

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA
CIÊNCIAS DA COMPUTAÇÃO

Caio Prá Silva

**Preenchimento de Dados Faltantes em Registros de Poços de Petróleo usando Redes
Neurais**

Florianópolis
10 de dezembro de 2025

Caio Prá Silva

Preenchimento de Dados Faltantes em Registros de Poços de Petróleo usando Redes Neurais

Trabalho de Conclusão de Curso submetido ao
Curso de Graduação em Ciências da Compu-
tação do Centro Tecnológico da Universidade
Federal de Santa Catarina como requisito para
obtenção do título de Bacharel em Ciências da
Computação.

Orientador: Prof. Mauro Roisenberg, Dr.

Coorientador: Bruno Barbosa Rodrigues, Dr.

Florianópolis
10 de dezembro de 2025

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Silva, Caio Prá

Preenchimento de dados faltantes em registros de poços de petróleo usando redes neurais / Caio Prá Silva ; orientador, Mauro Roisenberg, coorientador, Bruno Barbosa Rodrigues, 2025.

121 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Ciências da Computação, Florianópolis, 2025.

Inclui referências.

1. Ciências da Computação. 2. Machine Learning. 3. Preenchimento de dados faltantes. 4. Deep Learning. 5. Petrofísica. I. Roisenberg, Mauro. II. Rodrigues, Bruno Barbosa. III. Universidade Federal de Santa Catarina. Graduação em Ciências da Computação. IV. Título.

Caio Prá Silva

Preenchimento de Dados Faltantes em Registros de Poços de Petróleo usando Redes Neurais

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de Bacharel em Ciências da Computação e aprovado em sua forma final pelo curso de Graduação em Ciências da Computação.

Florianópolis, 10 de dezembro de 2025.

Banca Examinadora:

Prof. Mauro Roisenberg, Dr.
Orientador
Universidade Federal de Santa Catarina

Prof^a Jerusa Marchi, Dra.
Avaliador
Universidade Federal de Santa Catarina

Rafael de Santiago, Dr.
Avaliador
Universidade Federal de Santa Catarina

AGRADECIMENTOS

Agradeço a Deus por todas as oportunidades e bênçãos recebidas durante os anos em que participei da graduação e à Nossa Senhora, que iluminou meu caminho. Agradeço a UFSC e a Petrobras pela oportunidade de realizar esse trabalho, fornecendo a infraestrutura, dados e auxílio sempre que necessário através de seus profissionais.

Nada disso teria sido possível sem o apoio e incentivo do meu grande amor, Maria Isabel, que, em todos esses anos, me deu todo o suporte que precisava, foi meu amparo e não deixou que eu desistisse quando pensei que não daria conta, me dando coragem para enfrentar o que fosse necessário por ela. Minha família também foi crucial para que eu chegasse até aqui, com incentivo aos estudos ao longo de toda a minha vida, mostrando que a educação é um dos maiores bens e legados que podemos ter. Ao meu pai, por tudo que fez por mim ao longo da vida, seja me levando para todo lugar ou comprando o que eu precisava ou queria só de perceber; minha mãe, que sempre me auxiliou em todos os momentos, desde quando tive que aprender a cozinhar sozinho por morar sozinho, até ser minha conselheira; minha irmã, de quem tenho tanto orgulho, que foi motivo de muitas risadas e descontração quando mais precisei e sempre foi minha parceira; minha avó, que mesmo com 90 anos não se cansou de perguntar e tentar entender o que seu neto, que "faz pesquisa de Inteligência Artificial para a Petrobras", fazia e contava para todos, ela foi como uma segunda mãe na minha vida e até hoje tenho tanto zelo.

Diversos professores foram muito marcantes durante a graduação, mas agradeço em especial ao professor Mauro pela orientação nesse trabalho, ser inspiração em minha forma de trabalho e pesquisa, por todas as ideias e auxílio durante o percurso e ter me concedido a oportunidade de trabalhar junto de outros estudantes e professores no L3C (*Connectionism and Cognitive Science Lab*). Fica nítido que o que torna a UFSC uma das melhores universidades do país é a constituição de seu corpo docente, com diversos educadores e pesquisadores referências nacionais e mundiais em suas áreas de pesquisa, que nas mais adversas condições, priorizam a educação e formação profissional e acadêmica de excelência de seus alunos. Agradeço também ao Bruno pela coorientação e ajuda fundamental na área da petrofísica, crucial para o resultado encontrado nesse trabalho.

Por último, devo mencionar todos amigos que estiveram ao meu lado nessa longa caminhada, mas que fizeram ela mais leve e divertida. Em especial aos meus cunhados, Maria Elisa, Lucas e Maria Paula, que estiveram comigo em todos momentos do curso. Ao meu "irmão" da vida Martin, que nunca deixou que eu desanimasse, que tanto me animou e comemorou comigo cada passo ao longo dos anos; meus colegas de apartamento Arthur Santos e Arthur Vinícius pelo companheirismo nesses anos todos. Além deles, também os diversos amigos que fiz na graduação e que muitos são modelos para mim, Fillipi, João Paulo, João Victor, Leonardo, Rafael e diversos outros, os quais sua amizade e parceria foram extremamente valiosos para mim e foram verdadeiros "irmãos de farda".

RESUMO

Este trabalho aborda o preenchimento de dados faltantes em registros de poços de petróleo (*well logs*) utilizando redes neurais profundas, com foco na predição da curva de velocidade de onda cisalhante (VS). A completude e precisão desses registros são fundamentais para a caracterização petrofísica e modelagem de reservatórios, mas sua aquisição é complexa e sujeita a falhas, resultando em lacunas de dados, informações ruidosas ou ausência completa da curva. Métodos tradicionais de imputação, como interpolações e regressões lineares, mostram-se limitados diante da natureza não linear e altamente variável dos dados geológicos.

A pesquisa desenvolveu e comparou quatro arquiteturas de redes neurais profundas: uma rede *LSTM* utilizada como *baseline*, o modelo hierárquico *N-HITS* (*Neural Hierarchical Interpolation for Time Series*) e duas variações de modelos baseados em mecanismos de *Self-Attention*. A metodologia envolveu a análise de similaridade geológica entre poços, validação cruzada estratificada e avaliação quantitativa em poços com dados de referência reais fornecidos pela Petrobras.

Entre os modelos avaliados, o modelo baseado em *Self-Attention* apresentou os resultados mais promissores, demonstrando boa estabilidade e capacidade de generalização entre poços geologicamente distintos. Os resultados indicam que mecanismos de atenção, quando devidamente calibrados, constituem uma abordagem tecnicamente viável e eficaz para a predição de propriedades petrofísicas, estabelecendo um caminho promissor para futuras aplicações de aprendizado profundo na análise de *well logs*.

Palavras-chave: *Well logs*. *Machine Learning* (ML). Redes Neurais. Preenchimento de dados faltantes. Petrofísica.

ABSTRACT

This work addresses the problem of missing data imputation in petroleum well logs using deep neural networks, focusing on the prediction of the shear wave velocity curve (VS). The completeness and accuracy of well log data are essential for petrophysical characterization and reservoir modeling; however, their acquisition is complex and prone to failures, resulting in data gaps, noisy measurements or entire logs missing. Traditional imputation methods, such as interpolation and linear regression, are limited in capturing the nonlinear and highly variable nature of geological data.

The study developed and compared four deep neural network architectures: a *Long Short-Term Memory* (LSTM) network used as a baseline, the hierarchical *N-HITS* (*Neural Hierarchical Interpolation for Time Series*) model, and two variations of *Self-Attention*-based models. The methodology included geological similarity analysis between wells, stratified cross-validation, and quantitative evaluation on real well data provided by Petrobras.

Among the evaluated models, the *Self-Attention*-based model achieved the most promising results, demonstrating strong stability and generalization across geologically distinct wells. The findings indicate that attention-based mechanisms, when properly tuned, constitute an effective and technically sound approach for petrophysical property prediction, highlighting a promising direction for the application of deep learning in well log analysis.

Keywords. Well logs. Machine Learning (ML). Neural Networks. Missing data imputation. Petrophysic.

LISTA DE FIGURAS

Figura 1 – Exemplo de ferramenta para <i>logging</i> , demonstração de operação de coleta de dados e exemplo de curva de saída. Fonte: (BIDA; AL-SUDANI, 2021)	27
Figura 2 – Sistema petrolífero simplificado. Fonte: (GARCIA, 2022)	29
Figura 3 – Gráfico para volume de argila não linear. Fonte: (INTERNATIONAL, 1985)	31
Figura 4 – Exemplo de representação de um Perceptron. Adaptado de: (AHMAD, 2024)	36
Figura 5 – Estrutura interna de uma unidade <i>LSTM</i> . Fonte: (QUANTCONNECT, 2024)	41
Figura 6 – Estrutura interna de uma <i>GRU</i> . Fonte: (BRAZIL, n.d.)	42
Figura 7 – Arquitetura Transformer. Fonte: (VASWANI et al., 2023)	46
Figura 8 – <i>Heatmap</i> de distância estatística. Fonte: Autor.	60
Figura 9 – <i>Heatmap</i> de distância correlação. Fonte: Autor.	61
Figura 10 – <i>Heatmap</i> de distância <i>DTW</i> . Fonte: Autor.	63
Figura 11 – <i>Heatmap</i> de distância <i>DTW</i> . Fonte: Autor.	64
Figura 12 – <i>Heatmap</i> de distância combinada. Fonte: Autor.	65
Figura 13 – Projeção MDS da Distância Estatística (Wasserstein). Stress = 0.0009 (quase perfeito). Agrupa poços com distribuições similares de VP, RHO, POROSIDADE e SATURACAO. Fonte: Autor.	68
Figura 14 – Projeção MDS da Distância de Correlação. Stress = 0.0374 (excelente). Agrupa poços com padrões de correlação entre features similares. O Poço 1 apresenta-se extremamente isolado, indicando relações entre features fundamentalmente diferentes dos demais. Fonte: Autor.	68
Figura 15 – Projeção MDS da Distância DTW (Dynamic Time Warping). Stress = 0.0101 (excelente). Agrupa poços com sequências estratigráficas similares ao longo da profundidade. Fonte: Autor.	69
Figura 16 – Projeção MDS da Distância no Espaço de Features. Stress = 0.1573 (aceitável). Agrupa poços com propriedades médias globais similares. A separação clara entre dois grupos fornece evidência visual forte de tipos distintos de reservatório/formação. Fonte: Autor.	70
Figura 17 – Projeção MDS da Distância Combinada entre Poços. Cada ponto representa um poço. A proximidade espacial reflete similaridade geológica multidimensional integrando todos os aspectos analisados. Stress = 0.0515 (excelente). Fonte: Autor.	71
Figura 18 – Plot de DEPT x VS (real e previsto)	82
Figura 19 – Plot de VS real x VS previsto	83

Figura 20 – Distribuição de erros de predição do modelo base por poço. Nota-se que apenas o poço 2 apresenta distribuição centrada próxima a zero com baixa dispersão, enquanto os demais poços mostram vieses significativos.	88
Figura 21 – Gráficos de dispersão das predições versus valores reais do modelo base. Os poços 3 e 4 ($R^2 = 0,633$ e $0,553$) apresentam melhor alinhamento com a linha de predição perfeita, enquanto Well 1 ($R^2 = -1,662$) mostra dispersão extrema.	89
Figura 22 – Séries temporais de VS real versus predito do base. Well 0 mostra deslocamento sistemático (bias) entre predições e valores reais, enquanto Wells 1 e 2 apresentam falhas em capturar variações abruptas. Wells 3 e 4 demonstram melhor acompanhamento das tendências.	90
Figura 23 – Predições de VS para poços sem ground truth (base). Poço sem VS 1 apresenta predições mais conservadoras (menor variabilidade), enquanto Poço sem VS 2 mostra maior amplitude de variação, refletindo possivelmente maior heterogeneidade geológica.	91
Figura 24 – Distribuição de erros de predição do modelo de Self-Attention avançado por poço. Comparado ao modelo base, observa-se distribuições mais centradas em zero, especialmente para Wells 1, 2 e 3, indicando predições menos enviesadas.	93
Figura 25 – Gráficos de dispersão das predições versus valores reais do modelo avançado. Nota-se melhor alinhamento com a linha de predição perfeita em todos os poços comparado ao modelo base, com destaque para Well 3 ($R^2 = 0,736$) e Well 0 ($R^2 = 0,600$).	93
Figura 26 – Séries temporais de VS real versus predito do modelo baseado em Self-Attention avançado. O modelo avançado demonstra capacidade superior de acompanhar variações locais, particularmente em Wells 1 e 3, com menor defasagem em regiões de transição abrupta.	94
Figura 27 – Predições de VS para poços sem ground truth (avançado). Comparado ao modelo base, as predições do avançado apresentam maior variabilidade (desvio padrão $2,9\times$ e $2,2\times$ maiores para Wells 0 e 1, respectivamente), capturando mais detalhes da heterogeneidade geológica.	95

LISTA DE TABELAS

Tabela 1 – Comparativo: Perfilagem a Cabo (<i>Wireline Logging</i>) vs. Perfilagem Durante a Perfuração (LWD)	28
Tabela 2 – Comparação entre RNN/LSTM, Transformer e N-HiTS	49
Tabela 3 – Descrição do conjunto de dados bruto utilizado.	57
Tabela 4 – Descrição dos registros presentes em cada poço	58
Tabela 5 – dos poços em treino e teste para os cinco folds.	73
Tabela 6 – Hiperparâmetros testados e respectivas justificativas.	75
Tabela 7 – Hiperparâmetros testados para os modelos baseado em Transformer.	76
Tabela 8 – Métricas de teste dos melhores modelos N-HiTS por fold	84
Tabela 9 – Desempenho médio do modelo de Transformers base em validação cruzada (5 folds)	87
Tabela 10 – Desempenho do modelo base em poços individuais com ground truth	88
Tabela 11 – Desempenho médio do avançado em validação cruzada (5 folds)	92
Tabela 12 – Desempenho do avançado em poços individuais com ground truth	92
Tabela 13 – Comparação de desempenho entre modelo base e avançado	95
Tabela 14 – Comparação de desempenho entre N-HiTS, modelo de self-attention base e modelo de self-attention avançado	99

LISTA DE ABREVIACES

VS	Velocidade das Ondas S (<i>Shear-Wave Velocity</i>)
DT	Delta T / Tempo de Trnsito
RHO	Densidade
NPHI	Porosidade (Neutrnica)
LWD	<i>Logging While Drilling</i>
ML	<i>Machine Learning</i>
ANN	<i>Artificial Neural Network</i>
MLP	<i>Multilayer Perceptron</i>
RMSE	<i>Root Mean Squared Error</i>
MAE	<i>Mean Absolute Error</i>
MSE	<i>Mean Squared Error</i>
DNN	<i>Deep Neural Network</i>
RNN	<i>Recurrent Neural Network</i>
LSTM	<i>Long Short-Term Memory</i>
GRU	<i>Gated Recurrent Unit</i>
BiGRU	<i>Bidirectional Gated Recurrent Unit</i>
BiLSTM	<i>Bidirectional Long Short-Term Memory</i>
CNN	<i>Convolutional Neural Network</i>
ReLU	<i>Rectified Linear Unit</i>
N-HiTS	Neural Hierarchical Interpolation for Time Series
DTW	Dynamic Time Warping
LWD	Logging While Drilling (Perfilagem Durante a Perfurao)

SUMÁRIO

	Lista de Figuras	11
	Lista de tabelas	13
	Lista de Abreviações	15
1	INTRODUÇÃO	21
1.1	MOTIVAÇÃO	21
1.2	OBJETIVOS	22
1.2.1	Objetivos Específicos	22
1.3	ORGANIZAÇÃO DO TEXTO	23
2	REVISÃO BIBLIOGRÁFICA	25
2.1	FUNDAMENTAÇÃO DE GEOFÍSICA E PERFILAGEM DE POÇOS . . .	25
2.1.1	Métodos de aquisição	26
2.1.2	Perfis de poços (<i>well logs</i>)	28
2.1.3	Aplicação dos <i>logs</i> na indústria de petróleo	33
2.2	FUNDAMENTOS DE REDES NEURAIS E APRENDIZADO DE MÁ- QUINA	33
2.2.1	Fundamentos de redes neurais	34
2.2.2	Métricas de avaliação	34
2.2.3	Redes Neurais Artificiais (<i>ANN - Artificial Neural Networks</i>)	36
2.2.4	Funções de ativação, normalização e regularização	37
2.2.5	Redes Neurais Recorrentes (<i>RNN, LSTM, GRU</i>)	39
2.2.5.1	<i>LSTM (Long Short-Term Memory)</i>	39
2.2.5.2	<i>GRU (Gated Recurrent Unit)</i>	41
2.2.5.3	<i>Redes Bidirecionais (<i>BiLSTM, BiGRU</i>)</i>	42
2.2.6	Redes Neurais Convolucionais (<i>CNNs</i>)	43
2.2.7	Mecanismos de Atenção (<i>Attention Mechanisms</i>)	44
2.2.7.1	<i>Self-Attention (Auto-Atenção)</i>	44
2.2.7.2	<i>Multi-Head Attention</i>	45
2.2.7.3	<i>Positional Encoding</i>	45
2.2.7.4	<i>Transformer Block</i>	46
2.2.7.5	<i>Vantagens dos Mecanismos de Atenção</i>	47
2.2.7.6	<i>Limitações</i>	47
2.2.8	N-HITS: Neural Hierarchical Interpolation for Time Series	47
2.2.8.1	<i>Motivação e Conceito</i>	47
2.2.8.2	<i>Arquitetura Hierárquica</i>	48

2.2.8.3	<i>Multi-Scale Decomposition</i>	49
2.2.8.4	<i>Interpolação Hierárquica</i>	49
2.2.8.5	<i>Comparação com Outras Arquiteturas</i>	49
2.2.8.6	<i>Aplicação em Séries Temporais Complexas</i>	50
2.2.8.7	<i>Limitações</i>	50
3	TRABALHOS CORRELATOS	51
3.1	ABORDAGENS DE ESTIMATIVAS BASEADAS EM GEOFÍSICA	51
3.2	ABORDAGENS CLÁSSICAS DE <i>MACHINE LEARNING</i>	52
3.3	ABORDAGENS DE <i>DEEP LEARNING</i> E REDES RECORRENTES	53
3.3.1	Arquiteturas Híbridas com CNNs	54
4	MATERIAIS E MÉTODOS	57
4.1	PREPARAÇÃO DOS DADOS	57
4.1.1	<i>Dataset</i>	57
4.1.2	Pré-processamento dos Dados	58
4.1.3	Análise de Similaridade entre Poços	59
4.1.3.1	<i>Distância Estatística (Wasserstein Distance)</i>	59
4.1.3.2	<i>Distância de Correlação</i>	60
4.1.3.3	<i>Distância DTW (Dynamic Time Warping)</i>	62
4.1.3.4	<i>Distância no Espaço de Features</i>	63
4.1.3.5	<i>Distância Combinada</i>	64
4.1.4	Visualizações Multidimensionais	66
4.1.4.1	<i>Multidimensional Scaling (MDS)</i>	66
4.1.4.2	<i>Projeções MDS Individuais Detalhadas</i>	67
4.2	EXPERIMENTOS INICIAIS	71
4.2.1	Modelagem	72
4.3	EXPERIMENTOS COM ARQUITETURAS AVANÇADAS: N-HITS E <i>SELF-ATTENTION BASED</i>	73
4.3.1	Validação Cruzada <i>Leave-One-Well-Out</i> (LOWOCV)	73
4.3.2	<i>Grid Search</i> de Hiperparâmetros	74
4.3.3	Arquitetura <i>N-HITS</i>	74
4.3.4	Arquitetura Baseada em Self-Attention	75
4.3.4.1	<i>Arquitetura Base</i>	76
4.3.4.2	<i>Arquitetura Avançada</i>	77
4.3.5	Pipeline de Treinamento unificado	78
4.3.5.1	<i>Seleção do Melhor Modelo</i>	78
4.3.5.2	<i>Agregação de Métricas em Validação Cruzada</i>	79
5	RESULTADOS	81
5.1	MÉTRICAS E MÉTODOS DE AVALIAÇÃO E VALIDAÇÃO	81

5.2	AVALIAÇÃO DO DESEMPENHO DOS EXPERIMENTOS INICIAIS . . .	81
5.3	ANÁLISE DE RESULTADOS FINAIS	83
5.3.1	Resultados do Modelo N-HiTS	83
5.3.1.1	<i>Configuração e Hiperparâmetros</i>	84
5.3.1.2	<i>Desempenho por Fold</i>	84
5.3.1.3	<i>Análise de Configurações Exploradas</i>	85
5.3.1.4	<i>Discussão sobre a Inadequação do N-HiTS</i>	85
5.3.2	Resultados dos Modelos Baseados em Self-Attention	86
5.3.2.1	<i>Modelo Base</i>	86
5.3.2.1.1	Configuração do Modelo	86
5.3.2.1.2	Desempenho em Validação Cruzada	87
5.3.2.1.3	Validação em Poços com Ground Truth	87
5.3.2.1.4	Predições em Poços sem VS	87
5.3.2.1.5	Análise Qualitativa - Modelo base	88
5.3.2.2	<i>Modelo Avançado</i>	89
5.3.2.2.1	Configuração do Modelo	91
5.3.2.2.2	Desempenho em Validação Cruzada	92
5.3.2.2.3	Validação em Poços com Ground Truth	92
5.3.2.2.4	Predições em Poços sem VS	92
5.3.2.2.5	Análise Qualitativa - Modelo Self-Attention Avançado	93
5.3.2.3	<i>Comparação entre modelos baseados em Self-Attention (base e avançado) .</i>	93
5.3.2.3.1	Análise Comparativa Detalhada	93
5.3.2.4	<i>Limitações e Casos de Falha</i>	98
5.3.3	Síntese Comparativa dos Três Modelos	98
5.3.3.0.1	Conclusões da Análise Comparativa	98
6	CONCLUSÃO	101
6.1	SÍNTESE DOS RESULTADOS OBTIDOS	101
6.1.1	Fase Preliminar: Baseline LSTM	101
6.1.2	Fase Final: Modelos Avançados	101
6.1.2.1	<i>N-HiTS: Inadequação para Dados Geofísicos</i>	101
6.1.2.2	<i>Modelo Base de Self-Attention</i>	102
6.1.2.3	<i>Modelo Avançado de Self-Attention: Melhor Resultado</i>	102
6.1.3	Fatores Determinantes do Sucesso do Modelo Avançado	102
6.2	CONTRIBUIÇÕES DO TRABALHO	103
6.2.1	Contribuições Técnicas	103
6.2.2	Contribuições Científicas	103
6.2.3	Contribuições Práticas	104
6.3	LIMITAÇÕES E DESAFIOS REMANESCENTES	104
6.3.1	Limitações do Modelo	104

6.3.2	Limitações dos Dados	104
6.4	TRABALHOS FUTUROS E DIREÇÕES DE PESQUISA	105
6.4.1	Melhorias Arquiteturais	105
6.4.2	Expansão dos Dados	105
6.4.3	Validação e Interpretabilidade	106
6.4.4	Extensões do Problema	106
6.5	CONSIDERAÇÕES FINAIS	106
	REFERÊNCIAS	109
A	CÓDIGO-FONTE	113
B	ARTIGO NO PADRÃO SBC	115

1 INTRODUÇÃO

A análise e modelagem de reservatórios de petróleo dependem fortemente da qualidade, completude e precisão dos dados adquiridos durante a perfuração dos poços, sejam eles em terra (*onshore*) ou no mar (*offshore*). Dentre os dados obtidos, os *well logs*, registros obtidos utilizando ferramentas de medição específicas inseridas no interior do poço, depois ou durante a perfuração, são fundamentais para a estimativa de propriedades petrofísicas, classificação de litofácies, cálculos de saturação, porosidade e outras propriedades. Utilizando tais *logs*, especialistas são capazes de identificar regiões de interesse e estimar a chance de encontrar gás natural ou petróleo (AVSETH; MUKERJI; MAVKO, 2005).

Os instrumentos utilizados para a aquisição das curvas, tais como *DT* (sônico compressional), *DTS* (sônico cisalhante) e *RHO* (densidade), são extremamente sensíveis e se encontram em condições extremas, como altas pressões devido à profundidade, temperatura e presença de lama, líquido e gás, tornando o processo caro, demorado e, principalmente, sujeito a falhas. Problemas como *wellbore washouts* (alargamentos não intencionais na parede do poço, também conhecidos como "arrombamento" das paredes do poço), falhas de contato das ferramentas com a rocha e interferências mecânicas ou eletrônicas podem comprometer a coleta dos dados (GENG; WANG, 2020).

A ocorrência de dados faltantes nos registros de poços de petróleo compromete a continuidade e consistência das análises, afetando tanto a interpretação geofísica quanto as etapas de modelagem de reservatórios. Em muitas situações, não é economicamente viável ou tecnicamente possível repetir a medição nos poços. Assim, o preenchimento (ou *imputação*) dos dados ausentes torna-se um passo essencial na preparação dos dados para uso em ambiente de tomada de decisão.

1.1 MOTIVAÇÃO

Soluções clássicas para o preenchimento de dados ausentes incluem técnicas como interpolações lineares, regressões baseadas em um único poço ou modelagens geoestatísticas, como a krigagem (GRANA TAPAN MUKERJI, 2021). Embora úteis em cenários com baixa variabilidade ou lacunas pequenas, esses métodos assumem uma linearidade local que raramente se confirma em ambientes geológicos complexos, onde os dados possuem grande variabilidade e ruído, sem uma tendência explicitamente linear ou facilmente modelável. Além disso, técnicas tradicionais não conseguem capturar e explorar as correlações não lineares e espaciais entre variáveis distintas ou mesmo entre diferentes poços de uma mesma bacia. Outro fator não coberto, de forma geral, por tais técnicas é a possibilidade de perda total em uma propriedade de um poço, de forma que, usando um ou mais *logs*, não são capazes de estimar uma curva por completo (MCDONALD, 2021).

Nesse contexto, nos últimos anos, abordagens baseadas em inteligência artificial e aprendizado de máquina vêm se destacando pela capacidade de modelar séries temporais e

dados geofísicos de alta complexidade. Entre essas abordagens, as redes neurais profundas, especialmente arquiteturas como *LSTM* (*Long Short-Term Memory*), modelos hierárquicos como o *N-HITS* (*Neural Hierarchical Interpolation for Time Series*) e modelos baseados em mecanismos de *Self-Attention*, têm demonstrado desempenho superior em tarefas de imputação e previsão em séries temporais não lineares (WANG et al., 2022; LIN et al., 2023).

Apesar dos avanços, muitos trabalhos existentes ainda apresentam limitações de generalização, frequentemente treinando modelos em poços isolados e ignorando as diferenças geológicas entre regiões. Além disso, diversos estudos aplicam arquiteturas genéricas sem avaliar sua aplicabilidade prática no contexto da indústria de petróleo. Essa lacuna evidencia a necessidade de modelos mais robustos, capazes de generalizar entre poços geologicamente distintos e de incorporar informações estruturais e petrofísicas relevantes (PHAM; NAEINI, 2019).

1.2 OBJETIVOS

O objetivo geral deste trabalho é investigar o uso de redes neurais profundas para o preenchimento de dados faltantes em registros de poços de petróleo, com foco na predição da curva de velocidade da onda cisalhante (VS). Esse log é essencial para análises geomecânicas e inversões sísmicas, mas sua aquisição direta é cara e suscetível a falhas, o que o torna um candidato natural para predição via aprendizado profundo.

Para isso, foram desenvolvidas e comparadas quatro arquiteturas de aprendizado profundo, baseando-se em artigos que buscaram tratar também do tema de preenchimento de dados em *well logs* e em suposições e detalhes do funcionamento das redes, conforme a Seção 4: uma rede *LSTM* usada como *baseline*, o modelo hierárquico *N-HITS*, um modelo de *Self-Attention* básico e uma versão avançada baseada em *Transformers*. As avaliações utilizaram dados reais fornecidos pela Petrobras, e foram conduzidas por meio de validação cruzada *leave-one-well-out* e análises de similaridade geológica entre poços com base em métricas como distância de Wasserstein, correlação, *Dynamic Time Warping* (DTW) e distância no espaço de *features*.

1.2.1 Objetivos Específicos

Para alcançar o objetivo geral proposto, foram definidos os seguintes objetivos específicos:

- Fazer uma análise crítica das propostas existentes na literatura correlata para a escolha das melhores arquiteturas de redes neurais para imputação em *well logs*;
- Implementar e comparar arquiteturas de redes neurais profundas, incluindo *LSTM*, *N-HITS* e modelos baseados em *Self-Attention*;
- Definir modelos que sirvam como base para comparações quantitativas e qualitativas entre diferentes abordagens;

- Avaliar o desempenho dos modelos com métricas como *RMSE* (*Root Mean Squared Error*), *MAE* (*Mean Absolute Error*) e R^2 , tanto em poços individuais quanto em validação cruzada;
- Analisar a estabilidade e a capacidade de generalização dos modelos em diferentes poços e contextos geológicos;
- Desenvolver visualizações que permitam interpretar a coerência geológica das predições e os padrões aprendidos pelas redes neurais.

1.3 ORGANIZAÇÃO DO TEXTO

O restante deste documento apresenta a seguinte estrutura: o Capítulo 2 dedica-se à fundamentação teórica essencial para este trabalho, juntamente com uma análise aprofundada de estudos relacionados. No Capítulo 4, são detalhados os materiais e métodos empregados, incluindo o conjunto de dados, algoritmos e modelos testados. O Capítulo 5 apresenta os resultados obtidos, suas respectivas métricas e uma discussão crítica sobre os experimentos conduzidos. Por fim, o Capítulo 6 reúne as conclusões e propõe futuras linhas de pesquisa.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo será dividido em duas seções. A Seção 2.1 fornece as principais definições sobre geofísica e modelagem de reservatórios sísmicos (GRANA TAPAN MUKERJI, 2021), explicando principalmente sobre o perfilamento de poços (*well logs*), com os principais tipos de *logs* que serão mencionados e usados para o treinamento das redes neurais, bem como suas aplicações e principais formas de aquisição. Além disso, a Seção 2.2 aborda os principais conceitos de redes neurais e aprendizado de máquina (ML).

2.1 FUNDAMENTAÇÃO DE GEOFÍSICA E PERFILAGEM DE POÇOS

A física de rochas é um campo interdisciplinar que investiga as propriedades físicas das rochas e sua inter-relação com as condições geológicas. Ela constitui a base para a interpretação de dados geofísicos e a compreensão do comportamento do subsolo. A petrofísica é definida como o estudo das propriedades físicas das rochas e dos fluidos nelas contidos, com ênfase particular no sistema poroso, na distribuição de fluidos e nas características de fluxo. Essa disciplina é fundamental para determinar e avaliar reservas e fontes de hidrocarbonetos ou aquíferos. (COLLEY, 2015)

A interpretação de *logs* de poços busca estimar as propriedades das rochas *in situ* ao longo da trajetória do poço, como porosidade, saturação de água e permeabilidade. Essas estimativas são cruciais para a avaliação do volume de reservas, a previsão de produção e a tomada de decisões no desenvolvimento de reservatórios. Modelos de física de rochas são empregados no refinamento dos *logs* de poços adquiridos, com o objetivo de corrigir imprecisões ou prever segmentos de dados ausentes. Uma compreensão sólida da física de rochas é uma disciplina prática fundamental que sustenta toda a viabilidade econômica da exploração e produção de hidrocarbonetos. Ela permite a interpretação e correção precisas de dados cruciais de poços, e sem essa base, a interpretação de *logs* seria puramente empírica e não confiável, impactando diretamente decisões financeiras significativas. (LUCIA, 2007)

A caracterização de formações geológicas ao longo de um poço é um processo fundamental para a avaliação da viabilidade de exploração de hidrocarbonetos. Durante a perfuração, diferentes sensores são utilizados para coletar informações físicas das rochas perfuradas, gerando os chamados *well logs* ou perfis de poço. Esses registros fornecem medições contínuas ao longo da profundidade do poço de diferentes propriedades petrofísicas, tais como tempo de trânsito acústico (DT), densidade (RHO), resistividade (RES) e emissão gama natural (GR). Tais informações são fundamentais para inferir características do reservatório, como a litologia, características da porosidade, saturação de fluidos e a presença ou não de hidrocarbonetos, possibilitando sua modelagem mais precisa e auxiliando na decisão para a produção e desenvolvimento do campo.

Os *logs* dos poços são considerados a principal fonte de alta resolução vertical disponível em estudos de reservatórios, sendo também utilizados para calibração de dados sísmicos,

análise de *facies* e suporte ao planejamento de novos poços. Apesar disso, sua aquisição, além de cara, pode ser bastante complexa ou por vezes inviável, sujeita a diversos erros e ruídos, tornando essencial a compreensão de suas limitações e fundamentos. Especialmente, pode-se destacar problemas para a aquisição do log *DTS*, que representa a velocidade da onda de cisalhamento (*shear wave velocity*) ou velocidade da onda S, e é um dos valores cruciais para o correto perfilamento do poço, principalmente por conta do custo de sua aquisição, podendo chegar a centenas de milhares de dólares por dia, mas também pelas dificuldades operacionais em sua medição. As ferramentas utilizadas, por muitas vezes serem de grande porte, podem ter sua descida impedida por características do poço, como uma inclinação acentuada, que impossibilitam a passagem da sonda. Ao longo das próximas subseções, serão apresentados os principais tipos de *logs* usados neste trabalho, seus métodos de aquisição e aplicações no contexto da análise e preenchimento de dados faltantes.

2.1.1 Métodos de aquisição

A perfilagem de poços é um conjunto de técnicas geofísicas essenciais para avaliar formações rochosas em profundidade, fornecendo dados contínuos sobre propriedades como porosidade, saturação de fluidos e litologia. Essas medições são um pilar para a caracterização de reservatórios e para as decisões em exploração e produção de petróleo (ELLIS, 2007).

