulties in training, which can hinder their practical implementation. This survey addresses these limitations and provides insights into ongoing research aimed at mitigating these issues. In conclusion, we highlight recent advances in LSTM research and propose potential future directions that could lead to enhanced performance and broader applicability of LSTM networks. This survey serves as a foundational resource for researchers and practitioners seeking to understand the current landscape of LSTM technology and its future trajectory.

Keywords: LSTM; Long Short-Term Memory; deep learning; sequence modeling; natural language processing; time series forecasting; attention mechanisms; applications

check for updates

Academic Editor: Sufi Fahim

Received: 25 June 2025
Revised: 31 August 2025
Accepted: 1 September 2025
Published: 5 September 2025

Citation: Krichen, M.; Mihoub, A. Long Short-Term Memory Networks: A Comprehensive Survey. *AI* **2025**, *6*, 215. https://doi.org/10.3390/ai6090215

1. Introduction

The rapid evolution of deep learning has significantly impacted various domains, particularly those involving sequential data [1–4]. From natural language processing (NLP) to time series forecasting and speech recognition, the ability to model temporal dependencies is crucial for achieving high performance in these tasks. Traditional RNNs have been widely used for sequence modeling [5,6]; however, they are limited by their inability to effectively capture long-range dependencies due to the vanishing gradient problem. This problem arises during the training of RNNs, where gradients diminish exponentially through time, making it difficult for the network to learn from long sequences of data.

In response to these challenges, Long Short-Term Memory (LSTM) networks were introduced by Hochreiter and Schmidhuber in 1997 [7,8]. LSTMs are a specialized kind of RNN designed to remember information for long periods, making them particularly effective for tasks that require the retention of contextual information over time [9]. Their architecture incorporates memory cells and gating mechanisms, which regulate the flow of information, allowing the network to learn which data to keep or discard. This capability has made LSTMs a powerful tool for numerous applications, resulting in significant advancements in various fields.

LSTMs have been successfully applied in natural language processing, where they facilitate tasks such as language modeling, machine translation, and sentiment analysis [10,11]. By maintaining contextual information across sentences, LSTMs improve the quality of generated text and enhance understanding in conversational agents. In time series analysis, LSTMs excel at forecasting future events based on historical data, proving invaluable in industries like finance, healthcare, and meteorology [12,13]. Moreover, in speech recognition systems, LSTMs have improved the accuracy of automatic speech-to-text conversion, enabling more efficient communication and interaction with technology [14,15].

Despite their strengths, LSTMs are not without challenges. Training LSTM networks can be computationally intensive, requiring significant resources and large volumes of labeled data [16,17]. Additionally, issues such as overfitting and model complexity can hinder their performance in real-world applications. Researchers continue to explore various enhancements and adaptations of the LSTM architecture, including Bidirectional LSTMs that provide context from both past and future inputs, and the integration of Attention Mechanisms that allow the model to focus on relevant parts of the input sequence.

It is worth noting that this survey follows a narrative (descriptive) review approach, with the aim of synthesising and thematically organizing the most relevant research on LSTM architectures and applications rather than conducting a protocol-based systematic review. In line with this, the review adopts a broad perspective across multiple domains rather than focusing on a single use-case (e.g., finance or healthcare). The objective is to provide readers with a comprehensive overview that highlights the commonalities and differences of LSTM usage, thereby serving as a general entry point into the field. Hence, the key contributions of this paper are as follows:

1. Providing a clear and comprehensive explanation of the fundamental principles and architecture of LSTM networks, thereby facilitating a solid understanding for researchers and practitioners new to the topic.
2. Presenting a systematic review of LSTM applications and highlighting domains where LSTMs have demonstrated significant effectiveness.
3. Summarizing and comparing the different enhancements and variants of the LSTM architecture reported in the recent literature.
4. Identifying and critically analyzing the challenges and limitations that arise when LSTMs are used in practical implementations.
5. Highlighting recent research trends and outlining future research directions, thus offering guidance for subsequent work in the field.

The structure of this paper is organized as follows: Section 2 presents the fundamentals of LSTM networks, detailing their architecture and mechanisms that enable their unique capabilities. Section 3 explores the diverse applications of LSTMs across different domains, emphasizing their impact and effectiveness. In Section 4, we discuss various enhancements and adaptations of the LSTM framework, including Bidirectional LSTMs and Attention Mechanisms. Section 5 addresses the challenges and limitations associated with LSTM training and deployment, providing insights into Computational Complexity and Data Requirements. Recent advancements and trends in LSTM research are covered in Section 6, showcasing cutting-edge developments and emerging research areas. Finally, Section 7 summarizes the key findings and suggests potential future research directions, aiming to inspire further exploration in the field of LSTM networks. The structure of this paper is shown in Figure 1.

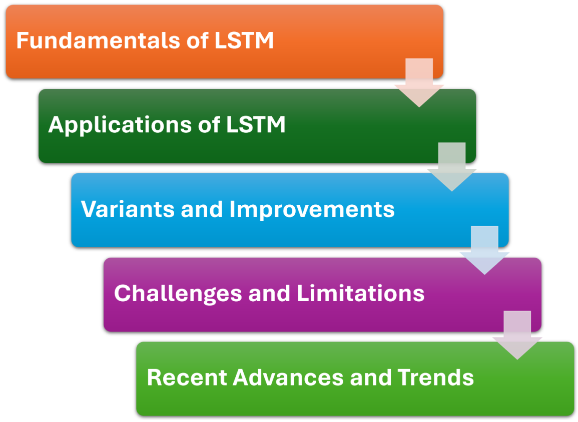

Figure 1. Structure of this paper.

2. Fundamentals of LSTM

LSTM networks are a specialized type of RNNs designed to effectively learn long-term dependencies in sequential data [18]. Traditional RNNs often struggle with the vanishing gradient problem, which limits their ability to capture information from distant elements in a sequence. LSTMs address this challenge through their unique architecture and gating mechanisms. This section provides a comprehensive overview of the LSTM architecture, the mechanisms that facilitate its functionality, and a comparison with traditional RNNs [19].

2.1. LSTM Architecture and Mechanism

The LSTM architecture is a sophisticated structure designed to manage and preserve information over long sequences. It consists of several key components that facilitate its enhanced performance:

- Cell State (C_t): The cell state serves as the memory unit of the LSTM, carrying relevant information throughout the sequence. It is updated at each time step, allowing the LSTM to retain information over long periods. The cell state is crucial for maintaining context, as it can store information from previous time steps without significant degradation. This attribute enables LSTMs to remember essential details over long sequences, making them suitable for applications like language modeling and time series prediction.
- Hidden State (h_t): The hidden state is the output of the LSTM at time step t and is used for making predictions. It encapsulates information about the input sequence thus far and is passed to subsequent LSTM cells. The hidden state can be interpreted as the filtered version of the cell state, representing the relevant information needed for the current prediction. This dynamic nature of the hidden state allows the model to adapt its outputs based on the evolving context of the input sequence.
- Gates: LSTMs utilize three types of gates to control the flow of information, each serving a distinct purpose:

– Input Gate (i_t): This gate determines how much of the new information from the current input should be added to the cell state. The input gate uses a sigmoid activation function to output values between 0 and 1, effectively acting as a filter. The formula for the input gate is

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i). \tag{1}$$

A value obtained by Formula (1) close to one indicates that the information should be fully added to the cell state, whereas a value close to zero implies that little to no information should be added. Here, i_t is the input gate activation, W_i is the weight matrix, h_{t-1} is the hidden state from the previous time step, x_t is the current input, and b_i is the bias term.

– Forget Gate (f_t): The forget gate decides which information from the cell state should be discarded. Similar to the input gate, it employs a sigmoid function:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \tag{2}$$

The forget gate obtained by Formula (2) enables the LSTM to remove irrelevant information, helping to prevent the cell state from becoming cluttered. This mechanism is essential for maintaining the model's ability to focus on relevant patterns over time. In this formula, f_t is the forget gate activation, W_f is the weight matrix, b_f is the bias term, and the other variables are as previously defined.

– Output Gate (o_t): This gate controls what part of the cell state will be output as the hidden state. It dictates the information passed to the next LSTM cell:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o). \tag{3}$$

The output gate calculated by applying Formula (3) ensures that the hidden state reflects only the most pertinent information from the cell state, which is crucial for making accurate predictions at each time step. Here, o_t is the output gate activation, W_o is the weight matrix, b_o is the bias term, and the other variables are as defined earlier.

The updates to the cell state and hidden state are governed by the following equations:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C). \tag{4}$$

In Formula (4), $\tilde{C}_t$ represents the candidate values for the cell state, W_C is the weight matrix, and b_C is the bias term.

