

ARQUITETURAS HÍBRIDAS EM DEEP LEARNING PARA PREVISÃO DE SÉRIES TEMPORAIS: UMA ANÁLISE DAS PRINCIPAIS ARQUITETURAS E ABORDAGENS

HYBRID DEEP LEARNING ARCHITECTURES FOR TIME SERIES
FORECASTING: AN ANALYSIS OF THE MAIN ARCHITECTURES AND
APPROACHES

Ciências Exatas e da Terra, Engenharias • 28/07/2026

REGISTRO DOI: [10.70773/revistatopicos/784956175](https://doi.org/10.70773/revistatopicos/784956175)

Eneida Glauce de Araújo Medeiros

RESUMO

A previsão de séries temporais desempenha papel estratégico em aplicações como sistemas energéticos, mercados financeiros, meteorologia, transporte e indústria, impulsionando o desenvolvimento de modelos capazes de representar a crescente complexidade dos dados temporais. Nesse contexto, arquiteturas híbridas em *Deep Learning* têm se consolidado como uma abordagem promissora ao integrar mecanismos complementares de aprendizagem para superar limitações observadas em modelos estatísticos e arquiteturas neurais empregadas de forma isolada. Este estudo analisa as principais arquiteturas híbridas aplicadas à previsão de séries temporais por meio de uma análise qualitativa fundamentada na literatura científica recente. São discutidos os fundamentos, a evolução, os princípios de funcionamento, as aplicações, as vantagens e as limitações de arquiteturas como CNN-LSTM, ARIMA-LSTM, LSTM-GARCH, LSTM-Transformer, CNN-Transformer, *Temporal Fusion Transformer* e xLSTM. A análise evidencia que a evolução dessas arquiteturas não decorre da substituição sucessiva de modelos, mas da integração progressiva de mecanismos especializados de aprendizagem, capazes de combinar extração automática de características, modelagem de dependências temporais, mecanismos de atenção e estratégias probabilísticas em uma mesma estrutura computacional. Os estudos analisados demonstram que essa integração amplia a capacidade preditiva, a robustez e a adaptabilidade dos modelos em diferentes domínios de aplicação, embora desafios relacionados ao custo computacional, à interpretabilidade, à generalização e à otimização de hiperparâmetros permaneçam relevantes. Conclui-se que o avanço da previsão de séries temporais está associado ao desenvolvimento de arquiteturas híbridas cada vez mais especializadas e complementares, cuja adoção deve considerar

simultaneamente as características dos dados, os objetivos da aplicação e os requisitos de desempenho.

Palavras-chave: Deep Learning; Previsão de séries temporais; Arquiteturas híbridas; Inteligência Artificial; Modelagem preditiva.

ABSTRACT

Time series forecasting plays a strategic role in applications such as energy systems, financial markets, meteorology, transportation, and industry, driving the development of models capable of addressing the increasing complexity of temporal data. In this context, hybrid Deep Learning architectures have emerged as a promising approach by integrating complementary learning mechanisms to overcome the limitations of traditional statistical models and standalone neural network architectures. This study analyzes the main hybrid architectures applied to time series forecasting through a qualitative analysis based on recent scientific literature. It discusses the foundations, evolution, operating principles, applications, advantages, and limitations of architectures such as CNN-LSTM, ARIMA-LSTM, LSTM-GARCH, LSTM-Transformer, CNN-Transformer, Temporal Fusion Transformer, and xLSTM. The analysis shows that the evolution of these architectures has been driven not by the successive replacement of existing models but by the progressive integration of specialized learning mechanisms capable of combining automatic feature extraction, temporal dependency modeling, attention mechanisms, and probabilistic strategies within a unified computational framework. The reviewed studies demonstrate that this integration enhances predictive performance, robustness, and adaptability across different application domains, although challenges related to computational cost, interpretability, generalization, and hyperparameter optimization remain significant. It is concluded that advances in time series forecasting are closely

associated with the development of increasingly specialized and complementary hybrid architectures, whose adoption should simultaneously consider data characteristics, application objectives, and performance requirements.

Keywords: Deep Learning; Time Series Forecasting; Hybrid Architectures; Artificial Intelligence; Predictive Modeling.

1. INTRODUÇÃO

A crescente digitalização de processos, a expansão da Internet das Coisas (*Internet of Things* – IoT) e a disseminação de sensores inteligentes ampliaram significativamente a produção de dados organizados ao longo do tempo. Em setores como energia, transporte, indústria, meteorologia e mercados financeiros, esses dados passaram a desempenhar papel estratégico no monitoramento de sistemas, no planejamento operacional e no apoio à tomada de decisão. Paralelamente, o aumento do volume, da diversidade e da complexidade dessas informações tornou a previsão de séries temporais um dos principais desafios da Inteligência Artificial, impulsionando o desenvolvimento de métodos capazes de representar padrões temporais cada vez mais complexos [2], [12], [18].

Durante décadas, a previsão de séries temporais esteve fundamentada principalmente em modelos estatísticos, entre os quais se destaca o *AutoRegressive Integrated Moving Average* (ARIMA). Embora esses métodos apresentem bom desempenho em séries caracterizadas por relações predominantemente lineares e comportamento relativamente estável, sua capacidade de modelagem torna-se limitada diante de fenômenos marcados por não linearidade, sazonalidade, elevada volatilidade, ruído e

interações complexas entre múltiplas variáveis. Nesse contexto, técnicas baseadas em *Deep Learning* passaram a representar uma alternativa capaz de aprender padrões diretamente dos dados e modelar dependências temporais de maior complexidade [2], [6], [7], [16].

Entre as arquiteturas de *Deep Learning* mais empregadas para esse propósito destacam-se as redes *Long Short-Term Memory* (LSTM), especializadas na representação de dependências temporais de longo prazo; as *Convolutional Neural Networks* (CNN), utilizadas para extração automática de características; e, mais recentemente, os modelos baseados em mecanismos de atenção, como os *Transformers*, capazes de capturar relações globais entre diferentes instantes da sequência temporal [9], [10], [13], [23]. Apesar dos avanços proporcionados por essas arquiteturas, estudos recentes indicam que sua utilização isolada nem sempre é suficiente para representar adequadamente a complexidade presente em aplicações reais, especialmente em cenários caracterizados por elevada variabilidade, comportamento não estacionário e múltiplas fontes de informação [5], [9], [14], [18].

Como resposta a essas limitações, a integração entre diferentes arquiteturas deu origem aos modelos híbridos em *Deep Learning*. Essas abordagens combinam mecanismos complementares de aprendizagem em uma mesma estrutura computacional, explorando simultaneamente a extração de características, a modelagem das dependências temporais e a identificação de relações de curto e longo prazo. Arquiteturas como CNN-LSTM, ARIMA-LSTM, LSTM-GARCH, LSTM-Transformer, CNN-Transformer, *Temporal Fusion Transformer* e xLSTM exemplificam essa evolução ao integrar diferentes estratégias de representação para ampliar a

capacidade preditiva em séries temporais complexas [1], [2], [6], [9], [10], [12], [17], [18], [23].

Embora a produção científica sobre arquiteturas híbridas tenha crescido de forma expressiva nos últimos anos, o conhecimento permanece disperso entre diferentes domínios de aplicação, propostas arquiteturais e estratégias de modelagem. Essa fragmentação dificulta a compreensão da evolução dessas arquiteturas, das motivações para sua combinação e das contribuições específicas de cada abordagem para diferentes tipos de séries temporais. Torna-se, portanto, relevante reunir e analisar criticamente as evidências disponíveis, identificando tendências, potencialidades e limitações que orientem tanto a pesquisa quanto o desenvolvimento de novas soluções [5], [9], [18].

Diante desse contexto, este artigo analisa as principais arquiteturas híbridas em *Deep Learning* aplicadas à previsão de séries temporais por meio de uma revisão bibliográfica de abordagem qualitativa. A partir da análise crítica de estudos recentes, busca-se sistematizar a evolução dessas arquiteturas, discutir suas principais aplicações, vantagens e limitações, bem como identificar tendências que vêm orientando o desenvolvimento de modelos preditivos mais robustos, eficientes e adaptáveis a diferentes contextos de aplicação.