Existem dois métodos principais para coletar esses dados. O primeiro, conhecido como perfilagem a cabo (*wireline logging*), utiliza ferramentas especializadas que são descidas no poço após a perfuração. Este método é reconhecido por fornecer dados de alta qualidade e resolução, mas exige a interrupção das operações, o que gera custos elevados e pode ser inviável em poços horizontais, aumentando o risco de incidentes como o aprisionamento de ferramentas (GENG; WANG, 2020). O segundo método, a perfilagem durante a perfuração (*logging while drilling* - LWD), integra as ferramentas à coluna de perfuração, permitindo a coleta de dados em tempo real. A principal vantagem do LWD é a tomada de decisão imediata, embora as severas condições de perfuração possam comprometer a qualidade dos dados e a baixa velocidade de transmissão de dados (telemetria) restrinja a complexidade das análises em tempo real (BOONEN, 2003). Um resumo comparativo dos dois métodos de perfilagem está descrito na Tabela 1

Independentemente do método, o processo é complexo, caro e sujeito a diversas limitações. As ferramentas operam em condições extremas de pressão e temperatura, e problemas como alargamentos na parede do poço (*washouts*) ou falhas de contato podem comprometer a qualidade dos dados, resultando em informações ruidosas, imprecisas ou ausentes. Destacam-se, em especial, os desafios na aquisição do log *DTS* (velocidade da onda de cisalhamento), cuja ausência compromete análises geomecânicas importantes e incentiva a busca por métodos de preenchimento de dados (BUUREN, 2018).

O custo substancial associado à perfilagem, particularmente em poços horizontais complexos, é um motor primário para a adoção de soluções baseadas em análises avançadas

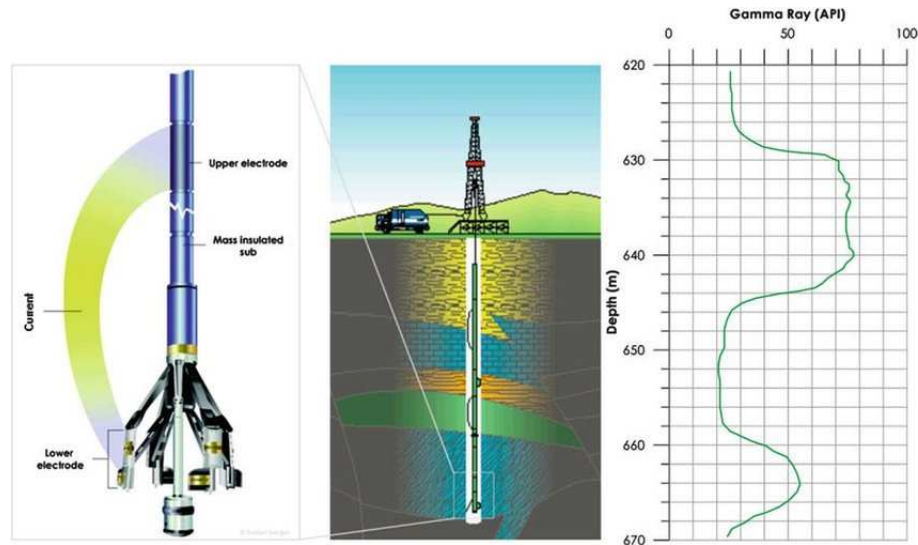


Figura 1 – Exemplo de ferramenta para *logging*, demonstração de operação de coleta de dados e exemplo de curva de saída. Fonte: (BIDA; AL-SUDANI, 2021)

de dados (PARSHALL, 2015). Essa tendência aponta para um futuro onde a modelagem preditiva pode otimizar as operações e diminuir a dependência de perfilagens caras em cada poço, favorecendo abordagens mais direcionadas e informadas por dados em todo o desenvolvimento do campo (SCHÖN, 2017).

Tabela 1 – Comparativo: Perfilagem a Cabo (*Wireline Logging*) vs. Perfilagem Durante a Perfuração (LWD)

Característica	Perfilagem a Cabo (<i>Wireline Logging</i>)	Perfilagem Durante a Perfuração (LWD)
Definição	Ferramentas de medição des-cidas por cabo após a perfuração ser interrompida.	Ferramentas de medição inte-gradas ao conjunto de fundo de poço, coletando dados du-rante a perfuração.
Tempo de Implanta-ção	Após a perfuração (requer pa-ralisação).	Durante a perfuração (contí-nuo, em tempo real).
Qualidade e Gama de Dados	Superior; ampla gama de pa-râmetros (resistividade, poro-sidade, densidade, sônica).	Geralmente mais restrita; ex-posta a condições severas que afetam precisão.
Resolução	Maior resolução (sem vibra-ções de perfuração).	Potencialmente menor reso-lução (vibrações, temperatu-ras, pressões).
Capacidade em Tempo Real	Não em tempo real; dados disponíveis após a operação.	Sim; dados em tempo real para decisões imediatas.
Adequação a Poços Complexos	Limitada em poços altamente desviados/horizontais (risco de ferramentas presas).	Particularmente benéfica em poços complexos/horizontais.
Impacto Operacional	Aumenta custos e tempo de perfuração (paralisação).	Minimiza tempo não produ-tivo; pode reduzir custos ge-rais.
Risco de Ferramentas Presas	Maior em poços complexos devido à dependência da gra-vidade.	Menor, mas falhas de ferra-mentas ainda podem causar atrasos.
Exposição a Condi-ções Severas	Menor (ferramentas não ope-ram durante a perfuração).	Alta (temperaturas, pressões, vibrações).
Efeitos Ambientais	Mais afetada por efeitos de poço e invasão (medição pos-terior).	Menos afetada por efeitos de poço e invasão (medição pre-coce).

2.1.2 Perfis de poços (*well logs*)

Os registros de perfilamento de poços, comumente chamados de *well logs*, constituem uma das principais fontes de dados para a caracterização petrofísica e geológica de formações subterrâneas. Esses registros são obtidos por meio de sensores acoplados a ferramentas que percorrem o interior do poço, fornecendo medições contínuas ao longo da sua profundidade. Cada tipo de *log* mede uma propriedade física ou resposta de formação específica, permitindo inferir características que serão usadas por profissionais da área da geologia e geofísica para a determinação da exploração ou não daquela região, tais como informações sobre a litologia,

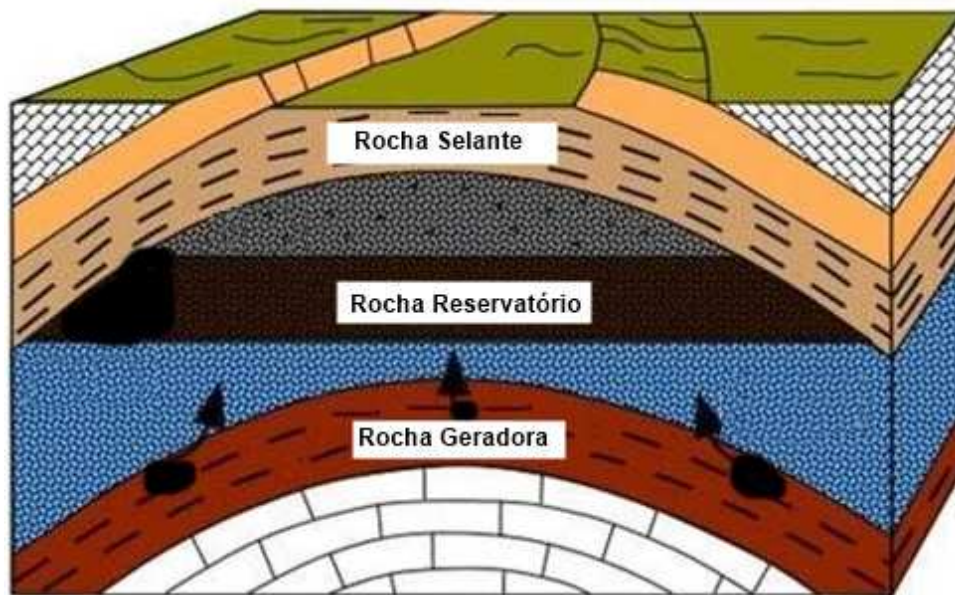


Figura 2 – Sistema petrolífero simplificado. Fonte: (GARCIA, 2022)

auxiliando na determinação do tipo de rocha das regiões ao longo do poço, características da porosidade das seções, como sobre o tamanho dos poros e o quanto estão conectados, conteúdo argiloso e a presença de hidrocarbonetos, que, se presentes e em condições favoráveis, podem ser explorados e extraídos.

No contexto da exploração de petróleo, o perfilamento ocorre após ou durante a perfuração, por meio da análise das variáveis obtidas pelas medições e também com inferências e modelagens. O objetivo é identificar se a formação sendo explorada apresenta características típicas de sistemas petrolíferos ativos, ou seja, com indícios de um conjunto de elementos geológicos que tornem a região produtiva. Entre os elementos que são observados para tentar determinar se há a presença de hidrocarbonetos na região, podendo ser vistos na representação da Figura 2, destacam-se:

- **Rocha geradora:** normalmente formada por sedimentos orgânicos ricos em carbono, como folhelhos, que tenham sido submetidos a condições adequadas de altas temperaturas e altas pressões ao longo do tempo geológico, gerando hidrocarbonetos como resultado da transformação da matéria orgânica depositada. Um longo período de tempo geológico é necessário para que essas rochas possam amadurecer e liberar os hidrocarbonetos, bem como para que, depois de formados, migrem para rochas reservatório e acumulem.
- **Rocha reservatório:** além de rochas que sejam capazes de gerar o fluido, é necessário que haja uma rocha altamente porosa para que ele possa ser armazenado, mas também que a mesma tenha alta permeabilidade, permitindo seu fluxo. De forma geral, arenitos e calcário fraturados são exemplos típicos desse tipo de rocha. Como o petróleo é menos denso que a água, migra pelos poros, então é também necessário que tais poros também sejam interconectados de forma suficiente que permita o fluxo entre si.

- **Rocha selante (ou selo):** acima da rocha que armazena, é necessário que haja uma camada impermeável de rocha, como argilitos ou evaporitos, impedindo a migração de hidrocarbonetos para a superfície, constituindo uma armadilha estrutural ou estratigráfica. Tais armadilhas são uma geometria estrutural que devem ser bastante eficazes, afim de causar o acúmulo do óleo ou gás, causando o depósito de uma quantidade que justifique a exploração daquela região economicamente.

A seguir, são descritos os principais *logs* utilizados nesse trabalho, incluindo aqueles que foram usados como entrada nos experimentos e, em especial, o *log VS*, que será a variável de saída dos modelos de aprendizado de máquina (LIU, 2017).

- **DT - Tempo de trânsito acústico (Delta-T):** o *log* de tempo de trânsito acústico mede o tempo necessário para que uma onda sônica percorra uma distância padrão ao longo da formação. Esse valor é relacionado com a velocidade compressional (*VP*) pela equação:

$$VP = \frac{1}{DT} \times k \quad (2.1)$$

onde *k* é uma constante de conversão.

O *DT* é um dos *logs* mais usados na petrofísica por sua relação com a porosidade, tipo de fluido e compactação da formação. Em formações mais porosas, há tendência de apresentar maior tempo de trânsito, devido à menor velocidade de propagação do som em materiais fluidos comparados com rochas consolidadas.

- **DTS - Velocidade da Onda de Cisalhamento (Shear Wave Velocity):** o *log* de *VS* representa a velocidade de propagação ao longo da formação. Geralmente é expresso em metros por segundo (*m/s*) e, diferentemente das ondas compressionais (*VP*), essas ondas não se propagam em fluidos, tornando essa propriedade particularmente sensível ao conteúdo de fluido da rocha e à matriz mineral. Sua obtenção direta requer ferramentas bastante específicas, como sondas dipolares, que nem sempre podem estar disponíveis ou são viáveis, devido seu elevado custo operacional, dimensão do equipamento ou restrições operacionais, pois características como a inclinação do poço e seu diâmetro interno irregular podem não permitir a passagem dos sensores ao longo de sua profundidade. Em muitos casos, a medição direta é substituída por estimativas empíricas baseadas em outras curvas, como o *VP*, por meio de equações como (CASTAGNA M. L. BATZLE, 1985):

$$VS = a \times VP + b \quad (2.2)$$

As limitações operacionais fazem de *VP* um alvo comum para modelagem e imputação por redes neurais ou métodos geostatísticos, com objetivo de reproduzir esse *log* a partir de outros mais facilmente obtidos. Essa propriedade é fundamental em aplicações geomecânicas, para o perfilamento do poço, inversões sísmicas e cálculo de módulos elásticos (como o módulo de cisalhamento $G = \rho \times VS^2$)

- **GR - Gamma Ray:** o log gama natural mede a emissão espontânea de radiação gama das formações geológicas. É utilizado principalmente para estimar a argilosidade das rochas, pois minerais argilosos, como illita e caulinita, são naturalmente radioativos. Em formações sedimentares, valores baixos de GR são geralmente associados a arenitos limpos, enquanto valores altos indicam presença de argila ou folhelhos.
- **VSH - Argilosidade ou volume de argila:** o volume de argila (V_{shale}) é derivado do log GR, utilizado para quantificar a fração de argila presente em uma formação. Um dos métodos mais comuns para seu cálculo é (GEOLOGISTS, 2022):

$$V_{sh} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}} \quad (2.3)$$

onde GR_{log} representa o valor mensurado para GR na profundidade desejada e GR_{max} e GR_{min} são valores conhecidos para essa propriedade em regiões sem xisto (ou folhelho) e 100% de xisto, respectivamente. Pode-se também estimar esse valor de forma não linear, inicialmente calculando usando a fórmula acima e, em seguida, usando um gráfico como na Figura 3.

- **Caliper:** medição do diâmetro interno do poço ao longo da profundidade, serve como diagnóstico da qualidade da parede do poço, sendo fundamental para interpretação da confiabilidade de outros logs. Desvios acentuados no caliper indicam zonas de colapsos ou *washouts*, onde sensores podem perder contato com a formação e produzir dados ruidosos ou inválidos, sendo necessário fazer correções pelo caliper ou então descartar os dados. Essa informação pode ser usada para mascarar regiões para treinamento de modelos de aprendizado de máquina ou como entrada para indicações de condições críticas.

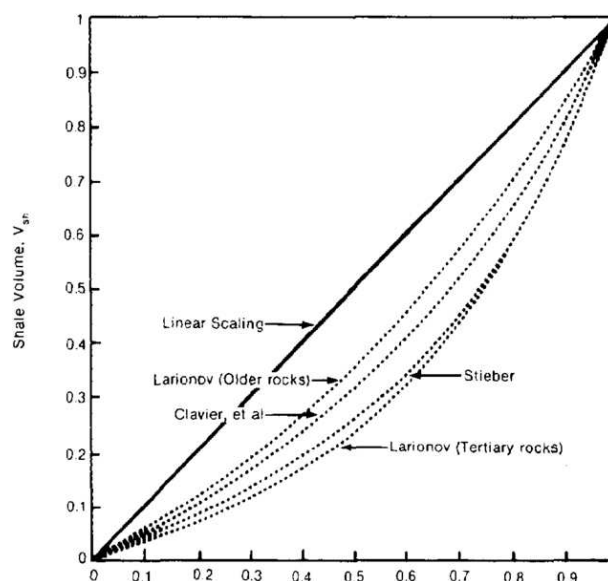


Figura 3 – Gráfico para volume de argila não linear. Fonte: (INTERNATIONAL, 1985)

- ***PHI* - Porosidade:** fração do volume da rocha ocupada por poros, podendo ser medida por logs como *NPHI* (porosidade neutrônica), ou estimada com equações empíricas que combinam *DT* e *RHO* (*Wyllie time-average equation*) (*GEOLOGISTS*, 2022). A porosidade efetiva considera somente poros interconectados que podem conter fluido móvel. É um dos principais indicadores de reservatórios produtivos e influencia diretamente na velocidade de propagação das ondas, pois em rochas mais porosas, a velocidade é menor.
- ***RHO* - Densidade:** mede a densidade aparente da formação, geralmente em g/cm^3 . É essencial para cálculo de outras propriedades como porosidade (combinando com densidade da matriz e dos fluidos) e módulos elásticos, como por exemplo para o cálculo $VP = \sqrt{\frac{K + \frac{4}{3}G}{\rho}}$. Valores mais baixos de *RHO* podem indicar zonas porosas ou então a presença de gás, enquanto valores altos indicam litologias mais densas, tais como calcários ou dolomitos.
- **Saturação de fluido:** saturação refere-se à porção de fluido (água, gás, óleo) presente nos poros da rocha. Embora não medida diretamente por um único *log*, pode ser estimada com modelos que utilizam essa propriedade em conjunto com a resistividade, porosidade e *VSH*. Influencia diretamente o comportamento acústico da rocha, pois zonas saturadas com óleo ou gás possuem velocidades diferentes das saturadas com água, o que afeta diretamente os *logs* de *DT* e *VS*.
- ***VP* - Velocidade da onda compressional:** geralmente derivada de *DT* com uma relação inversa, representa a velocidade da onda ao atravessar a formação. Junto de *VS* e *RHO*, é fundamental para calcular os módulos elásticos. Estes módulos são, essencialmente, medidas quantitativas da rigidez da rocha e de como ela se deforma sob a ação de forças, sendo cruciais para entender o comportamento mecânico do reservatório. Adicionalmente, essas mesmas variáveis permitem estimar as anisotropias sísmicas, que ocorrem quando a velocidade da onda varia conforme a direção em que ela se propaga. Essa variação direcional é causada por fatores como fraturas alinhadas, laminações finas ou o estado de tensão da rocha. Por fim, como as ondas compressionais e cisalhantes reagem de maneiras distintas aos fluidos, a razão *VP/VS* é comumente utilizada como um poderoso indicativo para diferenciar a litologia e o conteúdo de fluido nos poros.
- **Resistividade:** mede a capacidade da formação de resistir ao fluxo de corrente elétrica. Rochas com poros preenchidos por hidrocarbonetos (gás ou óleo) apresentam alta resistividade, enquanto rochas preenchidas com água salgada (salmoura) exibem baixa resistividade.

De maneira indireta, os *logs* permitem identificar essas condições. Pode-se exemplificar isso com o caso da leitura de resistividade, que pode indicar a presença de hidrocarbonetos na região, uma vez que valores altos de tal medição, por serem menos condutivos que a água salgada, causam o aumento da resistividade, enquanto *logs* de densidade e porosidade ajudam

na avaliação da qualidade da rocha do reservatório. O conteúdo argiloso, por sua vez, pode comprometer a permeabilidade da região e deve ser quantificado, como com o *log* de *GR* e o cálculo de volume de argila (*VSH*).

Dado o alto custo da perfuração e da aquisição de *logs*, extrair o máximo de informação dos dados registrados é essencial para reduzir riscos na exploração. Por isso, sua qualidade, continuidade e completude têm impacto direto sobre decisões técnicas e econômicas de toda a cadeia de produção petrolífera.

2.1.3 Aplicação dos *logs* na indústria de petróleo

Os *well logs* são fundamentais para a estimativa de propriedades petrofísicas, classificação de litofácies e cálculos de saturação, porosidade e outras propriedades, auxiliando especialistas a identificar áreas de interesse e estimar a probabilidade de encontrar gás natural ou petróleo. Dentre as aplicações principais, podem ser destacadas (ØDEGAARD; AVSETH, 2004):

- **Caracterização de reservatórios:** *logs* de poço são a principal fonte de dados para a caracterização petrofísica e geológica de formações subterrâneas, fornecendo medições contínuas de propriedades físicas das formações rochosas em profundidade. Isso é essencial para avaliar as formações e compreender o subsolo.
- **Classificação de litofácies:** os *logs* fornecem informações sobre a litologia, ajudando a determinar o tipo de rocha em regiões ao longo do poço. (SANTOS; ROISENBERG; NASCIMENTO, 2022)
- **Calibração de dados sísmicos:** registros de poço são utilizados para calibrar dados sísmicos, o que é crucial para a interpretação de dados sísmicos e a extrapolação de propriedades do poço para uma escala sísmica tridimensional.
- **Análise de fácies e suporte ao planejamento de novos poços:** são empregados na análise de fácies e no suporte ao planejamento de novos poços.
- **Avaliação geomecânica e inversões sísmicas:** *logs* como *DT* e, especialmente, *VS*, são importantes para a avaliação geomecânica da formação, essencial para a segurança da perfuração e completação. Essa avaliação também utiliza *logs* para inversões sísmicas, ajudando a extrapolar propriedades do poço que não podem ser medidas ou devem ser interpoladas.

2.2 FUNDAMENTOS DE REDES NEURAIS E APRENDIZADO DE MÁQUINA

A crescente complexidade dos dados de poços e a necessidade de lidar com dados faltantes têm impulsionado a aplicação de técnicas de aprendizado de máquina, especialmente redes neurais, para otimizar a análise e a interpretação de reservatórios.

2.2.1 Fundamentos de redes neurais

O aprendizado de máquina (*ML*) é um subconjunto da inteligência artificial que permite que sistemas aprendam e melhorem com base em dados, identifiquem padrões e tomem decisões ou façam previsões sem a necessidade de programação explícita para cada tarefa, continuamente aprimorando previsões e acurácia (RUSSEL, 2020).

Modelos de ML podem ser classificados quanto ao algoritmo usado, onde podem ou não depender do conhecimento prévio das respostas ou rótulos no conjunto de dados, ou seja, o valor correto que o modelo deve prever. No **aprendizado supervisionado**, modelos são treinados com dados rotulados, onde os valores de saída corretos são conhecidos. O algoritmo aprende a mapear entradas para saídas desejadas, ajustando-se iterativamente para minimizar erros. É adequado para tarefas de classificação (prever categorias) e regressão (prever valores numéricos contínuos). Já em modelos de aprendizado não supervisionado, o algoritmo recebe dados de entrada sem rótulos de saída. Ele trabalha independentemente para identificar padrões, estruturas e relacionamentos inerentes aos dados. É útil para análise exploratória de dados e tarefas de agrupamento, como detecção de anomalias ou segmentações.

Quanto aos dados, é importante para este trabalho o conhecimento sobre séries temporais, as quais são dados sequenciais, como palavras ou registros de poços, onde os componentes se inter-relacionam com base em regras complexas de semântica e sintaxe. A modelagem de séries temporais é uma técnica crucial que utiliza dados sequenciais para prever valores ou eventos futuros com base em dados históricos, analisando padrões, tendências e sazonalidade. A detecção de anomalias, por sua vez, identifica padrões incomuns nos dados de séries temporais. O conceito de dependência sequencial é fundamental, pois a ordem dos pontos de dados é crítica para fazer previsões precisas. Redes neurais recorrentes (*RNNs* 2.2.5) representam um passo pioneiro no processamento de dados sequenciais, mantendo uma forma de "memória" que lhes permite processar sequências de comprimentos variáveis. Os *logs* de poço, por sua natureza, representam dados sequenciais ao longo da profundidade, tornando os modelos de séries temporais particularmente adequados para sua análise e previsão.

2.2.2 Métricas de avaliação

A avaliação de modelos de regressão é uma etapa crucial para determinar o quão bem o modelo se ajusta aos dados e quão bem ele pode fazer previsões em novos dados. As métricas comuns para avaliar modelos de regressão incluem:

- **RMSE (Root Mean Squared Error - Raiz do Erro médio quadrático):** o RMSE é a raiz quadrada do Erro Quadrático Médio (*MSE*). Ele retorna uma métrica na mesma unidade da variável de destino, tornando-a mais interpretável do que o *MSE*. Penaliza erros maiores mais severamente devido ao termo quadrático. Valores mais próximos de zero

indicam melhor desempenho.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (2.4)$$

onde n é o número de observações, y_i o valor real da i -ésima observação e $\hat{y}_i$ o valor previsto pelo modelo.

Como vantagens, o valor calculado está na mesma unidade da variável de destino, facilitando a interpretação, bem como a penalização por erros grandes de forma mais acentuada pode ser desejável em alguns cenários, onde erros muito grandes são inaceitáveis. Entretanto, tal penalização também pode ser indesejada, pois torna o valor sensível a *outliers*, uma vez que o erro é quadrático, amplificando o impacto que valores discrepantes possuem sobre o valor final.

- **MAE (Mean Absolute Error - Erro médio absoluto):** mede a média das diferenças absolutas entre as previsões do modelo e os valores reais. É uma métrica simples e robusta a *outliers*, pois não eleva os erros ao quadrado. Representa o erro médio de previsão e está na mesma unidade da variável de destino. Valores mais baixos indicam um melhor ajuste.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.5)$$

onde n é o número de observações, y_i o valor real da i -ésima observação e $\hat{y}_i$ o valor previsto pelo modelo.

Essa métrica é de fácil implementação, entendimento e interpretação, pois também encontra na mesma unidade de medida da variável de saída, além de ser mais robusta a *outliers*, pois penaliza erros de forma linear, sem supervalorizar grandes desvios. Entretanto, por não tratar erros grandes de forma mais penalizante, como o *RMSE* 2.2.2 ou *MSE*, pode causar desvantagens caso erros altos sejam críticos.

- **R^2 - Coeficiente de determinação:** mede a proporção da variância na variável dependente que é previsível a partir das variáveis independentes. Ele indica o quão bem o modelo de regressão se ajusta aos dados observados. Um valor de R^2 mais próximo de 1 (ou 100%) indica que o modelo explica uma grande parte da variabilidade dos dados, sugerindo um bom ajuste.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (2.6)$$

onde n é o número de observações dos dados, y_i é o valor real da i -ésima observação, $\hat{y}_i$ o valor previsto da i -ésima observação e $\bar{y}$ é a média dos valores reais.

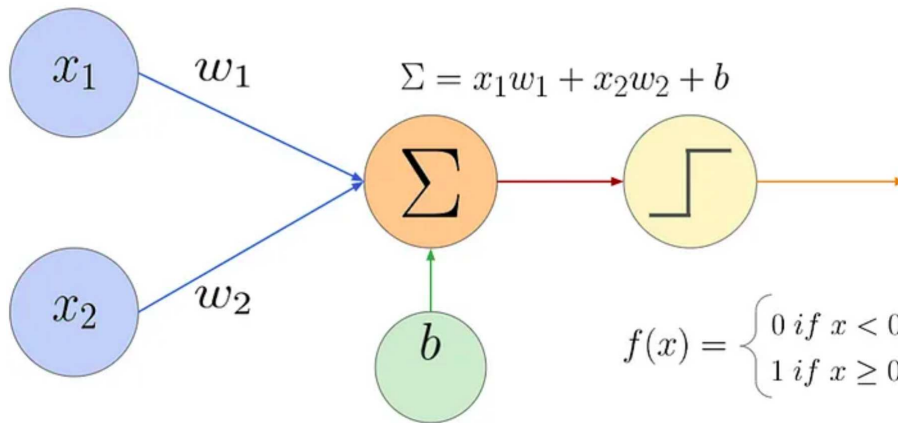


Figura 4 – Exemplo de representação de um Perceptron. Adaptado de: (AHMAD, 2024)

Fornece uma medida de quão bom está o ajuste do modelo, indicando a proporção da variância da variável dependente que é explicada pelas variáveis independentes, variando o valor de 0 a 1 (0% a 100%). A adição de novas variáveis preditoras sempre causa o aumento ou o permanecimento do valor de R^2 , mesmo que elas não sejam estatisticamente significativas ou não melhorem a capacidade preditiva real do modelo, o que pode causar problemas como o *overfitting* do modelo, o que pode-se tentar mitigar pelo uso de R^2 ajustado (considera o número de variáveis preditoras e o tamanho da amostra).

Essas métricas são para problemas de regressão, nos quais o objetivo é prever valores numéricos contínuos e não devem ser confundidas com métricas de classificação, tais como acurácia e precisão. A escolha de uma métrica de avaliação mais apropriada depende do problema específico, do conjunto de dados disponível e dos objetivos do projeto.

2.2.3 Redes Neurais Artificiais (ANN - *Artificial Neural Networks*)

Redes Neurais Artificiais (ANNs) são modelos computacionais inspirados no cérebro humano, capazes de aprender padrões complexos a partir de dados. O Perceptron é uma das arquiteturas mais simples de redes neurais artificiais, introduzida por Frank Rosenblatt (ROSENBLATT, 1958). Ele realiza classificação binária, mapeando características de entrada para uma decisão de saída binária (0 ou 1). Consiste em uma única camada de nós de entrada totalmente conectados a uma camada de nós de saída e é particularmente eficaz para aprender padrões linearmente separáveis. O Perceptron utiliza uma função de ativação de limiar (função degrau de Heaviside).

As Redes Feedforward, também conhecidas como Perceptrons Multicamadas (MLPs), são redes neurais onde a informação se move em uma única direção, do nó de entrada para o de saída, sem ciclos ou *loops* de *feedback*. As MLPs possuem duas ou mais camadas (camadas de entrada, uma ou mais camadas ocultas e uma camada de saída) e são capazes de distinguir dados que não são linearmente separáveis, ao contrário do Perceptron de camada única. Para

isso, utilizam funções de ativação não lineares, como a função sigmoide ou ReLU, em suas camadas ocultas.

A retropropagação (*backpropagation*) é o algoritmo de treinamento mais comum para redes neurais, tornando o gradiente descendente viável para redes neurais multicamadas. O processo envolve o cálculo da taxa de erro de uma propagação direta e a alimentação dessa perda de volta através das camadas da rede neural para ajustar os pesos. Utiliza a regra da cadeia para calcular como os pesos da rede contribuem para uma função de perda, ajustando-os para minimizar a diferença entre a saída prevista e a saída real.

A distinção entre redes neurais clássicas e Redes Neurais Profundas (*Deep Neural Networks* - *DNNs*) reside fundamentalmente na arquitetura e na capacidade de representação. Enquanto as redes neurais clássicas, como os (*MLPs*) tradicionais, geralmente contam com uma ou poucas camadas ocultas, as *DNNs* são caracterizadas por possuírem um grande número de camadas ocultas empilhadas. Essa profundidade permite que as *DNNs* aprendam hierarquias de características, onde cada camada subsequente constrói representações mais abstratas e complexas a partir das características aprendidas pela camada anterior. Isso as torna excepcionalmente eficazes na extração automática de características de dados brutos, uma tarefa que, em abordagens clássicas de *ML*, frequentemente requer engenharia de características manual e intensiva.

A complexidade e a profundidade das *DNNs* conferem-lhes uma capacidade superior de modelar relações não lineares e padrões intrincados em grandes volumes de dados, o que é um dos principais motivadores para o sucesso do *Deep Learning*. Contudo, essa capacidade vem acompanhada de desafios adicionais. O treinamento de redes muito profundas é computacionalmente intensivo e pode ser suscetível a problemas como o desaparecimento ou explosão de gradientes, que dificultam a propagação do erro através de muitas camadas. Para superar esses obstáculos, foram desenvolvidas diversas inovações arquitetônicas e técnicas de otimização, como as Unidades de Memória de Longo Prazo (*LSTMs*) e as Redes Convolucionais (*CNNs*), além de métodos de normalização e inicialização de pesos mais sofisticados.

Durante o treinamento de redes neurais, podem surgir problemas como gradientes evanescentes (*vanishing gradients*), onde os gradientes das camadas mais próximas da entrada se tornam muito pequenos, fazendo com que essas camadas treinem muito lentamente ou não treinem. A função de ativação ReLU pode ajudar a prevenir esse problema. O oposto são os gradientes explosivos (*exploding gradients*), que ocorrem se os pesos da rede forem muito grandes, levando a gradientes que se tornam excessivamente grandes para convergir. A normalização em lote (*batch normalization*) e a redução da taxa de aprendizado podem mitigar esse problema.

2.2.4 Funções de ativação, normalização e regularização

As funções de ativação são um elemento crucial das redes neurais, introduzindo não linearidades que permitem que a rede aprenda padrões complexos nos dados. Elas transformam o sinal de entrada de um nó em uma rede neural em um sinal de saída que é então passado para a próxima camada, decidindo se um neurônio deve ser ativado ou não. Sem funções de ativação,

as redes neurais seriam restritas a modelar apenas relações lineares. Dentre as mais utilizadas para redes neurais, destacam-se:

- **Função Sigmoid:** mapeia qualquer valor real de entrada para uma saída entre 0 e 1, é comumente usada na camada de saída para problemas de classificação binária, onde a saída pode ser interpretada como uma probabilidade. Devido sua formulação, seu uso pode causar o problema de *vanishing gradient* para valores de entrada muito grandes ou muito pequenos, tornando o gradiente muito próximo de zero, dificultando o aprendizado. É calculada por:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.7)$$

- **Função Tangente Hiperbólica (*Tanh*):** mapeia valores reais para uma saída entre -1 e 1. É uma versão centralizada da função Sigmoid, o que pode auxiliar no treinamento, ainda que seja também suscetível ao problema de *vanishing gradient*. Possui como fórmula:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.8)$$

- **ReLU - *Rectified Linear Unit*** (AGARAP, 2019): retorna o valor da entrada se ele for positivo e zero caso contrário, ou seja:

$$f(x) = \max(0, x) \quad (2.9)$$

Amplamente usada em camadas ocultas de redes neurais profundas devido sua simplicidade computacional e por ajudar a mitigar o problema de gradiente *vanishing* para valores positivos. Um problema que pode ocorrer com essa função de ativação é o de "neurônios mortos" (*dying ReLU*), onde neurônios param de aprender se suas entradas forem sempre negativas, resultando em gradiente zero. Para tentar solucionar tal problema, há a variação *Leaky ReLU*, na qual ao invés de retornar zero para entradas negativas, retorna um pequeno valor positivo, garantindo que o gradiente nunca seja zero e permitindo neurônios continuarem a aprender, mesmo para entradas negativas.

$$f(x) = \begin{cases} x, & \text{se } x \geq 0 \\ \alpha x, & \text{se } x < 0 \end{cases}$$

Além de ajustes nos parâmetros e na arquitetura da rede, para o adequado aprendizado de padrões pelos modelos, é de suma importância que sejam aplicadas técnicas de pré-processamento dos dados. Dentre essas técnicas, pode-se mencionar a normalização dos exemplos de entrada, que ajusta os valores das características para uma escala comum, tipicamente entre 0 e 1, sem distorcer as diferenças na faixa de valores. Isso garante que todas as características contribuam igualmente para o modelo, melhorando a convergência do algoritmo de aprendizado e o desempenho geral do modelo. A normalização em lote (*batch normalization*) é uma técnica comum que normaliza as ativações de uma camada para cada mini-lote durante o

treinamento. Outras técnicas comuns são as de regularização, as quais são técnicas usadas para prevenir o *overfitting* (quando o modelo aprende o ruído dos dados de treinamento e generaliza mal para dados novos) e melhorar a capacidade de generalização do modelo, minimizando a complexidade desnecessária e expondo a rede a dados mais diversos. Exemplos de técnicas de regularização são:

- **Dropout:** desativa aleatoriamente uma fração de neurônios em cada camada durante o treinamento, forçando a rede a aprender representações mais robustas.
- **Early Stopping:** interrompe o treinamento da rede neural em um ponto anterior, monitorando o desempenho do modelo em um conjunto de validação.
- **Regularização L1 e L2:** Adicionam um termo de penalidade à função de perda para desencorajar pesos grandes, promovendo modelos mais simples. L1 promove esparsidade (forçando pesos a zero), enquanto L2 impede que os pesos se tornem excessivamente grandes.
- **Aumento de Dados (*Data Augmentation*):** Aumenta a diversidade do conjunto de treinamento através de transformações nos dados existentes, útil quando há dados insuficientes.