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t. \tag{5}$$

In Equation (5), C_t is the updated cell state, C_{t-1} is the previous cell state, and $\otimes$ denotes element-wise multiplication.

$$h_t = o_t \otimes \tanh(C_t). \tag{6}$$

In Equation (6),

- h_t is the updated hidden state.
- σ is the sigmoid activation function, which outputs values in the range (0, 1).
- W_f, W_i, W_C, W_o are weight matrices, and b_f, b_i, b_C, b_o are bias vectors, which are learned during training.

The architecture of the LSTM cell is visualized in Figure 2, which illustrates the connectivity between the gates, cell state, and hidden state. To complement this, Figure 3 presents a schematic of three successive LSTM cells across time steps ($t-1$, t, $t+1$), highlighting how the input x, hidden state h, and cell state c are propagated through the sequence.

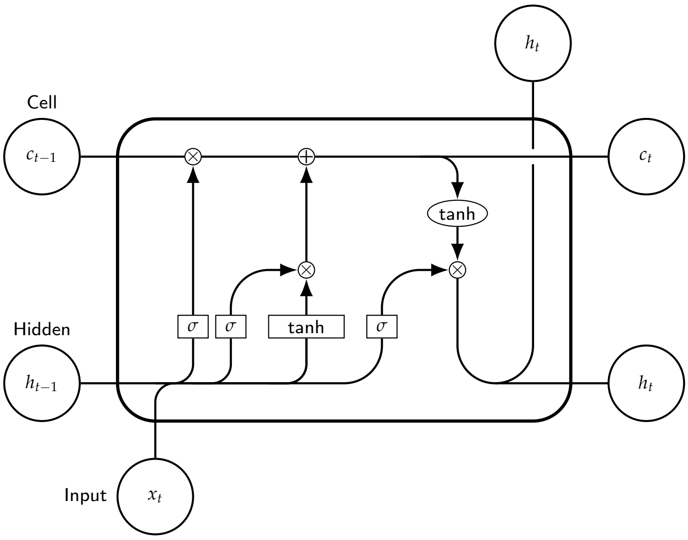

Figure 2. The architecture of the LSTM cell.

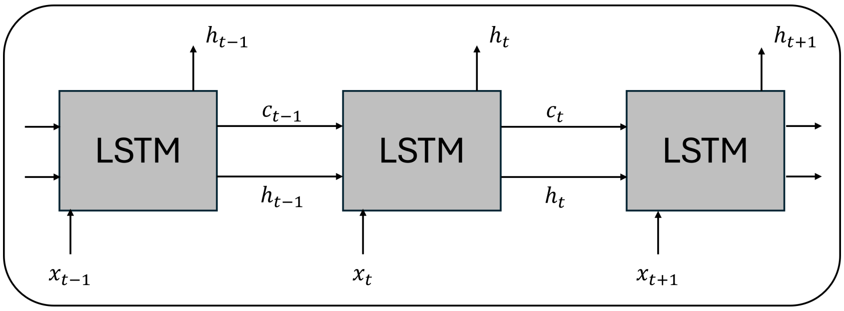

Figure 3. General schematic of an LSTM network with three successive cells ($t-1$, t, $t+1$), showing input x, hidden state h, and cell state c.

To further clarify how long-term information is preserved, we provide a more detailed explanation of the LSTM gating mechanisms, as each gate plays a crucial role in controlling the flow of information. The forget gate evaluates the previous cell state and the current input to decide what information to discard. This selective forgetting is essential to prevent the cell state from becoming cluttered with irrelevant data, which helps maintain the integrity of the learning process. The output of the forget gate can be interpreted as a mask that determines which parts of the cell state should be preserved. The input gate assesses the new input and the previous hidden state to determine how much new information to add to the cell state. This gate allows the model to dynamically incorporate relevant information while ignoring noise, thereby improving learning efficiency. New candidate values are created using a hyperbolic tangent function, which helps in scaling the new information and ensures that the LSTM can adapt to new information in a controlled manner. Finally, the output gate filters the cell state to generate the hidden state, which is used for predictions and passed to the next time step. This ensures that only the relevant information is exposed to the output, allowing the model to produce contextually relevant outputs based on the entire sequence.

2.2. Comparison with Traditional RNNs

To illustrate the advantages of LSTMs over traditional RNNs, we summarize the key differences in Table 1.

Table 1. Comparison of RNNs and LSTMs.

Feature	RNNs	LSTMs
Memory Retention	Short-term	Long-term
Vanishing Gradient Problem	Severe	Mitigated
Complexity	Simpler	More Complex
Gates	None	Three (Input, Forget, Output)
Learning Capability	Limited	Enhanced

The table highlights that while LSTMs are more complex, they are significantly better at retaining information over longer sequences compared to traditional RNNs. This capability makes LSTMs particularly suitable for tasks such as natural language processing, where context and meaning can span long distances in text. Furthermore, the enhanced learning capability of LSTMs is attributed to their ability to adaptively learn which information is relevant over time, whereas traditional RNNs often struggle to maintain this context. In practical applications, this translates to improved performance in tasks involving long sequences, such as language translation and speech recognition. The architecture and mechanisms of LSTMs enable them to effectively manage long-term dependencies, making them a robust choice for a variety of sequence modeling tasks. The unique properties of LSTMs have led to their adoption across various fields, demonstrating their versatility and effectiveness in handling complex sequential data. Although LSTM is a relatively mature architecture, it is still widely used in both applied and comparative studies due to its proven effectiveness and the availability of well-established implementations, which make it a strong and reliable baseline for evaluating more recent sequence modeling techniques.

2.3. Feature Selection and Hyperparameter Optimization

Feature selection (FS) and hyperparameter optimization (HPO) are central to achieving reliable LSTM performance, though their relative impact depends on data characteristics and task constraints [20,21]. Filter-style FS (e.g., mutual-information criteria) helps reduce redundancy and noise in high-dimensional sequences and has been shown to improve downstream modeling in time series settings [22,23]. Beyond filtering, representation-learning approaches such as autoencoders can provide compact, informative inputs for sequence models, effectively acting as data-driven feature selectors [24]. Attention Mechanisms likewise reweight informative temporal signals and have demonstrated benefits in multivariate time series forecasting [25].

On the optimization side, modern HPO encompasses Bayesian methods and evolutionary/population-based search, with recent studies documenting consistent gains over manual tuning and ad hoc heuristics [20,26]. Practical workflows often combine surrogate-based search with resource-aware scheduling to explore architectures and training settings efficiently (e.g., depth/width, look-back window, learning rate, dropout, batch size) [20]. Crucially, evaluation protocols must respect temporal order to avoid information leakage; recent analyses emphasize blocked/rolling validation and warn against random splits for time series [27,28]. In applied LSTM studies, pairing principled FS with systematic HPO has yielded measurable accuracy and robustness gains in forecasting tasks, underscoring the value of treating FS and HPO as first-class components of the modeling pipeline [29].

3. Applications of LSTM

LSTM networks have gained prominence due to their effectiveness in handling sequences and time-dependent data. This section explores various applications of LSTMs across different domains, highlighting their versatility and impact.

3.1. Natural Language Processing

LSTMs play a crucial role in Natural Language Processing (NLP), where understanding context and maintaining long-term dependencies are essential [30]. Key applications include the following:

- Machine Translation [31]: LSTMs are employed in translating text from one language to another. For instance, Google Translate utilizes LSTMs to process sequences of words effectively, improving translation quality across diverse languages. The model captures contextual relationships, allowing it to handle idiomatic expressions and complex grammatical structures, leading to translations that are not only accurate but also contextually relevant.
- Sentiment Analysis [32–34]: LSTMs analyze sentiments in text data, such as product reviews or social media posts, and are frequently evaluated on public benchmark datasets, such as the IMDB movie review corpus. Consider a scenario where a company analyzes Twitter data to gauge customer sentiment about a new product launch. By using an LSTM model, the company can accurately classify tweets as positive, negative, or neutral, enabling it to respond promptly to customer feedback and adjust marketing strategies accordingly.
- Text Generation [35,36]: In applications such as creative writing and chatbots, LSTMs are used to generate coherent and contextually relevant text. For example, early text generation systems relied on Stacked LSTM architectures to model long-range dependencies in language. More recent models, such as OpenAI's GPT-2, are based on the Transformer architecture but still highlight the importance of sequence learning for generating human-like text from a given prompt. Businesses can leverage these capabilities to create automated customer support chatbots that provide accurate and contextually appropriate responses to user inquiries.
- Named Entity Recognition [37–40]: LSTMs help identify and classify entities in text, such as names of people, organizations, and locations. For example, in healthcare, extracting patient information from clinical notes is crucial. An LSTM-based model can accurately identify and extract entities like drug names, dosages, and medical conditions, facilitating better data management and research.