2. DESENVOLVIMENTO

A evolução da previsão de séries temporais acompanha as transformações ocorridas tanto na natureza dos dados quanto nas estratégias de modelagem desenvolvidas para interpretá-los. O aumento do volume de informações, da heterogeneidade das variáveis e da ocorrência de padrões não lineares ampliou a

complexidade dos problemas de previsão, tornando insuficientes abordagens baseadas em um único paradigma de modelagem. Nesse contexto, observa-se uma evolução que parte dos modelos estatísticos tradicionais, avança para arquiteturas de *Deep Learning* e culmina no desenvolvimento de modelos híbridos capazes de integrar mecanismos complementares de aprendizagem para lidar com diferentes características das séries temporais [2], [5], [18].

2.1. Evolução da Previsão de Séries Temporais

A evolução da previsão de séries temporais não corresponde à substituição sucessiva de diferentes modelos, mas à incorporação de novas estratégias capazes de responder às limitações identificadas nas abordagens existentes. Cada avanço metodológico ampliou a capacidade de representação dos fenômenos temporais, preservando, em muitos casos, técnicas anteriores como componentes complementares de modelos mais sofisticados [2], [5], [18].

Durante muitos anos, os modelos estatísticos constituíram a principal alternativa para problemas de previsão, destacando-se a família *AutoRegressive Integrated Moving Average* (ARIMA). Essas abordagens demonstraram bom desempenho em séries caracterizadas por relações lineares, comportamento relativamente estável e padrões estatísticos bem definidos. Entretanto, sua capacidade de representação mostrou-se limitada em cenários marcados por não linearidade, elevada volatilidade, sazonalidades múltiplas e interações complexas entre variáveis, características frequentemente observadas em aplicações contemporâneas [2], [6], [7], [16].

A adoção de técnicas baseadas em *Deep Learning* ampliou significativamente a capacidade de modelagem ao permitir que representações relevantes fossem aprendidas diretamente dos dados. Nesse contexto, arquiteturas *Long Short-Term Memory* (LSTM) consolidaram-se como uma das principais alternativas para modelagem de dependências temporais de longo prazo, enquanto as *Convolutional Neural Networks* (CNN) passaram a ser empregadas na extração automática de padrões locais. Posteriormente, modelos baseados em mecanismos de atenção, especialmente os *Transformers*, expandiram essa capacidade ao representar relações globais entre diferentes instantes das sequências temporais, reduzindo limitações associadas às arquiteturas recorrentes tradicionais [9], [10], [13], [23].

À medida que diferentes arquiteturas passaram a ser aplicadas em domínios como mercados financeiros, sistemas energéticos, previsão meteorológica e mobilidade inteligente, tornou-se evidente que seu desempenho depende não apenas da arquitetura empregada, mas também das características da série temporal, do horizonte de previsão e do contexto da aplicação. Estudos recentes mostram que arquiteturas distintas apresentam vantagens específicas conforme a natureza do problema, evidenciando que não existe um modelo universalmente superior para todos os cenários de previsão [2], [9], [14], [15], [18], [23].

Essa constatação impulsionou uma mudança na forma de conceber sistemas preditivos. Em vez de concentrar esforços no desenvolvimento de arquiteturas isoladamente mais complexas, a pesquisa passou a explorar a integração de mecanismos complementares de aprendizagem. As arquiteturas híbridas surgem como resultado dessa evolução ao combinar diferentes paradigmas

de modelagem em uma mesma estrutura computacional, explorando de forma conjunta a extração de características, a modelagem de dependências temporais, os mecanismos de atenção e, em alguns casos, a complementaridade entre modelos estatísticos e redes neurais profundas. Essa estratégia permitiu ampliar a robustez e a capacidade de adaptação dos modelos diante da diversidade de comportamentos observados em séries temporais complexas [1], [2], [5], [6], [9], [10], [17], [18], [23].

Sob essa perspectiva, a evolução da previsão de séries temporais representa uma mudança de paradigma mais ampla do que a simples sucessão de novas arquiteturas. O foco deslocou-se da busca por modelos individualmente mais sofisticados para o desenvolvimento de soluções capazes de integrar diferentes mecanismos de aprendizagem de forma cooperativa, explorando suas características complementares para aumentar a capacidade preditiva em cenários cada vez mais complexos.

2.2. Arquiteturas Híbridas em Deep Learning

As arquiteturas híbridas em *Deep Learning* caracterizam-se pela integração de diferentes mecanismos de aprendizagem em uma mesma estrutura computacional, explorando as capacidades complementares de cada modelo para representar diferentes propriedades das séries temporais. Em vez de depender de uma única estratégia de processamento, essas arquiteturas combinam extração automática de características, modelagem de dependências temporais, representação de componentes lineares e identificação de relações globais por meio da integração entre modelos convolucionais, recorrentes, estatísticos e baseados em mecanismos de atenção. Essa abordagem ampliou

significativamente a capacidade de modelagem em problemas caracterizados por elevada complexidade, não linearidade e comportamento não estacionário, consolidando-se como uma das principais linhas de pesquisa em previsão de séries temporais [5], [9], [18].

2.2.1. Arquiteturas CNN-LSTM

A arquitetura CNN-LSTM combina dois mecanismos de aprendizagem que desempenham funções distintas durante o processo de previsão de séries temporais. Enquanto as *Convolutional Neural Networks* (CNN) são responsáveis pela identificação automática de padrões locais presentes nos dados de entrada, as redes *Long Short-Term Memory* (LSTM) modelam a evolução temporal dessas informações ao longo da sequência. A integração entre essas arquiteturas permite que o processo preditivo seja dividido em etapas complementares, nas quais a extração de características e a modelagem das dependências temporais ocorrem de forma sequencial. Essa organização tornou a CNN-LSTM uma das arquiteturas híbridas mais difundidas na literatura sobre previsão de séries temporais, servindo de base para o desenvolvimento de modelos híbridos mais complexos [5], [18].

Na arquitetura CNN-LSTM, a CNN atua como o primeiro componente de processamento, sendo responsável por identificar padrões locais presentes na série temporal. Esse processo ocorre por meio da convolução, operação matemática na qual pequenos filtros, também chamados de *kernels*, percorrem sequencialmente os dados de entrada para detectar combinações recorrentes de valores. Em séries temporais, esses filtros podem identificar variações de curto prazo, oscilações locais, tendências iniciais e padrões

repetitivos que poderiam não ser facilmente percebidos por uma análise direta dos dados brutos. À medida que os filtros percorrem a sequência, são geradas representações intermediárias que destacam as características mais relevantes da série. Em seguida, funções de ativação introduzem não linearidade ao modelo, enquanto operações de *pooling* sintetizam essas representações, reduzindo sua dimensionalidade, preservando as informações mais relevantes e reduzindo a complexidade computacional. Dessa forma, a CNN transforma a série temporal original em uma representação mais compacta e informativa, que servirá como entrada para a etapa recorrente da arquitetura [11], [13], [18].

As representações extraídas pela CNN são então encaminhadas para a rede *Long Short-Term Memory* (LSTM), responsável pela modelagem das dependências temporais existentes na série. Diferentemente das Redes Neurais Recorrentes tradicionais (*Recurrent Neural Networks* – RNN), as LSTM foram desenvolvidas para reduzir o problema do desaparecimento do gradiente (*vanishing gradient*), que dificulta a aprendizagem de relações de longo prazo durante o treinamento. Para isso, utilizam células de memória capazes de armazenar informações ao longo da sequência temporal e um conjunto de mecanismos de controle denominados portas (*gates*). A porta de entrada (*input gate*) determina quais informações da entrada atual devem ser incorporadas à memória; a porta de esquecimento (*forget gate*) decide quais informações previamente armazenadas devem ser descartadas; e a porta de saída (*output gate*) controla quais informações serão utilizadas na produção da saída da rede. Em conjunto, esses mecanismos regulam continuamente o fluxo de informações ao longo da sequência temporal, permitindo preservar dependências relevantes mesmo entre observações separadas por longos intervalos

temporais. Essa capacidade constitui um dos principais diferenciais da arquitetura em problemas de previsão de séries temporais [5], [11], [13], [18].

Na arquitetura híbrida CNN-LSTM, esses dois mecanismos atuam de forma complementar e sequencial. Inicialmente, a CNN transforma os dados brutos da série temporal em uma representação hierárquica composta pelas características mais relevantes identificadas durante o processo de convolução. Em seguida, essa representação é utilizada como entrada pela LSTM, que modela a evolução temporal dessas características ao longo da sequência e aprende as relações existentes entre diferentes instantes de tempo. Essa divisão de responsabilidades permite que cada componente execute a tarefa para a qual foi originalmente concebido, reduzindo limitações observadas quando CNN ou LSTM são empregadas isoladamente. O resultado é uma arquitetura na qual a representação dos dados e a modelagem temporal deixam de competir entre si e passam a atuar de forma integrada durante o processo preditivo [11], [13], [18].