2.2.5 Redes Neurais Recorrentes (*RNN*, *LSTM*, *GRU*)

Redes Neurais Recorrentes (*RNNs*) são modelos de aprendizado profundo projetados para processar e converter entradas de dados sequenciais em saídas sequenciais específicas. Dados sequenciais, como palavras ou séries temporais, possuem componentes que se inter-relacionam com base em regras complexas de semântica e sintaxe. Se distinguem de outras redes neurais pelo uso de *loops* de *feedback*, que permitem que a informação de estágios de processamento posteriores seja realimentada para estágios anteriores, o que é crucial para o processamento de sequências. Essa característica permite que a rede "lembre" dados passados ao processar a próxima saída, reconhecendo características sequenciais e ajudando a prever o próximo ponto de dados na sequência. As *RNNs* são capazes de lidar com entradas e saídas de comprimento variável e compartilham os mesmos pesos em várias etapas de tempo, o que generaliza o modelo para processar sequências de comprimento variável.

2.2.5.1 *LSTM* (*Long Short-Term Memory*)

Long Short-Term Memory é um tipo de rede neural recorrente que foi desenvolvida para mitigar o problema do gradiente evanescente, comum em *RNNs* tradicionais, e para capturar dependências de longo prazo em dados sequenciais. A principal motivação por trás da arquitetura *LSTM* é criar um módulo adicional em uma rede neural que aprende quando lembrar e quando esquecer informações pertinentes. São particularmente eficazes em modelar dependências de mais longo prazo, devido à introdução de um mecanismo interno que faz o controle do

fluxo de informações com o passar do tempo, mais complexo do que a memória interna básica das redes recorrentes (HOCHREITER; SCHMIDHUBER, 1997).

A principal novidade nesse tipo de rede é a introdução de células de memória e um conjunto de portas, responsáveis por regular o que deve ser armazenado, esquecido ou utilizado na próxima etapa pela rede. Uma célula c_t mantém um estado contínuo que pode ser atualizado de forma controlada, o que lhe dá capacidade de preservar informações importantes sobre os dados por um período mais prolongado, bem como remover informações de sua memória conforme o treinamento acontece.

Internamente, cada célula possui um valor c , que é propriamente a "memória de longo prazo" dela, que será atualizado ao longo do processamento dos dados e usado para auxiliar no treinamento e previsões. Esse mecanismo de memória também auxilia na solução dos problemas envolvendo gradiente, uma vez que não é modificado diretamente por pesos ou vieses e pode ser, portanto, usado pelas diferentes etapas de tempo. Há também um valor auxiliar para os cálculos, chamado "memória de curto prazo", que é também passado adiante durante as etapas, mas passível de modificações. Em cada etapa t , o funcionamento das LSTMs pode ser descrito pelas seguintes operações:

1. **Porta de esquecimento** (f_t) Responsável por decidir quais informações do estado da célula anterior (c_{t-1}) devem ser descartados.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.10)$$

onde σ é uma função de ativação que gera valores no intervalo $[0, 1]$ (como *sigmoide*) e $[h_{t-1}, x_t]$ são o estado oculto anterior ("curto prazo") e entrada atual, respectivamente.

O valor calculado é então usado como sendo a porcentagem da memória de longo prazo que deve ser lembrada, onde um valor de 0 irá reiniciar a memória de longo prazo e 1 irá manter exatamente o valor atual.

2. **Porta de entrada** (i_t) Determina quais novas informações devem ser armazenadas na célula de memória. Calcula-se a porcentagem de memória potencial a ser lembrada:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.11)$$

e a informação a ser lembrada, similarmente:

$$\bar{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.12)$$

Com ambos os valores, soma-se $i_t \cdot \bar{c}_t$ a memória de longo prazo, terminando de fazer sua atualização para a etapa temporal atual, t .

3. **Porta de saída** (o_t) É durante essa etapa que calcula-se o valor de saída (h_t) de toda essa unidade LSTM. Faz-se também a atualização da memória de curto prazo, que irá servir como saída e para alimentar a rede em próximas etapas.

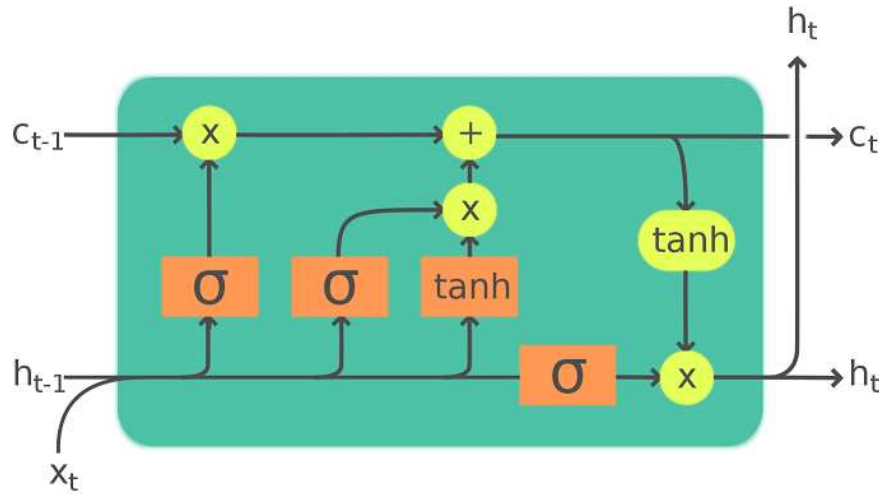


Figura 5 – Estrutura interna de uma unidade *LSTM*. Fonte: (QUANTCONNECT, 2024)

Calcula-se quais partes do estado da célula c_t devem ser usadas para gerar o estado oculto (memória de curto prazo) h_t :

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.13)$$

e atualiza-se o estado oculto, usando a memória de longo prazo (c_t):

$$h_t = o_t \cdot \tanh(c_t) \quad (2.14)$$

As LSTMs se destacam por mitigar os problemas com gradiente através de sua arquitetura, garantindo a estabilidade do treinamento ao controlar o fluxo de gradientes e permitir que a rede aprenda e utilize informações distantes ao longo das sequências de dados de entrada, modelando assim dependências de longo prazo. Essa flexibilidade, decorrente da combinação das portas de entrada, esquecimento e saída, torna as LSTMs adaptáveis a diversos padrões temporais em dados sequenciais, capacitando-as a aprender relações complexas, especialmente quando integradas a outras arquiteturas de redes neurais e técnicas de aprendizado de máquina.

Contudo, possuem limitações, como o alto custo computacional, devido ao aumento no número de parâmetros a serem aprendidos, resultando em modelos mais lentos para treinamento em comparação com RNNs básicas. Além disso, exigem um preparo rigoroso dos dados, incluindo *feature engineering* e ordenamento, e, apesar de sua robustez, ainda podem enfrentar desafios de otimização em sequências de dados de tamanhos extremos e com longas dependências temporais (AL-SELWI et al., 2023).

2.2.5.2 GRU (Gated Recurrent Unit)

As GRUs (CHO et al., 2014) são uma alternativa simplificada à LSTM, também projetadas para resolver o problema do gradiente evanescente em RNNs tradicionais. A principal

2.2.6 Redes Neurais Convolucionais (CNNs)

As Redes Neurais Convolucionais (CNNs) foram originalmente desenvolvidas para processamento de imagens (LECUN et al., 1998), mas têm demonstrado eficácia notável em dados sequenciais unidimensionais, como séries temporais e registros de poços (KIRANYAZ et al., 2021). Ao contrário das RNNs, que processam sequências de forma estritamente temporal, as CNNs aplicam filtros convolucionais que capturam padrões locais e características morfológicas ao longo da dimensão espacial ou temporal dos dados.

Em uma convolução unidimensional, um filtro, também chamado de *kernel* de tamanho k desliza ao longo da sequência de entrada, realizando o produto escalar entre o filtro e janelas locais dos dados. Para uma entrada $x \in \mathbb{R}^{L \times C}$, onde L é o comprimento da sequência e C o número de canais (features), a operação de convolução com um filtro $w \in \mathbb{R}^{k \times C \times C'}$ produz uma saída $y \in \mathbb{R}^{L' \times C'}$:

$$y_i^{c'} = \sum_{c=1}^C \sum_{j=0}^{k-1} w_{j,c,c'} \cdot x_{i+j}^c + b^{c'} \quad (2.15)$$

onde i é a posição na sequência, c e c' são os índices de canais de entrada e saída e $b^{c'}$ é o termo de vies.

As Redes Neurais Convolucionais (CNNs) são compostas por uma sequência de camadas que operam de forma cooperativa para extrair e processar representações hierárquicas dos dados de entrada. Os principais componentes incluem:

- **Camadas Convolucionais:** A camada convolucional é o núcleo da CNN. Ela aplica filtros (ou *kernels*) que extraem características locais da entrada.
 - **Tamanho do Kernel (*kernel size*):** Define a extensão do campo receptivo local. Kernels maiores capturam dependências de longo alcance, enquanto kernels menores enfatizam padrões locais e detalhes de curta duração (BAGNALL et al., 2016).
 - **Padding:** Controla o tratamento das bordas da entrada. Estratégias como *same padding* preservam o tamanho da saída, enquanto *valid padding* reduz sua dimensão.
 - **Stride:** Determina o deslocamento do filtro a cada operação de convolução. Um *stride* igual a 1 mantém a resolução original, enquanto valores maiores realizam uma forma de amostragem mais espaçada (*downsampling*).
- **Camadas de Pooling:** As camadas de *pooling* reduzem a dimensionalidade das representações intermediárias, diminuindo o número de parâmetros e a sensibilidade a pequenas variações espaciais ou temporais.
 - **Max Pooling:** Seleciona o valor máximo dentro de cada janela local, destacando as características mais relevantes.

- **Average Pooling:** Calcula a média dos valores dentro da janela, resultando em uma representação mais suavizada.
- **Camadas Totalmente Conectadas:** Após as operações convolucionais e de *pooling*, as representações extraídas podem ser passadas para camadas totalmente conectadas (*fully connected layers*), que combinam as características aprendidas para realizar tarefas como classificação, regressão ou reconstrução de dados.

Todas essas características das CNNs as tornam capazes de aprender representações hierárquicas, onde camadas iniciais capturam padrões simples e camadas mais profundas aprendem abstrações mais complexas (WANG; YAN; OATES, 2016). Ainda assim, essas redes são capazes de serem computacionalmente eficientes, já que as operações convolucionais são altamente paralelizáveis e otimizáveis em hardware moderno (BAGNALL et al., 2016). O uso de *pooling* garante invariância local, permitindo o reconhecimento de padrões independentemente da posição exata na entrada, reduzindo o ruído, pois tende a suavizar variações de alta frequência e instabilidades locais (FU et al., 2024).

2.2.7 Mecanismos de Atenção (*Attention Mechanisms*)

O mecanismo de atenção, introduzido por (BAHDANAU; CHO; BENGIO, 2016) no contexto de tradução neural e popularizado por (VASWANI et al., 2023) com a arquitetura Transformer, representa um avanço fundamental em aprendizado profundo para dados sequenciais. Diferentemente de RNNs e CNNs, que processam sequências de forma fixa (sequencial ou local), a atenção permite que o modelo foque dinamicamente em diferentes partes da sequência de entrada, aprendendo quais posições são mais relevantes para a tarefa.

2.2.7.1 Self-Attention (Auto-Atenção)

O mecanismo de *self-attention* calcula uma representação ponderada de cada elemento da sequência, onde os pesos são determinados pela relevância relativa entre todos os elementos. Para uma sequência de entrada $\mathbf{X} \in \mathbb{R}^{L \times d}$, onde L é o comprimento e d a dimensão das *features*, o *self-attention* opera através de três transformações lineares:

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{X}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{X}\mathbf{W}_V \quad (2.16)$$

onde $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{d \times d_k}$ são matrizes de pesos aprendidas, produzindo as representações de *Query* (consulta), *Key* (chave) e *Value* (valor).

A saída do *self-attention* é calculada como:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V} \quad (2.17)$$

O termo $\mathbf{QK}^T$ computa a similaridade (produto escalar) entre todas as posições da sequência, enquanto o fator de escala $\sqrt{d_k}$ estabiliza os gradientes. A função *softmax* normaliza os *scores* de atenção, produzindo pesos $\alpha_{ij} \in [0, 1]$ que somam 1, e a multiplicação por $\mathbf{V}$ gera a representação final ponderada. (VASWANI et al., 2023)

O peso α_{ij} indica o quanto a posição i deve “atender” à posição j ao computar sua representação. Isso permite que o modelo capture dependências arbitrariamente longas em uma única operação, ao contrário de RNNs, que requerem múltiplos passos de propagação (VASWANI et al., 2023).

2.2.7.2 Multi-Head Attention

Para aumentar a capacidade expressiva, Vaswani et al. (2023) propuseram o *multi-head attention*, que executa múltiplos mecanismos de atenção em paralelo com diferentes transformações lineares:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}_O \quad (2.18)$$

$$\text{head}_i = \text{Attention}(\mathbf{QW}_Q^i, \mathbf{KW}_K^i, \mathbf{VW}_V^i) \quad (2.19)$$

onde h é o número de cabeças de atenção e $\mathbf{W}_O$ é uma matriz de projeção final. Cada cabeça pode aprender diferentes aspectos da sequência, como dependências locais e globais, o que aumenta a capacidade representacional do modelo sem ampliar linearmente o custo computacional (VASWANI et al., 2023).

2.2.7.3 Positional Encoding

Como o mecanismo de atenção não possui noção inerente de ordem sequencial (é *permutation-invariant*), é necessário injetar informação posicional nos vetores de entrada. Para isso, foram propostas codificações posicionais baseadas em funções seno e cosseno (VASWANI et al., 2023):

$$PE_{(pos, 2i)} = \sin\left(\frac{pos}{10000^{2i/d}}\right) \quad (2.20)$$

$$PE_{(pos, 2i+1)} = \cos\left(\frac{pos}{10000^{2i/d}}\right) \quad (2.21)$$

onde pos é a posição na sequência e i é a dimensão. Alternativamente, podem ser utilizados *embeddings* posicionais aprendíveis, especialmente quando o comprimento da sequência é fixo (GEHRING et al., 2017).

2.2.7.4 Transformer Block

Um bloco *Transformer* completo combina *multi-head attention* com redes totalmente conectadas e conexões residuais (VASWANI et al., 2023):

$$\mathbf{Z} = \text{LayerNorm}(\mathbf{X} + \text{MultiHead}(\mathbf{X})) \quad (2.22)$$

$$\text{Output} = \text{LayerNorm}(\mathbf{Z} + \text{FFN}(\mathbf{Z})) \quad (2.23)$$

onde

- **LayerNorm:** Normalização por camada, que estabiliza o treinamento (BA; KIROU; HINTON, 2016);
- **FFN (Feed-Forward Network):** Rede totalmente conectada aplicada independentemente a cada posição;
- **Conexões residuais:** Facilitam o fluxo de gradientes em redes profundas (HE et al., 2015).

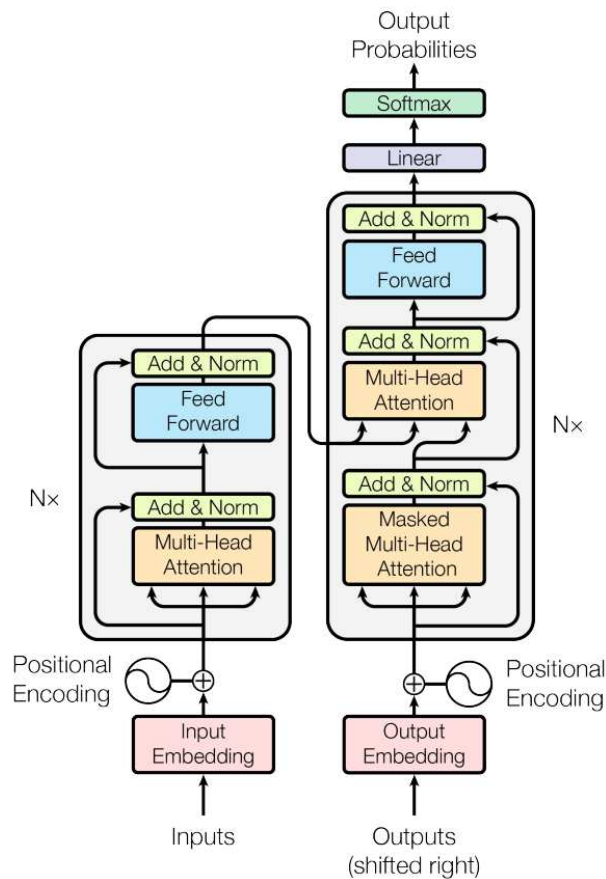


Figura 7 – Arquitetura Transformer. Fonte: (VASWANI et al., 2023)

2.2.7.5 Vantagens dos Mecanismos de Atenção

Os mecanismos de atenção apresentam diversas vantagens teóricas e práticas. Eles permitem capturar dependências de longo alcance entre quaisquer posições de uma sequência em uma única operação, superando limitações de janelas fixas de CNNs e da propagação sequencial das RNNs. Além disso, o processamento de todos os elementos da sequência ocorre em paralelo, o que aumenta significativamente a eficiência computacional durante o treinamento. Outra característica importante é a interpretabilidade: os pesos de atenção podem ser visualizados, permitindo compreender quais partes da entrada o modelo considera mais relevantes. Por fim, arquiteturas híbridas que combinam convoluções locais com atenção global têm se mostrado especialmente eficazes em diversas tarefas sequenciais (WANG et al., 2022).

2.2.7.6 Limitações

Apesar de seu sucesso, o uso de atenção apresenta desafios. Sua complexidade computacional é quadrática em relação ao comprimento da sequência, isto é, $\mathcal{O}(L^2 \cdot d)$, o que a torna custosa para entradas muito longas (TAY et al., 2022). Além disso, Transformers geralmente requerem grandes volumes de dados para generalizar de forma eficaz, em comparação com modelos baseados em RNNs ou CNNs (VASWANI et al., 2023).

2.2.8 N-HiTS: Neural Hierarchical Interpolation for Time Series

O *N-HiTS* (*Neural Hierarchical Interpolation for Time Series*) é uma arquitetura recente proposta por (CHALLU et al., 2022) para previsão de séries temporais, que combina amostragem multi-taxa (*multi-rate sampling*), decomposição hierárquica e interpolação para capturar padrões em múltiplas escalas temporais. Diferentemente de modelos puramente sequenciais, como RNNs, ou baseados em atenção, como Transformers, o N-HiTS explora explicitamente a estrutura hierárquica das séries temporais através de *stacks* especializados em diferentes bandas de frequência.

2.2.8.1 Motivação e Conceito

Séries temporais reais exibem padrões em múltiplas escalas com frequências variadas: alta frequência, para variações locais (como ruídos de medição), média frequência, em estruturas intermediárias e variações regionais e baixa frequência, em tendências de longo prazo e padrões globais.

Modelos tradicionais tratam todos os padrões uniformemente, enquanto o N-HiTS decompõe a tarefa em subproblemas especializados por frequência, inspirado em técnicas de processamento de sinais como *wavelets* (CHALLU et al., 2022).

2.2.8.2 Arquitetura Hierárquica

O N-HiTS consiste em múltiplos *stacks*, cada um especializado em uma escala temporal específica, por meio de diferentes taxas de *downsampling* (*pooling*). Cada *stack* contém múltiplos blocos que processam a informação de forma residual.

Para uma entrada $\mathbf{x} \in \mathbb{R}^{L \times F}$, onde L é o comprimento da sequência e F o número de *features*, cada bloco do N-HiTS realiza as seguintes etapas:

1. Multi-rate Sampling (Downsampling):

- Aplica-se *MaxPooling* com tamanho de janela p (e.g., $p = 8, 4, 2, 1$);
- O comprimento da sequência reduz-se para $L' = L/p$;
- Intuitivamente, um *pooling* maior foca em padrões de baixa frequência (tendências), enquanto valores menores capturam detalhes de alta frequência.

2. MLP Processing: A sequência reduzida é processada por uma rede totalmente conectada:

$$\mathbf{h} = \text{MLP}(\mathbf{x}_{pooled})$$

3. Geração de Coeficientes *Theta*: O bloco gera dois conjuntos de coeficientes:

- θ_b : para reconstrução do residual (*backcast*);
- θ_f : para previsão futura (*forecast*).

4. Interpolação Hierárquica: Os coeficientes são usados para reconstruir e prever por meio de funções de base:

$$\hat{\mathbf{x}} = \text{Interpolate}(\theta_b, L) \quad (2.24)$$

$$\hat{\mathbf{y}} = \text{Basis}(\theta_f) \quad (2.25)$$

5. Conexões Residuais: Os blocos dentro de cada *stack* operam sobre resíduos:

$$\mathbf{r}^{(0)} = \mathbf{x} \quad (2.26)$$

$$\mathbf{r}^{(i+1)} = \mathbf{r}^{(i)} - \text{Backcast}^{(i)}(\mathbf{r}^{(i)}) \quad (2.27)$$

A previsão final é obtida pela soma dos *forecasts* de todos os blocos e *stacks*:

$$\hat{\mathbf{y}} = \sum_{s=1}^S \sum_{b=1}^{B_s} \text{Forecast}_{s,b} \quad (2.28)$$

onde S é o número de *stacks* e B_s o número de blocos em cada um.

2.2.8.3 Multi-Scale Decomposition

Cada *stack* utiliza um tamanho de *pooling* distinto:

- **Stack 1 (pooling = 16):** padrões de altíssima frequência;
- **Stack 2 (pooling = 8):** padrões de alta-média frequência;
- **Stack 3 (pooling = 4):** padrões de média frequência;
- **Stack 4 (pooling = 2):** padrões de baixa-média frequência;
- **Stack 5 (pooling = 1):** tendências de baixa frequência.

Essa decomposição explícita por escala evita que um único modelo precise aprender padrões de diferentes frequências simultaneamente, reduzindo a interferência entre componentes durante o treinamento (CHALLU et al., 2022).

2.2.8.4 Interpolação Hierárquica

Diferentemente de modelos que produzem previsões diretas, o N-HiTS gera coeficientes θ , que são combinados por funções base (como polinômios ou *wavelets*) para compor a previsão final. Isso garante:

- **Extrapolação suave:** a interpolação assegura continuidade além do horizonte de treinamento;
- **Regularização implícita:** restringe o espaço de previsões possíveis, reduzindo o overfit;
- **Interpretabilidade:** os coeficientes θ representam contribuições por escala temporal.

2.2.8.5 Comparação com Outras Arquiteturas

Tabela 2 – Comparação entre RNN/LSTM, Transformer e N-HiTS

Aspecto	RNN/LSTM	Transformer	N-HiTS
Processamento temporal	Sequencial	Paralelo (atenção)	Hierárquico (multi-escala)
Escalas temporais	Uniforme	Uniforme	Explícita (via pooling)
Complexidade	$\mathcal{O}(L \cdot d^2)$	$\mathcal{O}(L^2 \cdot d)$	$\mathcal{O}(L \cdot d^2)$ (com pooling)
Interpretabilidade	Baixa	Média (pesos de atenção)	Alta (coeficientes por escala)
Dados necessários	Moderado	Alto	Moderado

2.2.8.6 Aplicação em Séries Temporais Complexas

O N-HiTS apresenta vantagens específicas para domínios com padrões multi-escala:

- **Captura de estruturas multi-escala:** separa variações locais e tendências globais;
- **Robustez a ruído:** o *pooling* em altas frequências suaviza ruído de medição;
- **Eficiência com dados limitados:** a decomposição hierárquica reduz a complexidade;
- **Flexibilidade arquitetural:** permite ajustar o número de *stacks* e taxas de *pooling* conforme o domínio.

2.2.8.7 Limitações

Apesar de seu desempenho expressivo, o N-HiTS apresenta algumas limitações:

- **Escolha de pooling sizes:** requer conhecimento do domínio sobre escalas relevantes (CHALLU et al., 2022);
- **Complexidade de hiperparâmetros:** muitos parâmetros por *stack* (e.g., número de blocos, tamanhos de camadas ocultas);
- **Perda de informação:** *downsampling* agressivo pode descartar eventos súbitos importantes.

3 TRABALHOS CORRELATOS

Esse capítulo apresenta estudos que mostram abordagens para o uso de redes neurais e algoritmos de aprendizado de máquina para o preenchimento de dados faltantes em registros de poços de petróleo. Em geral, usam diferentes conjuntos de dados, de acordo com a disponibilidade dos autores, enfatizando os resultados diferentes obtidos de acordo com os modelos e técnicas empregadas para fazer a *imputação* dos dados.

De maneira geral, pode-se notar a transição das abordagens clássicas, insuficientes para a complexidade e heterogeneidade dos dados geofísicos, para as técnicas avançadas de *Deep Learning* (DL) e *Ensemble Learning*. Serão detalhados os tipos de modelos utilizados (como RNNs, CNNs e suas variantes bidirecionais, além de métodos de *ensemble* como *Random Forest* (RF) e *XGBoost*), suas arquiteturas específicas, as metodologias de pré-processamento e validação de dados empregadas, as métricas de desempenho e as estratégias para mitigar o *overfitting* e quantificar a incerteza. A análise revelará que modelos híbridos e de *ensemble* frequentemente superam modelos singulares, oferecendo maior precisão e robustez em cenários de dados desafiadores.

3.1 ABORDAGENS DE ESTIMATIVAS BASEADAS EM GEOFÍSICA

Para abordar a tarefa de estimar dados de perfilagem ausentes, diversas pesquisas têm se dedicado a desvendar as complexas relações não lineares existentes entre diferentes perfis de poço. Historicamente, abordagens clássicas, como a estimativa manual, métodos empíricos baseados em relações geofísicas e estatísticas, regressão linear multivariável (MLR - *Multiple Linear Regression*) e até mesmo Redes Neurais Artificiais (ANNs) convencionais com poucas camadas, têm sido empregadas. Contudo, essas técnicas frequentemente apresentam limitações significativas, resultando em perfis estimados de qualidade questionável. Tais processos são influenciados por fatores subjetivos e simplificações inerentes aos modelos, comprometendo a precisão e a confiabilidade das previsões. (RIDER, 2011)

Modelos empíricos, como a equação de Gardner (GARDNER L. W. GARDNER, 1974) ou os métodos de Castagna, são baseados em relações predefinidas e geralmente dependem do tipo específico de rocha analisada. Por exemplo, o método de Greenberg-Castagna, embora seja uma abordagem convencional para estimar perfis sônicos de cisalhamento (VS) ausentes, é intrinsecamente limitado a rochas saturadas em salmoura, não sendo preciso em zonas com hidrocarbonetos. A calibração e a aplicação precisa desses métodos exigem tempo considerável e expertise humana aprofundada, além de não se adaptarem bem à variabilidade geológica encontrada em diferentes campos (GREENBERG; CASTAGNA, 1992).

3.2 ABORDAGENS CLÁSSICAS DE *MACHINE LEARNING*

Técnicas e modelos clássicos de Aprendizado de Máquina podem ser empregados para a previsão de dados faltantes em registros de poços. No entanto, devido à natureza intrinsecamente complexa, ruidosa e não linear dos dados geológicos, a capacidade de representação e generalização desses modelos pode ser limitada. Jian et al. (2020), por exemplo, observaram que a Regressão Linear Multivariável (MLR), devido à sua simplicidade e forte dependência da qualidade e quantidade dos dados de entrada, não se mostra uma opção robusta para o preenchimento de dados faltantes ou a previsão de curvas inteiras de perfis de poço. Uma contribuição importante desse trabalho foi demonstrar que, em comparação com modelos puramente empíricos (como os citados na Seção 3.1), mesmo os modelos de ML mais simples, sem o uso de *Deep Learning* ou RNNs, podem apresentar melhor desempenho em métricas como R^2 .

Outra categoria de modelos frequentemente referenciada na literatura para imputação de dados em *well logs* são os algoritmos baseados em árvores de decisão, notadamente *Random Forest* (RF) e *XGBoost* (CHEN; GUESTRIN, 2016). Feng et al. (2021) utilizaram RF em conjunto com análise de incerteza para o preenchimento do *log* DTS (FENG; GRANA; BALLING, 2021). Essa abordagem permite não apenas a previsão do conjunto de árvores do algoritmo, mas também a quantificação da incerteza do modelo sobre a previsão, fornecendo ao usuário informações valiosas sobre a confiabilidade do valor imputado. Isso pode ser particularmente útil para estimativas iniciais ou previsões em cenários com menor criticidade. Amir e Amir (2025) e Jian et al. (2020) também empregaram RF para a previsão, mas, de forma consistente com os achados de Feng et al. (2021), os resultados não se equipararam ao desempenho de modelos mais avançados. Em contrapartida, o uso de XGBoost, um modelo que emprega um conjunto de árvores de decisão com reforço gradativo (*gradient boosting*), demonstrou resultados significativamente melhores. Este algoritmo é reconhecido por sua capacidade de generalização robusta e eficiência, mesmo sendo relativamente mais simples que arquiteturas de Deep Learning.

Ainda no contexto de previsão de curvas como o DTS, Amir e Amir (2025) também exploraram o uso de MLPs, conforme detalhado na Seção 2.2.3. O MLP, com sua estrutura interna em camadas de neurônios interconectados, funções de ativação não lineares e processo iterativo de treinamento (*backpropagation*), representa um modelo de maior complexidade em comparação com a regressão linear ou árvores de decisão simples. Essa arquitetura confere ao MLP uma melhor capacidade de aprender padrões ruidosos e relações não lineares, características frequentemente encontradas em *logs* de poços. No entanto, a utilização de modelos com maior complexidade, como o MLP, também introduz desafios adicionais, como a necessidade de uma busca exaustiva por hiperparâmetros ótimos, incluindo configurações de camadas, taxas de aprendizado e a escolha de otimizadores, o que pode aumentar a demanda computacional e o tempo de desenvolvimento.

3.3 ABORDAGENS DE *DEEP LEARNING* E REDES RECORRENTES

A crescente complexidade e a natureza sequencial dos dados de perfilagem de poços têm impulsionado a adoção de técnicas de *Deep Learning* (DL), que superam as limitações dos métodos clássicos ao capturar relações não lineares e dependências temporais ou espaciais. RNNs e suas variantes, como LSTMs e GRUs, são particularmente adequadas para dados sequenciais, como os encontrados em perfis de poço, devido à sua capacidade de processar informações em função de passos de tempo anteriores.

Pham e Naeini (2019) propuseram um método para estimar *logs* ausentes (especificamente *logs* VS) utilizando uma rede LSTM bidirecional (BiLSTM) em cascata com uma *Fully Connected Neural Network* (FCNN). A abordagem se destaca por quantificar a incerteza do modelo através da adição de uma camada de *dropout*, que permite a criação de uma distribuição de Monte Carlo das curvas preditas, onde a variância representa a incerteza do modelo. Os autores aplicaram o método a 35 poços no Mar do Norte Central, demonstrando uma boa correspondência entre as previsões do modelo e os *logs* Vs registrados, bem como com estimativas baseadas em modelos de física de rocha. O coeficiente de correlação entre o *log* Vs predito e o medido foi de 94%. O diferencial desse trabalho é que o modelo não se limita à predição zona a zona, característica de alguns métodos de física de rocha, e pode ser estendido para prever outros tipos de *logs* ausentes.

Hussain et al. (2024) também empregaram uma rede BiLSTM combinada com uma *Self-Organizing Map* (SOM) para previsão de porosidade e litofácies. A SOM foi utilizada para agrupar intervalos de profundidade com fácies semelhantes em *clusters* com semelhança petrofísica. A rede BiLSTM, por sua vez, foi responsável por capturar componentes espaço-temporais abrangentes dos dados de *log*, garantindo que o modelo representasse com precisão os aspectos espaço-temporais. O treinamento foi feito com dados do campo de gás de Zamzama, localizado no Paquistão, e os resultados indicaram que o modelo recuperou com sucesso características significativas dos dados de *logging*, resultando em maior precisão de estimativa. A pesquisa destaca a eficácia do modelo BiLSTM na previsão de características físicas de reservatórios, oferecendo um novo método para a previsão precisa de parâmetros de caracterização de reservatórios.

