

Previsão de Handover usando TinyML na Borda

Gabriel Tadeu Gonzaga Diniz

Relatório Técnico - IC-PFG-26-30

Projeto Final de Graduação

2026 - Julho

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

The contents of this report are the sole responsibility of the authors.
O conteúdo deste relatório é de única responsabilidade dos autores.

Previsão de Handover usando TinyML na Borda

Gabriel Tadeu Gonzaga Diniz*

Resumo

Este trabalho propõe mitigar o impacto da latência e da perda de pacotes em redes veiculares por meio da predição antecipada de falhas de conexão. O estudo analisa a viabilidade de algoritmos de aprendizado de máquina otimizados para dispositivos com recursos limitados de memória e processamento, empregando o paradigma *Tiny Machine Learning* (TinyML). Os modelos foram treinados sem restrições iniciais de hardware, tendo seu consumo de memória e tempo de inferência avaliados a posteriori para assegurar a aplicabilidade na Borda (Unidade de Bordo / *On-Board Unit* - OBU). Utilizando dados empíricos de rádio no padrão *Intelligent Transport Systems* em 5 GHz (ITS-G5) e de rede celular, o sistema aplica o deslocamento de janelas no tempo (*time-shifting*) para prever quedas de sinal. Essa técnica permite o acionamento antecipado do *handover* preditivo. Cabe ressaltar que, embora os testes tenham sido conduzidos no contexto de transição entre o padrão *Wireless Access in Vehicular Environments* (WAVE) e as tecnologias celulares 4G/5G, o método proposto é agnóstico e aplicável a outros tipos de handover preditivo, tanto horizontal como vertical. Os resultados indicam que, enquanto modelos como o *Random Forest* são eficientes para previsões de curtíssimo prazo, a Rede Neural Convolucional 1D (CNN 1D) despontou como a solução *TinyML* mais equilibrada. Demandando apenas ≈ 2 KB de memória e apresentando inferência rápida (≈ 7.9 ms), a CNN 1D superou redes recorrentes mais complexas tais como *Long Short Term Memory* (LSTM) e *Gated Recurring Unit* (GRU), mantendo precisão satisfatória (acima de 75%) em horizontes de curto (1s) e médio prazo (10s) e apresentando limitações mais relevantes apenas em cenários de previsão de longo prazo (20s).

1 Introdução

A conectividade contínua é um requisito básico para aplicações veiculares voltadas à segurança, ao monitoramento e à coordenação do tráfego. Serviços como troca de alertas em tempo real, telemetria e apoio à condução conectada dependem de comunicação estável mesmo quando o veículo se desloca por áreas com cobertura heterogênea.

As redes veiculares, incluindo *Vehicular Ad Hoc Networks* (VANETs) [7] e sistemas *Vehicle-to-Everything* (V2X), sustentam essa comunicação veículo-a-veículo (*Vehicle-to-Vehicle* - V2V), veículo-a-infraestrutura (*Vehicle-to-Infrastructure* - V2I) e veículo-a-rede (*Vehicle-to-Network* - V2N). Para viabilizar esses cenários, diferentes tecnologias de acesso

*Instituto de Computação, UNICAMP, 13083-852 Campinas, SP. Orientador: Prof. Dr. Edmundo Roberto Mauro Madeira.

sem fio são comumente empregadas. As opções de curto e médio alcance, como os padrões do *Institute of Electrical and Electronics Engineers* (IEEE) 802.11p e do *European Telecommunications Standards Institute* (ETSI) ITS-G5 [11, 6], são ideais para trocas rápidas de dados locais. Simultaneamente, tecnologias celulares como C-V2X (apoiadas no 4G/5G) complementam o ecossistema provendo cobertura ampla e conexão com a nuvem. Estudos recentes demonstram, inclusive, a viabilidade técnica da coexistência prática e operação simultânea das tecnologias ITS-G5 e 5G NR-U (New Radio-Unlicensed) na mesma faixa de espectro [6, 5, 19].

Em ambientes urbanos dinâmicos, a área de cobertura de uma única antena ou tecnologia raramente abrange todo o trajeto. É comum que um veículo saia do raio de alcance de um ponto de acesso ITS-G5 localizado em um cruzamento, por exemplo, e precise transferir sua conexão para a rede celular 4G/5G para não perder conectividade. Esse processo de transferência da sessão em andamento entre antenas ou tecnologias distintas é conhecido como *handover*. Para que essa troca seja eficiente e a comunicação não sofra interrupções, o *handover* deve atender a requisitos rigorosos: precisa ocorrer no momento exato, com latência mínima e sem descarte de pacotes. Isso evita falhas em serviços críticos de segurança e garante uma transição transparente para as aplicações veiculares.