3.2. Time Series Analysis

LSTMs are particularly effective in analyzing time series data due to their ability to capture temporal dependencies [41]. Applications include the following:

- Financial Forecasting [42,43]: LSTMs are widely used in predicting stock prices and market trends by analyzing historical data. For instance, hedge funds employ LSTM models to forecast stock prices based on historical trading data, economic indicators, and news sentiment. By learning patterns from past data, LSTMs help traders make informed decisions about buying or selling stocks at optimal times.
- Weather Forecasting [44–46]: Meteorological models utilize LSTMs to predict weather conditions by analyzing historical weather data. An example is using LSTMs to forecast temperature and precipitation based on past climatic data, enabling more accurate predictions. This capability is crucial for agriculture, where farmers can better plan planting and harvesting times based on expected weather patterns.

- Anomaly Detection [47,48]: In fields such as manufacturing and finance, LSTMs can identify anomalies in time series data, such as fraudulent transactions or equipment failures. For example, a bank might use LSTMs to monitor transaction patterns and detect unusual activities that could indicate fraud. The system can automatically flag these transactions for further investigation, enhancing security and reducing losses.

3.3. Speech Recognition

LSTMs have significantly improved the accuracy of automatic speech recognition (ASR) systems. Their applications include the following:

- Voice Assistants [49]: Technologies like Siri and Google Assistant use LSTMs to understand and process spoken commands. For instance, when a user asks their voice assistant to set a reminder, the LSTM model processes the spoken input, maintaining context across the conversation to provide accurate responses. This enhances user experience by enabling more natural interactions.
- Transcription Services [50]: LSTMs facilitate the conversion of spoken language into written text. Companies like Otter.ai utilize LSTMs to provide real-time transcription services for meetings and lectures; in this field, LibriSpeech is one of the most commonly used benchmark datasets for training and evaluation. The model can handle various accents and speech patterns, ensuring that transcriptions are accurate and contextually relevant, which is invaluable for professionals needing precise records of discussions.
- Emotion Recognition [51,52]: By analyzing speech patterns and intonations, LSTMs can identify emotional states, which is useful in applications such as mental health monitoring and customer service. Understanding emotions can lead to improved user experiences and interactions, as customer service representatives can tailor their responses based on the detected emotional tone.

3.4. Other Emerging Applications

Beyond traditional applications, LSTMs are being explored in various emerging fields:

- Healthcare [53–56]: LSTMs analyze patient data over time, predicting health outcomes and personalizing treatment plans. For instance, hospitals can use LSTM models to track vital signs of patients in real-time, predicting potential health deteriorations. This enables timely interventions that can significantly improve patient outcomes, such as early detection of sepsis.
- Robotics [57–59]: In robotics, LSTMs help in path planning and decision-making by predicting future states based on past movements and sensory inputs. Consider autonomous vehicles, which use LSTMs to analyze traffic patterns and make real-time navigation decisions. This enhances the vehicle's ability to operate safely and efficiently in dynamic environments.
- Video Analysis [60–63]: LSTMs are used in analyzing video sequences for applications such as action recognition and event prediction. For example, security systems can utilize LSTMs to detect unusual activities in surveillance footage. By analyzing sequences of frames, the system can identify potential security threats, enhancing safety in public spaces.

Table 2 summarizes the various applications of LSTMs across different domains, highlighting their key features and benefits. To sum up, LSTMs have shown strong performance in a variety of application domains. For example, in speech processing and traffic prediction, their ability to capture temporal dependencies has yielded better results than traditional machine learning models. In financial forecasting, LSTMs demonstrated robustness to noisy and highly dynamic sequential data. However, several limitations have also been reported in the literature. In IoT scenarios, the high training and inference costs of LSTMs often make them unsuitable for deployment on resource-constrained devices. In healthcare, long and irregular time series sequences frequently cause performance degradation. Moreover, the limited interpretability of LSTM models has been identified as a critical issue in safety-critical domains, such as medical diagnosis. These observations suggest that further improvements are required, particularly to enhance model efficiency, increase robustness to long and irregular sequences, and integrate interpretability methods or hybrid architectures (e.g., attention-based extensions) into LSTM models.

Table 2. Applications of LSTM networks across different domains and their key benefits.

Application Domain	Key Tasks / Use-Cases	Benefits
Natural Language Processing	Machine Translation, Sentiment Analysis, Text Generation, Named Entity Recognition	Captures long-term dependencies in text and improves contextual understanding and generation quality
Time Series Analysis	Financial Forecasting, Weather Forecasting, Anomaly Detection	Learns temporal patterns for better forecasting and detects unusual behaviour in sequential data
Speech Processing	Voice Assistants, Speech-to-Text Transcription, Speech Emotion Recognition	Maintains context across spoken input and improves sequence-to-sequence prediction accuracy
Healthcare	Patient Monitoring, Outcome Prediction	Enables real-time prediction from evolving physiological signals and supports early intervention
Robotics	Path Planning, Navigation, Decision-Making	Predicts future states from sensor data and improves autonomous decision-making
Video Analysis	Action Recognition, Event Prediction	Models temporal dynamics in visual sequences to recognise complex activities

4. Variants and Improvements

While LSTM networks have proven to be highly effective for sequence modeling, several variants and improvements have emerged to enhance their performance in various applications. This section discusses three significant advancements: Bidirectional LSTMs, Stacked LSTMs, and Attention Mechanisms. It is worth noting that, unlike the application section above, this section provides a systematic overview of three main LSTM variants independently of their application domains, in order to clearly distinguish architectural differences from use-case specific deployments.

4.1. Bidirectional LSTMs

Bidirectional LSTMs (BiLSTMs) are an extension of the standard LSTM architecture that allow the network to learn from both past and future contexts in a sequence [64–67]. In a traditional LSTM, information flows in one direction, typically from past to future. However, BiLSTMs process the input sequence in both forward and backward directions:

- Architecture: A BiLSTM consists of two LSTM layers for each time step: one processes the sequence from the beginning to the end (forward) and the other processes it from the end to the beginning (backward). The outputs of both layers are concatenated or combined, providing a richer representation that captures context from both sides. This architecture is particularly useful in scenarios where the context surrounding a word is critical for understanding its meaning.
- Example: In a sentence like "The bank of the river," understanding the word "bank" requires context from both directions to determine whether it refers to a financial institution or the side of a river. A BiLSTM uses the forward layer to gather context from preceding words and the backward layer to gather context from subsequent words, enhancing its ability to disambiguate meaning effectively.
- Applications: This approach is particularly useful in tasks such as named entity recognition, where the model needs to identify proper nouns in context, and machine translation, where syntax and semantics from both directions are critical for accurate translations. For instance, BiLSTMs have been shown to improve the performance of systems like Google Translate, which need to understand entire phrases rather than isolated words.
- Benefits: By leveraging information from both past and future states, BiLSTMs improve the model's ability to make accurate predictions. This dual-context learning often results in better performance on tasks that require a comprehensive understanding of input sequences, leading to more reliable outputs in applications like sentiment analysis and text generation.

Paper [64] presents a bidirectional Long Short-Term Memory (BiDLSTM)-based intrusion detection system that significantly enhances detection accuracy while reducing false alarm rates for various network attacks, particularly User-to-Root (U2R) and Remote-to-Local (R2L) threats, when compared to conventional models. This model effectively learns long-term visual–language interactions, achieving competitive performance in caption generation and superior results in image–sentence retrieval through innovative hierarchical embeddings and data augmentation techniques. Moreover, article [65] develops and evaluates both unidirectional and Bidirectional LSTM models for short-term traffic prediction. The findings reveal that the Bidirectional LSTM outperforms its unidirectional counterpart, achieving over 90% accuracy in speed predictions for up to 60 min, as well as high accuracy in traffic flow predictions across various horizons, thereby providing reliable AI tools for road operators. In addition, the study by Huang et al. [66] introduces a novel prognostic method that utilizes Bidirectional LSTM networks to effectively integrate sensor and operational condition data for predicting the remaining useful life (RUL) of engineered systems. This approach demonstrates superior performance in a case study that focused on aircraft turbofan engines compared to existing AI-based methods. Finally, the article by Mahadevaswamy et al. [67] summarizes the application of a Bidirectional LSTM network for sentiment analysis, successfully classifying Amazon product reviews into positive and negative categories. This achievement is attributed to the network's capability to manage long-term dependencies, based on a comprehensive dataset of 104,975 reviews of mobile electronics.

4.2. Stacked LSTMs

Stacked LSTMs refer to the architecture where multiple LSTM layers are stacked on top of each other [68–71]. This deep architecture allows the model to learn more complex representations of the data:

- Architecture: In a Stacked LSTM, the output of one LSTM layer serves as the input to the next layer. This hierarchical structure enables the model to capture high-level features at deeper layers while retaining the temporal dynamics in the lower layers. Typically, the first few layers might focus on capturing simple patterns, while deeper layers can learn more complex relationships in the data.