As evidências reunidas na literatura indicam que a principal contribuição da arquitetura CNN-LSTM está na complementaridade funcional entre seus componentes, permitindo integrar a extração automática de características à modelagem das dependências temporais em um único fluxo de aprendizagem. Salman et al. [18] verificaram que essa arquitetura apresentou desempenho superior na previsão de geração de energia solar quando comparada a modelos individuais, atribuindo esse resultado à capacidade da CNN em representar padrões locais das variáveis meteorológicas e da LSTM em modelar sua evolução temporal. Resultados semelhantes foram observados por Li e Qian [11] na previsão de temperatura,

evidenciando redução consistente dos erros de previsão em relação a arquiteturas recorrentes convencionais. Em aplicações voltadas à previsão de carga elétrica, Lu et al. [12] também demonstraram que a integração entre mecanismos de extração de características e modelagem temporal aumenta a estabilidade das previsões em ambientes caracterizados por elevada variabilidade dos dados. Em conjunto, essas evidências indicam que a arquitetura CNN-LSTM apresenta desempenho consistente em diferentes problemas de previsão, evidenciando elevada capacidade de adaptação a distintos domínios de aplicação e robustez frente à complexidade e à variabilidade características de séries temporais reais.

O potencial de integração da arquitetura CNN-LSTM também favoreceu o desenvolvimento de novas variações capazes de incorporar mecanismos adicionais de aprendizagem. Nghiem et al. [13], por exemplo, integraram Inferência Bayesiana à arquitetura CNN-LSTM, permitindo estimar a incerteza associada às previsões e reduzir o risco de sobreajuste em bases de dados reduzidas. Essa abordagem amplia a confiabilidade do modelo, aspecto particularmente relevante em aplicações críticas. Em uma perspectiva mais ampla, Ganaie et al. [5] destacam que arquiteturas híbridas e modelos em *ensemble* vêm se consolidando como uma tendência no *Deep Learning*, justamente por explorarem a complementaridade entre diferentes algoritmos de aprendizagem. Esses resultados evidenciam que o sucesso da arquitetura CNN-LSTM não decorre apenas da combinação entre duas redes neurais, mas da possibilidade de integrar mecanismos adicionais capazes de aumentar sua robustez, capacidade de generalização e desempenho preditivo, estabelecendo as bases para o desenvolvimento de arquiteturas híbridas ainda mais sofisticadas.

2.2.2. Integração Entre Modelos Estatísticos e *Deep Learning*

A integração entre modelos estatísticos e arquiteturas de *Deep Learning* representa uma das principais estratégias empregadas para ampliar a capacidade preditiva em problemas de previsão de séries temporais. Diferentemente das arquiteturas híbridas formadas exclusivamente por redes neurais, essa abordagem fundamenta-se na utilização complementar de métodos desenvolvidos sob paradigmas distintos de modelagem. Modelos estatísticos, como *AutoRegressive Integrated Moving Average* (ARIMA) e *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH), apresentam elevado desempenho na representação de componentes lineares, tendências, sazonalidade e padrões de volatilidade, enquanto arquiteturas neurais destacam-se pela aprendizagem de relações não lineares e dependências temporais complexas diretamente dos dados. A motivação para integrar essas abordagens decorre da constatação de que diferentes séries temporais frequentemente apresentam, de forma simultânea, comportamentos lineares e não lineares, tornando insuficiente a utilização isolada de qualquer um desses modelos [2], [6], [7], [15], [16].

O funcionamento dessas arquiteturas híbridas baseia-se na divisão do processo de aprendizagem entre modelos especializados em representar diferentes componentes da série temporal. De modo geral, o processamento inicia-se pela aplicação de um modelo estatístico, responsável por identificar padrões lineares, tendências, sazonalidade ou volatilidade presentes nos dados. As informações produzidas nessa etapa, que podem corresponder às previsões do modelo estatístico, aos resíduos gerados durante a modelagem ou a parâmetros associados ao comportamento da série, são então utilizadas como entrada para uma arquitetura de *Deep Learning*,

geralmente baseada em redes *Long Short-Term Memory* (LSTM). Enquanto o modelo estatístico representa os padrões mais regulares da série, a rede neural concentra-se na aprendizagem das relações não lineares e das dependências temporais remanescentes, produzindo uma representação mais completa do comportamento da série temporal. Como consequência, a integração entre modelos estatísticos e *Deep Learning* não representa uma simples sobreposição de algoritmos, mas uma estratégia cooperativa de modelagem, na qual diferentes métodos atuam de forma complementar para representar a dinâmica temporal dos dados [2], [6], [7].

As evidências empíricas disponíveis indicam que a integração entre modelos estatísticos e *Deep Learning* produz ganhos consistentes de desempenho em séries temporais caracterizadas pela coexistência de padrões lineares e não lineares. Gasmi et al. [7] verificaram que a arquitetura ARIMA-LSTM apresentou maior precisão na previsão temporal ao combinar a capacidade do modelo ARIMA para representar tendências e sazonalidade com a habilidade da LSTM em aprender relações não lineares remanescentes. Resultados semelhantes foram obtidos por Dadi et al. [2], que aplicaram uma arquitetura híbrida ARIMA-LSTM na previsão do preço do Ethereum, evidenciando redução dos erros preditivos em comparação com modelos utilizados de forma isolada. Esses estudos indicam que a decomposição do problema de previsão em componentes lineares e não lineares permite uma representação mais completa da dinâmica das séries temporais, aumentando a robustez e a capacidade de adaptação dos modelos híbridos.

Em aplicações financeiras, a integração entre modelos estatísticos e redes neurais também se mostrou particularmente eficiente na

representação da volatilidade dos mercados. Gao, He e Kuruoglu [6] demonstraram que a combinação entre GARCH e LSTM melhora significativamente a modelagem das oscilações do Bitcoin ao utilizar o modelo GARCH para representar a volatilidade condicional e a LSTM para capturar padrões temporais não lineares. Essa perspectiva é reforçada pelos estudos de Ramani et al. [16], Dimitriadou e Gregoriou [4] e Raghavendran et al. [15], cujos resultados comparativos evidenciam que o desempenho dos modelos preditivos depende diretamente das características da série temporal analisada, não existindo uma abordagem universalmente superior. Em conjunto, essas evidências demonstram que o principal diferencial das arquiteturas híbridas não reside apenas na combinação de diferentes algoritmos, mas na capacidade de explorar, de forma complementar, paradigmas distintos de modelagem para representar fenômenos temporais complexos.

2.2.3. Arquiteturas Híbridas Baseadas em Mecanismos de Atenção

A incorporação de mecanismos de atenção (*Attention Mechanisms*) às arquiteturas híbridas em *Deep Learning* representa uma evolução natural dos modelos desenvolvidos para previsão de séries temporais. Embora arquiteturas como CNN-LSTM tenham apresentado resultados expressivos na representação de padrões locais e dependências temporais, estudos recentes evidenciaram limitações na modelagem de relações globais entre observações temporalmente distantes, especialmente em séries longas e altamente complexas. Nesse contexto, os mecanismos de atenção passaram a ser integrados às arquiteturas híbridas por permitirem que o modelo identifique, de forma dinâmica, quais instantes da

sequência exercem maior influência sobre a previsão, independentemente da distância temporal entre eles. Essa capacidade ampliou significativamente o potencial de representação dos modelos híbridos e impulsionou o desenvolvimento de arquiteturas capazes de integrar mecanismos convolucionais, recorrentes e de atenção em uma mesma estrutura computacional [9], [10], [17], [23].

O funcionamento dessas arquiteturas fundamenta-se no mecanismo de atenção (*Self-Attention*), desenvolvido para permitir que o modelo identifique automaticamente quais informações da sequência temporal são mais relevantes para a realização de uma previsão. Diferentemente das arquiteturas recorrentes, que processam os dados de forma sequencial, o mecanismo de atenção analisa simultaneamente todos os elementos da sequência, atribuindo pesos distintos a cada observação conforme sua contribuição para a tarefa preditiva. Esse processo é realizado por meio de três representações matemáticas denominadas *Query* (consulta), *Key* (chave) e *Value* (valor). A *Query* representa a informação para a qual o modelo procura contexto; a *Key* descreve as características de cada elemento da sequência; e a *Value* corresponde às informações efetivamente utilizadas durante o processo de aprendizagem. A partir da interação entre *Queries* e *Keys*, o modelo estima a relevância relativa de cada observação, concentrando maior atenção nas informações mais relevantes para gerar a previsão [9], [10], [23].