Em outro estudo, Zhou e Li (2024) propuseram uma LSTM com mecanismo de autoatenção (*self-attention*) (VASWANI et al., 2023) seletiva (ESA-LSTM) para a síntese de curvas de *well log*. O modelo visa superar as limitações das LSTMs tradicionais em lidar com dependências de longo prazo e alta complexidade computacional. A autoatenção é utilizada para analisar as dependências sequenciais dos dados e selecionar os resultados computacionais dominantes na LSTM, reduzindo a complexidade de $O(n^2)$ para $O(n \log n)$ e melhorando a eficiência do modelo. A camada de seleção introduzida no modelo permite que apenas as posições com alta pontuação de atenção influenciem a alteração dos parâmetros da camada LSTM, acelerando a convergência do modelo com impacto mínimo na precisão. Usando 20 poços da área de Nalینگou, na Bacia de Ordos, nordeste da China, os experimentos demonstraram que o método

proposto alcança maior precisão em comparação com métodos tradicionais de síntese de curvas baseados em FCNN e LSTM tradicional.

3.3.1 Arquiteturas Híbridas com CNNs

A combinação de *Convolutional Neural Networks* (CNNs) com RNNs tem se mostrado promissora para capturar tanto características espaciais quanto temporais em dados de *well log*. As CNNs são excelentes na extração de características locais e morfológicas dos dados, enquanto as RNNs são eficazes na modelagem de sequências e dependências temporais.

Hussain et al. (2024) empregaram uma arquitetura híbrida de CNN e BLSTM, onde a CNN extrai características espaciais dos dados de *logging* e a BiLSTM, alimentada por essas características, coleta componentes espaço-temporais abrangentes. Essa abordagem é crucial para representar com precisão os aspectos espaço-temporais dos dados de *log*. Os resultados de erro de previsão do modelo BiLSTM foram os menores em comparação com os modelos CNN e LSTM isolados, indicando a maior precisão alcançada pela abordagem híbrida (HUSSAIN et al., 2024).

Wang et al. (2022) desenvolveram um modelo de *Deep Learning* integrado (SAIDNN) que combina CNN, uma rede GRU bidirecional e um mecanismo de *self-attention* para prever *logs* ausentes. O modelo consiste em dois módulos principais: um módulo CNN para extrair características morfológicas locais dos dados de *logging*, e um módulo BiGRU para minerar a tendência de variação e informações de contexto dos dados com a profundidade, a partir das características de saída do módulo CNN. O mecanismo de autoatenção, por sua vez, permite que a rede aloque pesos para destacar informações relevantes, melhorando a precisão da previsão. Os resultados de aplicação em dados de campo reais de duas áreas diferentes, uma na Bacia de Ordos, na China, e outra do campo Volve, na Noruega (EQUINOR, 2018), demonstraram que o modelo proposto oferece resultados de previsão precisos e confiáveis, com alta viabilidade e praticidade. A autoatenção foi introduzida na CNN para extrair características morfológicas locais de pontos de amostragem adjacentes e na BiGRU para focar o modelo na aprendizagem de características de dados mais significativas, reduzindo a perda de informações históricas. Quantitativamente, o modelo SAIDNN reduziu significativamente o RMSE e o MAE em comparação com outros modelos como IDNN (CNN e BiGRU sem autoatenção), BiGRU, GRU e FCNN.

De forma semelhante, Ali e Al-Sudani (2021) propuseram uma abordagem para prever *logs* sísmicos de cisalhamento ausentes utilizando padrões de similaridade de poços com propriedades geofísicas semelhantes, combinando DNN com métricas de similaridade (Jaccard e Overlap). Embora o foco principal não seja uma arquitetura híbrida específica de CNN-RNN, o trabalho ressalta a importância da seleção de características (*feature selection*) e da otimização da rede. O *log* DT foi identificado como o mais importante para a predição do *log* DTS. A metodologia demonstra uma precisão de previsão superior em zonas com hidrocarbonetos, onde métodos convencionais como Greenberg-Castagna são limitados. Os experimentos foram

conduzidos com poços da formação *Lower Goru*, na Bacia do Indo Médio, Paquistão. Os resultados mostraram uma correlação de 98% entre o *log* DTS predito pela DNN e o original, superando a correlação de 75% do método de Greenberg-Castagna (PHAM; NAEINI, 2019).

4 MATERIAIS E MÉTODOS

Neste capítulo, serão detalhados os experimentos realizados neste trabalho, realizados após estudos sobre o conjunto de dados, seu pré-processamento, testes de diferentes modelos de *machine learning* e análise contínua dos resultados obtidos. Como objetivo principal do trabalho, deseja-se construir um modelo capaz de prever o *log* VS, de forma que, em casos de falha na sua aquisição, gerando perdas parciais, ou ausência no conjunto de dados, seja possível reconstruir tal curva, usando para isso outros *logs* disponíveis.

4.1 PREPARAÇÃO DOS DADOS

Para este trabalho, foi utilizado um conjunto de dados privado, fornecido pela Petrobras, composto de sete poços, com localização confidencial. As subseções seguintes descrevem as características dos poços, bem como as transformações realizadas para sua adequação para serem usados no treinamento e teste do modelo proposto.

4.1.1 Dataset

O conjunto de dados utilizado neste estudo é composto por *logs* de sete poços de petróleo fornecidos pela Petrobras, armazenados no formato *.las*. Desse conjunto de poços, quatro estão localizados em um bloco alto, ou seja, menos profundos em relação ao nível do mar, e os demais em um bloco baixo, separados por uma grande falha geológica, com um total de 10.236 amostras de dados, descritos na Tabela 3, junto da quantidade original de *logs* em cada. Por terem localizações e nomes confidenciais, os poços serão identificados por uma numeração. Os registros presentes em cada poço, com exceção dos que estão em apenas um (que foram removidos do conjunto final), estão descritos na Tabela 4, de forma que apenas *logs* comuns a todos os poços usados em um experimento são incluídos como entrada para o treinamento e teste.

Poço	Número de instâncias	Quantidade de <i>logs</i>
Poço 1	1.301	10
Poço 2	1,001	10
Poço 3	1,201	10
Poço 4	1,151	12
Poço 5	1,591	12
Poço 6	1,380	11
Poço 7	2,611	10

Tabela 3 – Descrição do conjunto de dados bruto utilizado.

Deve-se notar que os poços 1 e 7 não possuem a propriedade VS (poço 1 possui seu VS calculado a partir de VPVS, outra forma de aquisição do *log*), logo serão esses os poços utilizados para testes, após o treinamento e validação do modelo.

Poço	Logs
Poço 1	Argilosidade, CALIPER, DEPT, GR, MD, Porosidade, RHO, Saturação, VP, VS (calculado de VPVS)
Poço 2	Argilosidade, CALIPER, DEPT, GR, MD, Porosidade, RHO, Saturação, VP, VS
Poço 3	Argilosidade, CALIPER, DEPT, GR, MD, Porosidade, RHO, Saturação, VP, VS
Poço 4	Argilosidade, CALIPER, COTA, DEPT, DEPTH_TVDKB, GR, MD, Porosidade, RHO, Saturação, VP, VS
Poço 5	Argilosidade, CALIPER, COTA, DEPT, DEPTH_TVDKB, DEPTH_MD, GR, Porosidade, RHO, Saturação, TVD, VP, VS
Poço 6	Argilosidade, CALIPER, DEPT, GR, MD, Porosidade, RHO, Saturação, VP, VS
Poço 7	Argilosidade, CALIPER, COTA, DEPT, GR, Porosidade, RHO, Saturação, VP

Tabela 4 – Descrição dos registros presentes em cada poço

4.1.2 Pré-processamento dos Dados

A etapa de pré-processamento envolveu o escalonamento dos dados para mitigar os efeitos de diferentes escalas e *outliers*. O *RobustScaler* da biblioteca *Scikit-learn* foi utilizado, uma vez que é menos sensível a valores extremos em comparação com outros escaladores (AMORIM; CAVALCANTI; CRUZ, 2023). O *RobustScaler* usa quantis (Intervalo Interquartil, IQR) para o escalonamento, ao invés da média e desvio padrão, visando um modelo mais estável. Além disso, foi ajustado apenas nos dados dos poços de treinamento para evitar *data leakage* do conjunto de validação ou teste.

Valores ausentes nos *logs* foram preenchidos com um valor de máscara, um valor distintivo que pode ser tratado posteriormente no modelo, a fim de fazer o modelo não aprender trechos onde os dados estão faltando ou estão incorretos.

As RNNs exigem dados sequenciais como entrada para capturar as dependências temporais ou contextuais inerentes aos *logs* de poço. Para isso, buscou-se encontrar um valor para definir o tamanho da janela de observação, permitindo que o modelo aprenda padrões e tendências com base nos dados de profundidade anteriores. Cada sequência é composta pelos *logs* de entrada para os $n - 1$ passos de tempo anteriores e o *log* alvo, VS, no último passo de tempo. Sequências contendo o valor de máscara, indicando a presença de um *not a number* em qualquer uma das propriedades, foram removidas para garantir a integridade dos dados de treinamento e validação.

4.1.3 Análise de Similaridade entre Poços

Antes de definir a estratégia de validação cruzada e treinamento dos modelos, realizou-se uma análise quantitativa de similaridade entre os cinco poços que possuem dados de VS disponíveis (Poços 1, 2, 3, 4 e 5). Esta análise é fundamental para compreender a heterogeneidade geológica do conjunto de dados e estabelecer expectativas realistas sobre a capacidade de generalização dos modelos de machine learning.

A hipótese subjacente é que modelos treinados em poços geologicamente similares tendem a generalizar melhor para poços de características semelhantes, enquanto apresentam desempenho degradado quando aplicados a poços com propriedades petrofísicas substancialmente diferentes. Esta análise permite, portanto, prever qualitativamente o desempenho esperado dos modelos em diferentes cenários de validação cruzada. (REFAEILZADEH; TANG; LIU, 2009)

Para quantificar a similaridade entre poços, empregaram-se quatro métricas complementares, cada uma capturando diferentes aspectos das características geológicas e geofísicas: distância estatística (*Wasserstein Distance*), distância de correlação, distância DTW (*Dynamic Time Warping*) e distância no espaço de features, além de uma distância combinada, obtida como uma combinação das mesmas. Para cada métrica de distância, calculou-se uma matriz simétrica 5x5, onde valores menores indicam maior similaridade geológica.

4.1.3.1 Distância Estatística (*Wasserstein Distance*)

A distância de Wasserstein, também conhecida como Earth Mover's Distance (EMD), mede a dissimilaridade entre as distribuições de probabilidade das propriedades petrofísicas de dois poços. Para cada par de poços $\bar{i}$ e $\bar{j}$, a distância estatística é definida como: (RUBNER; TOMASI; GUIBAS, 2000)

$$d_{\text{stat}}(i, j) = \frac{1}{F} \sum_{f=1}^F W(\mathbb{P}_i^f, \mathbb{P}_j^f) \quad (4.1)$$

onde F é o número de features, $\mathbb{P}_i^f$ é a distribuição empírica da feature f no poço i e W é a distância de Wasserstein de primeira ordem.

Esta métrica captura diferenças nas faixas de valores e formas das distribuições. Valores baixos indicam que os poços apresentam estatísticas similares para suas propriedades petrofísicas, sugerindo condições geológicas comparáveis.

As estatísticas descritivas obtidas e demonstradas na Figura 8 indicam uma distância mínima de 0.051 (Poços 4–5), uma distância máxima de 0.263 (Poços 2–3) e uma distância média de 0.143 ± 0.057 .

O par mais similar, formado pelos Poços 4 e 5, apresenta distância de apenas 0.051, evidenciando distribuições de valores praticamente idênticas para todas as *features* analisadas. Isso indica faixas similares de VP, RHO, POROSIDADE e SATURAÇÃO, além de condições

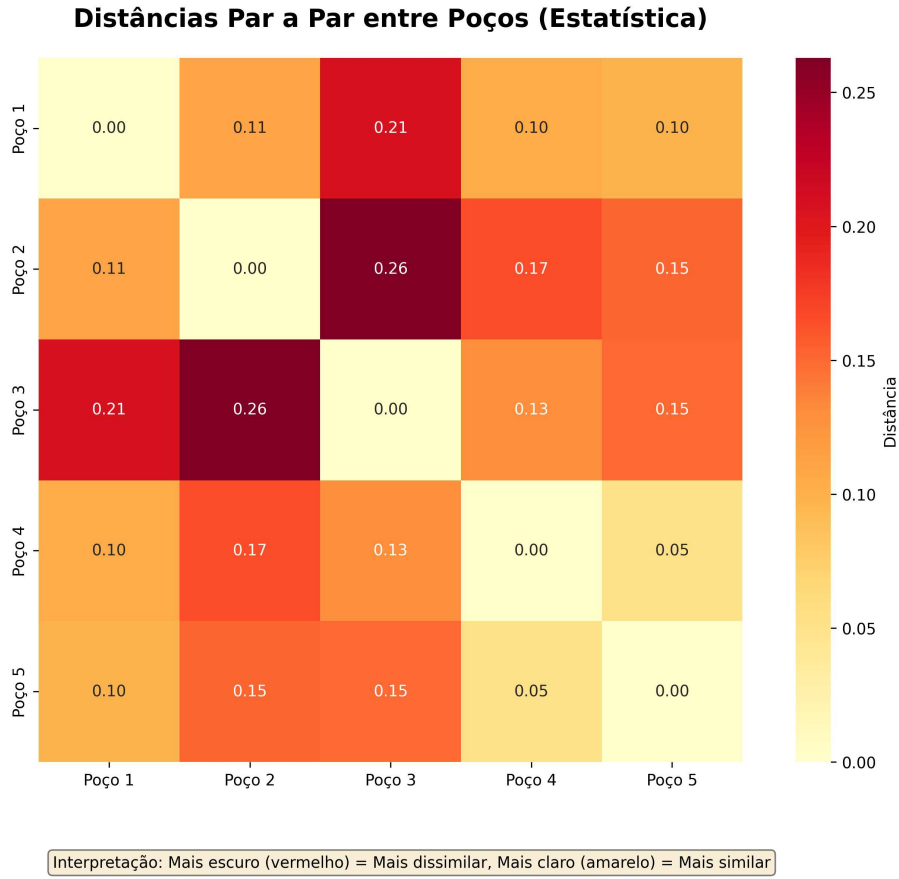


Figura 8 – *Heatmap* de distância estatística. Fonte: Autor.

geológicas e de reservatório comparáveis. Do ponto de vista de aprendizado de máquina, os dados desses poços podem ser considerados altamente intercambiáveis para fins de treinamento.

Por outro lado, o par mais dissimilar (Poços 2 e 3) apresenta uma distância de 0.263, sugerindo que as distribuições estatísticas das propriedades petrofísicas diferem substancialmente. O Poço 3 pode possuir faixas de valores sistematicamente deslocadas em relação ao Poço 2, possivelmente refletindo diferentes formações geológicas ou condições de deposição.

De modo geral, as distâncias estatísticas observadas são moderadas (máximo ≈ 0.26), indicando que, embora existam variações entre os poços, não há divergências extremas nas distribuições dos valores absolutos das propriedades petrofísicas.

4.1.3.2 Distância de Correlação

A distância de correlação quantifica diferenças nas relações entre features. Para cada poço, calcula-se a matriz de correlação de Pearson $C_i \in \mathbb{R}^{F \times F}$:

$$C_i[f_1, f_2] = \text{corr}(X_i^{f_1}, X_i^{f_2}) \quad (4.2)$$

A distância de correlação entre poços i e j é então definida como:

$$d_{\text{corr}}(i, j) = \|C_i - C_j\|_F \quad (4.3)$$

onde $\|\cdot\|_F$ denota a norma de Frobenius.

Esta métrica é bastante útil para modelos de machine learning, pois captura diferenças nas relações entre features (correlação VS-RHO, por exemplo). Modelos aprendem essas relações durante o treinamento, portanto, poços com correlações muito diferentes apresentarão padrões incompatíveis, dificultando a generalização.

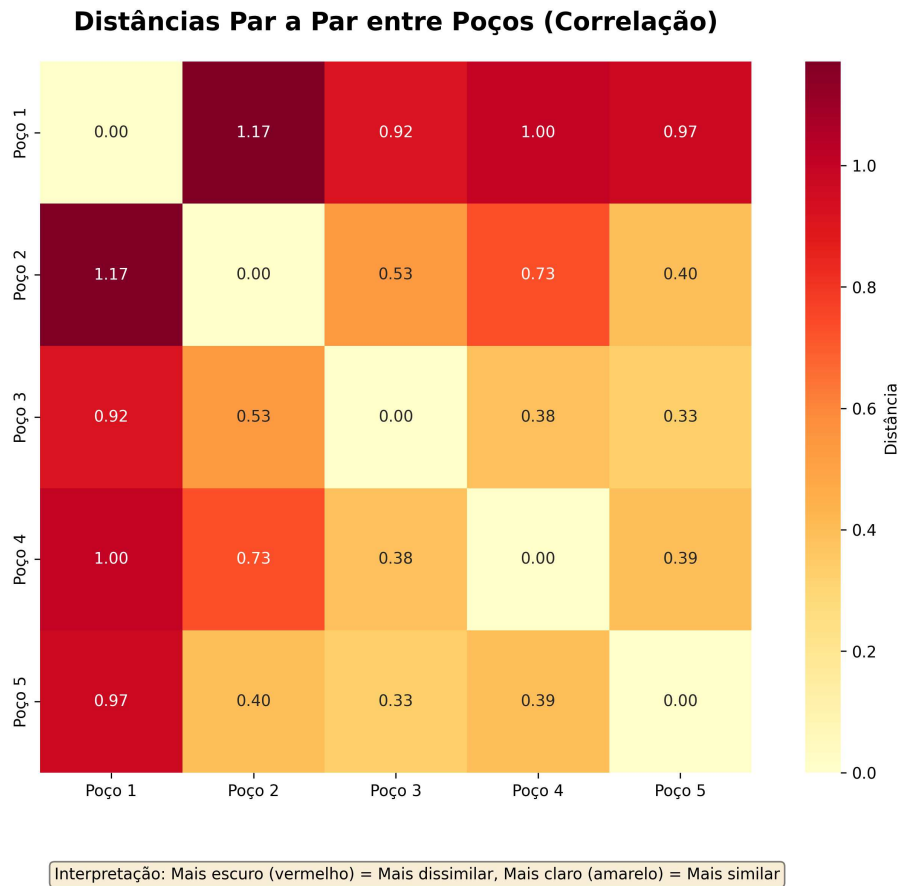


Figura 9 – *Heatmap* de distância correlação. Fonte: Autor.

As estatísticas descritivas na Figura 9 indicam uma distância mínima de 0.325 (Poços 3–5), uma distância máxima de 1.171 (Poços 1–2) e uma distância média de 0.682 ± 0.297 .

O principal achado desta análise é a identificação do Poço 1 como um *outlier* em termos de correlação. Este poço apresenta distância extremamente elevada (superior a 0.9) em relação a todos os demais, sendo que a distância de 1.171 observada entre os Poços 1 e 2 representa o maior valor entre todas as métricas avaliadas. Essa diferença indica que as relações entre as *features* no Poço 1 são fundamentalmente distintas das observadas nos outros poços. Por exemplo, enquanto a correlação entre VP e RHO pode ser de $r = 0.85$ no Poço 2, no Poço 1 ela pode cair para $r = 0.30$; de forma semelhante, a correlação entre VP e POROSIDADE pode ser negativa no Poço 1 e positiva nos demais.

Do ponto de vista de aprendizado de máquina, essa descoberta é particularmente relevante, pois modelos de regressão ou redes neurais aprendem relações do tipo “quando VP aumenta, RHO tende a aumentar”. Caso essas dependências variem fortemente entre poços,

como ocorre com o Poço 1, o modelo treinado em Poços 2–5 poderá aprender padrões que não se aplicam a ele, comprometendo a capacidade de generalização.

Por outro lado, observa-se um *cluster* de poços (3–5) com correlações similares, apresentando distâncias relativamente baixas entre si (entre 0.325 e 0.397). Isso indica que as relações entre as *features* são consistentes dentro desse grupo, tornando o treinamento conjunto desses poços uma estratégia promissora para modelos de aprendizado.

O Poço 2, por sua vez, apresenta distâncias moderadas em relação ao grupo 3–5 (entre 0.397 e 0.534), podendo atuar como um ponto intermediário ou “ponte” entre comportamentos distintos observados nos demais poços.

4.1.3.3 Distância DTW (Dynamic Time Warping)

Mede a similaridade entre sequências temporais (ou, neste caso, sequências de profundidade), permitindo alinhamentos não-lineares. Para sequências s_i e s_j representando as médias normalizadas das features ao longo da profundidade:(SAKOE; CHIBA, 1978)

$$d_{\text{DTW}}(i, j) = \min_{\pi} \sum_{(k,l) \in \pi} \|s_i[k] - s_j[l]\|_2 \quad (4.4)$$

onde π é um caminho de alinhamento ótimo.

DTW captura similaridades em padrões estratigráficos e sequências de formações, mesmo quando essas formações ocorrem em profundidades ligeiramente diferentes entre poços, ou seja, é robusto a variações locais de profundidade.

A análise das distâncias obtidas pela métrica DTW na Figura 10 apresentou valores variando entre 0.162 e 0.305, com média de 0.213 ± 0.049 . O menor valor (0.162) foi observado entre os Poços 2 e 5, indicando alta similaridade entre suas sequências estratigráficas. De forma consistente, o par 4–5 (0.163) também mostrou forte correspondência, reforçando a semelhança previamente identificada em outras métricas. Esses resultados sugerem que tais poços atravessam formações geológicas comparáveis, possivelmente dentro do mesmo contexto deposicional.

Por outro lado, o Poço 3 apresentou as maiores distâncias (0.288 com o Poço 1 e 0.305 com o Poço 2), destacando-se como o mais dissimilar. Essa diferença pode estar relacionada a variações no ambiente deposicional, à presença de estruturas tectônicas locais (como falhas ou dobras) ou à interceptação de uma coluna estratigráfica distinta.

De modo geral, a distribuição das distâncias DTW mostrou-se mais uniforme em comparação à métrica de correlação, sem presença de *outliers* acentuados. Isso indica que, embora haja certo grau de similaridade sequencial entre todos os poços, persistem diferenças relevantes nas suas estruturas verticais.

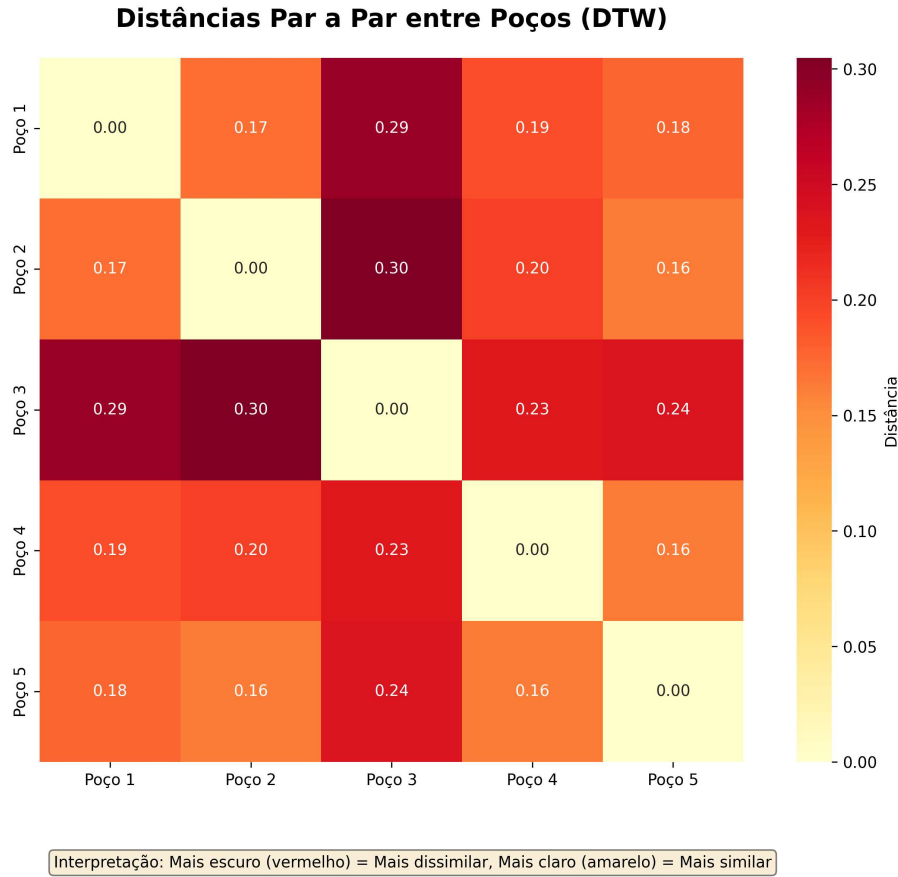


Figura 10 – *Heatmap* de distância *DTW*. Fonte: Autor.

4.1.3.4 Distância no Espaço de Features

A distância Euclidiana entre os vetores de features médias:

$$d_{\text{feat}}(i, j) = \|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|_2 \quad (4.5)$$

onde $\bar{\mathbf{x}}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} \mathbf{x}_i^{(n)}$ é o vetor de médias das features do poço i . eflète diferenças nas propriedades médias globais dos poços, indicando, por exemplo, se um poço é caracterizado por valores sistematicamente mais altos ou mais baixos de VP, RHO, etc.

A análise das distâncias obtidas e mostradas na Figura 11 revelou valores variando entre 0.940 e 4.559, com média de 2.964 ± 1.104 . O menor valor (0.940) foi observado entre os Poços 4 e 5, enquanto o maior (4.559) ocorreu entre os Poços 2 e 3. Essa ampla variação indica a presença de dois grupos bem distintos de poços em termos de suas características médias.

Identificaram-se dois agrupamentos principais. O **Grupo A**, composto pelos Poços 1 e 2, apresenta distância interna de 1.699 e caracteriza-se por valores médios mais elevados de propriedades como VP e RHO. O **Grupo B**, formado pelos Poços 3, 4 e 5, apresenta distâncias internas menores (0.940–2.432) e valores médios mais baixos dessas mesmas variáveis. As

Distâncias Par a Par entre Poços (Espaço de Características)

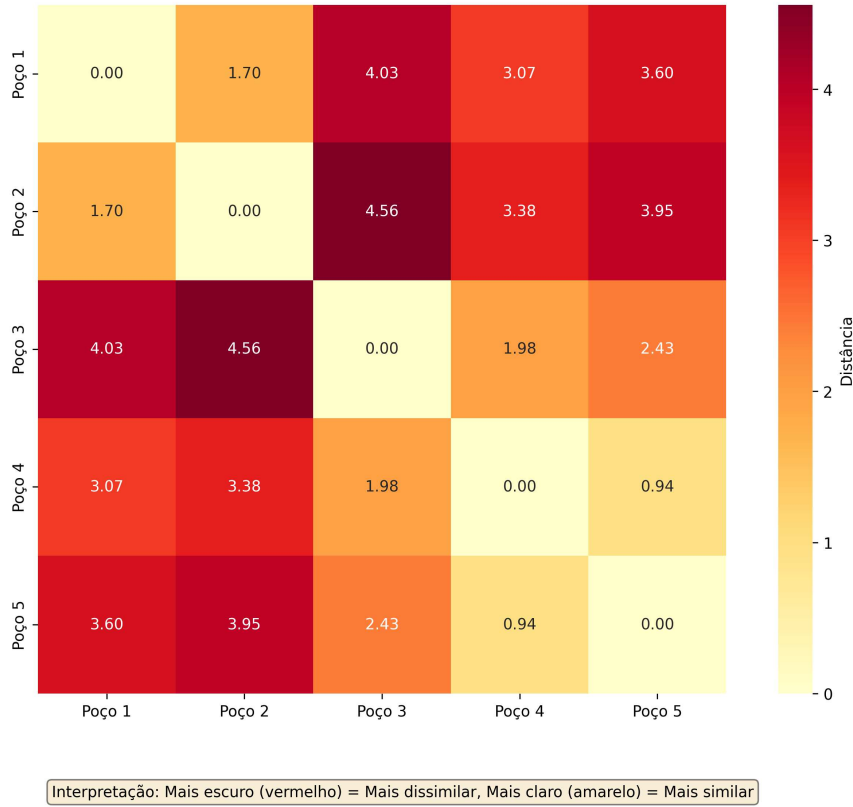


Figura 11 – *Heatmap* de distância DTW. Fonte: Autor.

distâncias entre os grupos são bastante elevadas (3.072–4.559), representando a maior separação observada entre quaisquer pares de poços em todas as métricas analisadas.

A magnitude dessas distâncias fornece a evidência mais clara da existência de dois tipos fundamentalmente distintos de reservatório ou formação. Possíveis explicações geológicas incluem: (i) diferença de profundidade, com o Grupo A possivelmente mais profundo e, portanto, sujeito a maiores pressões e velocidades (VP, RHO); (ii) variações litológicas, como maior consolidação ou cimentação no Grupo A; (iii) pertencimento a formações geológicas distintas; e (iv) diferenças nos tipos ou saturações de fluidos presentes.

Além disso, o par de Poços 4 e 5 apresentou a menor distância absoluta (0.940), confirmando que esses poços não apenas compartilham distribuições semelhantes, mas também exibem médias globais praticamente idênticas. Esse comportamento reforça a consistência interna do Grupo B e valida a separação observada entre os dois conjuntos principais de poços.

4.1.3.5 Distância Combinada

Para obter uma métrica única e interpretável, as quatro distâncias foram normalizadas para o intervalo [0, 1] através de min-max scaling e combinadas com pesos iguais:

$$d_{\text{combined}}(i, j) = \frac{1}{4} (d_{\text{stat}}^{\text{norm}}(i, j) + d_{\text{corr}}^{\text{norm}}(i, j) + d_{\text{DTW}}^{\text{norm}}(i, j) + d_{\text{feat}}^{\text{norm}}(i, j)) \quad (4.6)$$

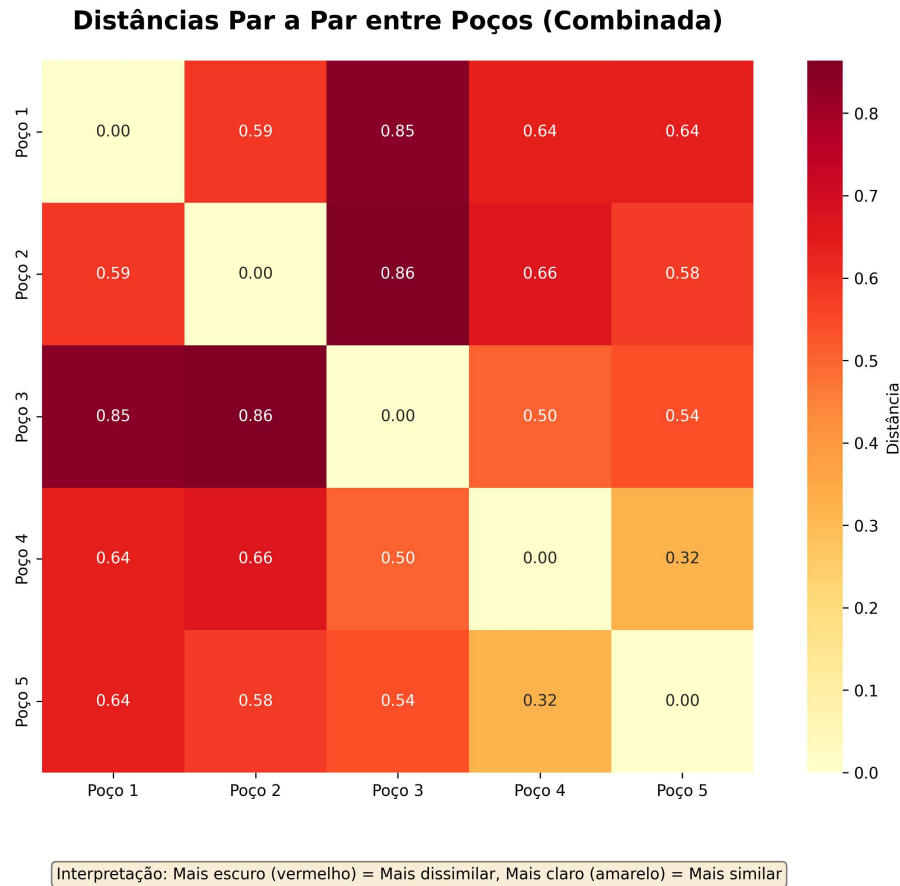


Figura 12 – *Heatmap* de distância combinada. Fonte: Autor.

Nota: A métrica final integra todos os aspectos de similaridade. Valores inferiores a 0.4 indicam alta similaridade, enquanto valores superiores a 0.7 indicam dissimilaridade crítica para aplicações de *machine learning*.

As estatísticas descritivas da métrica combinada indicam uma distância mínima de 0.318 (Poços 4–5), máxima de 0.864 (Poços 2–3) e média de 0.618 ± 0.151 . Esses valores refletem o comportamento agregado das diferentes métricas de similaridade, oferecendo uma visão global das relações entre os poços.

O **cluster de alta similaridade**, formado pelos Poços 4 e 5, apresenta a menor distância combinada (0.318), sendo consistentemente próximos em todas as métricas. Essa proximidade indica condições geológicas praticamente idênticas. Do ponto de vista de *machine learning*, há excelente potencial de generalização entre esses poços: um modelo treinado em um deles deve funcionar muito bem no outro, com desempenho esperado de R^2 entre 0.7 e 0.9.

O **grupo intermediário**, composto pelos Poços 3, 4 e 5, apresenta distâncias internas entre 0.318 e 0.540, formando um cluster moderadamente coeso. Isso sugere que o treinamento conjunto desses poços é viável, com generalização esperada entre $R^2 = 0.6$ –0.8.