- Example: In speech recognition, lower layers of a Stacked LSTM might focus on phonetic features, such as phonemes and intonation, while higher layers can capture linguistic structures and context. For instance, a model might first learn to recognize the sounds of words and then understand them in the context of sentences, leading to more accurate transcriptions.
- Applications: Stacked LSTMs are effective in tasks requiring intricate feature extraction, such as video analysis, natural language processing, and handwriting recognition. For example, in video analysis, lower layers might extract motion patterns, while higher layers could recognize actions like running or jumping. This capability allows for sophisticated applications, such as action recognition in sports or surveillance footage analysis.
- Benefits: The increased capacity of Stacked LSTMs allows them to model more complex relationships within the data, leading to improved performance on challenging tasks. However, this complexity requires careful tuning of hyperparameters to avoid overfitting, particularly when working with smaller datasets.

The study by Ma et al. [68] develops a Stacked LSTM neural network model designed to predict daily quality control records for radiotherapy machines with remarkable accuracy. This model not only outperforms the ARIMA model but also demonstrates robustness in forecasting potential machine failures, thereby facilitating preventive maintenance. In a comprehensive review, Ghimire et al. [69] present modeling techniques and introduce a hybrid SAELSTM framework that combines deep learning with Manta Ray Foraging Optimization. This innovative framework constructs accurate prediction intervals for daily Global Solar Radiation, showcasing superior performance compared to conventional machine learning models and significantly enhancing solar energy monitoring systems. Wang et al. [70] introduce a stacked residual LSTM model aimed at predicting sentiment intensity. By incorporating residual connections, this model effectively addresses the challenges associated with training deeper networks, resulting in improved optimization and prediction accuracy compared to traditional lexicon-based, regression, and neural network methods. Finally, Yu et al. [71] introduce a novel hierarchical fault diagnosis algorithm based on Stacked LSTM networks for rolling bearings, achieving an impressive accuracy of over 99%. This algorithm automatically extracts features from raw temporal signals without preprocessing, surpassing traditional methods and state-of-the-art intelligent fault diagnosis techniques.

4.3. Attention Mechanisms

Attention Mechanisms have revolutionized the way sequence-to-sequence models operate, particularly in conjunction with LSTMs [72–77]. These mechanisms allow the model to focus on specific parts of the input sequence when generating each part of the output sequence. These variants and improvements—Bidirectional LSTMs, Stacked LSTMs, and Attention Mechanisms—enhance the capabilities of traditional LSTM networks, making them more effective for a wider range of applications. By leveraging these advancements, researchers and practitioners can achieve better performance and more accurate results in complex sequence modeling tasks:

- Concept: Instead of treating all input tokens equally, Attention Mechanisms weigh the importance of different tokens for each output token. This selective focus enables the model to capture relevant context more effectively, which is particularly beneficial in long sequences where not all parts of the input are equally important for every output.

- Example: In machine translation, when translating a long sentence, the model can focus on the relevant words that contribute to the meaning of the current output word. For instance, in translating "The cat sat on the mat," the model might focus on "cat" when generating the word "it," improving the coherence of the translation. This capability helps in generating more fluent and accurate translations, particularly in complex sentences.
- Applications: Attention Mechanisms are widely used in tasks such as machine translation, text summarization, image captioning, and even video analysis. For example, in text summarization, attention helps the model identify key sentences in a document to create concise summaries that encapsulate the main ideas without losing important context.
- Benefits: The integration of attention with LSTMs leads to significant improvements in performance, particularly for long sequences. It reduces the burden on the LSTM to remember all previous states, allowing it to concentrate on relevant information dynamically. This has led to the development of models like the Transformer, which relies heavily on Attention Mechanisms to achieve state-of-the-art results in various NLP tasks.

The study by Sang et al. [72] introduces the Attention Mechanism Variant LSTM (AMV-LSTM), which significantly improves stock price prediction accuracy. This model enhances traditional LSTM frameworks by coupling the forget and input gates, adding a simple forget gate to the long-term state, and incorporating an Attention Mechanism. As a result, it effectively addresses stability and overfitting challenges commonly associated with conventional LSTM models. In a related effort, Lin et al. [73] present a dual-stage attention-based LSTM network tailored for short-term zonal electricity load probabilistic forecasting. This innovative approach enhances feature selection through a feature attention encoder and captures temporal dependencies via a decoder, yielding improved accuracy and generalization over existing models, particularly on the GEFCom2014 dataset. Moreover, Xiang et al. [74] propose a novel fault detection method for wind turbines that integrates a convolutional neural network (CNN) with an attention-enhanced Long Short-Term Memory network (LSTM). By utilizing SCADA data, this method effectively extracts dynamic changes, improves learning accuracy, and provides early warnings for anomalies, thereby enabling the prediction of early failures in wind turbine components. In the realm of environmental monitoring, Yang et al. [75] introduce a water quality prediction model known as CNN-LSTM with Attention (CLA). This model forecasts pH and NH3-N levels in the Beilun Estuary by employing linear interpolation for missing data and wavelet techniques for denoising. The hybrid CNN-LSTM architecture effectively addresses nonlinear time series prediction, with the Attention Mechanism enhancing long-term dependencies. Experimental findings demonstrate that CLA outperforms existing models and maintains stable predictions across various time lags. Furthermore, Ran et al. [76] propose an LSTM method augmented with an Attention Mechanism for travel time prediction. This model addresses the limitations of traditional LSTMs by incorporating departure time and modeling long-term dependencies through a tree structure. Utilizing datasets from Highways England, the experimental results indicate that the proposed model achieves superior accuracy compared to standard LSTM and other baseline methods, effectively leveraging the Attention Mechanism to enhance predictions. Finally, Liu et al. [77] introduce the attention-based bidirectional Long Short-Term Memory with a convolution layer (AC-BiLSTM) architecture. By combining a convolutional layer with BiLSTM and an Attention Mechanism, this model enhances text classification tasks in natural language processing. It excels at extracting higher-level phrase representations and capturing both local and global semantic features, demonstrating superior classification accuracy across six sentiment classification datasets and a question classification dataset, thereby outperforming existing state-of-the-art methods.

In Table 3, the various variants and improvements of LSTM networks are summarized, highlighting their key features, applications, and benefits.

Table 3. Variants and improvements of LSTM networks.

Variant	Key Features	Applications	Benefits
Bidirectional LSTMs	Processes sequences in both directions	Named Entity Recognition, Machine Translation	Improved context understanding, better disambiguation
Stacked LSTMs	Multiple LSTM layers for deeper learning	Speech Recognition, Video Analysis	Enhanced feature extraction, ability to model complex patterns
Attention Mechanisms	Focuses on specific input parts for each output	Machine Translation, Text Summarization	Better handling of long sequences, dynamic context focus

5. Challenges and Limitations

Despite the effectiveness of LSTM networks in various applications, they are not without their challenges and limitations. This section discusses three major issues: Computational Complexity, Data Requirements, and Training Difficulties.

5.1. Computational Complexity

LSTM networks are inherently complex due to their architecture, which includes multiple gates and memory cells. This complexity leads to several challenges:

- High Resource Consumption: LSTMs require significant computational resources, especially for large datasets and deep architectures. Training an LSTM model can demand substantial CPU or GPU time, making it less feasible for real-time applications. For instance, training a state-of-the-art LSTM for natural language processing tasks may require specialized hardware, like TPUs, which can be costly.
- Scalability Issues: As the size of the network increases, particularly when stacking multiple LSTM layers, scalability becomes a concern. The time taken to train and the memory required can grow exponentially, limiting the model's applicability in resource-constrained environments. This limitation can hinder the deployment of LSTM-based models in mobile or embedded systems where computational power is limited.
- Inference Speed: The complexity of LSTMs can also affect inference speed, which is critical in applications like real-time speech recognition or chatbots. Slower inference can hinder user experience and system performance, leading to delays that may frustrate users. For example, in a customer support chatbot, any lag in response time can diminish the perceived quality of the service.

5.2. Data Requirements

LSTMs require substantial amounts of high-quality data for effective training. This necessity poses several challenges:

- Large Datasets: LSTMs typically perform well when trained on large datasets, which may not always be available. In domains with limited data, such as specialized medical applications or niche markets, LSTMs can overfit, leading to poor generalization on

unseen data. This issue can result in models that perform well on training data but fail to deliver accurate predictions in real-world scenarios.
- Data Quality: The quality of the training data significantly impacts model performance. Noisy, unstructured, or biased data can result in suboptimal learning outcomes. For example, in natural language processing, poorly annotated datasets can lead to models that misunderstand context or semantics, thereby generating irrelevant or incorrect outputs.
- Domain-Specific Data: Training LSTMs for specialized tasks often requires domain-specific data, which can be difficult and time-consuming to collect. In fields like healthcare, obtaining annotated datasets can involve extensive collaboration with experts and adherence to regulatory standards, complicating the data collection process and extending project timelines.