Em arquiteturas como LSTM-Transformer, a LSTM realiza inicialmente a modelagem das dependências temporais, enquanto o Transformer identifica relações globais entre diferentes instantes da sequência. Em modelos CNN-Transformer, a CNN atua na

extração automática de padrões locais antes da aplicação do mecanismo de atenção. Já arquiteturas mais recentes, como CNN-LSTM-Transformer e *Temporal Fusion Transformer* (TFT), combinam convolução, memória temporal e atenção em uma única estrutura computacional, permitindo representar simultaneamente características locais, dependências de curto e longo prazo e interações entre múltiplas variáveis. Essa integração torna essas arquiteturas particularmente adequadas para séries temporais extensas, multivariadas e caracterizadas por elevada complexidade, nas quais diferentes mecanismos de aprendizagem contribuem para representar aspectos complementares da dinâmica temporal [2], [9], [17], [18], [23].

Em aplicações voltadas à previsão de geração de energia solar, Salman et al. [18] demonstraram que a arquitetura CNN-LSTM-Transformer superou modelos compostos por apenas uma ou duas dessas arquiteturas, evidenciando que a integração entre convolução, memória temporal e mecanismos de atenção permite representar, de forma simultânea, padrões locais, dependências sequenciais e relações globais presentes na série temporal. Resultados semelhantes foram obtidos por Kabir et al. [9], cuja arquitetura LSTM-Transformer apresentou elevada capacidade de modelar séries financeiras, destacando-se principalmente em cenários que exigem a representação de dependências temporais de longo alcance. Em outra perspectiva, Ranjbar e Rahimzadeh [17] verificaram que a combinação entre CNN, LSTM e Transformer também proporcionou melhorias significativas na previsão do consumo de combustíveis, reforçando que a complementaridade entre diferentes mecanismos de aprendizagem constitui um dos principais fatores responsáveis pelo desempenho dessas arquiteturas.

À medida que os mecanismos de atenção passaram a ser incorporados às arquiteturas híbridas, novas estratégias de modelagem começaram a surgir para atender problemas cada vez mais complexos. Tu [23] demonstrou que a arquitetura CNN-Transformer favorece a representação simultânea de padrões locais e relações globais em séries financeiras, enquanto Dadi et al. [2] destacam que o *Temporal Fusion Transformer* amplia essa capacidade ao incorporar mecanismos de seleção dinâmica de variáveis e atenção interpretável em problemas multivariados. Entretanto, a evolução da área não indica a substituição das arquiteturas recorrentes pelos Transformers. Nichani et al. [14], ao compararem modelos baseados em LSTM e Transformer, observaram que o desempenho permanece fortemente condicionado às características da série temporal e ao contexto de aplicação. Essa conclusão é reforçada por Alharthi e Mahmood [1], que demonstram a competitividade da xLSTM em previsões de longo horizonte temporal. Consideradas em conjunto, essas evidências indicam que o avanço recente da previsão de séries temporais está associado à integração de mecanismos complementares de aprendizagem, e não à predominância de uma arquitetura específica.

Apesar dos avanços proporcionados pelos mecanismos de atenção, essas arquiteturas ainda apresentam desafios importantes para sua aplicação em diferentes contextos. Um dos principais está relacionado ao elevado custo computacional do mecanismo de *Self-Attention*, cuja complexidade cresce à medida que aumenta o comprimento da sequência temporal, exigindo maior capacidade de processamento e memória. Além disso, modelos baseados em Transformers geralmente demandam grandes volumes de dados para treinamento e são sensíveis à escolha dos hiperparâmetros,

fatores que podem comprometer seu desempenho em bases reduzidas ou pouco representativas. Nichani et al. [14] ressaltam que a superioridade dessas arquiteturas depende das características da série temporal e do contexto de aplicação, não existindo um modelo universalmente superior. De forma semelhante, Alharthi e Mahmood [1] demonstram que arquiteturas recorrentes de nova geração, como a xLSTM, permanecem altamente competitivas em tarefas de previsão de longo horizonte temporal. Essas evidências indicam que os mecanismos de atenção representam um avanço significativo na modelagem de séries temporais, mas não eliminam os desafios relacionados ao custo computacional, à generalização dos modelos e à necessidade de estratégias mais eficientes para diferentes cenários de aplicação.

A análise integrada das arquiteturas baseadas em mecanismos de atenção evidencia que a evolução da previsão de séries temporais passou a ser orientada pela combinação de diferentes paradigmas de aprendizagem, em vez da busca por um modelo capaz de substituir seus predecessores. As evidências discutidas nesta seção mostram que CNN, LSTM, Transformers e suas variantes desempenham funções complementares, explorando diferentes níveis de representação dos dados para ampliar a capacidade preditiva dos modelos. Ao mesmo tempo, o surgimento de arquiteturas como o *Temporal Fusion Transformer* e a xLSTM demonstra que a pesquisa na área permanece em constante evolução, direcionando esforços para modelos cada vez mais eficientes, interpretáveis e adaptáveis a diferentes contextos de aplicação. Esse cenário reforça que o avanço da previsão de séries temporais está associado à integração de mecanismos complementares de aprendizagem e à especialização das arquiteturas híbridas, estabelecendo uma base sólida para o

desenvolvimento de soluções preditivas capazes de responder aos desafios impostos por séries temporais cada vez mais complexas.

2.2.4. Arquiteturas Híbridas Avançadas para Previsão de Séries Temporais

O avanço das arquiteturas híbridas para previsão de séries temporais tem sido acompanhado pelo desenvolvimento de modelos capazes de integrar mecanismos de aprendizagem cada vez mais especializados. Embora combinações entre CNN, LSTM e Transformers tenham ampliado significativamente a capacidade preditiva dos modelos, desafios relacionados à interpretabilidade, ao custo computacional, à modelagem de dependências de longo prazo, à quantificação das incertezas e à adaptação a diferentes características dos dados permanecem como importantes limitações. Nesse contexto, surgiram arquiteturas híbridas de nova geração que incorporam mecanismos adicionais, como memória expandida, inferência bayesiana, *autoencoders* e estratégias adaptativas de seleção de informações, buscando aumentar a robustez, a eficiência e a capacidade de generalização dos modelos preditivos [1], [5], [9], [10], [13], [23].

Essas arquiteturas avançadas caracterizam-se pela incorporação de mecanismos adicionais ao fluxo tradicional de processamento das arquiteturas híbridas, ampliando sua capacidade de aprendizagem e adaptação a diferentes tipos de séries temporais. Em vez de combinar apenas duas ou três arquiteturas neurais, esses modelos integram componentes especializados que desempenham funções específicas durante o processo preditivo. Entre esses mecanismos destacam-se arquiteturas de memória expandida, capazes de preservar informações relevantes por horizontes temporais mais

longos; *autoencoders*, utilizados para reduzir ruídos e gerar representações mais informativas dos dados; métodos de inferência bayesiana, empregados para estimar a incerteza associada às previsões; e mecanismos adaptativos de seleção de variáveis e atenção, que permitem ao modelo priorizar automaticamente as informações mais relevantes em cada etapa da previsão. Essa integração torna o processo de aprendizagem mais flexível e contribui para a construção de modelos capazes de lidar com séries temporais caracterizadas por elevada complexidade, não linearidade e variabilidade [1], [9], [10], [13], [17], [23].

A arquitetura xLSTM representa um dos principais avanços recentes na evolução das redes recorrentes ao ampliar o mecanismo de memória das LSTM tradicionais por meio de estruturas capazes de preservar informações relevantes por horizontes temporais mais extensos. Alharthi e Mahmood [1] demonstraram que essa arquitetura apresenta desempenho competitivo em tarefas de previsão de longo prazo, reduzindo limitações associadas ao esquecimento de informações observado em modelos recorrentes convencionais. Esses resultados evidenciam que a evolução das arquiteturas híbridas não está restrita à incorporação de mecanismos de atenção, mas também ao aprimoramento de componentes recorrentes, ampliando sua capacidade de representar dependências temporais complexas.