Quando a decisão de *handover* ocorre apenas após a degradação efetiva do enlace, surgem atrasos, perda de pacotes e instabilidade de comunicação. Esse problema é especialmente crítico em aplicações sensíveis a tempo, nas quais poucos milissegundos podem afetar a utilidade da mensagem transmitida. Por isso, abordagens recentes têm investigado o uso de inteligência artificial para antecipar o melhor momento de comutação [16, 8]. O desafio é que muitas dessas soluções dependem de processamento em nuvem ou em infraestrutura de Borda com maior capacidade computacional, o que pode introduzir atrasos adicionais ou elevar o custo de implantação.

Como alternativa, este trabalho investiga a viabilidade de embarcar a predição diretamente na Unidade de Bordo (*On-Board Unit* - OBU) do veículo. Como esses dispositivos possuem restrições severas de memória e processamento, a proposta baseia-se no paradigma *Tiny Machine Learning* (TinyML) [17]. Para capturar a dinâmica temporal das medições de rádio sem exigir modelos complexos em tempo de inferência, empregou-se a técnica de deslocamento temporal de rótulos (*time-shifting*) no treinamento.

Nesse contexto, foi desenvolvido um sistema para comparar diferentes modelos leves de predição, desde métodos estatísticos clássicos até uma Rede Neural Convolutiva 1D (CNN 1D), com o objetivo de acionar o *handover* de forma proativa. A validação foi realizada sobre uma base empírica de conectividade urbana coletada em Aveiro, Portugal [5], com medições reais de ITS-G5 e rede celular em condições de tráfego normal. Além de métricas clássicas de classificação, como Acurácia, F1-Score e a Área Sob a Curva (*Area Under the Curve* - AUC), a análise também considera aspectos centrais para execução embarcada, como uso de memória, tempo de inferência e comportamento em diferentes horizontes de previsão. Embora os experimentos tenham sido conduzidos no cenário de transição entre o padrão ITS-G5 e a rede celular — um caso de *handover* vertical, em que a sessão migra entre tecnologias distintas —, a metodologia proposta pode ser estendida a outros contextos de *handover* preditivo. Isso inclui o *handover* horizontal, no qual o veículo permanece na mesma tecnologia, mas troca de ponto de acesso, por exemplo ao passar de uma antena

ITS-G5 para outra ao longo da via.

Diferente das abordagens que delegam o processamento para a nuvem, o foco deste trabalho é executar a inferência na própria OBU. Embora as redes recorrentes (LSTM e GRU) tenham sido avaliadas como referência de comparação, o objetivo é verificar se arquiteturas sem memória sequencial explícita conseguem resultados competitivos, uma vez que manter estados ocultos ao longo do tempo exige mais RAM do que os microcontroladores típicos dispõem.

Para que esses modelos mais simples consigam prever eventos futuros, aplicou-se a técnica de *time-shifting* durante o treinamento. A ideia é deslocar os rótulos de falha W passos no tempo, de modo que o modelo aprenda a associar o estado atual da rede a uma queda que ocorrerá no futuro. Com isso, é possível usar classificadores estáticos para realizar previsões temporais sem a necessidade de manter memória entre amostras.

O restante deste relatório está organizado da seguinte forma. A Seção 2 apresenta a fundamentação teórica e os trabalhos relacionados. A Seção 3 descreve a arquitetura do sistema e a aplicação do *time-shifting*. A Seção 4 apresenta o dataset de Aveiro [22, 5], os cenários de avaliação e os resultados experimentais. Por fim, a Seção 5 apresenta as conclusões e possíveis desdobramentos do trabalho.

2 Fundamentação Teórica e Trabalhos Relacionados

Este capítulo apresenta a base teórica do trabalho e situa a proposta em relação à literatura. A Seção 2.1 resume os fundamentos de *TinyML* e os algoritmos avaliados. A Seção 2.2 discute trabalhos sobre redes V2X e sobre o uso de técnicas preditivas para gerência de *handover*.

2.1 Background Computacional

Em redes veiculares, determinar o momento exato (*timing*) da transição de rede é crítico. Um *handover* tardio resulta em perda de conexão e descarte de pacotes, enquanto uma transição prematura ou induzida por uma oscilação efêmera no sinal faz com que o sistema sofra com trocas sucessivas e desnecessárias entre redes — fenômeno conhecido como efeito *ping-pong* [15, 13].