Nevertheless, recent studies have shown that synthetic data can effectively compensate for limited real-world datasets in LSTM applications [78,79]. For example, GAN-generated time series [80–82] have been used to augment LSTM training in condition monitoring and predictive maintenance, and synthetic speech features have been employed to pre-train LSTM-based speaker recognition models. Similar approaches have also been applied in climate modeling by combining ConvLSTM [83,84] with synthetic sequence data to improve prediction accuracy in data-sparse regions. These examples demonstrate that LSTMs can still be applied in data-scarce scenarios when appropriate data augmentation techniques are used.

5.3. Training Difficulties

Training LSTM networks presents several challenges that can affect their effectiveness:

- Long Training Times: Due to their complexity, LSTMs can take a long time to train, especially with large datasets. This extended training period can be a barrier to rapid prototyping and experimentation. Researchers may find it difficult to iterate quickly on model designs, slowing down the overall development process and delaying the deployment of solutions.
- Hyperparameter Tuning: LSTMs have numerous hyperparameters (e.g., learning rate, number of layers, hidden units) that need to be carefully tuned for optimal performance. Finding the right combination can be challenging and often requires extensive experimentation. This tuning process can be resource-intensive, requiring multiple training runs and potentially consuming significant computational resources.
- Vanishing and Exploding Gradients: Although LSTMs are designed to mitigate the vanishing gradient problem, they are still susceptible to exploding gradients during training. This issue can lead to unstable learning and hinder the convergence of the model. If not properly managed, exploding gradients can cause the model to diverge, resulting in NaN values during training and necessitating the need for gradient clipping or other stabilizing techniques.

Table 4 summarizes the various challenges and limitations associated with LSTM networks, highlighting their implications on performance and applicability. While LSTM networks offer significant advantages in sequence modeling, they also face substantial challenges that can limit their effectiveness and applicability. Understanding these limitations is crucial for researchers and practitioners aiming to implement LSTMs successfully in various domains.

Table 4. Challenges and limitations of LSTM networks.

Challenge	Description	Implications
Computational Complexity	High resource consumption and scalability issues	Slower training and inference, limited applicability in real-time systems
Data Requirements	Need for large, high-quality datasets	Risk of overfitting, poor generalization, and limited domain applicability
Training Difficulties	Long training times and hyperparameter tuning challenges	Increased development time and potential instability in learning

6. Recent Advances and Trends

As the field of machine learning continues to evolve, recent advances and trends in LSTM networks reflect ongoing innovations in architectures, performance evaluations, and emerging research areas. This section explores these developments in detail.

6.1. Innovations in LSTM Architectures

Innovations in LSTM architectures have significantly enhanced their capabilities and performance in various applications. Researchers have introduced several modifications to the traditional LSTM design, aiming to address specific limitations and improve training efficiency. One notable innovation is the development of peephole connections, which allow LSTMs to access the cell state directly from the gates. This modification enables the model to make more informed decisions based on the current memory content, leading to improved learning dynamics. Another advancement is the Grid LSTM, which extends the LSTM architecture to handle multi-dimensional sequences. This architecture is particularly useful in applications like video processing, where spatial and temporal information must be captured simultaneously. By organizing the LSTM cells in a grid structure, the Grid LSTM can effectively learn complex relationships in high-dimensional data. Additionally, layer normalization techniques have been integrated into LSTM architectures to stabilize training and improve convergence speed. This approach normalizes the inputs across features rather than across the batch, making it particularly beneficial for recurrent networks where batch sizes can vary.

6.2. Performance Comparisons

As various LSTM architectures and variants have emerged, numerous studies have conducted performance comparisons to evaluate their effectiveness against traditional models and other sequence-based architectures [85–87]. Recent benchmarks indicate that LSTMs often outperform standard RNNs in tasks involving long-range dependencies [88–90]. However, newer architectures, such as the Transformer, have shown superior performance in many natural language processing tasks, prompting researchers to explore hybrid models that combine the strengths of LSTMs and Attention Mechanisms [91–93]. For instance, studies comparing LSTMs to Transformers in machine translation tasks have revealed that while LSTMs can handle sequence data effectively, Transformers often achieve better performance in terms of accuracy and training efficiency, mainly because Transformers process entire sequences simultaneously rather than sequentially [94–96]. Nevertheless, LSTMs remain valuable in several scenarios: in data-scarce or resource-limited settings, they often demonstrate better convergence and generalization properties due to their sequential inductive biasing, and in real-time or embedded applications (e.g., healthcare monitoring and industrial IoT), LSTM-based models have been shown to offer competitive performance and lower computational overhead on wearable sensor tasks [97–99]. Moreover, comparisons have highlighted that integrating Attention Mechanisms with LSTMs can enhance their performance significantly, particularly in tasks requiring contextual understanding over long sequences [100,101]. These hybrid models leverage the temporal strengths of LSTMs while benefiting from the contextual awareness provided by Attention Mechanisms [102,103].

6.3. Emerging Research Areas

Emerging research areas related to LSTMs reflect the ongoing exploration of their capabilities in various fields. One significant area is transfer learning, where pre-trained LSTM models are adapted to new tasks with limited data [104]. This approach has gained traction as it allows models to leverage prior knowledge, reducing the need for extensive training on datasets. Another promising area is the application of LSTMs in reinforcement learning. Researchers are investigating how LSTMs can be used to model environments with temporal dependencies, enabling agents to make better decisions based on past experiences. This integration has the potential to enhance the performance of reinforcement learning algorithms in complex and dynamic environments. Recent studies demonstrate this potential: for instance, LSTM networks have been incorporated into a TD3 reinforcement learning framework to improve robotic path planning in dynamic environments [105], while GT-LSTM integrates geospatial and temporal dependencies to enhance agent-based modeling in multimodal contexts [106]. Additionally, the adaptation of LSTMs for multimodal learning is gaining attention [107]. This involves combining LSTM networks with other data types, such as images and audio, to create models that can process and understand information across different modalities. Applications in this area include video analysis, where LSTMs can help analyze both visual and auditory components to interpret scenes more effectively. Finally, advancements in explainable AI (XAI) are prompting researchers to focus on making LSTM models more interpretable [108]. Understanding how LSTMs make decisions is crucial for their deployment in sensitive applications like healthcare and finance, where transparency is essential.

7. Conclusions

In conclusion, Long Short-Term Memory (LSTM) networks have established themselves as a cornerstone in the field of sequence modeling, offering powerful solutions for tasks involving temporal dependencies. While the most significant challenges currently limiting their broader adoption remain (i) the high computational cost and long training times, (ii) the tendency to overfit on limited data, and (iii) the lack of interpretability compared to more recent architectures, LSTMs nevertheless remain highly competitive in specific scenarios. These include applications where strict temporal ordering is essential (e.g., real-time sensor streams and speech processing), situations where training data is scarce, and deployments that require lightweight models for edge devices.

Recent innovations in LSTM architectures, such as peephole connections and Grid LSTMs, have enhanced their capabilities and broadened their applicability to more complex data. In parallel, performance comparisons with Transformer-based models highlight the evolving landscape of deep learning, where hybrid approaches that combine the strengths of LSTMs and Attention Mechanisms are gaining traction.

In terms of future research, emerging areas such as transfer learning, reinforcement learning, and multimodal learning indicate promising directions and demonstrate the adaptability of LSTM networks across various domains—from natural language processing to video analysis. Continued efforts to reduce computational overhead, improve generalization in low-data environments, and increase model interpretability will be central for advancing LSTM technology and maximizing its practical impact.