Outra estratégia amplamente explorada consiste na incorporação de mecanismos probabilísticos para representar a incerteza inerente às previsões. Nghiem et al. [13] integraram Inferência Bayesiana a uma arquitetura CNN-LSTM, permitindo estimar intervalos de confiança para as previsões e reduzir o risco de sobreajuste em conjuntos de dados limitados. Essa abordagem amplia a confiabilidade dos

modelos em aplicações nas quais a quantificação da incerteza possui relevância prática, como sistemas de apoio à decisão e aplicações críticas.

Além da incorporação de novos mecanismos de aprendizagem, observa-se uma tendência ao desenvolvimento de arquiteturas capazes de adaptar dinamicamente seu processamento às características dos dados. Modelos como o *Temporal Fusion Transformer* empregam mecanismos de seleção dinâmica de variáveis e atenção interpretável para identificar automaticamente quais informações exercem maior influência sobre cada etapa da previsão, favorecendo tanto o desempenho preditivo quanto a interpretabilidade do modelo [2], [9]. Em paralelo, arquiteturas híbridas que combinam CNN, LSTM e Transformers continuam sendo aprimoradas por meio da integração de componentes especializados, evidenciando que a evolução da área ocorre pela ampliação da complementaridade entre diferentes mecanismos de aprendizagem, e não pela substituição de arquiteturas previamente consolidadas [17], [18], [23].

Em conjunto, essas evidências demonstram que as arquiteturas híbridas avançadas representam uma nova etapa na evolução da previsão de séries temporais. Mais do que aumentar a complexidade estrutural dos modelos, essas abordagens procuram incorporar mecanismos especializados capazes de enfrentar limitações específicas relacionadas à memória de longo prazo, à interpretabilidade, à quantificação das incertezas e à adaptação às características dos dados. Como resultado, observa-se uma tendência ao desenvolvimento de modelos preditivos cada vez mais robustos, flexíveis e capazes de atender às exigências de aplicações em cenários temporais complexos.

CNN-LSTM-Bayes

Outra evolução observada na literatura consiste na incorporação de métodos probabilísticos às arquiteturas híbridas. Na arquitetura CNN-LSTM-Bayes, proposta por **Nghiem et al. [13]**, a CNN permanece responsável pela extração automática de características da série temporal, enquanto a LSTM modela as dependências temporais. O diferencial dessa abordagem está na integração da Inferência Bayesiana ao processo de aprendizagem, permitindo estimar explicitamente a incerteza associada às previsões. Diferentemente das arquiteturas determinísticas, que produzem apenas um valor previsto, esse modelo fornece também uma medida de confiança para cada estimativa, aspecto particularmente relevante em aplicações nas quais a confiabilidade das previsões constitui um requisito fundamental, como sistemas energéticos, transporte e monitoramento ambiental.

Ao incorporar informações probabilísticas ao processo preditivo, a arquitetura amplia a capacidade do modelo de representar a incerteza inerente aos dados temporais, reduzindo os efeitos do sobreajuste em bases de dados limitadas e fornecendo informações adicionais para apoiar a tomada de decisão. Essa característica diferencia a CNN-LSTM-Bayes das arquiteturas híbridas convencionais, cujo foco está predominantemente na maximização da precisão preditiva, ao acrescentar uma dimensão probabilística que contribui para a interpretação e a confiabilidade dos resultados **[13]**.

Bi-LSTM + Denoising Autoencoder + Transformer

Arquiteturas ainda mais recentes passaram a integrar mecanismos de pré-processamento, memória bidirecional e atenção em uma única estrutura computacional. O modelo proposto por Koohfar e Woldemariam [10] combina um *Denoising Autoencoder*, responsável por remover ruídos e reconstruir representações mais informativas da série temporal, com uma rede *Bidirectional Long Short-Term Memory* (Bi-LSTM), capaz de analisar simultaneamente as dependências temporais nos sentidos passado–futuro e futuro–passado. Em seguida, um Transformer aplica mecanismos de *Self-Attention* para identificar relações globais entre diferentes instantes da sequência, produzindo uma representação mais abrangente da dinâmica temporal. A integração desses componentes permite que cada etapa do processamento desempenhe uma função complementar — redução de ruídos, modelagem bidirecional e identificação de relações globais — resultando em modelos mais robustos para séries temporais complexas e sujeitas a elevado grau de variabilidade.

A arquitetura proposta evidencia uma tendência recente na evolução dos modelos híbridos: a incorporação de mecanismos especializados capazes de atuar em diferentes etapas do processo preditivo. Nesse contexto, o *Denoising Autoencoder* melhora a qualidade das representações de entrada ao reduzir o impacto dos ruídos presentes nos dados, a Bi-LSTM amplia a capacidade de modelar dependências temporais em ambas as direções da sequência, e o Transformer complementa esse processamento ao identificar relações globais que dificilmente seriam capturadas por arquiteturas recorrentes utilizadas isoladamente [10].

Essa combinação evidencia uma mudança importante na concepção das arquiteturas híbridas para previsão de séries

temporais. O desenvolvimento de modelos mais recentes passou a privilegiar não apenas o aumento da precisão das previsões, mas também aspectos relacionados à robustez, à interpretabilidade, à capacidade de generalização e à adaptação a diferentes características dos dados. Como consequência, observa-se uma evolução em direção a arquiteturas cada vez mais modulares e especializadas, nas quais diferentes mecanismos de aprendizagem atuam de forma cooperativa para representar a dinâmica temporal em múltiplos níveis. Essa tendência reforça que o avanço da área está associado à integração de estratégias complementares de modelagem, ampliando o potencial de aplicação dessas arquiteturas em problemas caracterizados por elevada complexidade temporal.

xLSTM

Entre as arquiteturas híbridas de nova geração, a *Extended Long Short-Term Memory* (xLSTM) representa uma evolução das redes LSTM tradicionais, desenvolvida para ampliar sua capacidade de modelar dependências temporais de longo alcance. Enquanto a arquitetura LSTM convencional utiliza células de memória controladas por portas de entrada, esquecimento e saída, a xLSTM incorpora mecanismos de memória exponencial (*Exponential Gating*) e estruturas de memória escaláveis (*Scalar Memory* e *Matrix Memory*), permitindo preservar informações relevantes por intervalos temporais mais extensos e representar relações complexas com maior eficiência computacional. Essas modificações reduzem limitações relacionadas à propagação de informações em sequências muito longas e favorecem a aprendizagem de padrões temporais persistentes, sem descaracterizar a estrutura recorrente

que tornou as LSTM amplamente empregadas na previsão de séries temporais [1].

Segundo Alharthi e Mahmood [1], a arquitetura xLSTM apresenta desempenho competitivo em tarefas de previsão de longo horizonte temporal ao ampliar a capacidade de retenção das informações relevantes sem aumentar significativamente a complexidade estrutural do modelo. Essa característica torna a arquitetura particularmente adequada para aplicações nas quais dependências temporais distantes exercem influência significativa sobre as previsões, reduzindo limitações observadas em arquiteturas recorrentes convencionais.

Mais do que representar uma substituição das LSTM tradicionais, a xLSTM evidencia uma tendência recente de aperfeiçoamento das arquiteturas recorrentes por meio da incorporação de mecanismos especializados de memória. Essa evolução demonstra que o avanço das arquiteturas híbridas não ocorre apenas pela integração de diferentes paradigmas de aprendizagem, como CNN, Transformers e modelos estatísticos, mas também pelo refinamento dos próprios componentes recorrentes, ampliando sua capacidade de adaptação às exigências impostas por séries temporais cada vez mais extensas e complexas.

2.3. Principais Aplicações das Arquiteturas Híbridas em Deep Learning

Tabela 1. Principais domínios de aplicação das arquiteturas híbridas em Deep Learning para predição de séries temporais

Domínio de aplicação	Principais arquiteturas	Exemplos encontrados na	Principais contribuições
----------------------	-------------------------	-------------------------	--------------------------

		literatura	
Sistemas energéticos	CNN-LSTM, CNN-LSTM-Transformer, Transformer	Geração de energia solar, previsão de carga elétrica	Maior precisão na previsão da geração e otimização do gerenciamento energético.
Mercado financeiro e criptomoedas	ARIMA-LSTM, LSTM-GARCH, LSTM-Transformer, CNN-Transformer, Temporal Fusion Transformer	Bitcoin, Ethereum, índices financeiros e ações	Melhor modelagem da volatilidade, das relações não lineares e das dependências temporais de longo prazo.
Meteorologia	CNN-LSTM	Temperatura, precipitação e previsão climática	Extração eficiente de padrões locais e melhoria da previsão de eventos meteorológicos.
Transporte e mobilidade	Bi-LSTM, Transformer, Autoencoder	Demanda de carregamento de veículos elétricos, consumo de combustíveis	Representação de séries temporais complexas e redução dos efeitos do ruído nos dados.
Aplicações multivariadas	xLSTM, CNN-LSTM-Transformer, Transformer	Séries temporais multivariadas em diferentes domínios	Maior capacidade de generalização, escalabilidade e previsão de longo horizonte temporal.