Para solucionar esse impasse e antecipar o *handover*, adotam-se algoritmos preditivos. A implementação desses algoritmos, no entanto, exige a definição prévia de onde o processamento dos dados ocorrerá. As principais estratégias arquiteturais dividem-se em três níveis: processamento centralizado na Nuvem (*Cloud*), processamento intermediário na infraestrutura de Borda (*Edge*, como nas antenas das vias) ou processamento local no próprio dispositivo móvel (o veículo). Alocar a tomada de decisão em servidores em nuvem oferece amplo poder computacional, mas adiciona um atraso de comunicação de rede que pode inviabilizar a utilidade temporal da previsão. A computação de Borda mitiga esse atraso aproximando a inferência física do usuário, mas ainda assim mantém a dependência de um enlace sem fio estável para retornar o comando de transição ao carro. Dessa forma, para garantir respostas estritamente ágeis e autônomas, eliminando a sobrecarga de rede, o pro-

cessamento da inteligência artificial deve ocorrer localmente, de forma descentralizada, na própria Unidade de Bordo (OBU) do veículo.

É nesse ponto que se insere o Tiny Machine Learning (TinyML). Formalizado por Warden e Situnayake [24], o TinyML engloba a aplicação de inteligência artificial em sistemas embarcados de baixa complexidade e consumo energético restrito. A premissa é otimizar algoritmos para que possam ser executados em dispositivos de baixa capacidade (possuindo poucos Kilobytes de memória RAM – *Random Access Memory* – e Flash) [17, 23]. Dessa forma, viabiliza-se a inferência preditiva em milissegundos localmente no dispositivo, mitigando a dependência da nuvem e preservando a privacidade dos dados do veículo.

Para prever a degradação do canal de comunicação, foram testadas duas categorias de algoritmos. A primeira abrange o Aprendizado Estatístico Clássico:

- **Regressão Logística [10]:** Trata-se de um modelo linear de baixa complexidade que estima a probabilidade de falha com base nas métricas de rádio. Embora exija recursos computacionais mínimos (sendo ideal para hardware embarcado), sua simplicidade impõe limitações na modelagem de oscilações não-lineares, típicas de ambientes urbanos densos.
- **Árvores de Decisão [21]:** Estabelece partições sequenciais no espaço de características. São interpretáveis e possuem inferência rápida, mas são particularmente suscetíveis ao sobreajuste (*overfitting*), correndo o risco de apenas memorizar o trajeto de treinamento em vez de generalizar o comportamento da atenuação do sinal.
- **Random Forest [2]:** Utiliza um conjunto de árvores de decisão (*ensemble*) que chegam a uma predição combinando as predições individuais de cada árvore. Essa estratégia atenua o *overfitting* e demonstra alta robustez em cenários urbanos imprevisíveis. Em contrapartida, o modelo consome significativamente mais memória, configurando um gargalo crítico em aplicações TinyML.

A segunda categoria abrange o *Deep Learning* (Redes Neurais Profundas). Essas arquiteturas, concebidas inicialmente para infraestruturas de alto desempenho, exigem otimização drástica para viabilizar sua operação sob as restrições de um dispositivo de baixa capacidade:

- **LSTM [9] e GRU [4]:** As redes *Long Short-Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU) são topologias consolidadas para o processamento de sequências temporais, pois mantêm estados internos (*hidden states*) sobre o histórico de oscilações do rádio. Entretanto, a atualização contínua dessas matrizes de estado demanda elevado consumo de memória RAM, sendo testadas neste estudo primordialmente para estabelecer um limite superior de sobrecarga suportável pela Borda.
- **CNN 1D [14]:** Esta arquitetura propõe uma abstração distinta: em vez de manter memória recorrente sobre o longo histórico temporal, ela extrai características de uma janela de tempo específica através de filtros convolucionais, tratando as métricas como um padrão estático. Ao prescindir de estados ocultos persistentes, a CNN 1D reduz drasticamente o consumo de memória e processamento, despontando como uma arquitetura promissora para execução embarcada.

2.2 Trabalhos Relacionados

A Tabela 1 apresenta um resumo consolidador dos principais artigos investigados no contexto de gerência de *handover*. O levantamento abrange desde trabalhos seminais, focados em estratégias de decisão em redes heterogêneas, até abordagens mais recentes que exploram o aprendizado de máquina para previsão e transição proativa em ambientes veiculares.