Por outro lado, observa-se **alta dissimilaridade** entre os Poços 1–2 e o Poço 3, com distâncias de 0.850 (Poços 1–3) e 0.864 (Poços 2–3), os maiores valores observados. Essa diferença reflete uma heterogeneidade geológica significativa. Em termos de *machine learning*,

isso implica que modelos treinados nos Poços 1–2 provavelmente falharão ao prever o comportamento do Poço 3, e vice-versa. O desempenho esperado para essas combinações é baixo, com R^2 entre 0.1 e 0.3, podendo até ser negativo.

O padrão geral confirma a existência de **dois grupos principais**:

- **Grupo A:** Poços 1 e 2, com distância interna de 0.589;
- **Grupo B:** Poços 3, 4 e 5, com distâncias internas entre 0.318 e 0.540.

As distâncias entre grupos variam de 0.578 a 0.864, reforçando a separação entre os dois conjuntos e sugerindo que representam formações geológicas distintas.

O **ranking de similaridade** mostra os pares mais e menos semelhantes:

- **Pares mais similares:** Poços 4–5 (0.318, excelente similaridade), Poços 3–4 (0.504, boa similaridade) e Poços 3–5 (0.540, boa similaridade);
- **Pares mais dissimilares:** Poços 2–3 (0.864, alta dissimilaridade), Poços 1–3 (0.850, alta dissimilaridade) e Poços 2–4 (0.663, dissimilaridade moderada-alta).

Esses resultados reforçam o agrupamento observado em todas as análises anteriores, com destaque para a forte similaridade entre os Poços 4 e 5 e a marcante diferença entre os Poços 1–2 e 3, o que tem implicações diretas na transferibilidade de modelos de aprendizado de máquina entre diferentes regiões geológicas.

4.1.4 Visualizações Multidimensionais

Para visualizar as relações de similaridade da Subseção 4.1.3 em espaço bidimensional e validar os padrões identificados nas matrizes de distância, aplicaram-se técnicas de redução de dimensionalidade e *clustering* hierárquico.

4.1.4.1 Multidimensional Scaling (MDS)

O *Multidimensional Scaling* é uma técnica de redução de dimensionalidade que projeta objetos de um espaço de alta dimensionalidade (ou, neste caso, uma matriz de distâncias) em um espaço de menor dimensionalidade (tipicamente 2D ou 3D) de forma a preservar ao máximo as distâncias relativas entre os objetos. Diferentemente de técnicas como PCA (*Principal Component Analysis*), que requer dados numéricos brutos, o MDS opera diretamente sobre uma matriz de dissimilaridade, tornando-o ideal para nossa aplicação onde as distâncias já foram calculadas por múltiplas métricas. (KRUSKAL, 1964a)

Dada uma matriz de distâncias $\mathbf{D} \in \mathbb{R}^{n \times n}$, onde D_{ij} representa a distância entre os poços i e j , o MDS busca encontrar coordenadas $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^{n \times 2}$ no espaço bidimensional que minimizem a seguinte função de *stress*:

$$\text{Stress} = \sqrt{\frac{\sum_{i < j} (D_{ij} - d_{ij})^2}{\sum_{i < j} D_{ij}^2}} \quad (4.7)$$

onde $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$ é a distância Euclidiana entre os pontos i e j no espaço 2D projetado.

O *stress value* quantifica a qualidade da representação: (KRUSKAL, 1964a)

- Stress < 0.05: 5: Excelente representação (quase perfeita)
- Stress 0.05-0.10: Boa representação (confiável)
- Stress 0.10-0.20: Representação aceitável (interpretação cautelosa)
- Stress > 0.20: Representação pobre (não recomendada)

O uso do MDS para análise de poços é particularmente vantajoso porque ele opera diretamente sobre distâncias, não exigindo os dados brutos, apenas a matriz de distâncias previamente calculada. Além disso, o método preserva as relações originais, de modo que poços próximos no espaço bidimensional também apresentam distâncias pequenas na matriz original. Outra vantagem é a interpretabilidade, já que o MDS fornece uma visualização intuitiva de agrupamentos e outliers, e sua flexibilidade permite aplicação a diferentes métricas de distância, como Wasserstein, DTW, correlação, entre outras. (KRUSKAL, 1964b)

Cada métrica de distância captura um aspecto distinto da similaridade geológica; portanto, aplicar o MDS separadamente a cada métrica possibilita visualizar diferentes dimensões da relação entre poços. A convergência de padrões observada entre múltiplas métricas constitui um forte indicativo da robustez dos agrupamentos identificados, sugerindo que as formações geológicas compartilhadas entre poços próximos no espaço do MDS são consistentes sob diferentes perspectivas métricas.

4.1.4.2 Projeções MDS Individuais Detalhadas

Para análise mais aprofundada de como cada métrica contribui para o padrão geral de similaridade, as Figuras 13 a 16 apresentam cada projeção MDS individual em detalhe.

Interpretação da Figura 13: o stress excepcionalmente baixo (0.0009) indica que as distâncias de Wasserstein possuem estrutura bem definida em 2D. Observa-se a proximidade extrema entre Poços 4-5 e o relativo afastamento do Poço 3, sugerindo que suas distribuições estatísticas (faixas de valores, médias, variâncias) diferem dos demais poços. A configuração espacial indica que a variabilidade estatística das propriedades petrofísicas é relativamente bem comportada, sem outliers extremos em termos de distribuições.

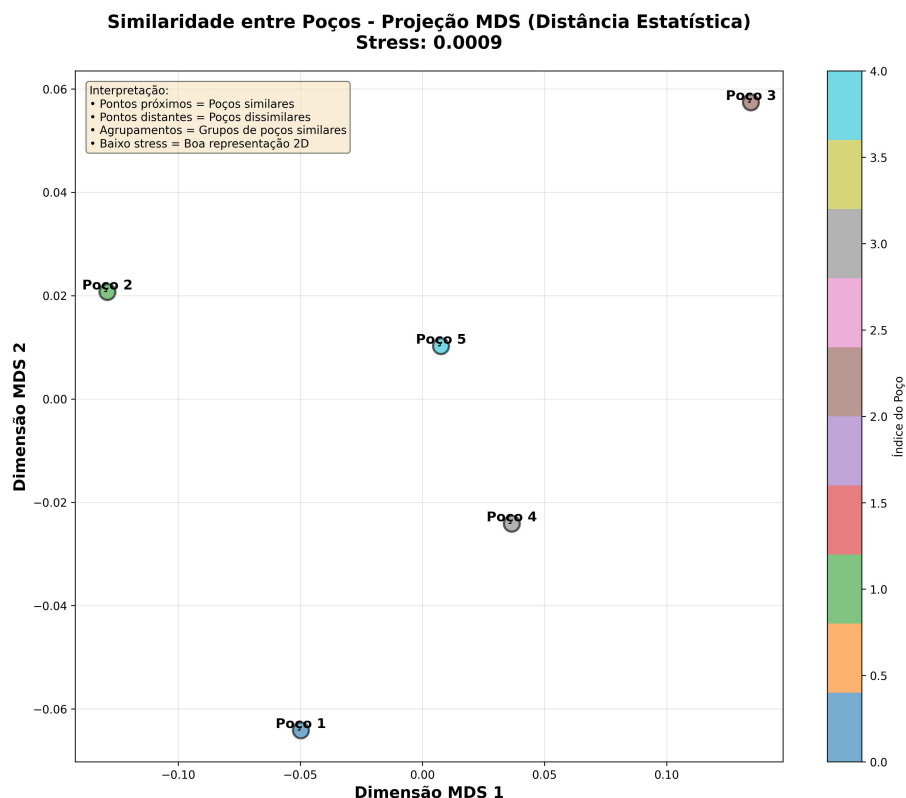


Figura 13 – **Projeção MDS da Distância Estatística (Wasserstein)**. Stress = 0.0009 (quase perfeito). Agrupa poços com distribuições similares de VP, RHO, POROSIDADE e SATURACAO. Fonte: Autor.

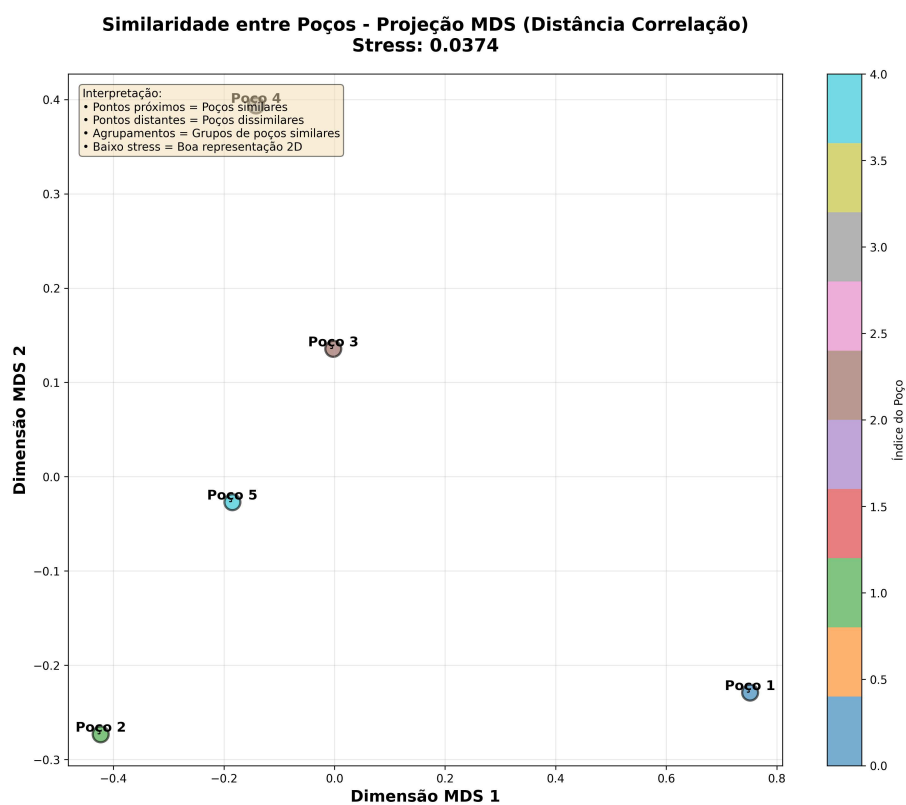


Figura 14 – **Projeção MDS da Distância de Correlação**. Stress = 0.0374 (excelente). Agrupa poços com padrões de correlação entre features similares. O Poço 1 apresenta-se extremamente isolado, indicando relações entre features fundamentalmente diferentes dos demais. Fonte: Autor.

Interpretação da Figura 14: esta projeção revela o achado mais crítico da análise: o Poço 1 está dramaticamente separado de todos os outros, confirmando visualmente as distâncias numéricas extremas (> 0.9) observadas na matriz de correlação. Para modelos de machine learning, isto é especialmente problemático, pois indica que as relações fundamentais entre features (e.g., como VP e RHO co-variam) diferem substancialmente neste poço. Um modelo treinado nos Poços 2-5 aprenderá padrões de co-variação que não se aplicam ao Poço 1. O stress de 0.0374 (excelente) valida que a separação observada é real, não artefato da projeção.

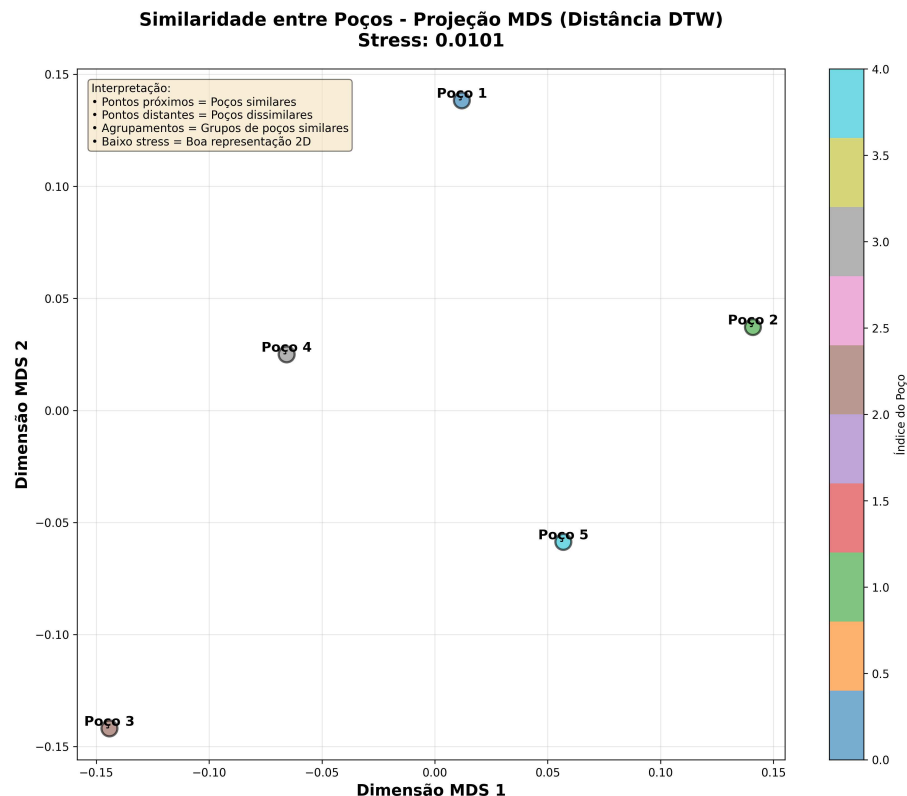


Figura 15 – **Projeção MDS da Distância DTW (Dynamic Time Warping)**. Stress = 0.0101 (excelente). Agrupa poços com sequências estratigráficas similares ao longo da profundidade. Fonte: Autor.

Interpretação da Figura 15: diferentemente das outras métricas, a projeção DTW mostra uma distribuição espacial mais uniforme, sem outliers extremos. Isto sugere que, embora os poços difiram em valores absolutos (estatística) e relações entre features (correlação), as sequências verticais de formações mantêm algum grau de similaridade entre todos os poços. Uma possível interpretação geológica é que todos os poços atravessam sequências deposicionais relacionadas (e.g., mesma bacia sedimentar), mas com variações locais em litologia e propriedades. Os Poços 4-5 mantêm-se próximos, confirmando o padrão de alta similaridade observado em outras métricas.

Interpretação da Figura 16: Esta projeção apresenta o stress mais alto entre as métricas (0.1573), ainda dentro do aceitável (< 0.2), refletindo a grande magnitude das distâncias no espaço de features (valores de 0.94 a 4.56). Apesar do stress ligeiramente elevado, o padrão de dois grupos espacialmente separados é evidente: Grupo A (Poços 1-2) claramente distinto

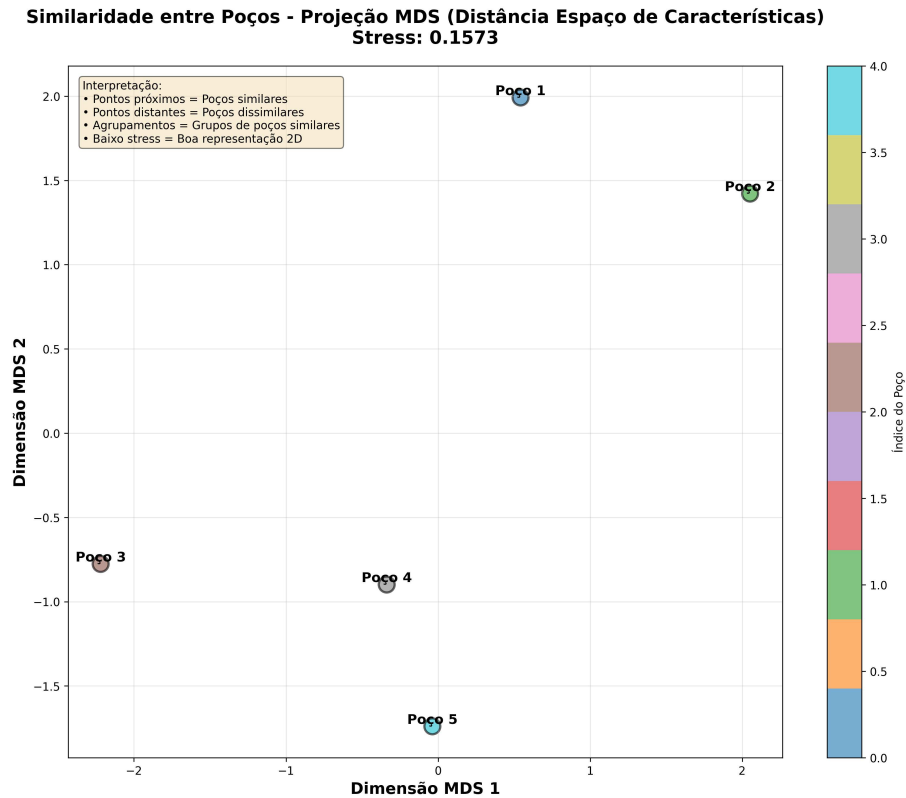


Figura 16 – **Projeção MDS da Distância no Espaço de Features**. Stress = 0.1573 (aceitável). Agrupa poços com propriedades médias globais similares. A separação clara entre dois grupos fornece evidência visual forte de tipos distintos de reservatório/formação. Fonte: Autor.

do Grupo B (Poços 3-4-5). Esta separação sugere que os grupos possuem valores médios de VP, RHO, POROSIDADE e SATURACAO substancialmente diferentes, consistente com a interpretação de diferentes profundidades (pressão), graus de consolidação/cimentação, ou tipos de formação geológica.

A Figura 17 apresenta a projeção MDS da matriz de distância combinada. A projeção sintetiza os padrões observados nas métricas individuais, revelando claramente dois grupos principais: Grupo A (Poços 1-2) e Grupo B (Poços 3-4-5), com separação espacial pronunciada. Dentro do Grupo B, os Poços 4-5 formam um cluster extremamente coeso, confirmando visualmente a menor distância combinada do dataset (0.318). O Poço 3 ocupa posição intermediária, mais afastado do cluster 4-5. O stress de 0.0515 (excelente) valida que a projeção 2D preserva fielmente as distâncias originais, conferindo alta confiabilidade à interpretação geológica.

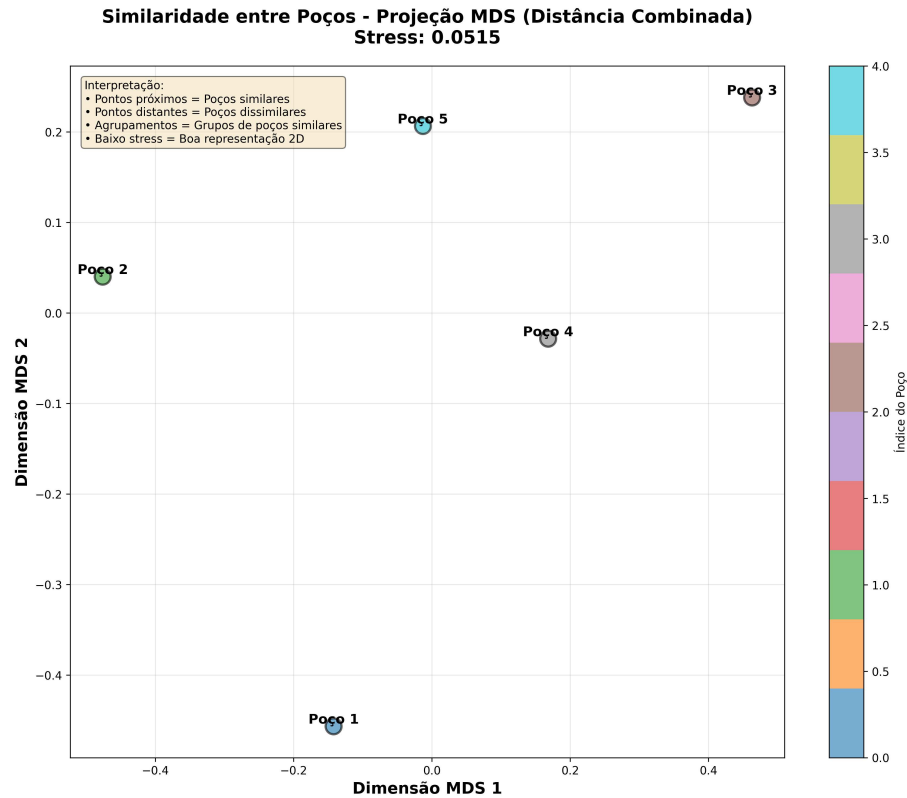


Figura 17 – **Projeção MDS da Distância Combinada entre Poços**. Cada ponto representa um poço. A proximidade espacial reflete similaridade geológica multidimensional integrando todos os aspectos analisados. Stress = 0.0515 (excelente). Fonte: Autor.

4.2 EXPERIMENTOS INICIAIS

Em um primeiro momento, para análise do potencial dos modelos elaborados, o conjunto de dados total foi dividido em conjuntos de treinamento, validação e teste para avaliar o desempenho do modelo de forma robusta e evitar *overfitting*. Para isso, os poços de número 3 a 6 foram usados como entrada para o treinamento, otimizando os pesos do modelo; o poço 2 foi usado como validação, servindo para monitorar o desempenho do modelo durante o treinamento, ajustar hiperparâmetros e implementar o *early stopping*; por último, os testes do modelo final foram feitos nos poços 1 e 7 pois, como mencionado na Seção 4.1.1, eles não possuem a propriedade VS.

A fase de experimentação teve como objetivo principal otimizar o desempenho do modelo baseado em LSTMs na predição de VS. Para isso, foram realizadas buscas iterativas tanto para encontrar a melhor combinação de *features* de entrada quanto para um primeiro refinamento dos hiperparâmetros do modelo. Uma lista de diferentes combinações de *features* de entrada foi definida para serem testadas de forma iterativa. Cada combinação representa um experimento distinto, permitindo avaliar a influência de diferentes conjuntos de *logs* no desempenho preditivo. As combinações testadas foram:

- VS, VP, RHO, GR, caliper, porosidade, saturação

- VS, VP, RHO, GR, caliper, porosidade, argilosidade
- VS, VP, RHO, porosidade, argilosidade
- VS, VP, RHO, GR, porosidade
- VS, VP, RHO, argilosidade
- VS, VP, RHO, GR
- VS, VP, RHO, porosidade

4.2.1 Modelagem

Para este trabalho, foi escolhida uma arquitetura baseada em LSTM. Como discutido acerca dos artigos citados no Capítulo 3, o uso de redes LSTM para a previsão de curvas de *logs* representa uma abordagem inovadora e que explora as características de continuidade e correlação dos dados dos registros. A forma com que as redes LSTM são capazes de armazenar um estado interno para capturar as dependências temporais de longo prazo a torna uma boa candidata para aprender e prever padrões complexos de formação dos registros.

O modelo é composto por camadas LSTM e camadas densas (*fully connected*). A parte LSTM é responsável por aprender as dependências de longo prazo nas sequências de dados de entrada, recebendo o número de *logs* sendo usados para a entrada (*input size*) e os processando com 64 neurônios (*hidden size*) em suas camadas ocultas, distribuídos em 2 camadas, no qual, entre cada uma, usa-se um *dropout* de 30%, para aplicar regularização e prevenir *overfitting*. Usou-se também processamento em lotes, com dimensão definida também pela quantidade de *features* sendo utilizadas. Usando a saída da camada LSTM é passada para a primeira camada densa, com 64 neurônios na entrada e 32 na saída e função de ativação ReLU. Por fim, mais uma camada de *droupout* de 30% é utilizada antes de seguir para a última camada densa, que mapeia a saída para um único neurônio, que irá representar a predição final de VS da rede.

Para auxiliar no treinamento, usou-se *early stopping* com paciência de 15 épocas, monitorando a perda na validação, auxiliando a não ocorrer *overfitting* e manter os melhores pesos da rede. Foi usado também um *scheduler* do *learning rate*, com paciência de 10 épocas e fator de 0.5, reduzindo o fator de aprendizado da rede conforme ela aprende os padrões, auxiliando a fazer ajustes finos nos pesos. Além disso, a rede e demais partes do código foram estruturadas para execução otimizada com GPU, reduzindo consideravelmente o tempo de treinamento e previsão, especialmente importante para redes com mais camadas e usando mais *logs* na entrada.

4.3 EXPERIMENTOS COM ARQUITETURAS AVANÇADAS: N-HITS E *SELF-ATTENTION BASED*

Após os experimentos preliminares com modelos baseados em LSTMs, que estabeleceram uma baseline de desempenho, prosseguiu-se com a avaliação de arquiteturas mais sofisticadas: N-HITS (Neural Hierarchical Interpolation for Time Series) e redes de Deep Learning baseadas no mecanismo de Self-Attention. Estas arquiteturas foram projetadas para capturar padrões multi-escala e dependências de longo alcance, potencialmente superando as limitações das redes recorrentes convencionais. (VASWANI et al., 2023; KRUSKAL, 1964b)

Diferentemente dos experimentos preliminares com LSTM, onde configurações mais simples foram testadas, as arquiteturas mais complexas foram submetidas a busca extensiva de hiperparâmetros (*grid search*) para identificar configurações ótimas. Esta abordagem sistemática visou:

1. **Explorar espaço de hiperparâmetros:** Testar múltiplas combinações de parâmetros arquiteturais e de treinamento
2. **Evitar viés de configuração:** Garantir que comparações entre modelos não sejam confundidas por escolhas arbitrárias de hiperparâmetros
3. **Identificar configurações robustas:** Selecionar modelos que generalizem bem através de múltiplos *folds*

4.3.1 Validação Cruzada *Leave-One-Well-Out (LOWOCV)*

Para avaliar robustamente a capacidade de generalização cross-well dos modelos, empregou-se a técnica de *Leave-One-Well-Out Cross-Validation (LOWOCV)*. Nesta estratégia, cada poço com VS disponível (Poços 2 a 6) serve como conjunto de teste em um fold, enquanto os demais quatro poços são utilizados para treinamento. Durante o treinamento, 20% do conjunto de dados é reservado para a validação. Portanto, a configuração final dos *folds* está na Tabela 5.

Tabela 5 – dos poços em treino e teste para os cinco folds.

Fold	Treino	Teste
0	Poços 3, 4, 5, 6	Poço 2
1	Poços 2, 4, 5, 6	Poço 3
2	Poços 2, 3, 5, 6	Poço 4
3	Poços 2, 3, 4, 6	Poço 5
4	Poços 2, 3, 4, 5	Poço 6

Essa metodologia foi adotada por oferecer uma avaliação mais realista do desempenho dos modelos em cenários de aplicação prática. Essa abordagem simula a situação de uso

real, em que o modelo é treinado exclusivamente em poços conhecidos e utilizado para prever propriedades em um poço completamente novo, sem qualquer dado deste poço no treinamento ou então com buracos e trechos faltantes. Além disso, ela permite avaliar a capacidade de generalização geológica do modelo, ao testar se o modelo realmente aprendeu relações físicas transferíveis entre diferentes regiões do reservatório. (KOHAVI, 1995)

Outra vantagem é a compatibilidade com análises de similaridade geológica, possibilitando correlacionar o desempenho do modelo com a distância entre o conjunto de treino e o poço de teste, conforme discutido na Subseção 4.1.3. Por fim, a estratégia também maximiza o uso de dados disponíveis: com cinco poços ao todo, cada fold utiliza quatro deles para treinamento e um distinto para teste, garantindo que todos os poços sejam avaliados exatamente uma vez. (VARMA; SIMON, 2006)

Apesar de suas vantagens, o método apresenta limitações conhecidas. Cada fold treina o modelo com dados de apenas quatro poços, o que pode ser insuficiente para arquiteturas muito complexas. Além disso, é esperada uma alta variância de desempenho entre os folds devido à heterogeneidade geológica identificada na análise de similaridade. Por fim, há desbalanceamento entre os poços, pois alguns contêm significativamente mais amostras que outros, o que pode influenciar o processo de treinamento e avaliação. (CAWLEY; TALBOT, 2010)

4.3.2 *Grid Search* de Hiperparâmetros

Assim como nos experimentos preliminares com LSTM usou-se a técnica de grid search para a busca de hiperparâmetros, também usou-se nos modelos mais complexos, com a diferença de que mais parâmetros e em faixas de valores maiores foram usados, além de buscas mais inteligentes, pois os modelos são computacionalmente mais exigentes. Com a busca extensiva de hiperparâmetros, visou-se:

1. **Explorar espaço de hiperparâmetros:** testar múltiplas combinações de parâmetros arquiteturas e de treinamento;
2. **Evitar viés de configuração:** garantir que comparações entre modelos não sejam confundidas por escolhas arbitrárias de hiperparâmetros;
3. **Identificar configurações robustas:** selecionar modelos que generalizem bem através de múltiplos folds

4.3.3 Arquitetura *N-HITS*

A Tabela 6 apresenta o espaço de busca utilizado nos experimentos com N-HITS, totalizando aproximadamente 150-200 combinações.

Número Total de Configurações Testadas: ~150–200 combinações.

Configuração do Melhor Modelo:

Tabela 6 – Hiperparâmetros testados e respectivas justificativas.

Hiperparâmetro	Valores Testados	Justificativa
Configuração de Stacks	[16, 8, 4, 2, 1], [8, 4, 1], [4, 2, 1, 1]	Diferentes níveis de decomposição hierárquica temporal
Blocos por Stack	[1, 1, 2, 2, 2], [2, 2, 2], [1, 1, 1, 1]	Profundidade de processamento por escala
Hidden Sizes	[256, 256], [512, 512], [128, 256]	Capacidade representacional das MLPs
Theta Sizes	[2, 2, 4, 4, 4], [4, 4, 4], [2, 2, 2, 2]	Dimensionalidade dos coeficientes de forecast
Learning Rate	1×10^{-3} , 5×10^{-4} , 3×10^{-4}	Taxa de convergência
Batch Size	16, 32, 64	Balanco entre velocidade e estabilidade
Dropout	0.1, 0.2, 0.3	Regularização contra overfitting
Features	4, 5, 6, 7 features	Conjuntos de entrada (ver Tabela 4)

Após *grid search* extensivo através dos 5 folds de validação cruzada, a configuração que apresentou melhor desempenho médio foi:

- **Stack types:** [8, 4, 1] (3 stacks para alta, média e baixa frequência)
- **N blocks:** [2, 2, 2] (2 blocos MLP por stack)
- **Hidden sizes:** [[512, 512], [512, 512], [512, 512]]
- **Theta sizes:** [4, 4, 4] (4 coeficientes de forecast por stack)
- **Dropout:** 0.2
- **Learning rate:** 5×10^{-4}
- **Batch size:** 32
- **Optimizer:** AdamW (com *weight decay* de 0.01)
- **Scheduler:** ReduceLROnPlateau
- **Loss:** Huber Loss ($\delta = 1.0$)

4.3.4 Arquitetura Baseada em Self-Attention

Nas variantes baseadas em CNN com Self-Attention que foram desenvolvidas e avaliadas, uso-se os parâmetros da Tabela 7. Usou-se um modelo híbrido com a premissa de que combinar as capacidades de extração local das CNNs com a capacidade de armazenar e entender relações a longo prazo dos blocos Transformer, resultaria em um modelo robusto e capaz de entender o funcionamento das propriedades geológicas.

Número Total de Configurações Testadas: ~100–120 combinações.

Tabela 7 – Hiperparâmetros testados para os modelos baseado em Transformer.

Hiperparâmetro	Valores Testados	Justificativa
Embed Dimension	64, 128, 256	Dimensionalidade das representações
Num Heads	4, 8, 16	Número de cabeças de atenção
Num Blocks	2, 3, 4	Profundidade da rede Transformer
Learning Rate	1×10^{-3} , 5×10^{-4} , 3×10^{-4}	Taxa de convergência
Batch Size	16, 32, 64	Balanco entre velocidade e estabilidade
Dropout	0.1, 0.2, 0.3	Regularização
Features	4, 5, 6 features	Conjuntos de entrada

4.3.4.1 Arquitetura Base

Em um primeiro momento, utilizou-se uma arquitetura menos complexa abrangendo a combinação de CNN e Transformers, :

- CNN Feature Extractor: duas camadas convolucionais 1D (4-64-128) com tamanho de kernel 3
- Positional Encoding: Encoding sinusoidal adicionado após a convolução
- Attention Blocks: 3 blocos Transformer padrão com Post-LayerNorm
- Output: droptou seguido de camada linear

Configuração do Melhor Modelo Base:

- **Embed dim:** 128
- **Num heads:** 8 (cada cabeça processa 16 dimensões)
- **Num blocks:** 3 blocos Transformer
- **Dropout:** 0.2
- **Learning rate:** 5×10^{-4}
- **Batch size:** 32
- **Optimizer:** Adam
- **Positional encoding:** Sinusoidal (fixo, não-treinável)
- **Sequence aggregation:** Último token
- **Output:** Linear simples ($128 \rightarrow 1$)

Número de Parâmetros: ~420K parâmetros treináveis.

4.3.4.2 Arquitetura Avançada

A segunda variante incorpora melhorias arquiteturais modernas com o objetivo de potencialmente aprimorar o desempenho, ainda que à custa de maior complexidade.

Principais Diferenças em Relação à Versão Base:

- **CNNs com conexões residuais:** três blocos convolucionais com *skip connections*.
- **Positional embeddings aprendíveis:** substituição do encoding sinusoidal fixo por parâmetros treináveis.
- **Pre-Layer Normalization:** normalização aplicada antes das subcamadas, em vez de depois.
- **Attention pooling:** agregação ponderada de toda a sequência, ao invés de considerar apenas o último token.
- **MLP de saída profundo:** três camadas ($128 \rightarrow 256 \rightarrow 128 \rightarrow 1$) no lugar de uma projeção linear simples.
- **Layer normalization após CNNs:** adicionada para estabilização adicional das ativações.