Author Contributions: Conceptualization, M.K.; formal analysis, M.K.; methodology, M.K.; validation, A.M.; writing—original draft, M.K.; writing—review and editing, M.K. and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Janiesch, C.; Zschech, P.; Heinrich, K. Machine learning and deep learning. *Electron. Mark.* **2021**, *31*, 685–695. [CrossRef]
2. Taye, M.M. Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions. *Computers* **2023**, *12*, 91. [CrossRef]
3. Gasmi, K.; Ben Ltaifa, I.; Eltoum Abdalrahman, A.; Hamid, O.; Othman Altaieb, M.; Ali, S.; Ben Ammar, L.; Mrabet, M. Hybrid Feature and Optimized Deep Learning Model Fusion for Detecting Hateful Arabic Content. *IEEE Access* **2025**, *13*, 131411–131431. [CrossRef]
4. Gasmi, K.; Hrizi, O.; Ben Aoun, N.; Alrashdi, I.; Alqazzaz, A.; Hamid, O.; Othman Altaieb, M.; Abdalrahman, A.E.M.; Ben Ammar, L.; Mrabet, M.; et al. Enhanced Multimodal Physiological Signal Analysis for Pain Assessment Using Optimized Ensemble Deep Learning. *Comput. Model. Eng. Sci.* **2025**, *143*, 2459–2489. [CrossRef]
5. Mienye, I.D.; Swart, T.G.; Obaido, G. Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information* **2024**, *15*, 517. [CrossRef]
6. Ullah, I.; Mahmoud, Q.H. Design and development of RNN anomaly detection model for IoT networks. *IEEE Access* **2022**, *10*, 62722–62750. [CrossRef]
7. Van Houdt, G.; Mosquera, C.; Nápoles, G. A review on the long short-term memory model. *Artif. Intell. Rev.* **2020**, *53*, 5929–5955. [CrossRef]
8. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [CrossRef]
9. Halbouni, A.; Gunawan, T.S.; Habaebi, M.H.; Halbouni, M.; Kartiwi, M.; Ahmad, R. Machine learning and deep learning approaches for cybersecurity: A review. *IEEE Access* **2022**, *10*, 19572–19585. [CrossRef]
10. Behera, R.K.; Jena, M.; Rath, S.K.; Misra, S. Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. *Inf. Process. Manag.* **2021**, *58*, 102435. [CrossRef]
11. Edara, D.C.; Vanukuri, L.P.; Sistla, V.; Kolli, V.K.K. Sentiment analysis and text categorization of cancer medical records with LSTM. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 5309–5325. [CrossRef]
12. Lindemann, B.; Maschler, B.; Sahlab, N.; Weyrich, M. A survey on anomaly detection for technical systems using LSTM networks. *Comput. Ind.* **2021**, *131*, 103498. [CrossRef]
13. Hua, Y.; Zhao, Z.; Li, R.; Chen, X.; Liu, Z.; Zhang, H. Deep learning with long short-term memory for time series prediction. *IEEE Commun. Mag.* **2019**, *57*, 114–119. [CrossRef]
14. Oruh, J.; Viriri, S.; Adegun, A. Long short-term memory recurrent neural network for automatic speech recognition. *IEEE Access* **2022**, *10*, 30069–30079. [CrossRef]
15. Agarwal, P.; Kumar, S. Electroencephalography-based imagined speech recognition using deep long short-term memory network. *ETRI J.* **2022**, *44*, 672–685. [CrossRef]
16. Fang, K.; Kifer, D.; Lawson, K.; Shen, C. Evaluating the potential and challenges of an uncertainty quantification method for long short-term memory models for soil moisture predictions. *Water Resour. Res.* **2020**, *56*, e2020WR028095. [CrossRef]
17. Zhang, J.; Wang, P.; Yan, R.; Gao, R.X. Long short-term memory for machine remaining life prediction. *J. Manuf. Syst.* **2018**, *48*, 78–86. [CrossRef]
18. Fernando, T.; Denman, S.; McFadyen, A.; Sridharan, S.; Fookes, C. Tree memory networks for modelling long-term temporal dependencies. *Neurocomputing* **2018**, *304*, 64–81. [CrossRef]
19. Muhuri, P.S.; Chatterjee, P.; Yuan, X.; Roy, K.; Esterline, A. Using a long short-term memory recurrent neural network (LSTM-RNN) to classify network attacks. *Information* **2020**, *11*, 243. [CrossRef]
20. Bischl, B.; Binder, M.; Lang, M.; Pielok, T.; Richter, J.; Coors, S.; Thomas, J.; Ullmann, T.; Becker, M.; Boulesteix, A.L.; et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Min. Knowl. Discov.* **2023**, *13*, e1484. [CrossRef]

21. Kazemi, F.; Asgarkhani, N.; Shafighfard, T.; Jankowski, R.; Yoo, D.Y. Machine-Learning Methods for Estimating Performance of Structural Concrete Members Reinforced with Fiber-Reinforced Polymers. *Arch. Comput. Methods Eng.* **2025**, *32*, 571–603. [CrossRef]
22. Huang, L.; Zhou, X.; Shi, L.; Gong, L. Time Series Feature Selection Method Based on Mutual Information. *Appl. Sci.* **2024**, *14*, 1960. [CrossRef]
23. Alalhareth, M.; Hong, S.C. An Improved Mutual Information Feature Selection Technique for Intrusion Detection Systems in the Internet of Medical Things. *Sensors* **2023**, *23*, 4971. [CrossRef]
24. Berahmand, K.; Daneshfar, F.; Salehi, E.S.; Li, Y.; Xu, Y. Autoencoders and their applications in machine learning: A survey. *Artif. Intell. Rev.* **2024**, *57*, 28. [CrossRef]
25. Sakib, S.; Mahadi, M.K.; Abir, S.R.; Moon, A.M.; Shafiullah, A.; Ali, S.; Faisal, F.; Nishat, M.M. Attention-Based Models for Multivariate Time Series Forecasting: Multi-step Solar Irradiation Prediction. *Heliyon* **2024**, *10*, e27795. [CrossRef]
26. Vincent, A.M.; Jidesh, P. An improved hyperparameter optimization framework for AutoML systems using evolutionary algorithms. *Sci. Rep.* **2023**, *13*, 4737. [CrossRef] [PubMed]
27. Yang, X.; Li, J.; Jiang, X. Research on information leakage in time series prediction based on empirical mode decomposition. *Sci. Rep.* **2024**, *14*, 28362. [CrossRef] [PubMed]
28. Liu, S.; Zhou, D.J. Using cross-validation methods to select time series models: Promises and pitfalls. *Br. J. Math. Stat. Psychol.* **2024**, *77*, 337–355. [CrossRef]
29. Dhake, H.; Kashyap, Y.; Kosmopoulos, P. Algorithms for Hyperparameter Tuning of LSTMs for Time Series Forecasting. *Remote Sens.* **2023**, *15*, 2076. [CrossRef]
30. Purba, M.R.; Akter, M.; Ferdows, R.; Ahmed, F. A hybrid convolutional long short-term memory (CNN-LSTM) based natural language processing (NLP) model for sentiment analysis of customer product reviews in Bangla. *J. Discret. Math. Sci. Cryptogr.* **2022**, *25*, 2111–2120. [CrossRef]
31. Su, C.; Huang, H.; Shi, S.; Jian, P.; Shi, X. Neural machine translation with Gumbel Tree-LSTM based encoder. *J. Vis. Commun. Image Represent.* **2020**, *71*, 102811. [CrossRef]
32. Gondhi, N.K.; Chaahat.; Sharma, E.; Alharbi, A.H.; Verma, R.; Shah, M.A. Efficient Long Short-Term Memory-Based Sentiment Analysis of E-Commerce Reviews. *Comput. Intell. Neurosci.* **2022**, *2022*, 3464524. [CrossRef] [PubMed]
33. Al-Smadi, M.; Talafha, B.; Al-Ayyoub, M.; Jararweh, Y. Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2163–2175. [CrossRef]
34. Gandhi, U.D.; Malarvizhi Kumar, P.; Chandra Babu, G.; Karthick, G. Sentiment analysis on twitter data by using convolutional neural network (CNN) and long short term memory (LSTM). *Wirel. Pers. Commun.* **2021**, 1–10. [CrossRef]
35. Perez-Castro, A.; Martínez-Torres, M.d.R.; Toral, S. Efficiency of automatic text generators for online review content generation. *Technol. Forecast. Soc. Change* **2023**, *189*, 122380. [CrossRef]
36. Hajipoor, O.; Nickabadi, A.; Homayounpour, M.M. GPTGAN: Utilizing the GPT language model and GAN to enhance adversarial text generation. *Neurocomputing* **2025**, *617*, 128865. [CrossRef]
37. Ma, P.; Jiang, B.; Lu, Z.; Li, N.; Jiang, Z. Cybersecurity named entity recognition using bidirectional long short-term memory with conditional random fields. *Tsinghua Sci. Technol.* **2020**, *26*, 259–265. [CrossRef]
38. Khalifa, M.; Shaalan, K. Character convolutions for Arabic named entity recognition with long short-term memory networks. *Comput. Speech Lang.* **2019**, *58*, 335–346. [CrossRef]
39. Santoso, J.; Setiawan, E.I.; Purwanto, C.N.; Yuniarno, E.M.; Hariadi, M.; Purnomo, M.H. Named entity recognition for extracting concept in ontology building on Indonesian language using end-to-end bidirectional long short term memory. *Expert Syst. Appl.* **2021**, *176*, 114856. [CrossRef]
40. Lyu, C.; Chen, B.; Ren, Y.; Ji, D. Long short-term memory RNN for biomedical named entity recognition. *BMC Bioinform.* **2017**, *18*, 1–11. [CrossRef]
41. Cheng, Q.; Chen, Y.; Xiao, Y.; Yin, H.; Liu, W. A dual-stage attention-based Bi-LSTM network for multivariate time series prediction. *J. Supercomput.* **2022**, *78*, 16214–16235. [CrossRef]
42. Bukhari, A.H.; Raja, M.A.Z.; Sulaiman, M.; Islam, S.; Shoaib, M.; Kumam, P. Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. *IEEE Access* **2020**, *8*, 71326–71338. [CrossRef]
43. Cao, J.; Li, Z.; Li, J. Financial time series forecasting model based on CEEMDAN and LSTM. *Phys. A Stat. Mech. Its Appl.* **2019**, *519*, 127–139. [CrossRef]
44. Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* **2018**, *148*, 461–468. [CrossRef]
45. Tukymbekov, D.; Saymbetov, A.; Nurgaliyev, M.; Kuttybay, N.; Dosymbetova, G.; Svanbayev, Y. Intelligent autonomous street lighting system based on weather forecast using LSTM. *Energy* **2021**, *231*, 120902. [CrossRef]
46. Hossain, M.S.; Mahmood, H. Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *IEEE Access* **2020**, *8*, 172524–172533. [CrossRef]