Tabela 1: Resumo dos artigos citados sobre *handover*.

Referência	Tipo	Tecnologia / cenário
<i>An overview of vertical handover decision strategies in heterogeneous wireless networks</i> [12]	Vertical (survey)	Redes sem fio heterogêneas
<i>Cross-Layer-Based Adaptive Vertical Handoff With Predictive RSS in Heterogeneous Wireless Networks</i> [3]	Vertical	WiMAX / WiFi / B3G / 4G
<i>Predictive link trigger mechanism for seamless handovers in heterogeneous wireless networks</i> [26]	Horizontal e vertical	Redes heterogêneas; IEEE 802.21 MIH
<i>Ping-Pong Avoidance Algorithm for Vertical Handover in Wireless Overlay Networks</i> [13]	Vertical	Redes overlay sem fio
<i>Theoretical analysis of handover failure and ping-pong rates for heterogeneous networks</i> [15]	Vertical	Redes heterogêneas
<i>A two-tier machine learning-based handover management scheme for intelligent vehicular networks</i> [1]	Vertical / veicular	Redes veiculares inteligentes
<i>Machine Learning-Based Handovers for Sub-6 GHz and mmWave Integrated Vehicular Networks</i> [25]	Vertical / veicular	Rede veicular integrada sub-6 GHz + mmWave
<i>Federated Learning Based Proactive Handover in Millimeter-wave Vehicular Networks</i> [20]	Veicular / proativo	Redes veiculares mmWave
<i>Mobility Management Paradigm Shift: From Reactive to Proactive Handover Using AI/ML</i> [18]	Survey	5G e além
<i>Deep Learning-Based Handover Prediction for 5G and Beyond Networks</i> [16]	Horizontal / celular	5G e beyond; dados reais + simulação
<i>Machine Learning-Based Handover Performance Prediction for Beyond 5G Communication Networks</i> [8]	Horizontal / celular	B5G
<i>Deep Learning Based Handover for High-Speed Connected Vehicles in Ultra-Dense Networks</i> [27]	Horizontal / celular / veicular	Ultra-dense networks com veículos conectados

2.2.1 Redes Veiculares V2X e Gerência de Handover

O interesse por redes veiculares V2X cresceu com a necessidade de manter comunicação confiável para aplicações de segurança, coordenação de tráfego e sensoriamento em mobilidade [6]. Nesse cenário, a gerência de *handover* torna-se um problema central, sobretudo quando o veículo precisa alternar entre tecnologias com coberturas e características distintas. Em redes heterogêneas, a literatura consolidou uma base importante para a decisão de

handover vertical, levando em conta critérios como intensidade de sinal, QoS (Quality of Service - Qualidade de Serviço), custo e contexto de mobilidade [12].

As primeiras abordagens preditivas buscavam antecipar a troca de rede a partir da evolução do *Received Signal Strength* (RSS) e de gatilhos adaptativos, reduzindo perdas e atrasos durante a transição [3, 26]. Mais recentemente, o foco passou de mecanismos reativos para estratégias proativas apoiadas por inteligência artificial e aprendizado de máquina [18, 16, 8]. No contexto veicular, Yan et al. [25] utilizam previsão de posição e histórico de *handovers* para acelerar a comutação entre sub-6 GHz e mmWave. Aljeri e Boukerche [1] combinam predição de RSS com seleção do próximo ponto de acesso em redes veiculares inteligentes. Qi et al. [20] exploram aprendizado federado para *handover* proativo, enquanto Zhang, Hou e Sun [27] avaliam CNN 1D em cenários de alta mobilidade.

Entre as variáveis usadas para disparar a decisão, o *Received Signal Strength Indicator* (RSSI) continua sendo uma das mais comuns, por indicar de forma direta a degradação do enlace e a proximidade da perda de cobertura [3, 26]. Em trabalhos mais recentes, porém, ele costuma ser combinado com informações adicionais, como posição, histórico de mobilidade, qualidade do canal e métricas de rede, para reduzir transições desnecessárias e melhorar a continuidade do serviço [25, 1, 8].

Apesar desses avanços, a maior parte das soluções da literatura pressupõe execução em nuvem, em infraestrutura de Borda mais robusta ou em plataformas veiculares com maior capacidade computacional. A contribuição deste trabalho difere desse foco ao investigar a viabilidade de realizar a predição diretamente na OBU com modelos leves, avaliados também sob consumo de memória e latência de inferência. Em vez de buscar apenas alta acurácia de predição, a proposta prioriza o equilíbrio entre desempenho e viabilidade embarcada, o que aproxima a discussão do contexto prático de *TinyML* para *handover* preditivo.