Fonte: Elaborado pela autora (2026)

A diversidade de arquiteturas híbridas desenvolvidas para previsão de séries temporais reflete a ampla variedade de problemas encontrados em aplicações reais. Conforme sintetizado na Tabela 1, essas arquiteturas vêm sendo empregadas em domínios como sistemas energéticos, mercados financeiros, meteorologia, transporte e aplicações multivariadas, cujas séries apresentam características distintas quanto à sazonalidade, volatilidade, dimensionalidade e nível de complexidade. Essa diversidade evidencia que a escolha da arquitetura mais adequada depende das propriedades dos dados e dos objetivos da aplicação, e não da existência de um modelo universalmente superior [1], [2], [6], [9], [17], [18], [23].

Nos sistemas energéticos e meteorológicos, as séries temporais são fortemente influenciadas por fatores climáticos, sazonalidade e variáveis ambientais interdependentes, exigindo modelos capazes de representar simultaneamente padrões locais e dependências temporais de diferentes escalas. Nesse contexto, arquiteturas como CNN-LSTM e CNN-LSTM-Transformer têm apresentado resultados expressivos na previsão da geração de energia solar, da carga elétrica e de variáveis meteorológicas, combinando extração automática de características, modelagem sequencial e mecanismos de atenção para aumentar a precisão das previsões e apoiar o gerenciamento de sistemas energéticos inteligentes [11], [18].

Em mercados financeiros e criptomoedas, por outro lado, predominam séries temporais caracterizadas por elevada volatilidade, comportamento não estacionário e relações não lineares. Essas características favorecem a utilização de arquiteturas híbridas que integram modelos estatísticos, redes recorrentes e

mecanismos de atenção, como ARIMA-LSTM, LSTM-GARCH, LSTM-Transformer e *Temporal Fusion Transformer*, ampliando a capacidade de representar tendências, volatilidade e dependências temporais de longo alcance. Estratégias semelhantes também vêm sendo empregadas em aplicações relacionadas ao transporte inteligente e à mobilidade elétrica, nas quais modelos que combinam *Denoising Autoencoder*, Bi-LSTM e Transformer demonstraram maior robustez para lidar com dados multivariados, ruidosos e sujeitos a mudanças frequentes de comportamento [2], [6], [10], [16], [17].

De forma mais ampla, observa-se que o avanço das arquiteturas híbridas está diretamente relacionado ao crescimento das aplicações multivariadas, nas quais informações provenientes de diferentes fontes precisam ser processadas de maneira integrada. Arquiteturas como CNN-LSTM-Transformer, *Temporal Fusion Transformer* e xLSTM destacam-se por incorporar mecanismos complementares de aprendizagem capazes de representar relações locais, dependências temporais de curto e longo prazo e interações globais entre variáveis. Em conjunto, essas evidências reforçam que a evolução da previsão de séries temporais está associada ao desenvolvimento de arquiteturas cada vez mais especializadas e adaptáveis às características de cada problema, consolidando a integração de diferentes mecanismos de aprendizagem como a principal tendência da área [1], [2], [9], [18], [23].

2.4. Vantagens e Limitações das Arquiteturas Híbridas em Deep Learning

A obtenção de previsões mais precisas representa apenas um dos critérios utilizados para avaliar o desempenho das arquiteturas

híbridas em *Deep Learning*. À medida que esses modelos evoluíram, aspectos como capacidade de modelagem, custo computacional, interpretabilidade e generalização passaram a exercer papel igualmente relevante na seleção da arquitetura mais adequada. Conforme sintetizado na Tabela 2, as vantagens e limitações desses modelos devem ser analisadas de forma integrada, considerando os compromissos estabelecidos entre desempenho preditivo, complexidade computacional e aplicabilidade em cenários reais. Sob essa perspectiva, a escolha da arquitetura deixa de depender exclusivamente da precisão alcançada e passa a considerar as características do problema, a natureza dos dados disponíveis e os requisitos específicos de cada aplicação [1], [2], [6], [9], [13], [14], [18], [23].

Tabela 2. Principais vantagens e limitações das arquiteturas híbridas em Deep Learning para previsão de séries temporais

Aspecto analisado	Vantagens identificadas	Limitações identificadas
Precisão preditiva	Melhor desempenho em comparação com modelos estatísticos e arquiteturas individuais.	O desempenho depende fortemente da arquitetura utilizada e do conjunto de dados.
Modelagem temporal	Capacidade de representar dependências de curto e longo prazo.	Algumas arquiteturas apresentam dificuldades em séries muito longas ou altamente não estacionárias.
Extração de características	Aprendizagem automática de padrões complexos, reduzindo a necessidade de engenharia manual de atributos.	Elevada dependência da qualidade e quantidade dos dados disponíveis.

Integração de modelos	Combinação de diferentes mecanismos de aprendizagem aumenta a robustez e a capacidade de generalização.	A integração aumenta a complexidade arquitetural e dificulta a otimização dos hiperparâmetros.
Custo computacional	Melhor aproveitamento das características de cada arquitetura.	Necessidade de maior poder computacional e maior tempo de treinamento.
Interpretabilidade	Modelos recentes incorporam mecanismos de atenção e estimativas de incerteza, aumentando parcialmente a interpretabilidade.	A maioria das arquiteturas continua sendo considerada uma "caixa-preta", dificultando a interpretação das previsões.
Generalização	Boa adaptação a diferentes domínios de aplicação.	Ainda existem dificuldades de generalização entre diferentes bases de dados e cenários reais.

Fonte: Elaborado pela autora (2026)

As características das séries temporais exercem influência decisiva sobre o desempenho das arquiteturas híbridas. Séries marcadas por elevada não linearidade, múltiplas escalas temporais, sazonalidade ou grande quantidade de variáveis exigem modelos capazes de integrar diferentes mecanismos de aprendizagem para representar aspectos complementares da dinâmica dos dados. Nessas situações, a combinação entre modelos estatísticos, redes convolucionais, arquiteturas recorrentes e mecanismos de atenção amplia significativamente a capacidade preditiva quando comparada ao emprego isolado dessas abordagens. Em contrapartida, problemas menos complexos nem sempre justificam a utilização de arquiteturas compostas por múltiplos componentes, evidenciando

que a escolha do modelo deve considerar o equilíbrio entre a complexidade da série temporal e os benefícios efetivamente obtidos com a integração das diferentes estratégias de aprendizagem [2], [6], [9], [18], [23].

Entretanto, a incorporação de mecanismos complementares de aprendizagem aumenta a complexidade estrutural dos modelos e amplia a quantidade de parâmetros a serem ajustados durante o treinamento. Como consequência, cresce a demanda por recursos computacionais, memória e tempo de processamento, além de tornar mais sensível a definição de hiperparâmetros, estratégias de otimização e procedimentos de validação. Esses fatores podem limitar a aplicação de arquiteturas híbridas em ambientes com infraestrutura computacional restrita ou em cenários que exigem respostas em tempo reduzido, tornando necessário ponderar os ganhos obtidos em desempenho preditivo em relação ao custo computacional associado à sua implementação [9], [14], [18], [23].

Além da complexidade computacional, outro desafio relevante refere-se à interpretabilidade dos modelos. A integração de diferentes mecanismos de aprendizagem dificulta a identificação da contribuição individual de cada componente para a geração das previsões, reduzindo a transparência do processo decisório. Essa limitação é particularmente importante em aplicações nas quais a confiabilidade das previsões constitui requisito fundamental, como sistemas financeiros, energéticos e de apoio à decisão. Nesse contexto, mecanismos de atenção interpretáveis e abordagens probabilísticas, como aquelas baseadas em Inferência Bayesiana, representam alternativas promissoras para ampliar a compreensão do comportamento dos modelos sem comprometer seu desempenho preditivo [2], [13], [18], [23].