47. Nguyen, H.D.; Tran, K.P.; Thomassey, S.; Hamad, M. Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *Int. J. Inf. Manag.* **2021**, *57*, 102282. [CrossRef]
48. Ergen, T.; Kozat, S.S. Unsupervised anomaly detection with LSTM neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 3127–3141. [CrossRef]
49. Wubet, Y.A.; Lian, K.Y. Voice conversion based augmentation and a hybrid CNN-LSTM model for improving speaker-independent keyword recognition on limited datasets. *IEEE Access* **2022**, *10*, 89170–89180. [CrossRef]
50. Orosoo, M.; Raash, N.; Treve, M.; Lahza, H.F.M.; Alshammry, N.; Ramesh, J.V.N.; Rengarajan, M. Transforming English language learning: Advanced speech recognition with MLP-LSTM for personalized education. *Alex. Eng. J.* **2025**, *111*, 21–32. [CrossRef]
51. Zhao, J.; Mao, X.; Chen, L. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomed. Signal Process. Control* **2019**, *47*, 312–323.
52. Du, X.; Ma, C.; Zhang, G.; Li, J.; Lai, Y.K.; Zhao, G.; Deng, X.; Liu, Y.J.; Wang, H. An efficient LSTM network for emotion recognition from multichannel EEG signals. *IEEE Trans. Affect. Comput.* **2020**, *13*, 1528–1540. [CrossRef]
53. Guo, A.; Beheshti, R.; Khan, Y.M.; Langabeer, J.R.; Foraker, R.E. Predicting cardiovascular health trajectories in time-series electronic health records with LSTM models. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 5. [CrossRef] [PubMed]
54. Yin, J.; Han, J.; Xie, R.; Wang, C.; Duan, X.; Rong, Y.; Zeng, X.; Tao, J. MC-LSTM: Real-time 3D human action detection system for intelligent healthcare applications. *IEEE Trans. Biomed. Circuits Syst.* **2021**, *15*, 259–269. [CrossRef]
55. Srivastava, S.; Sharma, S.; Tanwar, P.; Dubey, G.; Memoria, M. Improving Health Care Analytics: LSTM Networks for Enhanced Risk Assessment. *Procedia Comput. Sci.* **2025**, *259*, 11–22. [CrossRef]
56. Rashid, T.A.; Hassan, M.K.; Mohammadi, M.; Fraser, K. Improvement of variant adaptable LSTM trained with metaheuristic algorithms for healthcare analysis. In *Research Anthology on Artificial Intelligence Applications in Security*; IGI Global: Hershey, PA, USA, 2021; pp. 1031–1051.
57. Aslan, S.N.; Özalp, R.; Uçar, A.; Güzeliş, C. New CNN and hybrid CNN-LSTM models for learning object manipulation of humanoid robots from demonstration. *Clust. Comput.* **2022**, *25*, 1575–1590. [CrossRef]
58. Molina-Leal, A.; Gómez-Espinosa, A.; Escobedo Cabello, J.A.; Cuan-Urquizo, E.; Cruz-Ramírez, S.R. Trajectory planning for a mobile robot in a dynamic environment using an LSTM neural network. *Appl. Sci.* **2021**, *11*, 10689. [CrossRef]
59. Park, D.; Hoshi, Y.; Kemp, C.C. A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. *IEEE Robot. Autom. Lett.* **2018**, *3*, 1544–1551. [CrossRef]
60. Bin, Y.; Yang, Y.; Shen, F.; Xie, N.; Shen, H.T.; Li, X. Describing video with attention-based bidirectional LSTM. *IEEE Trans. Cybern.* **2018**, *49*, 2631–2641. [CrossRef]
61. Ge, H.; Yan, Z.; Yu, W.; Sun, L. An attention mechanism based convolutional LSTM network for video action recognition. *Multimed. Tools Appl.* **2019**, *78*, 20533–20556. [CrossRef]
62. Hussain, T.; Muhammad, K.; Ullah, A.; Cao, Z.; Baik, S.W.; De Albuquerque, V.H.C. Cloud-assisted multiview video summarization using CNN and bidirectional LSTM. *IEEE Trans. Ind. Inform.* **2019**, *16*, 77–86. [CrossRef]
63. Ullah, A.; Ahmad, J.; Muhammad, K.; Sajjad, M.; Baik, S.W. Action recognition in video sequences using deep bi-directional LSTM with CNN features. *IEEE Access* **2017**, *6*, 1155–1166. [CrossRef]
64. Imrana, Y.; Xiang, Y.; Ali, L.; Abdul-Rauf, Z. A bidirectional LSTM deep learning approach for intrusion detection. *Expert Syst. Appl.* **2021**, *185*, 115524. [CrossRef]
65. Abduljabbar, R.L.; Dia, H.; Tsai, P.W. Unidirectional and bidirectional LSTM models for short-term traffic prediction. *J. Adv. Transp.* **2021**, *2021*, 5589075. [CrossRef]
66. Huang, C.G.; Huang, H.Z.; Li, Y.F. A bidirectional LSTM prognostics method under multiple operational conditions. *IEEE Trans. Ind. Electron.* **2019**, *66*, 8792–8802. [CrossRef]
67. Mahadevaswamy, U.; Swathi, P. Sentiment analysis using bidirectional LSTM network. *Procedia Comput. Sci.* **2023**, *218*, 45–56. [CrossRef]
68. Ma, M.; Liu, C.; Wei, R.; Liang, B.; Dai, J. Predicting machine's performance record using the stacked long short-term memory (LSTM) neural networks. *J. Appl. Clin. Med. Phys.* **2022**, *23*, e13558. [CrossRef]
69. Ghimire, S.; Deo, R.C.; Wang, H.; Al-Musaylh, M.S.; Casillas-Pérez, D.; Salcedo-Sanz, S. Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: A review and new modeling results. *Energies* **2022**, *15*, 1061. [CrossRef]
70. Wang, J.; Peng, B.; Zhang, X. Using a stacked residual LSTM model for sentiment intensity prediction. *Neurocomputing* **2018**, *322*, 93–101. [CrossRef]
71. Yu, L.; Qu, J.; Gao, F.; Tian, Y. A novel hierarchical algorithm for bearing fault diagnosis based on stacked LSTM. *Shock Vib.* **2019**, *2019*, 2756284. [CrossRef]
72. Sang, S.; Li, L. A novel variant of LSTM stock prediction method incorporating attention mechanism. *Mathematics* **2024**, *12*, 945. [CrossRef]