2.2.2 Coexistência entre Tecnologias: ITS-G5 e 5G NR-U

Os veículos autônomos e conectados lidam com necessidades diferentes de comunicação. Para alertas de segurança no trânsito, é necessário o envio muito rápido de mensagens curtas (papel em que a rede ITS-G5 é excelente). Por outro lado, serviços modernos exigem uma conexão que suporte um grande volume de dados (como transmissão de vídeos de sensores), algo que as redes celulares privadas 5G NR-U podem oferecer [19].

Para reunir esses benefícios, uma solução promissora é utilizar as duas tecnologias de forma combinada. No entanto, fazer o ITS-G5 e o 5G operarem juntos na mesma faixa de rádio (a frequência de 5.9 GHz) gera o risco de interferências. Para que as duas redes compartilhem o mesmo canal de forma justa e sem colisões, elas utilizam um mecanismo chamado *Listen-Before-Talk* (LBT), que exige que o rádio “escute” o meio de transmissão para confirmar que ele está livre antes de enviar qualquer dado.

O trabalho recente de Peixoto et al. [19] testou essa convivência na prática, usando equipamentos reais em laboratório. Os autores comprovaram que a coexistência é totalmente viável: mesmo quando a rede 5G estava enviando muitos dados e ocupando bastante o canal, as mensagens críticas de segurança do ITS-G5 quase não foram afetadas. O atraso gerado foi de apenas 0,7 milissegundos e praticamente não houve perda de pacotes.

Esse estudo prático reforça a motivação do nosso trabalho. Sabendo que é possível

manter as tecnologias ITS-G5 e 5G ativas e operando juntas no mesmo veículo, o mecanismo de predição de *handover* torna-se fundamental. Em uma rede veicular híbrida (que possui as duas tecnologias), antecipar quando o sinal de uma rede vai cair ou ficar congestionado permite que o sistema migre a comunicação para a outra tecnologia a tempo, garantindo que o veículo nunca perca sua conexão com o ambiente.

3 Arquitetura Proposta

Esta seção descreve a estrutura do sistema desenvolvido, desde o carregamento e pré-processamento dos dados até o treinamento dos modelos e a análise de viabilidade para execução embarcada. É importante ressaltar que o escopo deste trabalho se concentra na fase de treinamento e avaliação offline. O objetivo é validar a abordagem algorítmica e verificar se os modelos atendem às restrições de hardware da OBU.

A Figura 1 ilustra o fluxo sequencial de execução da metodologia, desde o tratamento dos dados até a avaliação de desempenho e custo computacional:

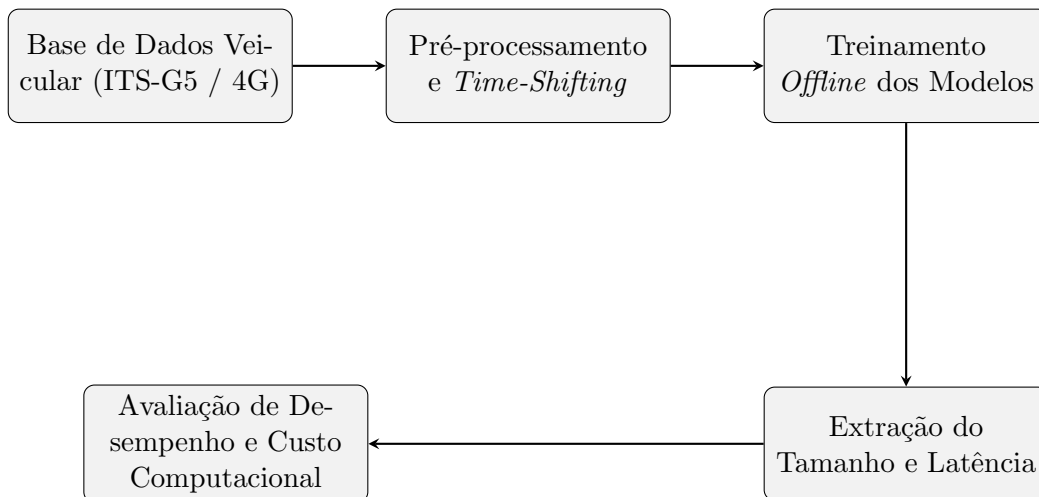


Figura 1: Diagrama de *Pipeline*: fluxo sequencial de execução do sistema de treinamento e validação.