Configuração da Arquitetura Avançada:

Devido ao custo computacional adicional, esta variante foi avaliada com um espaço de busca reduzido, adotando a seguinte configuração experimental:

- **Embed dim:** 160 (maior que a versão base)
- **Número de cabeças de atenção:** 10
- **Número de blocos Transformer:** 3
- **Dropout:** 0.2
- **Learning rate:** 5×10^{-4}
- **Batch size:** 32
- **Otimizador:** AdamW (com *weight decay*)
- **Attention pooling:** habilitado

Número de Parâmetros: aproximadamente 650K (um aumento de cerca de 55% em relação à versão base).

Trade-off: maior capacidade representacional em troca de um risco ampliado de *overfitting* e de maior tempo computacional (cerca de 56% mais lento por época).

4.3.5 Pipeline de Treinamento unificado

Todos os modelos com arquiteturas avançadas (N-HiTS e variantes baseadas em self-attention) seguiram o mesmo pipeline de treinamento, garantindo comparabilidade.

O treinamento foi conduzido de forma independente para cada *fold*, seguindo um procedimento padronizado de preparação, configuração e avaliação. Primeiramente, aplicou-se um *RobustScaler* específico para cada *fold*, ajustado exclusivamente sobre os dados de treino, garantindo a ausência de vazamento de informação. As amostras foram organizadas em janelas deslizantes de tamanho 10 e passo 1 (*window_size=10, stride=1*), removendo-se as sequências que continham valores mascarados.

Em cada treinamento, utilizou-se quatro poços para o conjunto de treino e validação e um poço para testes. O processo foi executado por até 200 épocas, com *early stopping* configurado com paciência de 15 épocas, interrompendo o treinamento caso não houvesse melhora no erro de validação. Um *scheduler* com paciência de 10 épocas e fator 0.5 foi empregado para reduzir a taxa de aprendizado de forma adaptativa, também monitorando a perda de validação.

Durante o treinamento, as métricas *train_loss*, *val_loss*, *val_R²* e *val_RMSE* foram registradas a cada época, permitindo o acompanhamento da convergência e do desempenho do modelo. O modelo correspondente ao menor valor de *val_loss* foi salvo automaticamente, e todo o histórico de treinamento foi exportado para posterior análise quantitativa e visual.

4.3.5.1 Seleção do Melhor Modelo

Para cada arquitetura, o modelo final foi selecionado com base em um processo sistemático de avaliação em cinco *folds* (correspondentes aos cinco poços disponíveis). O principal critério de seleção foi o coeficiente de determinação médio (R^2) obtido na validação cruzada, utilizado como métrica primária de desempenho. Como critérios adicionais, consideraram-se a estabilidade entre os *folds* (avaliada pelo desvio padrão de R^2), o erro quadrático médio (*RMSE*) médio, a ausência de sinais de sobreajuste (analisando a diferença entre as perdas de treino e validação) e, por fim, a complexidade do modelo. Entre modelos com desempenho semelhante (diferença inferior a 5% em R^2), optou-se por arquiteturas mais simples, seguindo o princípio da parcimônia.

O processo sistemático de seleção envolveu o treinamento de todas as configurações do *grid search* em cada um dos cinco *folds*, seguido do cálculo das métricas agregadas: média e desvio padrão de R^2 , *RMSE*, *MAE* e *MSE*. As configurações foram então ordenadas de forma decrescente pelo R^2 médio, e as cinco melhores foram inspecionadas manualmente para detectar possíveis indícios de sobreajuste severo. A configuração final foi escolhida considerando o melhor equilíbrio entre desempenho médio, estabilidade entre *folds* e complexidade arquitetural.

Observou-se uma alta variabilidade do R^2 entre os *folds* (desvio padrão em torno de 0.4–0.5) em todas as configurações testadas. Essa variação consistente indica que a principal

fonte de instabilidade está relacionada à heterogeneidade geológica entre os poços, como discutido anteriormente, e não à instabilidade de configurações específicas. Tal resultado reforça a hipótese de que a generalização de um modelo global único é desafiada pela variabilidade intrínseca dos dados.

4.3.5.2 Agregação de Métricas em Validação Cruzada

Para cada modelo, foram reportadas estatísticas individuais e agregadas sobre as métricas de desempenho. As métricas por *fold* incluem R^2 , $RMSE$, MAE e MSE . As métricas agregadas foram calculadas por meio da média aritmética ($\bar{R}^2$, $\overline{RMSE}$, $\overline{MAE}$, $\overline{MSE}$) e do desvio padrão correspondente (σ_{R^2} , σ_{RMSE} , σ_{MAE} , σ_{MSE}).

A média representa o desempenho esperado do modelo em um cenário típico, enquanto o desvio padrão expressa sua estabilidade e robustez entre diferentes poços. Assim, modelos que apresentam desempenho ligeiramente inferior, mas com menor variabilidade entre *folds*, podem ser preferíveis a modelos mais instáveis, especialmente quando o objetivo é a generalização para novos poços.

Os resultados detalhados, incluindo as análises *fold*-por-*fold*, visualizações de predições, distribuições de erro e correlações com a similaridade geológica, são apresentados integralmente no Capítulo 5.

5 RESULTADOS

Após a realização dos experimentos, foram feitas diversas análises dos resultados, a fim de garantir que o modelo obtido faz previsões consistentes e de que possui boa generalização, mas também que seus resultados sejam geologicamente condizentes, demonstrando ser capaz de ser utilizado em contextos de aplicações reais.

5.1 MÉTRICAS E MÉTODOS DE AVALIAÇÃO E VALIDAÇÃO

A avaliação do desempenho do modelo proposto foi feita utilizando as métricas R^2 , RMSE e MAE, todas apresentadas na Seção 2.2.2. A escolha dessas métricas específicas justifica-se pela natureza do problema de regressão e pelas informações complementares que cada uma oferece sobre a qualidade das previsões.

O R^2 é uma métrica fundamental que indica a proporção da variância na variável dependente (o *log* alvo, VS, neste caso) que é previsível a partir das variáveis independentes (*logs* de entrada). Um valor de R^2 próximo de 1 sugere que o modelo explica uma grande parte da variabilidade dos dados reais, indicando um bom ajuste.

O RMSE e o MAE são métricas de erro que quantificam a magnitude dos desvios entre os valores preditos e os valores reais. O MSE calcula a média dos quadrados das diferenças, penalizando erros maiores de forma mais significativa. A RMSE é a raiz quadrada do MSE, colocando o erro na mesma unidade da variável predita, tornando-a mais interpretável. O MAE, por sua vez, calcula a média dos valores absolutos das diferenças, sendo menos sensível a *outliers*.

Em conjunto, essas métricas fornecem uma visão abrangente do desempenho do modelo: o R^2 avalia a capacidade de explicação da variância, o RMSE quantifica o erro com sensibilidade a desvios grandes, além de dar uma métrica na mesma unidade de medida da variável de saída, e o MAE oferece uma estimativa do erro médio menos influenciada por *outliers*. Esta combinação é essencial para avaliar a precisão e a confiabilidade das previsões em um contexto geológico.

5.2 AVALIAÇÃO DO DESEMPENHO DOS EXPERIMENTOS INICIAIS

Após a realização dos experimentos descritos na Seção 4.2, a combinação de *features* que produziu o melhor resultado foi utilizando VP, RHO, GR, caliper, porosidade e saturação, para a qual se obtiveram as métricas:

- **RMSE:** 0,1265
- **MSE:** 0,0160
- **R^2 :** 0,6008

Utilizando o poço de testes para esse modelo, a fim de poder comparar sua previsão com um poço que possui dados de VS presentes, realizou-se o *plot* da Figura 18, na qual o eixo *y* apresenta a profundidade e o eixo *x* contém o valor predito e o valor real de VS.

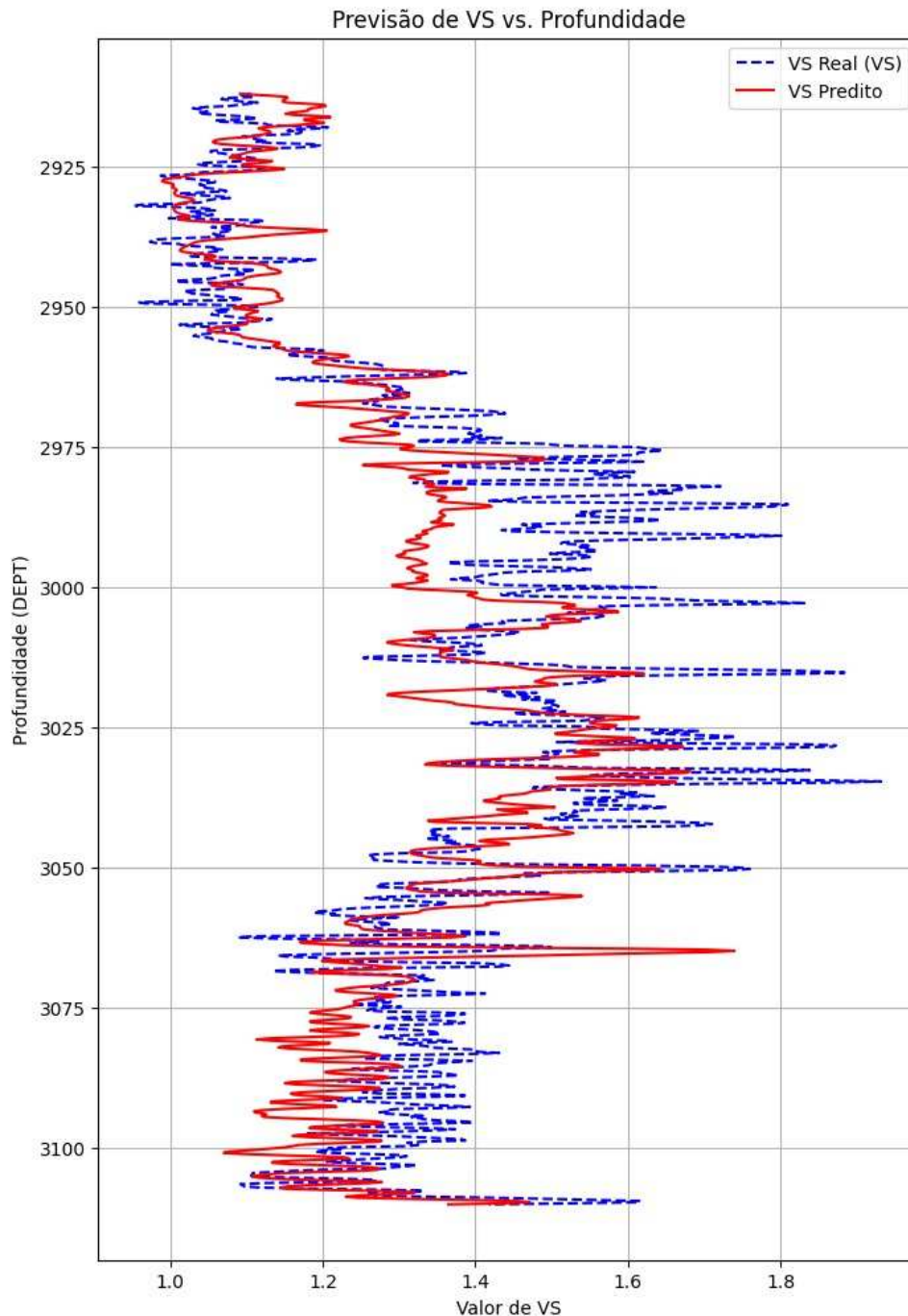


Figura 18 – *Plot* de DEPT x VS (real e previsto)

Além disso, outro gráfico que pode apresentar satisfatoriamente a comparação entre os valores preditos e os valores reais do *log* é utilizar os valores da previsão no eixo *y* e os valores reais no eixo *x*, com o qual obtém-se a Figura 19.

Durante os experimentos, notou-se que a rede pode apresentar instabilidades em diferentes iterações, ou seja, treinar repetidas vezes o modelo resulta em variações consideráveis

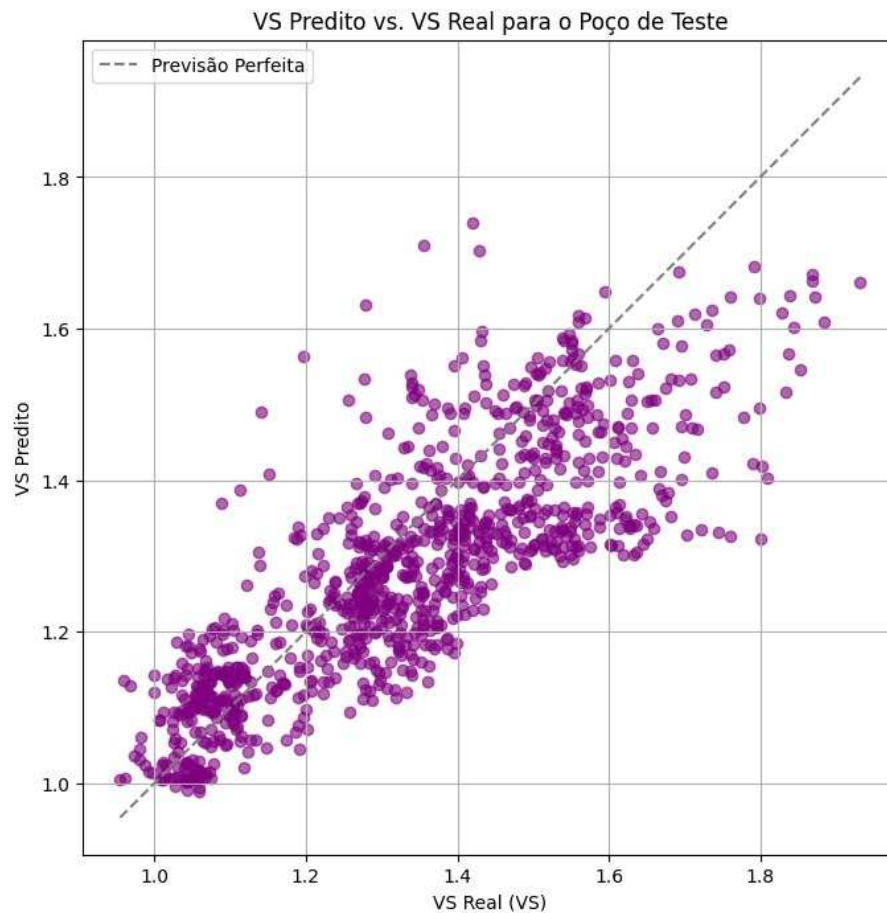


Figura 19 – *Plot de VS real x VS previsto*

na qualidade dos modelos obtidos. Isso pode ocorrer por conta da aleatoriedade inerente aos modelos de ML, bem como pode mostrar pontos a serem considerados sobre a capacidade ou não de generalização satisfatória da arquitetura atualmente adotada.

5.3 ANÁLISE DE RESULTADOS FINAIS

Após os experimentos preliminares com o modelo LSTM, foram conduzidos experimentos mais abrangentes com arquiteturas baseadas em mecanismos de atenção e modelos especializados em séries temporais. Esta seção apresenta os resultados obtidos com os modelos N-HiTS e modelos baseados em Transssformers, este último em duas configurações: uma versão base e uma versão avançada.

5.3.1 Resultados do Modelo N-HiTS

O modelo N-HiTS foi avaliado através de uma estratégia de validação cruzada com 5 folds, permitindo uma análise robusta da capacidade de generalização do modelo em diferentes subconjuntos dos dados.

5.3.1.1 Configuração e Hiperparâmetros

Os experimentos com N-HiTS exploraram diferentes combinações de hiperparâmetros e conjuntos de features, resultando em configurações distintas para cada fold da validação cruzada. As principais variações incluíram:

- **Conjuntos de features:** Variando entre 4 e 7 variáveis de entrada, incluindo VP (velocidade de onda P), RHO (densidade), GR (raios gama), CALIPER, POROSIDADE, SATURACAO e ARGILOSIDADE
- **Arquitetura hierárquica:** Empilhamentos (stacks) com diferentes tamanhos de interpolação, variando de configurações mais profundas (16-8-4-2-1) até mais rasas (8-4-1)
- **Capacidade representacional:** Camadas ocultas de 256 a 1024 neurônios, com 1 a 3 blocos por stack
- **Estratégias de regularização:** Dropout entre 0,2 e 0,3
- **Otimização:** Taxa de aprendizado de 0,0003 a 0,0005, batch sizes de 16 a 32, utilizando Adam ou AdamW

5.3.1.2 Desempenho por Fold

A Tabela 8 apresenta as métricas de teste para o melhor modelo de cada fold, evidenciando a variabilidade significativa no desempenho entre diferentes partições dos dados.

Tabela 8 – Métricas de teste dos melhores modelos N-HiTS por fold

Fold	Features	R ²	RMSE	MSE	MAE
0	VP, RHO, POR, SAT	0,4175	0,1526	0,0233	0,1230
1	VP, RHO, GR, CAL, POR, SAT	-0,8771	0,1185	0,0141	0,1052
2	VP, RHO, GR, CAL, POR, SAT, ARG	0,0490	0,1148	0,0132	0,0794
3	VP, RHO, GR, POR, SAT	0,6393	0,1103	0,0122	0,0843
4	VP, RHO, GR, POR, SAT	0,4248	0,1099	0,0121	0,0835
Média	-	0,1307	0,1212	0,0150	0,0951
Desvio Padrão	-	0,5604	0,0177	0,0049	0,0185

Observações principais:

1. **Variabilidade extrema no R²:** O coeficiente de determinação variou drasticamente entre os folds, desde -0,8771 (Fold 1) até 0,6393 (Fold 3), resultando em uma média de apenas

0,1307 com desvio padrão de 0,5604. Valores negativos de R^2 indicam que o modelo teve desempenho inferior à predição pela média dos dados.

2. **Melhor desempenho:** O Fold 3 apresentou o melhor desempenho geral, com $R^2 = 0,6393$ e o menor RMSE (0,1103), utilizando 5 features (VP, RHO, GR, POROSIDADE, SATURACAO) e uma arquitetura hierárquica profunda (16-8-4-2-1).
3. **Instabilidade com mais features:** O Fold 1, que utilizou 6 features, apresentou o pior R^2 (-0,8771), sugerindo que a adição de features não necessariamente melhora o desempenho e pode levar a overfitting ou dificuldades de generalização.
4. **Consistência em métricas de erro absoluto:** Apesar da variação no R^2 , as métricas RMSE e MAE mostraram-se mais estáveis entre os folds (desvio padrão de 0,0177 e 0,0185, respectivamente), indicando que o modelo mantém uma magnitude de erro relativamente consistente, mesmo quando falha em capturar a variância dos dados.

5.3.1.3 *Análise de Configurações Exploradas*

A análise dos resultados médios de diferentes configurações de hiperparâmetros testadas revelou padrões importantes sobre a adequação do modelo N-HiTS para este problema:

- **Desempenho geral insatisfatório:** Das 15 configurações principais testadas, 13 apresentaram R^2 médio negativo, indicando desempenho sistematicamente inferior à baseline (predição pela média)
- **Melhor configuração média:** A configuração com features VP, RHO, GR, CALIPER, POROSIDADE e SATURACAO, utilizando stacks [4,2,1,1] e otimizador Adam, alcançou R^2 médio de -0,045, o melhor resultado médio entre todas as configurações, mas ainda negativo
- **Impacto da complexidade da arquitetura:** Arquiteturas mais profundas (5 stacks) e com maior capacidade (hidden sizes de 1024) tenderam a apresentar desempenhos piores, com R^2 médio chegando a -0,95, sugerindo problemas de overfitting
- **Efeito do conjunto de features:** Configurações com 4 features (VP, RHO, POROSIDADE, SATURACAO) apresentaram desempenho altamente variável, com R^2 médio entre -0,28 e -0,95, dependendo dos demais hiperparâmetros

5.3.1.4 *Discussão sobre a Inadequação do N-HiTS*

Os resultados evidenciam que o modelo N-HiTS, apesar de sua arquitetura especializada para séries temporais, não se mostrou adequado para o problema de predição de velocidade de onda cisalhante a partir de perfis de poços. Diversos fatores podem contribuir para este resultado:

1. **Natureza dos dados:** Os perfis de poços representam medições contínuas em profundidade, caracterizando-se mais como dados espaciais univariados do que como séries temporais tradicionais com dependências sazonais ou tendências temporais que o N-HiTS foi projetado para capturar.
2. **Pressuposições arquiteturais:** O N-HiTS utiliza interpolação hierárquica e decomposição em múltiplas escalas temporais, pressupondo padrões periódicos ou sazonais que podem não estar presentes nos dados geofísicos.
3. **Heterogeneidade geológica:** A alta variabilidade entre folds sugere que diferentes poços ou formações geológicas apresentam características muito distintas, dificultando a captura de padrões generalizáveis pelo modelo hierárquico. Tal resultado pode ser explicado pelas conclusões da análise das métricas de similaridade entre os poços na Subseção 4.1.3, onde foram identificados grupos de poços, nos quais indicava-se ser possível que modelos treinados para uma configuração geral não seriam capazes de generalizar bem poços com propriedades diferentes das observadas durante o treinamento.
4. **Limitações na captura de relações complexas:** As relações físicas entre as propriedades petrofísicas (densidade, porosidade, saturação) e a velocidade de onda cisalhante podem ser altamente não-lineares e dependentes de contextos geológicos específicos, requerendo mecanismos mais sofisticados de modelagem de dependências.

Essa análise motivou a investigação de arquiteturas alternativas baseadas em mecanismos de self-attention, capazes de modelar dependências de longo alcance sem as pressuposições específicas de séries temporais do N-HiTS.

5.3.2 Resultados dos Modelos Baseados em Self-Attention

Os modelos baseados em Self-Attention foram desenvolvidos para aproveitar mecanismos de atenção na captura de relações complexas entre as diferentes propriedades petrofísicas. Duas versões foram avaliadas: uma arquitetura base e uma arquitetura avançada.

5.3.2.1 *Modelo Base*

5.3.2.1.1 Configuração do Modelo

O modelo base representa uma implementação simplificada da arquitetura self-attention, com as seguintes características:

- **Features utilizadas:** VP, RHO, POROSIDADE, ARGILOSIDADE (4 variáveis)
- **Arquitetura:**

- Dimensão de embedding: 32
- Número de cabeças de atenção: 4
- Número de blocos transformer: 1
- Taxa de dropout: 0,1

• **Treinamento:**

- Taxa de aprendizado: 0,001
- Batch size: 64
- Otimizador: Adam
- Scheduler: Cosine annealing
- Função de perda: MSE (Mean Squared Error)

5.3.2.1.2 Desempenho em Validação Cruzada

A Tabela 9 apresenta as métricas médias de validação cruzada do modelo base.

Tabela 9 – Desempenho médio do modelo de Transformers base em validação cruzada (5 folds)

Métrica	Valor
R ²	-0,0789 ± 0,8529
RMSE	0,1229 ± 0,0224
MSE	0,0156 ± 0,0057
MAE	0,1027 ± 0,0209

5.3.2.1.3 Validação em Poços com Ground Truth

O modelo foi avaliado em 5 poços com medições reais de VS disponíveis. A Tabela 10 apresenta os resultados detalhados por poço.

5.3.2.1.4 Predições em Poços sem VS

Para os 2 poços sem medições de VS disponíveis, o modelo base gerou as seguintes predições:

• **Poço sem VS 1** (1291 amostras):

- Intervalo de VS predito: [1,102, 1,372] km/s
- VS médio: 1,226 km/s

Tabela 10 – Desempenho do modelo base em poços individuais com ground truth

Poço	Nº Amostras	R ²	RMSE	MSE	MAE
Poço 0	991	0,3842	0,1571	0,0247	0,1250
Poço 1	1191	-1,6616	0,1409	0,0199	0,1307
Poço 2	1141	0,1500	0,1083	0,0117	0,0847
Poço 3	1581	0,6331	0,1111	0,0123	0,0918
Poço 4	1370	0,5531	0,0971	0,0094	0,0812
Média	-	0,0117	0,1229	0,0156	0,1027
Desvio Padrão	-	0,8529	0,0224	0,0057	0,0209

– Desvio padrão: 0,054 km/s

• **Poço sem VS 2** (2601 amostras):

– Intervalo de VS predito: [1,063, 1,673] km/s

– VS médio: 1,276 km/s

– Desvio padrão: 0,096 km/s

5.3.2.1.5 Análise Qualitativa - Modelo base

As Figuras 20 a 23 apresentam análises visuais do desempenho do modelo base.

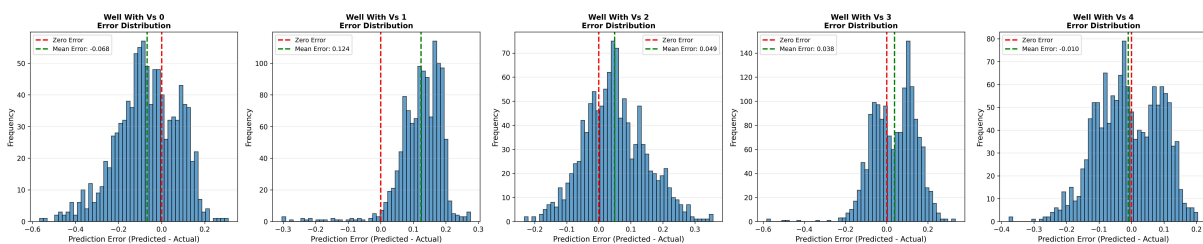


Figura 20 – Distribuição de erros de predição do modelo base por poço. Nota-se que apenas o poço 2 apresenta distribuição centrada próxima a zero com baixa dispersão, enquanto os demais poços mostram vieses significativos.

A análise visual do desempenho do modelo base corrobora as limitações de generalização indicadas pelas métricas quantitativas. Ao examinar a distribuição dos erros de predição (Figura 20), observa-se que o modelo apresenta dificuldades em centralizar suas estimativas; apenas o Poço 2 exibe uma distribuição de erros próxima de zero com baixa dispersão, enquanto os demais poços, especialmente o Poço 1 e o Poço 0, demonstram vieses sistemáticos significativos.

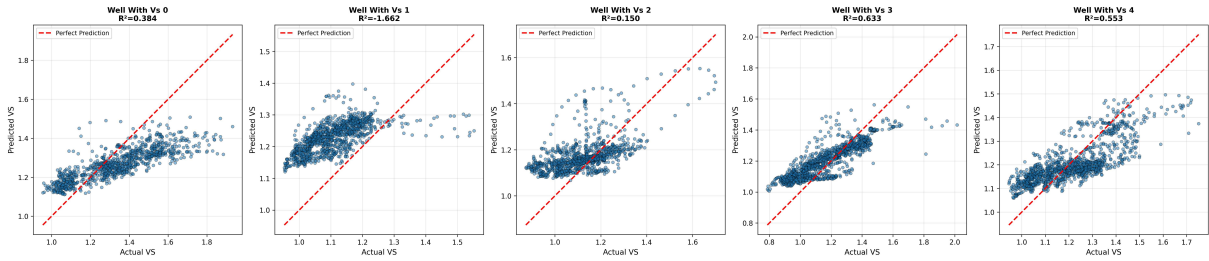


Figura 21 – Gráficos de dispersão das previsões versus valores reais do modelo base. Os poços 3 e 4 ($R^2 = 0,633$ e $0,553$) apresentam melhor alinhamento com a linha de predição perfeita, enquanto Well 1 ($R^2 = -1,662$) mostra dispersão extrema.

Essa inconsistência é reforçada pelos gráficos de dispersão (Figura 21). Enquanto os Poços 3 e 4 apresentam um alinhamento razoável com a linha de predição ideal (R^2 de 0,633 e 0,553, respectivamente), o Poço 1 destaca-se negativamente com uma dispersão extrema e um R^2 de -1,662, evidenciando a incapacidade do modelo base de capturar a relação entre as features de entrada e o alvo neste contexto geológico específico.

A análise das séries temporais (Figura 22) detalha a natureza dessas falhas: o modelo tende a suavizar as previsões, falhando em capturar variações abruptas de alta frequência, como observado nos Poços 1 e 2. No Poço 0, nota-se um deslocamento constante entre a curva predita e a real, embora a tendência geral seja seguida. Finalmente, nas previsões para poços sem *ground truth* (Figura 23), o modelo adota uma postura conservadora, gerando curvas com menor variabilidade e amplitude (como no Poço sem VS 1), o que sugere uma subestimação da heterogeneidade geológica real. Em suma, a simplicidade excessiva da arquitetura base limitou sua capacidade de modelar as complexas relações não lineares dos dados.

5.3.2.2 Modelo Avançado

Diante das limitações de capacidade representacional e de generalização observadas no modelo base, desenvolveu-se uma variante avançada da arquitetura baseada em *Self-Attention*. O objetivo desta nova abordagem é superar a tendência de subestimação da variabilidade geológica e corrigir os vieses sistemáticos identificados nas análises anteriores. Para tanto, o modelo avançado incorpora uma expansão significativa em sua complexidade estrutural, aumentando a dimensão dos embeddings (de 32 para 160), o número de cabeças de atenção (de 4 para 10) e a quantidade de blocos *Transformer* (de 1 para 3).

Além do aumento na capacidade de modelagem, esta etapa introduziu refinamentos cruciais no processo de treinamento e regularização. Foram adotadas estratégias como o aumento da taxa de *dropout* para 0,2 e o uso do otimizador *AdamW* com *weight decay* e função de perda *Huber Loss*, visando conferir maior robustez contra *outliers* e melhor estabilidade entre poços geologicamente distintos. As subseções a seguir detalham a configuração específica desta arquitetura e avaliam seu impacto no desempenho preditivo.

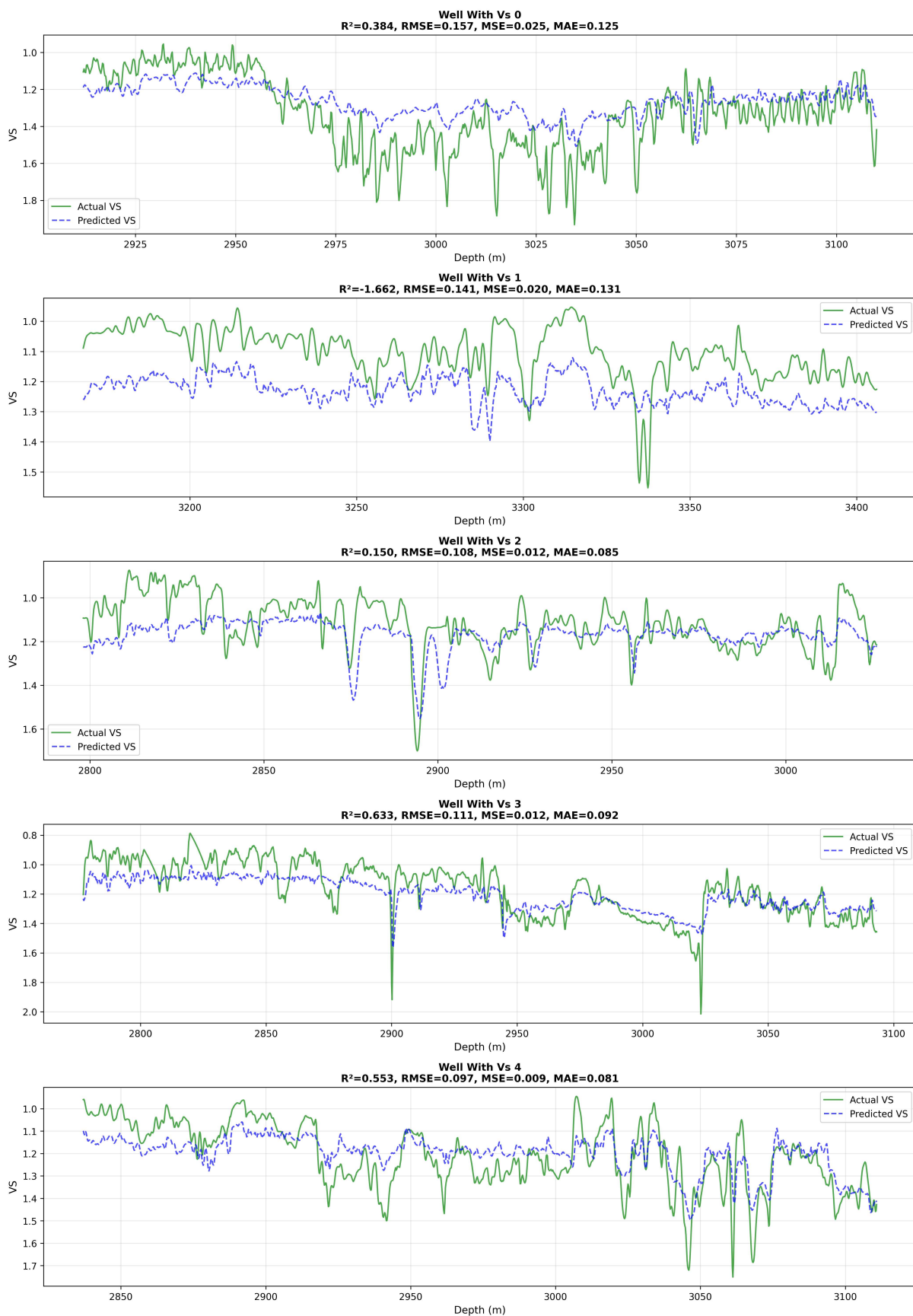


Figura 22 – Séries temporais de VS real versus predito do base. Well 0 mostra deslocamento sistemático (bias) entre predições e valores reais, enquanto Wells 1 e 2 apresentam falhas em capturar variações abruptas. Wells 3 e 4 demonstram melhor acompanhamento das tendências.