Outro aspecto diretamente relacionado ao desempenho desses modelos diz respeito à sua capacidade de generalização. Embora a literatura apresente resultados consistentes em diferentes domínios de aplicação, o desempenho das arquiteturas híbridas permanece fortemente condicionado às características dos dados utilizados em seu treinamento e validação. Alterações na distribuição das séries temporais, no conjunto de variáveis disponíveis ou no horizonte de previsão frequentemente exigem adaptações na arquitetura, novos processos de treinamento e ajustes de hiperparâmetros. Essas evidências reforçam que modelos desenvolvidos para um contexto específico nem sempre podem ser transferidos diretamente para outros cenários, destacando a importância de estratégias de adaptação e validação compatíveis com cada problema [1], [14], [23].

Em conjunto, essas evidências demonstram que a avaliação das arquiteturas híbridas não deve estar fundamentada em um único indicador de desempenho, mas no equilíbrio entre precisão preditiva, capacidade de modelagem, custo computacional, interpretabilidade e generalização. Embora modelos compostos por múltiplos mecanismos de aprendizagem ampliem significativamente a capacidade de representar séries temporais complexas, sua adoção envolve compromissos que variam conforme as características dos dados, os objetivos da aplicação e os recursos computacionais disponíveis. Assim, mais do que identificar uma arquitetura universalmente superior, a literatura aponta para o desenvolvimento de modelos cada vez mais especializados, eficientes e adaptáveis às demandas específicas de diferentes cenários de previsão de séries temporais.

3. CONSIDERAÇÕES FINAIS

A presente revisão analisou a evolução das arquiteturas híbridas em *Deep Learning* aplicadas à previsão de séries temporais, evidenciando como a integração entre diferentes mecanismos de aprendizagem tem ampliado a capacidade de modelar séries caracterizadas por elevada complexidade, não linearidade e múltiplas escalas temporais. A análise da literatura demonstrou que a evolução dessas arquiteturas não está associada à simples substituição de modelos previamente consolidados, mas à incorporação progressiva de mecanismos complementares — como redes convolucionais, arquiteturas recorrentes, modelos estatísticos, mecanismos de atenção, inferência probabilística e estruturas avançadas de memória — capazes de explorar diferentes aspectos da dinâmica temporal em uma mesma estrutura computacional.

Os estudos analisados evidenciam que não existe uma arquitetura universalmente superior para todos os problemas de previsão de séries temporais. O desempenho dos modelos permanece fortemente condicionado às características dos dados, ao horizonte de previsão, à quantidade de variáveis disponíveis e aos requisitos específicos de cada aplicação. Nesse contexto, a combinação entre diferentes mecanismos de aprendizagem mostrou-se mais relevante do que a adoção isolada de arquiteturas cada vez mais complexas, permitindo ampliar a capacidade de representação dos dados e favorecer a obtenção de previsões mais robustas em domínios como sistemas energéticos, mercados financeiros, meteorologia, transporte e aplicações multivariadas.

A revisão também demonstrou que os avanços recentes da área extrapolam o aumento da precisão preditiva. Aspectos relacionados à interpretabilidade, à capacidade de generalização, à eficiência computacional e à quantificação das incertezas passaram a ocupar

posição central no desenvolvimento de novas arquiteturas híbridas. Nesse cenário, modelos como *Temporal Fusion Transformer*, xLSTM, CNN-Transformer e arquiteturas que incorporam *autoencoders*, memória bidirecional e inferência bayesiana representam uma nova geração de soluções, desenvolvidas para integrar mecanismos especializados de aprendizagem e adaptar o processamento às características dos dados, ampliando simultaneamente a robustez, a flexibilidade e a confiabilidade das previsões.

Apesar dos avanços observados, permanecem desafios importantes relacionados ao elevado custo computacional, à dependência de grandes volumes de dados para treinamento, à limitada interpretabilidade de modelos altamente complexos e às dificuldades de generalização entre diferentes contextos de aplicação. Esses aspectos evidenciam que a evolução das arquiteturas híbridas deve ser compreendida como um processo contínuo de equilíbrio entre desempenho preditivo, eficiência computacional, transparência e capacidade de adaptação, e não apenas como uma busca por modelos estruturalmente mais sofisticados.

Como limitação, este trabalho fundamentou-se exclusivamente na análise da literatura científica, não contemplando avaliações experimentais ou comparações quantitativas entre as arquiteturas revisadas. Ainda assim, a diversidade dos estudos analisados permitiu identificar padrões consistentes de evolução, aplicações recorrentes e desafios que caracterizam o estado atual da pesquisa em previsão de séries temporais.

Por fim, observa-se que as tendências atuais apontam para o desenvolvimento de arquiteturas híbridas cada vez mais

especializadas, capazes de integrar múltiplos mecanismos de aprendizagem, incorporar informações provenientes de diferentes fontes de dados, adaptar-se dinamicamente às mudanças das séries temporais e reduzir a necessidade de intervenção humana durante o treinamento e a atualização dos modelos. Nesse contexto, pesquisas futuras poderão explorar estratégias de aprendizagem adaptativa, integração de dados multimodais, otimização automática de hiperparâmetros e técnicas de Inteligência Artificial Explicável, contribuindo para o desenvolvimento de modelos mais eficientes, transparentes, escaláveis e aplicáveis a cenários reais de elevada complexidade. Dessa forma, espera-se que esta revisão contribua para a consolidação do conhecimento sobre arquiteturas híbridas em *Deep Learning* e ofereça subsídios para o desenvolvimento de novas pesquisas e soluções preditivas voltadas aos desafios emergentes da previsão de séries temporais.

REFERÊNCIAS BIBLIOGRÁFICAS

1. ALHARTHI, Musleh; MAHMOOD, Ausif. **xLSTMTime: Long-term Time Series Forecasting With xLSTM**. *Preprints*, 2024. DOI: 10.20944/preprints202407.1219.v1. Disponível em: <https://doi.org/10.20944/preprints202407.1219.v1>. Acesso em: 16 jun. 2026.
2. DADI, Ramesh; NARSINGOJU, Chandramouli; SRAVANI, K.; SRIDHAR, K.; MOUNIKA, Vemula; SHABANA. **Temporal Fusion and Hybrid ARIMA-LSTM Models for Robust Time Series Forecasting: A Case Study on Ethereum Price Dynamics**. In: INTERNATIONAL CONFERENCE ON COGNITIVE ROBOTICS AND INTELLIGENT SYSTEMS (ICC-ROBINS), 2., 2025. *Proceedings [...]* Piscataway: IEEE, 2025. p. 622–627. DOI:

<https://doi.org/10.1109/ICC-ROBINS64345.2025.11086123>.

Disponível em: <https://ieeexplore.ieee.org/document/11086123>.

Acesso em: 16 jun. 2026.

3. DAVID S., Alex; BELINDA M. J., Carmel Mary; NAVEENA N., Ruth; BEGUM, Almas; HEMALATHA, D. **Bitcoin Price Forecasting: A Comparison of LSTM and Feedforward Neural Network**. In: 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE). Piscataway: IEEE, 2023. DOI:

<https://doi.org/10.1109/RMKMATE59243.2023.10368720>.

Disponível em: <https://ieeexplore.ieee.org/document/10368720>.

Acesso em: 16 jun. 2026.

4. DIMITRIADOU, Athanasia; GREGORIOU, Andros. **Predicting Bitcoin Prices Using Machine Learning**. *Entropy*, Basel, v. 25, n. 5, art. 777, 2023. DOI: <https://doi.org/10.3390/e25050777>. Disponível em: <https://www.mdpi.com/1099-4300/25/5/777>.

Acesso em: 16 jun. 2026.

5. GANAIE, M. A.; HU, Minghui; MALIK, A. K.; TANVEER, M.; SUGANTHAN, P. N. **Ensemble deep learning: a review**. *Engineering Applications of Artificial Intelligence*, v. 115, art. 105151, 2022. DOI: <https://doi.org/10.1016/j.engappai.2022.105151>.

Disponível em: <https://www.sciencedirect.com/science/article/pii/S095219762200269X>. Acesso em: 16 jun. 2026.

6. GAO, Zidi; HE, Yiwen; KURUOGLU, Ercan Engin. *A Hybrid Model Integrating LSTM and GARCH for Bitcoin Price Prediction*. In:

2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP). Gold Coast, Australia: IEEE, 2021.

DOI: <https://doi.org/10.1109/MLSP52302.2021.9596429>.

Disponível em: <https://ieeexplore.ieee.org/document/9596429>.

Acesso em: 16 jun. 2026.