A Figura 2 complementa a visão acima, apresentando a arquitetura do sistema sob a perspectiva dos seus módulos funcionais e das responsabilidades de cada componente. Enquanto o diagrama anterior descreve *como* os dados fluem, este diagrama descreve *o que* cada bloco do sistema é e faz:

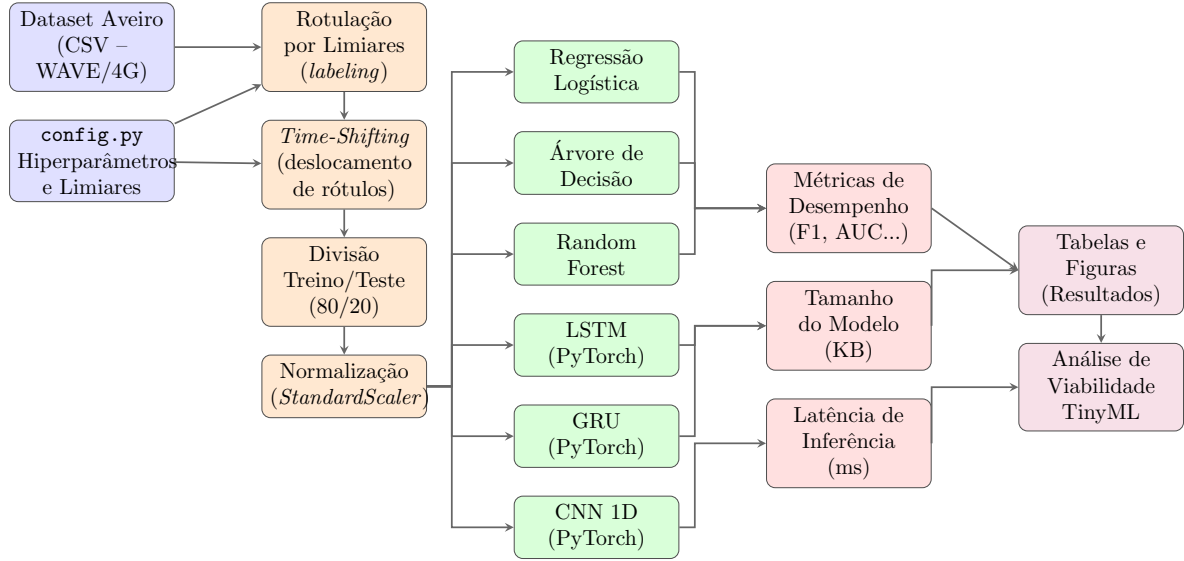


Figura 2: Diagrama de Blocos do Sistema: módulos funcionais e suas responsabilidades no processo de predição de *handover* via TinyML.

O fluxo de processamento (*pipeline*) inicia-se com a leitura dos dados empíricos coletados em Aveiro. Conforme detalhado por Dias et al. [5], a coleta foi realizada utilizando 8 veículos instrumentados com OBUs, que transitaram de forma aleatória pela malha urbana durante meses. Respeitando o limite de velocidade local (50 km/h), os veículos injetaram tráfego contínuo (através da ferramenta *iPerf3*) em direção a 11 Unidades de Beira de Estrada (*Roadside Units* - RSUs) distribuídas pelo município. Durante os percursos, o sistema registrou as coordenadas geográficas (GPS) e aferiu o desempenho da rede em uma frequência de 10 Hz, compondo uma base de dados massiva de telemetria ITS-G5.

A Tabela 2 elenca as variáveis (*features*) utilizadas para o treinamento dos modelos preditivos, incluindo os dados originais aferidos no canal de rádio e as características sintéticas derivadas no pré-processamento:

Tabela 2: Descrição das características de telemetria e rede obtidas do Dataset de Aveiro (Dias et al., 2023).