73. Lin, J.; Ma, J.; Zhu, J.; Cui, Y. Short-term load forecasting based on LSTM networks considering attention mechanism. *Int. J. Electr. Power Energy Syst.* **2022**, *137*, 107818. [CrossRef]
74. Xiang, L.; Wang, P.; Yang, X.; Hu, A.; Su, H. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. *Measurement* **2021**, *175*, 109094. [CrossRef]
75. Yang, Y.; Xiong, Q.; Wu, C.; Zou, Q.; Yu, Y.; Yi, H.; Gao, M. A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. *Environ. Sci. Pollut. Res.* **2021**, *28*, 55129–55139. [CrossRef] [PubMed]
76. Ran, X.; Shan, Z.; Fang, Y.; Lin, C. An LSTM-based method with attention mechanism for travel time prediction. *Sensors* **2019**, *19*, 861. [CrossRef] [PubMed]
77. Liu, G.; Guo, J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing* **2019**, *337*, 325–338. [CrossRef]
78. Wei, L.; Chen, S.; Lin, J.; Shi, L. Enhancing return forecasting using LSTM with agent-based synthetic data. *Decis. Support Syst.* **2025**, *193*, 114452. [CrossRef]
79. Azkue, M.; Miguel, E.; Martinez-Laserna, E.; Oca, L.; Iraola, U. Creating a Robust SoC Estimation Algorithm Based on LSTM Units and Trained with Synthetic Data. *World Electr. Veh. J.* **2023**, *14*, 197. [CrossRef]
80. Brophy, E.; Wang, Z.; She, Q.; Ward, T. Generative Adversarial Networks in Time Series: A Systematic Literature Review. *ACM Comput. Surv.* **2023**, *55*, 1–31. [CrossRef]
81. Takahashi, S.; Chen, Y.; Tanaka-Ishii, K. Modeling financial time-series with generative adversarial networks. *Phys. A Stat. Mech. Its Appl.* **2019**, *527*, 121261. [CrossRef]
82. Chatterjee, S.; Hazra, D.; Byun, Y.C. GAN-based synthetic time-series data generation for improving prediction of demand for electric vehicles. *Expert Syst. Appl.* **2025**, *264*, 125838. [CrossRef]
83. Deng, F.; Chen, Z.; Liu, Y.; Yang, S.; Hao, R.; Lyu, L. A Novel Combination Neural Network Based on ConvLSTM-Transformer for Bearing Remaining Useful Life Prediction. *Machines* **2022**, *10*, 1226. [CrossRef]
84. Bounoua, I.; Saidi, Y.; Yaagoubi, R.; Bouziani, M. Deep Learning Approaches for Water Stress Forecasting in Arboriculture Using Time Series of Remote Sensing Images: Comparative Study between ConvLSTM and CNN-LSTM Models. *Technologies* **2024**, *12*, 77. [CrossRef]
85. da Silva, D.G.; Geller, M.T.B.; dos Santos Moura, M.S.; de Moura Meneses, A.A. Performance evaluation of LSTM neural networks for consumption prediction. *e-Prime Adv. Electr. Eng. Electron. Energy* **2022**, *2*, 100030. [CrossRef]
86. Haque, S.T.; Debnath, M.; Yasmin, A.; Mahmud, T.; Ngu, A.H.H. Experimental Study of Long Short-Term Memory and Transformer Models for Fall Detection on Smartwatches. *Sensors* **2024**, *24*, 6235. [CrossRef] [PubMed]
87. Trujillo-Guerrero, M.F.; Román-Niemes, S.; Jaén-Vargas, M.; Cadiz, A.; Fonseca, R.; Serrano-Olmedo, J.J. Accuracy Comparison of CNN, LSTM, and Transformer for Activity Recognition Using IMU and Visual Markers. *IEEE Access* **2023**, *11*, 106650–106669. [CrossRef]
88. Curreri, F.; Patanè, L.; Xibilia, M.G. RNN- and LSTM-Based Soft Sensors Transferability for an Industrial Process. *Sensors* **2021**, *21*, 823. [CrossRef]
89. Yunita, A.; Pratama, M.I.; Almuzakki, M.Z.; Ramadhan, H.; Akhir, E.A.P.; Firdausiah Mansur, A.B.; Basori, A.H. Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models. *MethodsX* **2025**, *15*, 103462. [CrossRef]
90. Crisóstomo de Castro Filho, H.; Abílio de Carvalho Júnior, O.; Ferreira de Carvalho, O.L.; Pozzobon de Bem, P.; dos Santos de Moura, R.; Olino de Albuquerque, A.; Rosa Silva, C.; Guimarães Ferreira, P.H.; Fontes Guimarães, R.; Trancoso Gomes, R.A. Rice Crop Detection Using LSTM, Bi-LSTM, and Machine Learning Models from Sentinel-1 Time Series. *Remote Sens.* **2020**, *12*, 2655. [CrossRef]
91. Jiang, J.; Wan, X.; Zhu, F.; Xiang, D.; Hu, Z.; Mu, S. A deep learning framework integrating Transformer and LSTM architectures for pipeline corrosion rate forecasting. *Comput. Chem. Eng.* **2025**, 109365. [CrossRef]
92. Gao, S.; Chen, M.; Yang, W.; Zhu, H. TFF-TL: Transformer based on temporal feature fusion and LSTM for dynamic soft sensor modeling of industrial processes. *J. Taiwan Inst. Chem. Eng.* **2025**, *176*, 106328. [CrossRef]
93. Pantopoulou, S.; Cilliers, A.; Tsoukalas, L.H.; Heifetz, A. Transformers and Long Short-Term Memory Transfer Learning for GenIV Reactor Temperature Time Series Forecasting. *Energies* **2025**, *18*, 2286. [CrossRef]
94. Casola, S.; Lauriola, I.; Lavelli, A. Pre-trained transformers: An empirical comparison. *Mach. Learn. Appl.* **2022**, *9*, 100334. [CrossRef]
95. Papa, L.; Russo, P.; Amerini, I.; Zhou, L. A Survey on Efficient Vision Transformers: Algorithms, Techniques, and Performance Benchmarking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 7682–7700. [CrossRef] [PubMed]
96. Wang, R.; Ma, L.; He, G.; Johnson, B.A.; Yan, Z.; Chang, M.; Liang, Y. Transformers for Remote Sensing: A Systematic Review and Analysis. *Sensors* **2024**, *24*, 3495. [CrossRef] [PubMed]
97. Zhang, N.; Song, Y.; Fang, D.; Gao, Z.; Yan, Y. An Improved Deep Convolutional LSTM for Human Activity Recognition Using Wearable Sensors. *IEEE Sens. J.* **2024**, *24*, 1717–1729. [CrossRef]

98. Khatun, M.A.; Yousuf, M.A.; Ahmed, S.; Uddin, M.Z.; Alyami, S.A.; Al-Ashhab, S.; Akhdar, H.F.; Khan, A.; Azad, A.; Moni, M.A. Deep CNN-LSTM With Self-Attention Model for Human Activity Recognition Using Wearable Sensor. *IEEE J. Transl. Eng. Health Med.* **2022**, *10*, 1–16. [CrossRef]
99. Zhao, Y.; Yang, R.; Chevalier, G.; Xu, X.; Zhang, Z. Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors. *Math. Probl. Eng.* **2018**, *2018*, 7316954. [CrossRef]
100. Wu, Z.; Hu, P.; Liu, S.; Pang, T. Attention Mechanism and LSTM Network for Fingerprint-Based Indoor Location System. *Sensors* **2024**, *24*, 1398. [CrossRef]
101. Cao, M.; Yao, R.; Xia, J.; Jia, K.; Wang, H. LSTM Attention Neural-Network-Based Signal Detection for Hybrid Modulated Faster-Than-Nyquist Optical Wireless Communications. *Sensors* **2022**, *22*, 8992. [CrossRef]
102. Chen, H.; Yang, J.; Fu, X.; Zheng, Q.; Song, X.; Fu, Z.; Wang, J.; Liang, Y.; Yin, H.; Liu, Z.; et al. Water Quality Prediction Based on LSTM and Attention Mechanism: A Case Study of the Burnett River, Australia. *Sustainability* **2022**, *14*, 13231. [CrossRef]
103. Xie, T.; Ding, W.; Zhang, J.; Wan, X.; Wang, J. Bi-LS-AttM: A Bidirectional LSTM and Attention Mechanism Model for Improving Image Captioning. *Appl. Sci.* **2023**, *13*, 7916. [CrossRef]
104. Hassan, N.; Miah, A.; Shin, J. A Deep Bidirectional LSTM Model Enhanced by Transfer-Learning-Based Feature Extraction for Dynamic Human Activity Recognition. *Appl. Sci.* **2024**, *14*, 603. [CrossRef]
105. Lin, Y.; Zhang, Z.; Tan, Y.; Fu, H.; Min, H. Efficient TD3-based path planning of mobile robot in dynamic environments using prioritized experience replay and LSTM. *Sci. Rep.* **2025**, *15*, 18331. [CrossRef]
106. S.K.B., S.; Mathivanan, S.K.; Rajadurai, H.; Cho, J.; Easwaramoorthy, S.V. A multi-modal geospatial–temporal LSTM based deep learning framework for predictive modeling of urban mobility patterns. *Sci. Rep.* **2024**, *14*, 31579. [CrossRef]
107. Zhou, H.; Zhao, Y.; Liu, Y.; Lu, S.; An, X.; Liu, Q. Multi-Sensor Data Fusion and CNN-LSTM Model for Human Activity Recognition System. *Sensors* **2023**, *23*, 4750. [CrossRef]
108. Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* **2018**, *51*, 1–42. [CrossRef]