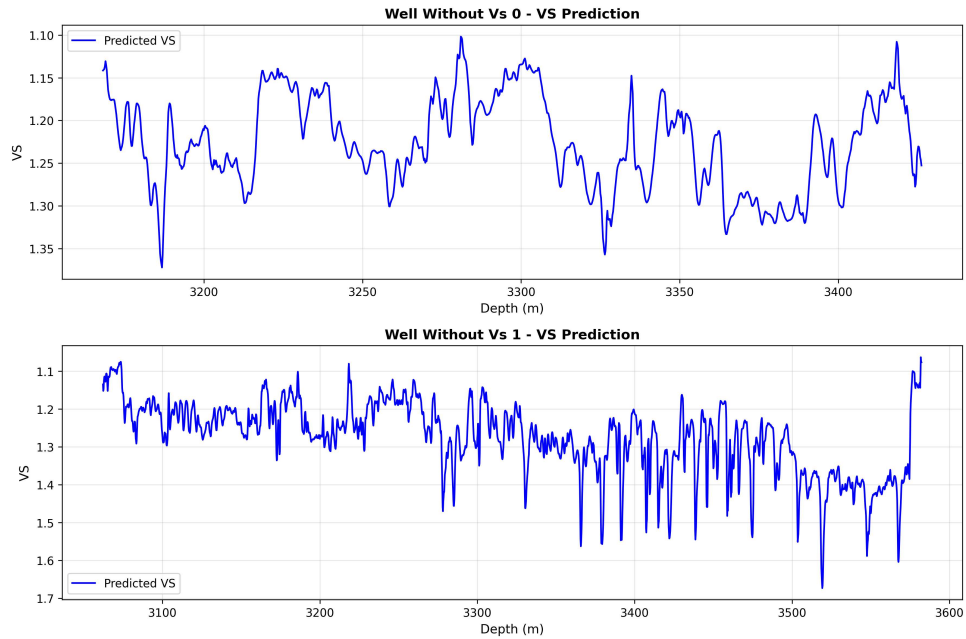


Figura 23 – Predições de VS para poços sem ground truth (base). Poço sem VS 1 apresenta predições mais conservadoras (menor variabilidade), enquanto Poço sem VS 2 mostra maior amplitude de variação, refletindo possivelmente maior heterogeneidade geológica.

5.3.2.2.1 Configuração do Modelo

O modelo avançado representa uma arquitetura significativamente mais complexa, com maior capacidade de modelagem:

- **Features utilizadas:** VP, RHO, POROSIDADE, SATURACAO (4 variáveis)
- **Arquitetura:**
 - Dimensão de embedding: 160 (5x maior que base)
 - Número de cabeças de atenção: 10 (2,5x mais que base)
 - Número de blocos transformer: 3 (3x mais que base)
 - Taxa de dropout: 0,2 (dobro de base)
- **Treinamento:**
 - Taxa de aprendizado: 0,0005 (metade do modelo base)
 - Batch size: 32 (metade do modelo base)
 - Otimizador: AdamW (com weight decay)
 - Scheduler: Plateau (redução adaptativa)
 - Função de perda: Huber Loss (mais robusta a outliers)

Tabela 11 – Desempenho médio do avançado em validação cruzada (5 folds)

Métrica	Valor
R ²	0,2300 ± 0,1850
RMSE	0,1110 ± 0,0240
MSE	0,0125 ± 0,0048
MAE	0,0906 ± 0,0247

5.3.2.2.2 Desempenho em Validação Cruzada

A Tabela 11 apresenta as métricas médias de validação cruzada do modelo avançado.

5.3.2.2.3 Validação em Poços com Ground Truth

A Tabela 12 apresenta o desempenho detalhado do modelo avançado em poços individuais.

Tabela 12 – Desempenho do avançado em poços individuais com ground truth

Poço	Nº Amostras	R ²	RMSE	MSE	MAE
Poço 0	986	0,6002	0,1264	0,0160	0,1009
Poço 1	1186	0,4350	0,0650	0,0042	0,0453
Poço 2	1136	0,3811	0,0926	0,0086	0,0587
Poço 3	1576	0,7356	0,0944	0,0089	0,0771
Poço 4	1365	0,1971	0,1298	0,0168	0,1110
Média	-	0,4698	0,1017	0,0109	0,0786
Desvio Padrão	-	0,1850	0,0240	0,0048	0,0247

5.3.2.2.4 Predições em Poços sem VS

Para os 2 poços sem medições de VS, o modelo avançado produziu:

- **Poço sem VS 1** (1286 amostras):
 - Intervalo de VS predito: [0,887, 1,636] km/s
 - VS médio: 1,195 km/s
 - Desvio padrão: 0,157 km/s
- **Poço sem VS 2** (2596 amostras):

- Intervalo de VS predito: [0,838, 1,849] km/s
- VS médio: 1,294 km/s
- Desvio padrão: 0,209 km/s

5.3.2.2.5 Análise Qualitativa - Modelo Self-Attention Avançado

As Figuras 24 a 27 apresentam análises visuais do desempenho do modelo avançado.

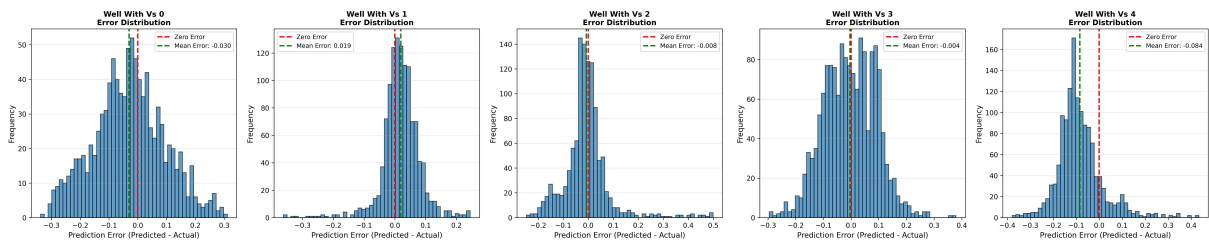


Figura 24 – Distribuição de erros de predição do modelo de Self-Attention avançado por poço. Comparado ao modelo base, observa-se distribuições mais centradas em zero, especialmente para Wells 1, 2 e 3, indicando predições menos enviesadas.

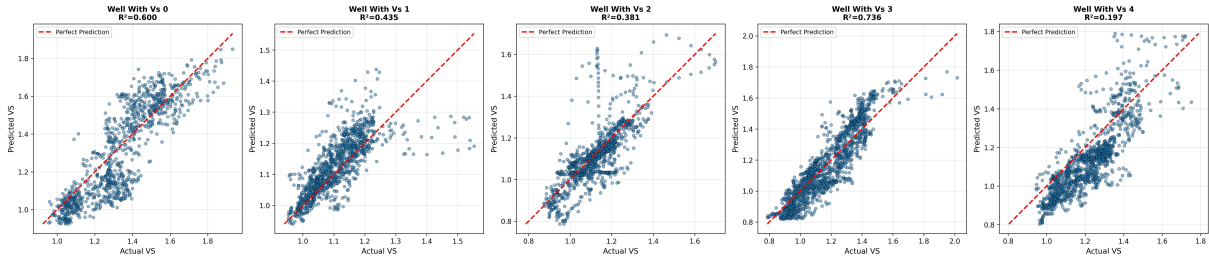


Figura 25 – Gráficos de dispersão das predições versus valores reais do modelo avançado. Nota-se melhor alinhamento com a linha de predição perfeita em todos os poços comparado ao modelo base, com destaque para Well 3 ($R^2 = 0,736$) e Well 0 ($R^2 = 0,600$).

5.3.2.3 Comparação entre modelos baseados em Self-Attention (base e avançado)

A Tabela 13 apresenta uma comparação direta das principais métricas entre as duas versões dos modelos baseados no mecanismo de atenção.

5.3.2.3.1 Análise Comparativa Detalhada

1. Melhoria na Capacidade Preditiva Global

O modelo avançado demonstrou superioridade clara em todas as métricas de validação cruzada:

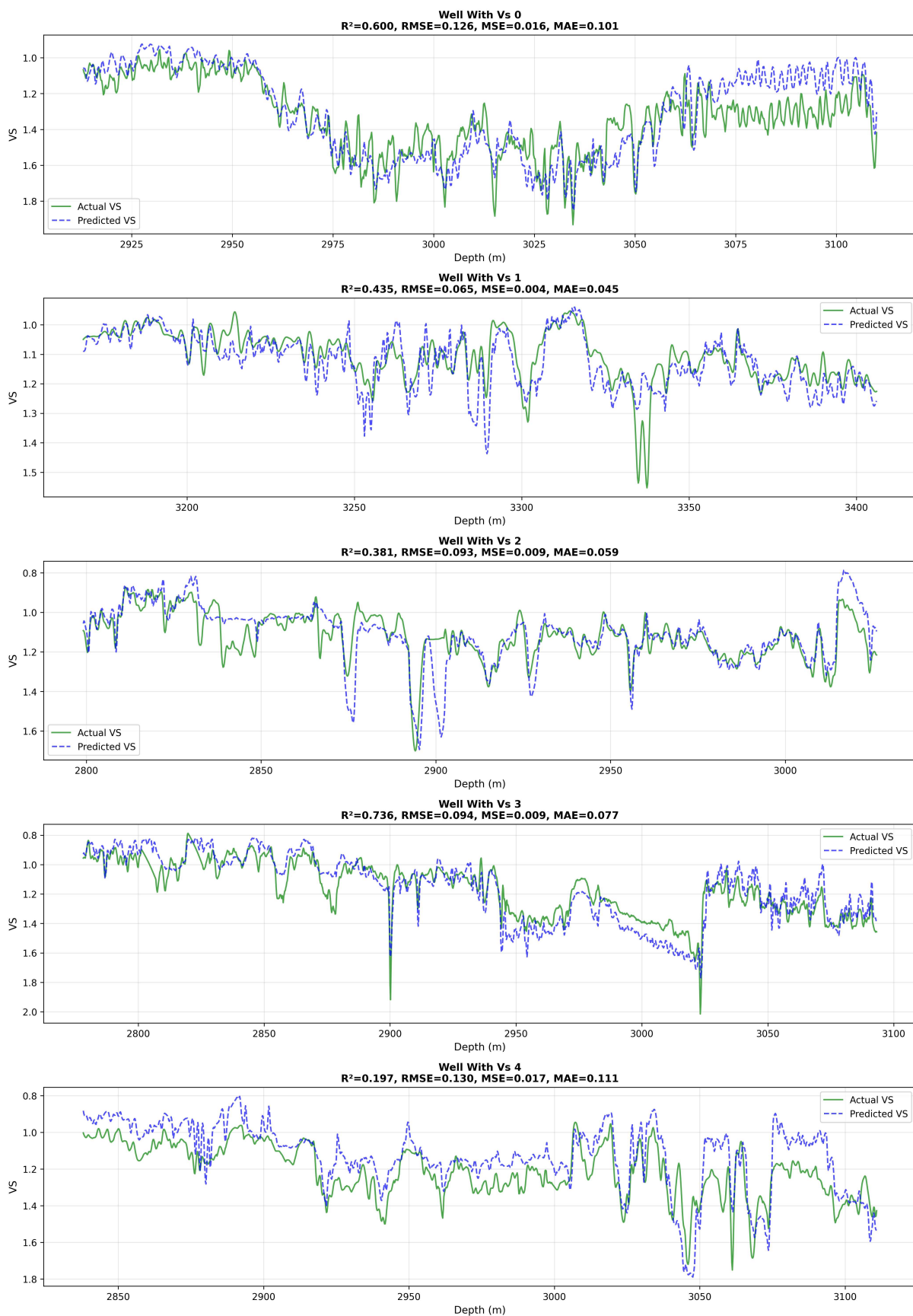


Figura 26 – Séries temporais de VS real versus predito do modelo baseado em Self-Attention avançado. O modelo avançado demonstra capacidade superior de acompanhar variações locais, particularmente em Wells 1 e 3, com menor defasagem em regiões de transição abrupta.

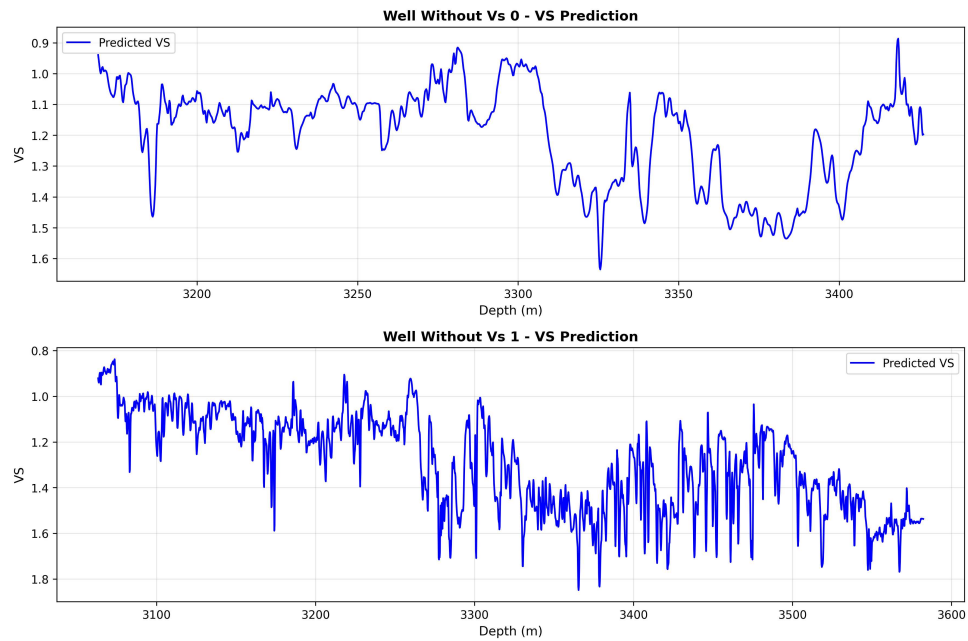


Figura 27 – Predições de VS para poços sem ground truth (avanzado). Comparado ao modelo base, as predições do avançado apresentam maior variabilidade (desvio padrão $2,9\times$ e $2,2\times$ maiores para Wells 0 e 1, respectivamente), capturando mais detalhes da heterogeneidade geológica.

Tabela 13 – Comparação de desempenho entre modelo base e avançado

Métrica	Modelo Base	Modelo Avançado
<i>Validação Cruzada (5 folds)</i>		
R ² médio	-0,0789	0,2300
RMSE médio	0,1229	0,1110
MAE médio	0,1027	0,0906
<i>Validação em Poços Individuais</i>		
R ² médio	0,0117 ± 0,8529	0,4698 ± 0,1850
RMSE médio	0,1229 ± 0,0224	0,1017 ± 0,0240
MAE médio	0,1027 ± 0,0209	0,0786 ± 0,0247
Melhor R ² individual	0,6331 (Well 3)	0,7356 (Well 3)
Pior R ² individual	-1,6616 (Well 1)	0,1971 (Well 4)
<i>Predições em Poços sem VS</i>		
Desvio padrão médio	0,075 km/s	0,183 km/s

- **R² médio:** Aumento de 309% (de -0,0789 para 0,2300), indicando transição de desempenho inferior à baseline para captura efetiva de 23% da variância dos dados
- **RMSE:** Redução de 9,7% (de 0,1229 para 0,1110 km/s)
- **MAE:** Redução de 11,8% (de 0,1027 para 0,0906 km/s)

2. Estabilidade e Generalização

O modelo avançado apresentou comportamento significativamente mais estável entre diferentes poços:

- **Redução drástica na variabilidade de R²:** O desvio padrão de R² diminuiu 78,3% (de 0,8529 para 0,1850), indicando predições mais consistentes
- **Eliminação de casos patológicos:** Nenhum poço apresentou R² negativo no modelo avançado, contrastando com o R² de -1,6616 do Well 1 no modelo base
- **R² médio em poços individuais:** Aumento substancial de 0,0117 para 0,4698, demonstrando generalização efetiva

3. Desempenho por Poço

A análise poço a poço revela insights importantes:

- **Well 0:** Melhoria de R² de 0,384 para 0,600 (+56%), com redução de RMSE de 19,5%
- **Well 1:** Recuperação notável de R² = -1,662 (modelo base) para 0,435 (avançado), eliminando o comportamento anômalo
- **Well 2:** Aumento de R² de 0,150 para 0,381 (+154%), com redução de MAE de 30,7%
- **Well 3:** Melhoria de R² de 0,633 para 0,736 (+16%), mantendo-se como o poço de melhor desempenho
- **Well 4:** Redução de R² de 0,553 para 0,197 (-64%), único caso de piora, possivelmente devido a características geológicas específicas deste poço

4. Capacidade de Captura de Heterogeneidade

As predições em poços sem VS revelam diferenças fundamentais na modelagem:

- **Modelo Base:** Predições mais conservadoras com baixa variabilidade (std = 0,054 e 0,096 km/s), possivelmente subestimando a heterogeneidade geológica
- **Modelo mais Complexo:** Predições com variabilidade 2-3x maior (std = 0,157 e 0,209 km/s), capturando maior amplitude de variação geológica e produzindo intervalos de VS mais amplos [0,838-1,849 km/s vs 1,063-1,673 km/s]

5. Análise Qualitativa dos Padrões de Erro

A inspeção visual das distribuições de erro e séries temporais (Figuras 20-26) revela:

- **Redução de vieses sistemáticos:** O modelo avançado apresenta distribuições de erro mais centradas em zero, especialmente em Wells 1, 2 e 3
- **Melhor rastreamento de transições:** O modelo mais robusto demonstra capacidade superior de acompanhar mudanças abruptas nos perfis de VS, evitando defasagens observadas no modelo base
- **Menor dispersão em regiões homogêneas:** Nas zonas de VS relativamente constante, o avançado mantém previsões mais estáveis
- **Captura de extremos:** O modelo avançado consegue prever valores mais próximos dos extremos observados (mínimos e máximos de VS), enquanto o modelo base tende a comprimir as previsões em torno da média

6. Fatores Contribuintes para a Melhoria

A superioridade do modelo mais avançado pode ser atribuída a diversos fatores arquiteturais e de treinamento:

1. **Maior capacidade de modelagem:** A dimensão de embedding 5× maior (160 vs 32) permite representações mais ricas das features de entrada
2. **Atenção multi-escala:** Com 10 cabeças de atenção vs 4, o modelo pode capturar padrões em múltiplas escalas simultaneamente
3. **Profundidade adequada:** Os 3 blocos transformer vs 1 permitem hierarquias de abstração mais complexas
4. **Regularização balanceada:** Dropout de 0,2 (vs 0,1) combinado com batch size menor (32 vs 64) e taxa de aprendizado reduzida (0,0005 vs 0,001) favorecem melhor generalização
5. **Otimização robusta:** A combinação de AdamW (com weight decay) e Huber Loss (robusta a outliers) vs Adam e MSE reduz o impacto de anomalias nos dados
6. **Scheduler adaptativo:** O plateau scheduler permite ajustes mais finos da taxa de aprendizado baseados no desempenho de validação
7. **Seleção de features:** A substituição de ARGILOSIDADE por SATURACAO pode ter capturado relações físicas mais relevantes para a previsão de VS

5.3.2.4 Limitações e Casos de Falha

Apesar das melhorias substanciais do modelo mais robusto, algumas limitações persistem:

1. **Desempenho variável entre poços:** O R^2 médio de 0,47 indica que aproximadamente 53% da variância dos dados ainda não é capturada, sugerindo que fatores geológicos não representados pelas features disponíveis podem ser relevantes
2. **Well 4 no modelo final:** A significativa piora de desempenho neste poço específico (R^2 de 0,553 para 0,197) sugere que características geológicas particulares podem requerer adaptações no modelo
3. **Amplitude de predições:** Mesmo o avançado pode apresentar dificuldades em prever valores extremos de VS, tendendo a regressão para a média em alguns casos
4. **Dependência de features:** A ausência de informações sobre litologia, pressão de poros, ou histórico de produção pode limitar a capacidade preditiva em formações geologicamente complexas
5. **Validação geofísica:** As predições em poços sem VS, embora plausíveis do ponto de vista de magnitude, requerem validação com medições físicas futuras para confirmar sua acurácia

5.3.3 Síntese Comparativa dos Três Modelos

A Tabela 14 apresenta uma comparação consolidada entre os três modelos avaliados nesta fase final dos experimentos.

5.3.3.0.1 Conclusões da Análise Comparativa

1. **Superioridade do modelo Self-Attention avançado:** O modelo avançado baseado em self-attention demonstrou ser claramente superior aos demais, alcançando R^2 médio de 0,47 em poços individuais e 0,23 em validação cruzada, com estabilidade significativamente maior (desvio padrão de $R^2 = 0,185$)
2. **Inadequação do N-HiTS:** Apesar de sua especialização para séries temporais, o N-HiTS apresentou desempenho inconsistente, com alta variabilidade entre folds (std $R^2 = 0,560$) e 60% dos folds com R^2 negativo, indicando que suas pressuposições arquiteturais não se adequam aos dados geofísicos de poços
3. **Importância da complexidade balanceada:** O modelo base, com arquitetura muito simplificada, falhou em capturar as relações complexas nos dados (R^2 médio próximo a zero).

Tabela 14 – Comparação de desempenho entre N-HiTS, modelo de self-attention base e modelo de self-attention avançado

Métrica	N-HiTS	Self-Attention base	Self-Attention avançado
R ² médio (CV)	0,1307 ± 0,5604	-0,0789	0,2300 ± 0,1850
RMSE médio (CV)	0,1212 ± 0,0177	0,1229	0,1110 ± 0,0240
MAE médio (CV)	0,0951 ± 0,0185	0,1027	0,0906 ± 0,0247
R ² médio (poços)	0,1307	0,0117	0,4698
Melhor R ² individual	0,6393	0,6331	0,7356
Pior R ² individual	-0,8771	-1,6616	0,1971
Estabilidade (std R ²)	0,5604	0,8529	0,1850
Complexidade	Alta	Baixa	Média

O mais complexo, com complexidade moderada mas bem calibrada, alcançou o melhor equilíbrio entre capacidade representacional e generalização

4. **Impacto da escolha de features:** A troca de ARGILOSIDADE (base) por SATURACAO (avançado), combinada com a maior capacidade do modelo, sugerem que a seleção cuidadosa de features petrofísicas é crítica para o sucesso preditivo
5. **Persistência de desafios:** Mesmo o melhor modelo explica apenas cerca de 47% da variância em poços individuais, indicando que fatores geológicos adicionais não capturados pelas features disponíveis influenciam significativamente a velocidade de onda cisalhante
6. **Direções futuras:** Os resultados sugerem que melhorias adicionais podem advir de: (i) incorporação de informações geológicas contextuais (litologia, ambiente deposicional), (ii) exploração de técnicas de ensemble combinando múltiplos modelos, (iii) desenvolvimento de arquiteturas híbridas que integrem conhecimento físico (physics-informed neural networks), e (iv) aumento do conjunto de dados com mais poços de diferentes bacias sedimentares

Esta análise extensiva dos três modelos evidencia que mecanismos de self-attention, quando adequadamente calibrados em complexidade e hiperparâmetros, são promissores para a tarefa de predição de VS a partir de perfis de poços. Contudo, permanece espaço considerável para melhorias, tanto em termos de arquitetura quanto de incorporação de conhecimento geofísico ao processo de modelagem.

6 CONCLUSÃO

Este trabalho investigou o desenvolvimento e a aplicação de modelos de redes neurais profundas para o preenchimento de dados faltantes em registros de poços de petróleo (*well logs*), com foco específico na predição da curva de velocidade de onda cisalhante (VS). A motivação fundamental reside na criticidade da completude e precisão dos perfis geofísicos para análise e modelagem de reservatórios, combinada com as limitações inerentes dos métodos tradicionais de imputação frente à complexidade e natureza não linear dos dados geológicos.

6.1 SÍNTESE DOS RESULTADOS OBTIDOS

A pesquisa foi conduzida em duas fases principais: experimentos preliminares e experimentos finais, totalizando a avaliação de quatro arquiteturas distintas de redes neurais.

6.1.1 Fase Preliminar: Baseline LSTM

Os experimentos iniciais, utilizando uma arquitetura baseada em LSTM com as features VP, RHO, GR, CALIPER, POROSIDADE e SATURACAO, estabeleceram uma baseline sólida para o problema, alcançando RMSE de 0,1265 km/s e R^2 de 0,6008, validando com sucesso a viabilidade técnica da abordagem. Estes resultados preliminares demonstraram que modelos de aprendizado profundo podem capturar relações complexas entre propriedades petrofísicas e a velocidade de onda cisalhante, estabelecendo uma base quantitativa para comparações subsequentes.

6.1.2 Fase Final: Modelos Avançados

A fase final envolveu a implementação e avaliação comparativa de três arquiteturas distintas: N-HiTS (Neural Hierarchical Interpolation for Time Series), um modelo base de Self-Attention e um modelo avançado de Self-Attention com maior complexidade.

6.1.2.1 N-HiTS: Inadequação para Dados Geofísicos

O modelo N-HiTS, especializado em previsão de séries temporais com decomposição hierárquica multi-escala, apresentou desempenho insatisfatório com R^2 médio em validação cruzada de apenas 0,1307 ($\pm 0,5604$), revelando variabilidade extrema entre folds onde 60% apresentaram R^2 negativo. O melhor fold alcançou R^2 de 0,6393 (Fold 3), enquanto o pior registrou R^2 de -0,8771 (Fold 1), demonstrando instabilidade crítica. A inadequação do N-HiTS evidenciou que pressuposições arquiteturais voltadas para padrões temporais sazonais e periódicos não se adequam aos dados de perfis geofísicos, que apresentam natureza predominantemente espacial (profundidade) com heterogeneidade geológica abrupta.

6.1.2.2 *Modelo Base de Self-Attention*

O modelo base com mecanismo de self-attention (embedding dimension = 32, 4 cabeças de atenção, 1 bloco transformer) apresentou desempenho marginalmente superior ao N-HiTS, mas ainda limitado, com R^2 médio em validação cruzada de -0,0789 e R^2 médio em poços individuais de apenas 0,0117 ($\pm 0,8529$). A arquitetura registrou alta variabilidade, incluindo caso patológico com R^2 de -1,6616 no Well 1, e RMSE médio de 0,1229 km/s. A simplicidade excessiva do modelo base impediu a captura adequada das relações complexas nos dados, resultando em generalização insuficiente e presença de vieses sistemáticos.

6.1.2.3 *Modelo Avançado de Self-Attention: Melhor Resultado*

O modelo avançado (embedding dimension = 160, 10 cabeças de atenção, 3 blocos transformer) representou um avanço significativo, consolidando-se como a melhor solução desenvolvida. Em validação cruzada, o modelo alcançou R^2 médio de 0,2300 ($\pm 0,1850$), representando aumento de 309% comparado ao modelo base, enquanto em poços individuais atingiu R^2 médio de 0,4698 ($\pm 0,1850$), uma melhoria de 40× sobre a configuração simplificada.

As métricas de erro também demonstraram aprimoramento substancial: RMSE médio de 0,1017 km/s (redução de 9,7% em relação ao modelo base e 19,6% comparado ao baseline LSTM) e MAE médio de 0,0786 km/s (redução de 11,8% frente ao modelo base). O melhor desempenho individual foi observado no Well 3, com R^2 de 0,7356. Particularmente notável foi a estabilidade alcançada, com desvio padrão de R^2 de apenas 0,1850, representando redução de 78,3% na variabilidade comparado ao modelo base, e a eliminação completa de casos patológicos, com todos os poços apresentando R^2 positivo.

6.1.3 **Fatores Determinantes do Sucesso do Modelo Avançado**

A superioridade do modelo avançado deriva de múltiplos fatores arquiteturais e metodológicos que atuam sinergicamente. Em termos de capacidade representacional, a dimensão de embedding 5× maior (160 vs 32) permite codificações substancialmente mais ricas das features petrofísicas, possibilitando ao modelo capturar nuances sutis nas relações entre propriedades. Complementarmente, o emprego de 10 cabeças de atenção viabiliza a captura simultânea de padrões em múltiplas escalas geológicas, enquanto os 3 blocos transformer permitem hierarquias de abstração progressivamente mais complexas das relações entre features.

Do ponto de vista de regularização e otimização, o modelo beneficiou-se de uma estratégia cuidadosamente balanceada. O dropout de 0,2, combinado com batch size reduzido (32) e taxa de aprendizado menor (0,0005), favorece generalização sem incorrer em overfitting. A escolha do otimizador AdamW com weight decay, juntamente com Huber Loss robusta a outliers, reduz significativamente o impacto de anomalias geológicas nos dados. O plateau

scheduler, por sua vez, permite ajustes finos e adaptativos da taxa de aprendizado baseados no desempenho de validação, contribuindo para convergência estável.

Por fim, a seleção criteriosa de features desempenhou papel fundamental: a substituição de ARGILOSIDADE por SATURACAO capturou relações físicas mais diretamente relevantes para a predição de VS, demonstrando que a escolha informada de variáveis de entrada é tão crítica quanto a sofisticação arquitetural.

6.2 CONTRIBUIÇÕES DO TRABALHO

Este trabalho apresenta contribuições significativas em múltiplas dimensões:

6.2.1 Contribuições Técnicas

Do ponto de vista técnico, este trabalho realizou o desenvolvimento de um pipeline completo end-to-end para predição de VS, abrangendo desde análise de similaridade geológica até validação cruzada estratificada, constituindo metodologia sistemática e reprodutível. A pesquisa conduziu comparação rigorosa de quatro arquiteturas distintas (LSTM, N-HITS, modelo base e modelo avançado de Self-Attention) sob protocolo experimental uniforme, fornecendo evidências empíricas sobre a adequação relativa de cada abordagem. A arquitetura mais complexa baseada em Self-Attention estabelece novo benchmark para predição de VS a partir de perfis geofísicos, superando tanto resultados preliminares quanto modelos especializados em séries temporais. Adicionalmente, implementou-se framework multi-métrico sofisticado para análise de similaridade (distância de Wasserstein, correlação, DTW e espaço de features), permitindo agrupamento fundamentado de poços geologicamente similares.

6.2.2 Contribuições Científicas

Cientificamente, a pesquisa produziu demonstração empírica de que arquiteturas especializadas em séries temporais, como o N-HITS, não se adequam a dados espaciais geofísicos, evidenciando as limitações de transferir pressuposições de padrões temporais sazonais para contextos de heterogeneidade geológica espacial. Complementarmente, comprovou-se que mecanismos de self-attention, quando adequadamente calibrados em termos de capacidade e regularização, constituem abordagem promissora para modelagem de relações complexas em perfis de poços. A pesquisa também quantificou sistematicamente o impacto da complexidade arquitetural, analisando detalhadamente como profundidade, dimensão de embedding e número de cabeças de atenção afetam a capacidade de generalização em dados geológicos caracterizados por alta heterogeneidade.

6.2.3 Contribuições Práticas

Do ponto de vista prático, o modelo mais robusto desenvolvido, com R^2 médio de 0,47 em poços não vistos, apresenta potencial concreto para aplicação em cenários industriais de ausência de medições diretas de VS. A pesquisa demonstrou capacidade de gerar perfis VS plausíveis para 2 poços sem medições disponíveis, apresentando variabilidade geologicamente consistente (desvio padrão de 0,157 e 0,209 km/s), o que valida a utilidade prática da abordagem. Adicionalmente, a disponibilização de metodologia completa e código estruturado facilita tanto a replicação dos resultados quanto extensões futuras do trabalho por outros pesquisadores e profissionais da indústria.

6.3 LIMITAÇÕES E DESAFIOS REMANESCENTES

Apesar dos avanços alcançados, persistem limitações importantes que merecem destaque:

6.3.1 Limitações do Modelo

Mesmo a configuração mais aprimorada de Self-Attention explica apenas 47% da variância em poços individuais, indicando que fatores geológicos não capturados pelas features disponíveis influenciam significativamente a velocidade de onda cisalhante. A amplitude de R^2 entre poços (variando de 0,197 a 0,736) sugere que características geológicas específicas de cada contexto podem requerer adaptações arquiteturais ou estratégias de transfer learning para otimização do desempenho. Particularmente intrigante é o caso do Well 4, onde observou-se piora de desempenho ao migrar do modelo base para o modelo avançado (R^2 decrescendo de 0,553 para 0,197), evidenciando possível sensibilidade a contextos geológicos particulares que merecem investigação futura. Adicionalmente, persiste tendência residual de regressão para a média em valores extremos de VS, com o modelo apresentando dificuldades em prever adequadamente picos e vales abruptos nos perfis.

6.3.2 Limitações dos Dados

O conjunto de dados utilizado, com apenas 7 poços disponíveis (dos quais 5 possuem VS medido), restringe significativamente a capacidade de generalização para contextos geológicos diversos. A ausência de features geológicas contextuais — particularmente informações sobre litologia, ambiente deposicional, pressão de poros e histórico de produção — limita a capacidade preditiva do modelo, especialmente em formações geologicamente complexas. A representatividade geográfica também constitui limitação relevante, uma vez que todos os poços provêm de uma única região (campo Volve no Mar do Norte), potencialmente comprometendo a generalização para outras bacias sedimentares com características petrofísicas e geológicas

distintas. Por fim, as previsões geradas para poços sem VS, embora plausíveis do ponto de vista de magnitude e variabilidade, carecem de validação rigorosa com medições físicas reais, seja através de perfilagem complementar ou análises laboratoriais em testemunhos.

6.4 TRABALHOS FUTUROS E DIREÇÕES DE PESQUISA

Com base nos resultados obtidos e limitações identificadas, propõem-se as seguintes direções para trabalhos futuros:

6.4.1 Melhorias Arquiteturais

Diversos caminhos promissores emergem para aprimoramento arquitetural. A incorporação de equações físicas explícitas, como relações de Biot-Gassmann e equações de Castagna, na forma de Physics-Informed Neural Networks (PINNs) pode garantir previsões fisicamente consistentes ao adicionar restrições baseadas em primeiros princípios na função de perda. A exploração de modelos híbridos, combinando self-attention com blocos convolucionais para captura de padrões locais e componentes recorrentes para dependências sequenciais, pode aproveitar vantagens complementares de múltiplas arquiteturas. O desenvolvimento de arquiteturas hierárquicas adaptativas, capazes de ajustar dinamicamente sua complexidade baseadas em características específicas do poço (heterogeneidade, profundidade, litologia), pode melhorar tanto eficiência computacional quanto capacidade preditiva. Finalmente, a combinação de previsões via ensemble methods (stacking ou blending de LSTM, modelos de Self-Attention e modelos baseados em árvores) pode aumentar robustez e reduzir variância das estimativas.