Feature	Unidade	Descrição
obu_id / rsu_id	Numérico	Identificadores do veículo (OBU) e da antena de Borda (RSU) atrelados ao enlace.
latitude / longitude	Graus	Coordenadas geográficas exatas do veículo (interpoladas da frequência do GPS).
heading	Graus	Direção cinemática de deslocamento do veículo em relação ao norte geográfico.
dist_rsu	Metros	Distância euclidiana instantânea entre o veículo em movimento e a antena conectada.
rss_i_obu / rss_i_rsu	dBm	<i>Received Signal Strength Indicator</i> : potência do sinal recebido sob a perspectiva do veículo (OBU) e da antena (RSU), respectivamente. No <i>pipeline</i> , ambas são fornecidas como <i>features</i> independentes. Contudo, é o <code>rss_i_obu</code> que define o critério de falha: leituras abaixo de -87 dBm (do ponto de vista do veículo) ativam o rótulo de desconexão iminente.
bitrate	Mbps	Taxa de transferência de dados efetiva atingida no canal de rádio ITS-G5.
loss	%	Percentual de pacotes <i>User Datagram Protocol</i> (UDP) perdidos na comunicação (<i>Packet Loss</i>).
jitter	ms	Variação estatística do atraso de entrega na transmissão contínua de pacotes.
rtt	ms	<i>Round-Trip Time</i> : latência total de ida e volta aferida na transmissão.
label	Binário (0/1)	Rótulo sintético gerado por nossa etapa de <i>time-shifting</i> : 1 indica iminência preditiva de desconexão da rede.

No pré-processamento, definiu-se o critério de falha de conexão: amostras com perda de pacotes acima de 10% ou RSSI abaixo de -87 dBm recebem o rótulo positivo (falha iminente). Esses limiares não são fornecidos pelo dataset de Aveiro — que apenas disponibiliza as métricas brutas de qualidade do enlace — e foram adotados neste trabalho como critério operacional de rotulação. A escolha de 10% de perda alinha-se à prática dos padrões IEEE 802.11p / ITS-G5, nos quais a sensibilidade do receptor é especificada sob a condição de taxa de erro de pacotes (*Packet Error Rate* - PER) de até 10% [11, 6]. Já o limiar de -87 dBm situa-se na região típica de limite de recepção confiável para enlaces ITS-G5 em canal de 10 MHz, indicando que o sinal já se encontra próximo da borda em que a comunicação deixa de ser estável. Em seguida, aplica-se o *time-shifting*: os rótulos são deslocados W passos no passado, de forma que o modelo aprenda a prever, a partir do estado atual da rede, uma falha que ocorrerá no futuro.

O treinamento foi realizado offline, sem restrições impostas ao hardware durante o ajuste dos modelos. A avaliação, por sua vez, contemplou tanto o desempenho preditivo (Acurácia, F1-Score e demais métricas) quanto o custo computacional de cada solução, medido pelo tempo de inferência por amostra e pela estimativa de uso de memória RAM na OBU.

4 Avaliação de Resultados

Esta seção apresenta a avaliação empírica dos modelos. A Subseção 4.1 descreve o dataset de Aveiro, as métricas e os três horizontes de antecipação. As Subseções 4.2, 4.3 e 4.4 analisam, respectivamente, os resultados em curto ($W = 10$), médio ($W = 100$) e longo prazo ($W = 200$). Por fim, a Subseção 4.5 discute a importância das variáveis na predição.

4.1 Dataset de Aveiro e Cenários de Avaliação

Para validar o sistema empiricamente, utilizou-se a base de dados coletada por Dias et al. [5], resultado de campanhas reais de medição conduzidas na cidade de Aveiro, Portugal, onde está instalado o *Aveiro Tech City Living Lab* (ATCLL) [22].

A principal vantagem dessa base é que ela não se limita a simulações. O dataset foi gerado em um ambiente urbano real, com 44 pontos de acesso multimodal (4G, 5G, Wi-Fi e ITS-G5) e frotas de veículos públicos equipados com OBUs percorrendo a cidade. Os dados de conectividade foram coletados de forma contínua, em condições normais de tráfego [22].

Como as falhas de conexão ocorreram por causa de obstáculos físicos reais (prédios, cruzamentos, variações de tráfego), o dataset captura de forma direta a relação entre a posição do veículo (via *Global Positioning System* - GPS) e o comportamento da rede (RSSI, perda de pacotes, *jitter*). Isso o torna adequado para treinar e avaliar modelos preditivos em condições realistas.

Os experimentos utilizam leituras do estado do canal (`rss_i_obu`, `rss_i_rsu`, perda, *jitter*, *bitrate*, latência) combinadas ao posicionamento do veículo (`latitude`, `longitude`, distância, direção). O conjunto foi particionado na proporção 80/20 (treino/teste), preservando-se rigorosamente a ordem temporal das amostras para evitar o vazamento de informações futuras para o treinamento (*data leakage*). A validação baseou-se nas seguintes métricas: **Acurácia** (taxa global de acertos), **Precisão** (proporção de alertas de *handover* que eram genuinamente necessários), **Recall** (capacidade de detectar as quedas reais), **F1-Score** (média harmônica entre Precisão e Recall, sendo o principal indicador neste estudo) e **AUC** (área sob a curva *Receiver Operating Characteristic* - ROC, resumindo a qualidade discriminativa geral).