7. GASMI, Laid; KABOU, Salheddine; LAICHE, Nabil; NICHANI, Rabia. **Time series forecasting using deep learning hybrid model (ARIMA-LSTM).** Studies in Engineering and Exact Sciences, Curitiba, v. 5, n. 2, p. 1–15, 2024. DOI: <https://doi.org/10.54021/seesv5n2-125>. Disponível em: <https://periodicos.newsciencepubl.com/sees/article/view/125>. Acesso em: 16 jun. 2026.

8. JAVADIMASOUDIAN, Soudeh; KATHURIA, Paras M.; GOWDA, Nisha S.; KHAN, Talha Ali. **Bitcoin Price Prediction using Long Short Term Memory Neural Networks.** In: 2022 International Conference Automatics and Informatics (ICAI). Varna, Bulgaria: IEEE, 2022. p. 69–72. DOI: <https://doi.org/10.1109/ICAI55857.2022.9960055>. Disponível em: <https://ieeexplore.ieee.org/document/9960055>. Acesso em: 16 jun. 2026.

9. KABIR, Md R.; BHADRA, Dipayan; RIDOY, Moinul; MILANOVA, Mariofanna. **LSTM-Transformer-Based Robust Hybrid Deep Learning Model for Financial Time Series Forecasting.** *Sci*, v. 7, n. 1, art. 7, 2025. DOI: <https://doi.org/10.3390/sci7010007>. Disponível em: <https://www.mdpi.com/2413-4155/7/1/7>. Acesso em: 16 jun. 2026.

10. KOOHFAR, Sahar; WOLDEMARIAM, Wubeshet. ***Time Series Forecasting Using a Hybrid Deep Learning Method: A Bi-LSTM Embedding Denoising Auto Encoder Transformer***. CoRR (Computing Research Repository), arXiv:2509.17165, 2025. Disponível em: <https://arxiv.org/abs/2509.17165>. Acesso em: 16 jun. 2026.
11. LI, Bangyu; QIAN, Yang. ***Weather Prediction Using CNN-LSTM for Time Series Analysis: A Case Study on Delhi Temperature Data***. 2025. Disponível em: https://www.researchgate.net/publication/390649734_Weather_Prediction_Using_CNN-LSTM_for_Time_Series_Analysis_A_Case_Study_on_Delhi_Temperature_Data. Acesso em: 16 jun. 2026.
12. LU, Nan; OUYANG, Quan; LI, Yang; ZOU, Changfu. ***Electrical Load Forecasting Model Using Hybrid LSTM Neural Networks with Online Correction***. arXiv, 2024. Disponível em: <https://arxiv.org/abs/2403.03898>. Acesso em: 16 jun. 2026.
13. NGHIEM, Thi-Lich; LE, Viet-Duc; LE, Thi-Lan; MARÉCHAL, Pierre; DELAHAYE, Daniel; VIDOSAVLJEVIC, Andrija. ***Applying Bayesian Inference in a Hybrid CNN-LSTM Model for Time Series Prediction***. In: *International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*. Phu Quoc, Vietnam, 2022. HAL Id: hal-04056437, 2023. Disponível em: <https://hal.science/hal-04056437>. Acesso em: 16 jun. 2026.
14. NICHANI, Rabia; GASMI, Laid; KABOU, Salheddine; LAICHE, Nabil. ***Novel insights on the comparative study between LSTM and Transformer models for financial time series***

prediction. Journal of Engineering and Technology for Industrial Applications, Manaus, v. 11, n. 55, p. 212–221, set./out. 2025. DOI: <https://doi.org/10.5935/jetia.v11n55.2658>. Disponível em: <https://itegam-jetia.org/journal/index.php/jetia/article/view/2658>. Acesso em: 16 jun. 2026.

15. RAGHAVENDRAN, Ch. V.; MOULI, K. Chandra; HAJARI, Manu; REDDY, A. Anil Kumar; REDDY, V. Shiva Narayana; PERUMALLA, Janaki Ramulu. **Ethereum Price Prediction Using Time Series and Deep Learning Techniques.** *Applied Computational Intelligence and Soft Computing*, v. 2025, Article ID 6167862, 13 p., 2025. DOI: <https://doi.org/10.1155/acis/6167862>. Disponível em: <https://doi.org/10.1155/acis/6167862>. Acesso em: 16 jun. 2026.
16. RAMANI, K.; VENKATACHAKRAVARTHI, P.; JAHNAVI, M.; MEGHANATH, P.; JAGADEESH REDDY, P.; IMRAN, S. K. **Prediction of Bitcoin Price through LSTM, ARIMA, XGBoost, Prophet and Sentiment Analysis on Dynamic Streaming Data.** In: 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS). Piscataway: IEEE, 2023. DOI: <https://doi.org/10.1109/ICACCS57279.2023.10113014>. Disponível em: <https://ieeexplore.ieee.org/document/10113014>. Acesso em: 16 jun. 2026.
17. RANJBAR, Mahmoud; RAHIMZADEH, Mohammad. **Advancing Gasoline Consumption Forecasting: A Novel Hybrid Model Integrating Transformers, LSTM, and CNN.** 2025. Disponível em:

https://www.researchgate.net/publication/389424018_Advancing_Gasoline_Consumption_Forecasting_A_Novel_Hybrid_Model_Integrating_Transformers_LSTM_and_CNN. Acesso em: 16 jun. 2026.

18. SALMAN, Diaa; DIREKOGLU, Cem; KUSAF, Mehmet; FAHRIOGLU, Murat. **Hybrid deep learning models for time series forecasting of solar power.** *Neural Computing and Applications*, v. 36, p. 9095–9112, 2024. DOI: <https://doi.org/10.1007/s00521-024-09558-5>. Disponível em: <https://link.springer.com/article/10.1007/s00521-024-09558-5>. Acesso em: 16 jun. 2026.
19. SCANLON, Brian; QUILLE, Keith; JAISWAL, Rajesh. **Optimizing Bitcoin Price Prediction: Multivariate LSTM Triumphs.** *International Journal of Combinatorial Optimization Problems and Informatics*, v. 16, n. 1, p. 164–176, Jan./Apr. 2025. DOI: <https://doi.org/10.61467/2007.1558.2025.v16i1.552>. Disponível em: <https://doi.org/10.61467/2007.1558.2025.v16i1.552>. Acesso em: 16 jun. 2026.
20. SINGH, Japjeet; THULASIRAM, Ruppa; THAVANESWARAN, Aerambamoorthy. **LSTM based algorithmic trading model for Bitcoin.** In: IEEE Symposium Series on Computational Intelligence (SSCI), 2022. *Proceedings [...]* Piscataway: IEEE, 2022. p. 344–351. DOI: <https://doi.org/10.1109/SSCI51031.2022.10022021>. Disponível em: <https://ieeexplore.ieee.org/document/10022021>. Acesso em: 16 jun. 2026.

21. SRIMAN, B.; TAMIL INIYAL, T.; THASMIYA, J.; TAARIQ ZIYAADH, J.; SRI RANJANI, I.; THOMAS, S. Teju. **Price Prediction of Bitcoin and Ethereum – A Machine Learning Approach**. In: 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF). Piscataway: IEEE, 2023. DOI: <https://doi.org/10.1109/ICECONF57129.2023.10084003>. Disponível em: <https://ieeexplore.ieee.org/document/10084003>. Acesso em: 16 jun. 2026.
22. TODOROVIC, Mihailo; PETROVIC, Aleksandar; TOSKOVIC, Ana; ZIVKOVIC, Miodrag; JOVANOVIC, Luka; BACANIN, Nebojsa. Multivariate **Bitcoin price prediction based on LSTM tuned by hybrid reptile search algorithm**. In: INTERNATIONAL CONFERENCE ON ADVANCED TECHNOLOGIES, SYSTEMS AND SERVICES IN TELECOMMUNICATIONS (TELSIKS), 16., 2023. *Proceedings [...]* Piscataway: IEEE, 2023. p. 195–198. DOI: <https://doi.org/10.1109/TELSIKS57806.2023.10316108>. Disponível em: <https://ieeexplore.ieee.org/document/10316108>. Acesso em: 16 jun. 2026.
23. TU, Tiantian. **Bridging Short- and Long-Term Dependencies: A CNN-Transformer Hybrid for Financial Time Series Forecasting**. arXiv, 2025. arXiv:2504.19309. Disponível em: <https://arxiv.org/abs/2504.19309>. Acesso em: 16 jun. 2026.