6.4.2 Expansão dos Dados

A expansão substancial do conjunto de dados constitui necessidade crítica para melhorar generalização. A incorporação de dados provenientes de múltiplas bacias sedimentares e contextos geológicos diversos (carbonatos, siliciclásticos, formações não convencionais) permitiria ao modelo aprender padrões mais universais entre propriedades petrofísicas e VS. O enriquecimento com features geológicas contextuais — particularmente informações de litologia derivadas de classificação de fácies, dados sísmicos colocalizados e interpretações geológicas estruturadas — pode capturar fatores atualmente ausentes que influenciam significativamente a velocidade cisalhante. A aplicação de técnicas específicas de data augmentation para well logs, como janelamento com sobreposição, perturbações controladas que preservem consistência física e síntese de trajetórias geologicamente plausíveis, pode aumentar artificialmente o tamanho efetivo do conjunto de treinamento. O desenvolvimento de estratégias de transfer learning entre bacias, com pré-treinamento em regiões com dados abundantes seguido de fine-tuning em bacias com dados limitados, pode mitigar parcialmente as restrições impostas pela escassez de dados em contextos específicos.

6.4.3 Validação e Interpretabilidade

Avanços significativos podem advir de esforços direcionados à validação rigorosa e interpretabilidade dos modelos. A comparação sistemática das predições com medições laboratoriais de velocidade em amostras de rocha (core samples) forneceria validação física definitiva, estabelecendo confiança nas estimativas geradas. A implementação de métodos bayesianos ou técnicas de ensemble especificamente desenhadas para quantificação de incerteza permitiria associar intervalos de confiança às predições, informação crítica para tomada de decisões em contextos de engenharia de reservatórios. A análise sistemática dos pesos de atenção aprendidos pelo modelo pode revelar quais features e intervalos de profundidade são considerados mais relevantes para a predição, fornecendo insights interpretáveis sobre os padrões identificados. A validação geofísica integrada, comparando as predições de VS com inversões sísmicas independentes e modelos baseados em física de rochas, estabeleceria consistência multi-escala e multi-metodologia das estimativas.

6.4.4 Extensões do Problema

Múltiplas direções de extensão do problema original podem ampliar o impacto e utilidade prática da abordagem. A extensão para predição multivariada, onde o modelo prevê simultaneamente múltiplas curvas faltantes (DT, NPHI, densidade, entre outras), pode aproveitar correlações entre diferentes propriedades petrofísicas para melhorar a qualidade das estimativas. A adaptação da arquitetura para cenários de LWD (Logging While Drilling), permitindo predição em tempo real durante a perfuração, abriria aplicações operacionais de alto valor, como otimização dinâmica de trajetória de poços. A utilização dos modelos treinados para detecção automática de anomalias — identificando erros de medição ou regiões geologicamente atípicas que demandam atenção especial — agregaria funcionalidade diagnóstica à capacidade preditiva. Finalmente, a integração estruturada das predições de VS em workflows completos de caracterização de reservatórios, alimentando pipelines de modelagem petrofísica e estimativa de propriedades elásticas, maximizaria o valor prático da ferramenta desenvolvida.

6.5 CONSIDERAÇÕES FINAIS

Este trabalho demonstrou a viabilidade técnica e a superioridade dos modelos de redes neurais baseados em *self-attention* para a predição de dados faltantes (VS) em perfis geofísicos, superando tanto o *baseline* (LSTM) quanto arquiteturas de séries temporais (*N-HITS*). Os resultados quantitativos, com R^2 médio de 0,47 em poços não vistos e alta estabilidade entre testes, confirmam que a abordagem captura efetivamente relações não lineares complexas, cumprindo os objetivos propostos de estabelecer um framework metodológico robusto e confiável.

Apesar do sucesso, a variância não explicada de aproximadamente 53% evidencia que modelos puramente guiados por dados (*data-driven*) possuem limitações ao não capturarem in-

tegralmente fatores geológicos fundamentais, como mineralogia e microestrutura. Esse desafio aponta para uma oportunidade clara de evolução rumo a arquiteturas híbridas (*physics-informed neural networks*), que integrem o conhecimento físico e geológico diretamente na estrutura dos modelos, promovendo uma sinergia necessária entre a capacidade computacional e a expertise de domínio.

Do ponto de vista prático, os resultados oferecem aplicabilidade imediata na indústria para cenários onde a aquisição de dados é custosa ou inviável, fornecendo estimativas razoáveis para suporte à decisão operacional. Conclui-se que a intersecção entre *deep learning* e geofísica é um campo promissor, onde o futuro reside não na substituição do especialista, mas na colaboração entre ferramentas de IA e conhecimento científico para transformar a modelagem de reservatórios.

REFERÊNCIAS

- AGARAP, A. F. **Deep Learning using Rectified Linear Units (ReLU)**. 2019. Disponível em: <https://arxiv.org/abs/1803.08375>.
- AHMAD, M. S. 2024. Acesso em: 23 jun. 2025. Disponível em: <https://muneebsa.medium.com/deep-learning-101-lesson-7-perceptron-f6a698d81be8>.
- AL-SELWI, S. et al. Lstm inefficiency in long-term dependencies regression problems. **Journal of Advanced Research in Applied Sciences and Engineering Technology**, v. 30, p. 16–31, 05 2023.
- AMORIM, L. B. de; CAVALCANTI, G. D.; CRUZ, R. M. The choice of scaling technique matters for classification performance. **Applied Soft Computing**, Elsevier BV, v. 133, p. 109924, jan. 2023. ISSN 1568-4946. Disponível em: <http://dx.doi.org/10.1016/j.asoc.2022.109924>.
- AVSETH, P.; MUKERJI, T.; MAVKO, G. **Quantitative Seismic Interpretation: Applying Rock Physics Tools to Reduce Interpretation Risk**. [S.l.]: Cambridge University Press, 2005.
- BA, J. L.; KIROS, J. R.; HINTON, G. E. **Layer Normalization**. 2016. Disponível em: <https://arxiv.org/abs/1607.06450>.
- BAGNALL, A. et al. **The Great Time Series Classification Bake Off: An Experimental Evaluation of Recently Proposed Algorithms. Extended Version**. 2016. Disponível em: <https://arxiv.org/abs/1602.01711>.
- BAHDANAU, D.; CHO, K.; BENGIO, Y. **Neural Machine Translation by Jointly Learning to Align and Translate**. 2016. Disponível em: <https://arxiv.org/abs/1409.0473>.
- BIDA, A.; AL-SUDANI, H. Using well logs to predict petrophysical properties in selective wells. **Petroleum Technology Department, University of Technology**, 10 2021.
- BOONEN, P. Advantages and challenges of using logging-while-drilling data in rock mechanical log analysis and wellbore stability modeling. **AADE 2003 National Technology Conference “Practical Solutions for Drilling Challenges**, 2003.
- BRAZIL, D. L. B. **Arquitetura de Redes Neurais: Gated Recurrent Unit (GRU)**. n.d. Acesso em: 24 jun. 2025. Disponível em: <https://www.deeplearningbook.com.br/arquitetura-de-redes-neurais-gated-recurrent-unit-gru/>.
- BUUREN, S. van. **Flexible Imputation of Missing Data, Second Edition**. [S.l.]: Chapman and Hall/CRC, 2018.
- CASTAGNA M. L. BATZLE, R. L. E. J. P. Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks. **GEOPHYSICS**, 1985.
- CAWLEY, G. C.; TALBOT, N. L. C. On over-fitting in model selection and subsequent selection bias in performance evaluation. **Journal of Machine Learning Research**, v. 11, p. 2079–2107, jul. 2010.
- CHALLU, C. et al. **N-HiTS: Neural Hierarchical Interpolation for Time Series Forecasting**. 2022. Disponível em: <https://arxiv.org/abs/2201.12886>.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. **CoRR**, abs/1603.02754, 2016. Disponível em: <http://arxiv.org/abs/1603.02754>.

CHO, K. et al. **Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation**. 2014. Acesso em: 24 jun. 2025. Disponível em: <https://arxiv.org/abs/1406.1078>.

COLLEY, N. **An introduction to Petrophysics**. 2015. Acesso em: 2 jul. 2025. Disponível em: <https://opc.co.uk/news/an-introduction-to-petrophysics/>.

ELLIS, J. M. S. D. V. **Well Logging for Earth Scientists**. [S.l.]: Springer Dordrecht, 2007.

EQUINOR, E. Volve data village. **Equinor Data Portal Beta**, 2018.

FENG, R.; GRANA, D.; BALLING, N. Imputation of missing well log data by random forest and its uncertainty analysis. **Computers & Geosciences**, v. 152, p. 104763, 03 2021.

FU, J. et al. Deep learning applications for well log correlation. In: . [S.l.: s.n.], 2024.

GARCIA, K. **O sistema petrolífero**. 2022. Acesso em: 22 jun. 2025. Disponível em: <https://www.desgeoeduorg.com.br/post/o-sistema-petrolifero>.

GARDNER L. W. GARDNER, A. R. G. G. H. F. Formation velocity and density - the diagnostic basis for stratigraphic traps. **Society of Exploration Geophysicists**, 1974.

GEHRING, J. et al. **Convolutional Sequence to Sequence Learning**. 2017. Disponível em: <https://arxiv.org/abs/1705.03122>.

GENG, Z.; WANG, Y. Physics-guided deep learning for predicting geological drilling risk of wellbore instability using seismic attributes data. **Engineering Geology**, v. 279, p. 105857, 2020. ISSN 0013-7952. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0013795220317543>.

GEOLOGISTS, A. A. of P. **Standard interpretation**. 2022. Acesso em: 26 jun. 2025. Disponível em: https://wiki.aapg.org/Standard_interpretation.

GRANA TAPAN MUKERJI, P. D. D. **Seismic Reservoir Modeling: Theory, Examples, and Algorithms**. [S.l.]: Wiley-Blackwell, 2021.

GREENBERG, M. L.; CASTAGNA, J. P. Shear-wave velocity estimation in porous rocks: Theoretical formulation, preliminary verification and applications. **Geophysical Prospecting**, v. 40, n. 2, p. 195–209, 1992. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2478.1992.tb00371.x>.

HE, K. et al. **Deep Residual Learning for Image Recognition**. 2015. Disponível em: <https://arxiv.org/abs/1512.03385>.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, v. 9, p. 1735–80, 12 1997.

HUSSAIN, M. et al. Application of deep learning for reservoir porosity prediction and self organizing map for lithofacies prediction. **Journal of Applied Geophysics**, v. 230, p. 105502, 2024. ISSN 0926-9851. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0926985124002180>.

INTERNATIONAL, W. A. **File:Standard-interpretation fig1.png**. 1985. Acesso em: 22 jun. 2025. Disponível em: https://wiki.aapg.org/File:Standard-interpretation_fig1.png#filelinks.

KIRANYAZ, S. et al. 1d convolutional neural networks and applications: A survey. **Mechanical Systems and Signal Processing**, v. 151, p. 107398, 2021. ISSN 0888-3270. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0888327020307846>.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)**. [S.l.: s.n.], 1995. p. 1137–1145.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. **Psychometrika**, v. 29, n. 1, p. 1–27, 1964. Disponível em: <https://doi.org/10.1007/BF02289565>.

KRUSKAL, J. B. Nonmetric multidimensional scaling: A numerical method. **Psychometrika**, v. 29, n. 2, p. 115–129, 1964. Disponível em: <https://doi.org/10.1007/BF02289694>.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998.

LIN, L. et al. Missing well log reconstruction using a sequence self-attention deep learning framework. **GEOPHYSICS**, v. 88, p. 1–60, 08 2023.

LIU, H. **Principles and Applications of Well Logging**. [S.l.]: Springer Berlin, Heidelberg, 2017.

LUCIA, F. J. **Carbonate Reservoir Characterization**. [S.l.]: Springer, 2007.

MCDONALD, A. Data quality considerations for petrophysical machine-learning models. **Petrophysics - The SPWLA Journal**, v. 62, n. 06, p. 585–613, 12 2021. ISSN 1529-9074. Disponível em: <https://doi.org/10.30632/PJV62N6-2021a1>.

PARSHALL, J. **Drilling Data Provide Solution to Horizontal Well Log Costs**. 2015. Acesso em: 23 jun. 2025. Disponível em: <https://jpt.spe.org/drilling-data-provide-solution-horizontal-well-log-costs>.

PHAM, N.; NAEINI, E. Z. Missing well log prediction using deep recurrent neural networks. *European Association of Geoscientists & Engineers*, v. 2019, n. 1, p. 1–5, 2019. ISSN 2214-4609. Disponível em: <https://www.earthdoc.org/content/papers/10.3997/2214-4609.201901612>.

QUANTCONNECT. **Applying Research: Long Short-Term Memory**. 2024. Acesso em: 24 jun. 2025. Disponível em: <https://www.quantconnect.com/docs/v2/research-environment/applying-research/long-short-term-memory>.

REFAEILZADEH, P.; TANG, L.; LIU, H. Cross-validation. In: _____. **Encyclopedia of Database Systems**. Boston, MA: Springer US, 2009. p. 532–538. ISBN 978-0-387-39940-9. Disponível em: https://doi.org/10.1007/978-0-387-39940-9_565.

RIDER, M. **The Geological Interpretation of Well Logs**. [S.l.]: Rider-French Consulting Limite, 2011.

ROSENBLATT, F. F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, 1958.

RUBNER, Y.; TOMASI, C.; GUIBAS, L. J. The earth mover's distance as a metric for image retrieval. **International Journal of Computer Vision**, v. 40, n. 2, p. 99–121, 2000. ISSN 1573-1405. Disponível em: <https://doi.org/10.1023/A:1026543900054>.

RUSSEL, P. N. S. **Artificial Intelligence: A Modern Approach**. [S.l.]: Pearson, 2020.

SAKOE, H.; CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 26, n. 1, p. 43–49, 1978.

SANTOS, D. T. d.; ROISENBERG, M.; NASCIMENTO, M. d. S. Deep recurrent neural networks approach to sedimentary facies classification using well logs. **IEEE Geoscience and Remote Sensing Letters**, v. 19, p. 1–5, 2022.

SCHÖN, J. **Basic Well Logging and Formation Evaluation**. [S.l.]: Bookboon, 2017.

TAY, Y. et al. **Efficient Transformers: A Survey**. 2022. Disponível em: <https://arxiv.org/abs/2009.06732>.

VARMA, S.; SIMON, R. Bias in error estimation when using cross-validation for model selection. **BMC Bioinformatics**, BioMed Central, v. 7, n. 1, p. 91, 2006.

VASWANI, A. et al. **Attention Is All You Need**. 2023. Disponível em: <https://arxiv.org/abs/1706.03762>.

WANG, J. et al. Missing well logs prediction using deep learning integrated neural network with the self-attention mechanism. **Energy**, v. 261, p. 125270, 2022. ISSN 0360-5442. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0360544222021557>.

WANG, Z.; YAN, W.; OATES, T. **Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline**. 2016. Disponível em: <https://arxiv.org/abs/1611.06455>.

ØDEGAARD, E.; AVSETH, P. Well log and seismic data analysis using rock physics templates. **First Break**, v. 22, p. 37–43, 10 2004.

A CÓDIGO-FONTE

Os códigos produzidos e que geraram os resultados gerados nesse trabalho encontram-se disponíveis em:

- Repositório oficial da UFSC: <https://codigos.ufsc.br/caio.pra/tcc#>
- Repositório pessoal no Github: <https://github.com/caioopra/tcc-missing-data-well-logs>

B ARTIGO NO PADRÃO SBC

Imputação de Dados Faltantes em Perfis de Poços de Petróleo: Uma Abordagem Comparativa com Redes Neurais Profundas e Mecanismos de Atenção

Caio Prá Silva, Bruno Barbosa Rodrigues e Mauro Roisenberg
Departamento de Informática e Estatística (INE)
Universidade Federal de Santa Catarina (UFSC), Brasil
{caio.pra, bruno.rodrigues, mauro.roisenberg}@ufsc.br

Resumo—A completude e precisão dos registros de poços de petróleo (*well logs*) são fundamentais para a caracterização petrofísica e modelagem de reservatórios. No entanto, a aquisição de perfis como a velocidade da onda cisalhante (VS) é onerosa e suscetível a falhas, resultando em lacunas de dados. Este trabalho investiga o uso de redes neurais profundas para o preenchimento de dados faltantes de VS, superando as limitações de linearidade dos métodos tradicionais. Foram desenvolvidas e comparadas quatro arquiteturas: uma rede LSTM (*baseline*), o modelo hierárquico N-HITS e duas variações baseadas em mecanismos de *Self-Attention* (Base e Avançada). A metodologia incluiu uma análise robusta de similaridade geológica entre poços e validação cruzada *Leave-One-Well-Out*. Os resultados indicaram que o modelo N-HITS é inadequado para dados de profundidade sem sazonalidade. Em contrapartida, o modelo de *Self-Attention* Avançado obteve o melhor desempenho, com R^2 médio de 0,47 em poços de teste, demonstrando que mecanismos de atenção bem calibrados capturam eficazmente as relações não lineares entre propriedades petrofísicas.

Index Terms—Well logs, Imputação de Dados, Deep Learning, Self-Attention, Onda Cisalhante (VS), N-HITS.

I. INTRODUÇÃO

A caracterização de reservatórios de petróleo depende intrinsecamente da qualidade dos dados adquiridos durante a perfuração. Dentre estes, os *well logs* (perfis de poço) fornecem medições contínuas de propriedades físicas das rochas, essenciais para estimar porosidade, saturação e litologia [1]. Em particular, a velocidade da onda de cisalhamento (VS) é crítica para análises geomecânicas e inversões sísmicas.

Contudo, a aquisição do perfil VS é técnica e financeiramente desafiadora. Ferramentas de medição operam em condições extremas de pressão e temperatura, estando sujeitas a falhas como o aprisionamento da sonda ou perda de contato com a parede do poço (*washouts*) [2]. A reaquisição desses dados é muitas vezes inviável economicamente, tornando a imputação computacional uma etapa essencial no fluxo de trabalho exploratório.

Métodos tradicionais, como regressões lineares multivariadas ou fórmulas empíricas (e.g., equação de Castagna [3]), frequentemente falham em capturar a alta não-linearidade e heterogeneidade dos dados geológicos. Recentemente, abordagens de *Deep Learning* têm mostrado superioridade na modelagem de séries temporais complexas [4], [5].

Este trabalho propõe e avalia arquiteturas de redes neurais profundas para a predição de VS a partir de outros perfis convencionais (como Raios Gama, Resistividade e Densidade). As principais contribuições são:

- Uma análise comparativa rigorosa entre arquiteturas recorrentes (LSTM), hierárquicas (N-HITS) e baseadas em atenção (*Transformers*);
- A demonstração empírica da inadequação de modelos focados em séries temporais sazonais (N-HITS) para dados espaciais de poços;
- O desenvolvimento de um modelo baseado em *Self-Attention* capaz de generalizar predições para poços geologicamente distintos.

II. TRABALHOS RELACIONADOS

Historicamente, a estimativa de logs faltantes baseava-se em relações empíricas físicas, como as propostas por Gardner et al. [6] e Greenberg e Castagna [7]. Embora fundamentadas na física das rochas, essas equações requerem calibração local específica e perdem precisão em zonas complexas ou com presença de gás.

A introdução de algoritmos de *Machine Learning* clássico, como *Random Forest* e XGBoost, permitiu modelar relações não lineares com maior flexibilidade [8]. Entretanto, esses modelos tratam cada ponto de profundidade independentemente, ignorando a natureza sequencial e contextual das formações geológicas.

Mais recentemente, o uso de *Deep Learning* permitiu capturar dependências espaciais ao longo do poço. Pham e Naeini [9] demonstraram a eficácia de LSTMs bidirecionais na predição de logs. Wang et al. [4] propuseram a arquitetura SAIDNN, combinando CNNs para extração de características locais e mecanismos de atenção para relevância global. Este trabalho avança nessa direção ao comparar sistematicamente modelos de estado da arte em um conjunto de dados heterogêneo fornecido pela Petrobras.

III. MATERIAIS E MÉTODOS

A. Conjunto de Dados e Pré-processamento

Foram utilizados dados de sete poços de petróleo, totalizando mais de 10.000 amostras. Cinco poços possuem a curva VS (utilizados para treino/validação) e dois não possuem

(utilizados para inferência final). As variáveis de entrada (*features*) selecionadas incluíram: Tempo de Trânsito (DT/VP), Densidade (RHO), Raios Gama (GR), Caliper, Porosidade e Saturação.

O pré-processamento envolveu o uso de *RobustScaler* para mitigar o impacto de *outliers* e a criação de janelas deslizantes para fornecer contexto sequencial às redes neurais.

B. Análise de Similaridade Geológica

Para entender a capacidade de generalização dos modelos, realizou-se uma análise de similaridade entre os poços utilizando quatro métricas: Distância de Wasserstein (distribuição estatística), Correlação de Pearson (relações entre features), *Dynamic Time Warping* (DTW - alinhamento estratigráfico) e Distância Euclidiana no espaço de features.

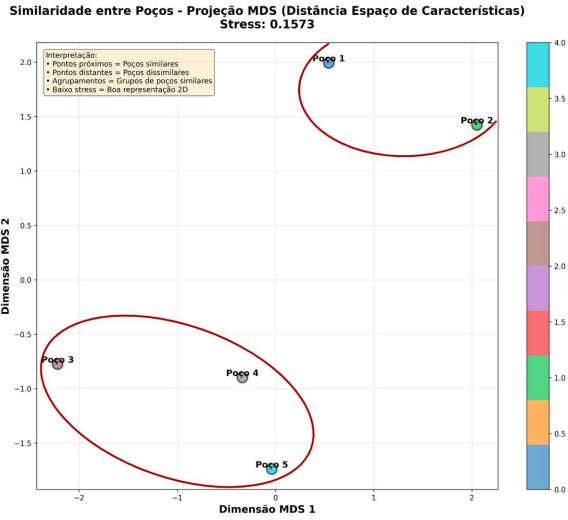


Figura 1. Projeção MDS da Distância Estatística entre poços. Observa-se a formação de dois grupos distintos (Poços 1-2 e Poços 3-5), indicando heterogeneidade geológica que desafia a generalização dos modelos.

A Figura 1 apresenta a projeção MDS (*Multidimensional Scaling*) da distância combinada. Identificou-se uma clara separação entre dois grupos de poços, sugerindo que modelos treinados em um grupo poderiam ter dificuldade de generalização no outro, o que motivou a estratégia de validação rigorosa.

C. Arquiteturas Avaliadas

1) *LSTM (Baseline)*: Uma rede recorrente *Long Short-Term Memory* foi utilizada como linha de base, composta por duas camadas LSTM com 64 neurônios e *dropout* de 30%, seguida de camadas densas.

2) *N-HITS*: O *Neural Hierarchical Interpolation for Time Series* [10] é um modelo de estado da arte para previsão de séries temporais longas. Ele utiliza decomposição multi-escala e interpolação hierárquica. Testamos configurações com 3 a 5 *stacks* e diferentes taxas de amostragem para capturar tendências de "frequência" variadas nos dados de profundidade.

3) *Self-Attention (Base e Avançado)*: Implementamos arquiteturas baseadas no mecanismo de atenção de Vaswani et al. [11], que permitem capturar dependências de longo alcance sem as limitações sequenciais das RNNs.

- **Modelo Base**: Arquitetura simplificada com 1 bloco Transformer, 4 cabeças de atenção e dimensão de embedding de 32.
- **Modelo Avançado**: Arquitetura robusta combinando extração de features via CNN com 3 blocos Transformer, 10 cabeças de atenção, dimensão de embedding de 160 e *attention pooling*. Inclui *positional encodings* treináveis e otimização com AdamW e *Huber Loss*.

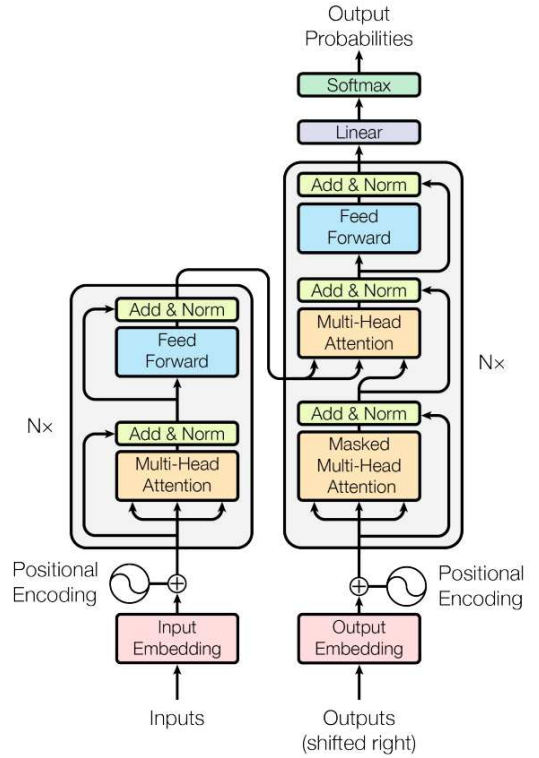


Figura 2. Representação simplificada da arquitetura baseada em Transformers utilizada nos experimentos (adaptado de [11]).

D. Estratégia de Validação

Utilizou-se a validação cruzada *Leave-One-Well-Out* (LOWOCV). Neste método, o modelo é treinado em $N - 1$ poços e testado no poço restante, simulando o cenário real de aplicação em um poço novo desconhecido.

IV. RESULTADOS E DISCUSSÃO

A. Experimentos Iniciais e N-HITS

O modelo LSTM obteve um desempenho razoável ($R^2 \approx 0,60$) nos experimentos iniciais. Entretanto, o modelo N-HITS apresentou resultados insatisfatórios e instáveis.

Como mostra a Tabela I, o N-HITS teve um R^2 médio de apenas 0,13 na validação cruzada, com alta variabilidade ($\pm 0,56$). A análise sugere que a premissa do N-HITS de

decomposição de frequência e sazonalidade não se adequa à natureza dos dados de perfis de poço, que são espaciais (profundidade) e não possuem as periodicidades típicas de séries temporais cronológicas.

B. Comparação dos Modelos de Atenção

A principal comparação se deu entre as abordagens de *Self-Attention*. O Modelo Base mostrou-se incapaz de capturar a complexidade dos dados ($R^2 \approx 0,01$).

O **Modelo Avançado**, por sua vez, superou significativamente todas as outras abordagens. A Figura 3 ilustra a dispersão entre valores reais e preditos para os poços de teste usando o modelo avançado.

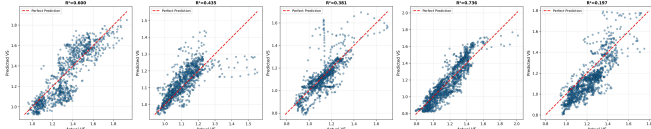


Figura 3. Gráficos de dispersão (Real vs. Predito) para o modelo Self-Atenção Avançado. Nota-se um alinhamento consistente com a diagonal principal, especialmente no Poço 3 ($R^2 = 0,736$).

A Tabela I resume a comparação final. O modelo avançado não apenas melhorou as métricas de erro (RMSE e MAE), mas principalmente a estabilidade da generalização (menor desvio padrão no R^2).

Tabela I
COMPARAÇÃO DE DESEMPENHO (MÉDIA $\pm$ DESVIO PADRÃO NA VALIDAÇÃO CRUZADA)

Métrica	N-HITS	SA Base	SA Avançado
R^2 Médio	0,13 $\pm$ 0,56	-0,08 $\pm$ 0,85	0,23 $\pm$ 0,18
RMSE	0,121 $\pm$ 0,018	0,123 $\pm$ 0,022	0,111 $\pm$ 0,024
MAE	0,095 $\pm$ 0,018	0,103 $\pm$ 0,021	0,091 $\pm$ 0,025
<i>Desempenho em Poços Individuais (Média)</i>			
R^2 Poços	0,1307	0,0117	0,4698
Melhor R^2	0,6393	0,6331	0,7356

C. Análise Qualitativa

A Figura 4 apresenta a reconstrução do perfil VS ao longo da profundidade para os poços de teste. O modelo avançado (linha tracejada) consegue acompanhar as tendências de alta frequência e variações abruptas da litologia com maior fidelidade que o modelo base, que tende a suavizar excessivamente as previsões (regressão à média).

A superioridade do modelo avançado atribui-se à maior capacidade de representação (dimensão de *embedding* 160 vs 32) e ao uso de múltiplas cabeças de atenção, permitindo que a rede foque simultaneamente em diferentes escalas de correlação geológica ao longo do poço.

V. CONCLUSÃO

Este estudo demonstrou a viabilidade técnica do uso de redes neurais para imputação de dados de onda cisalhante

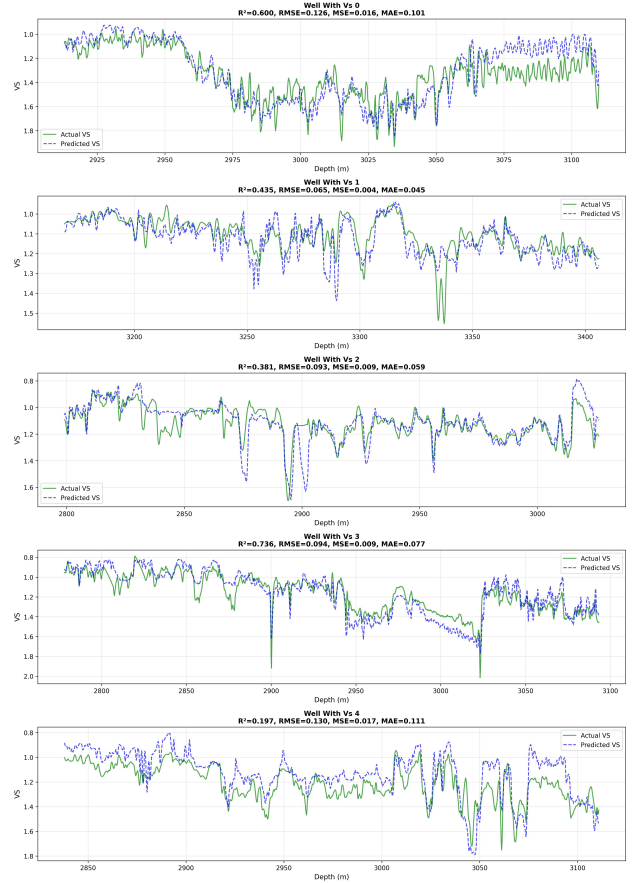


Figura 4. Comparação visual entre VS Real (linha contínua) e VS Predito (tracejada) pelo modelo Avançado em diferentes poços. O modelo captura transições geológicas abruptas com boa precisão.

em perfis de poços. A comparação rigorosa revelou que arquiteturas projetadas especificamente para séries temporais sazonais, como o N-HITS, não são adequadas para dados de profundidade geológica.

O modelo baseado em *Self-Attention* Avançado estabeleceu-se como a melhor solução, alcançando um R^2 médio de 0,47 em poços não vistos durante o treinamento. A estabilidade das previsões e a ausência de casos patológicos (R^2 negativos) validam sua robustez.

Trabalhos futuros focarão na incorporação de restrições físicas na função de perda (*Physics-Informed Neural Networks*) e na expansão do conjunto de dados para incluir diferentes bacias sedimentares, visando aumentar a generalização para contextos geológicos diversos.

AGRADECIMENTOS

Agradecemos à Petrobras pelo fornecimento dos dados e suporte técnico, e à Universidade Federal de Santa Catarina pela infraestrutura de pesquisa.

REFERÊNCIAS

- [1] P. Avseth, T. Mukerji, and G. Mavko, *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press, 2005.
- [2] Z. Geng and Y. Wang, “Physics-guided deep learning for predicting geological drilling risk of wellbore instability using seismic attributes data,” *Engineering Geology*, vol. 279, p. 105857, 2020.
- [3] J. P. Castagna, M. L. Batzle, and R. L. Eastwood, “Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks,” *Geophysics*, vol. 50, no. 4, pp. 571–581, 1985.
- [4] J. Wang *et al.*, “Missing well logs prediction using deep learning integrated neural network with the self-attention mechanism,” *Energy*, vol. 261, p. 125270, 2022.
- [5] L. Lin *et al.*, “Missing well log reconstruction using a sequence self-attention deep learning framework,” *Geophysics*, vol. 88, pp. 1–60, 2023.
- [6] G. H. F. Gardner, L. W. Gardner, and A. R. Gregory, “Formation velocity and density—the diagnostic basis for stratigraphic traps,” *Geophysics*, vol. 39, no. 6, pp. 770–780, 1974.
- [7] M. L. Greenberg and J. P. Castagna, “Shear-wave velocity estimation in porous rocks: Theoretical formulation, preliminary verification and applications,” *Geophysical Prospecting*, vol. 40, no. 2, pp. 195–209, 1992.
- [8] R. Feng, D. Grana, and N. Balling, “Imputation of missing well log data by random forest and its uncertainty analysis,” *Computers & Geosciences*, vol. 152, p. 104763, 2021.
- [9] N. Pham and E. Z. Naeini, “Missing well log prediction using deep recurrent neural networks,” *First Break*, vol. 37, 2019.
- [10] C. Challu *et al.*, “N-hits: Neural hierarchical interpolation for time series forecasting,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.
- [11] A. Vaswani *et al.*, “Attention is all you need,” in *Advances in neural information processing systems*, 2017, pp. 5998–6008.