Para avaliar a robustez dos modelos, definiram-se três cenários de aplicação a partir da variação do parâmetro W (*time-shifting*): $W = 10$ (curto prazo, ≈ 1 segundo de antecedência), $W = 100$ (médio prazo, ≈ 10 segundos) e $W = 200$ (longo prazo, ≈ 20 segundos). Como a frequência de amostragem da telemetria é de 10 Hz [5], o deslocamento de 10 passos nos dados equivale à antecipação temporal de 1 segundo.

Esses horizontes não foram tomados de um limiar experimental fixo da literatura; foram escolhidos neste trabalho para cobrir regimes distintos de antecipação, à luz da mobilidade observada na coleta. Nos testes de Aveiro, os veículos respeitaram o limite urbano de

aproximadamente 50 km/h [5]. Nessa velocidade média, o deslocamento espacial em cada horizonte fica claro: em 1 segundo o veículo percorre cerca de 14 metros; em 10 segundos, cerca de 140 metros; e em 20 segundos, quase 280 metros. O horizonte de 1 segundo mantém o carro praticamente na mesma vizinhança de cobertura, o que é tempo suficiente para acionar um *handover* ágil e reduzir perdas imediatas, sem exigir uma previsão de longo alcance. O horizonte de 10 segundos já cobre uma distância típica de quarteirão urbano — conversões, saída da área de uma antena ou entrada em sombra de prédios — e, por isso, oferece uma margem útil para planejamento da troca de rede. Já o horizonte de 20 segundos empurra a previsão para uma escala em que o trajeto pode incluir semáforos, mudanças de via e passagem por várias zonas de cobertura; ele serve, portanto, para testar o limite prático do método frente à imprevisibilidade da mobilidade urbana.

4.2 Resultados em Curto Prazo ($W = 10$)

A Tabela 3 consolida os resultados no cenário mais favorável à previsão (antecedência de 1 segundo). Dado que o deslocamento espacial do veículo é mínimo nesse intervalo, a degradação atual do rádio apresenta alta correlação com a falha iminente. Consequentemente, todos os modelos obtiveram desempenho excelente. A análise central recai, portanto, sobre a relação de custo-benefício computacional: ponderar a carga de memória e inferência exigida por cada algoritmo em troca de ganhos marginais na eficácia preditiva.

Tabela 3: Valores Analíticos Consolidados - Antecipação de 1 segundo ($W = 10$).

Modelo	Tempo de Treino (s)	Tempo Infer. (s)	Tam. (KB)	Acurácia	Precisão	Recall	F1-Score	AUC
Regressão Logística	0.0770	0.0020	0.74	0.8912	0.8820	0.9562	0.9176	0.9198
Árvore de Decisão	0.1883	0.0018	6.00	0.9049	0.8966	0.9607	0.9276	0.9046
Random Forest	0.1211	0.0377	78.76	0.9039	0.8913	0.9661	0.9272	0.9262
LSTM Neural Network	40.9273	0.0276	22.13	0.8987	0.8776	0.9763	0.9243	0.9269
GRU Neural Network	54.2732	0.0604	16.63	0.9089	0.8990	0.9646	0.9307	0.9319
CNN 1D	18.5307	0.0079	2.00	0.9058	0.9006	0.9569	0.9279	0.9242

A Figura 3 ilustra visualmente as Matrizes de Confusão e as Curvas ROC dos modelos clássicos em $W = 10$. A Matriz de Confusão permite visualizar os quatro cenários de predição: Verdadeiros Positivos (quedas corretamente alertadas), Verdadeiros Negativos (conexões estáveis corretamente confirmadas), Falsos Positivos (alarmes desnecessários de *handover*) e Falsos Negativos (quedas reais que passaram despercebidas — o erro mais crítico em termos de continuidade de serviço). A Curva ROC, por sua vez, traça o balanço entre sensibilidade e especificidade para todos os limiares de decisão possíveis, com a AUC quantificando a área sob essa curva.

Entre os modelos clássicos, o *Random Forest* obteve a maior AUC (0.926) e o melhor Recall (0.966), indicando que rara foi a queda de conexão que passou sem ser detectada. A *Árvore de Decisão* ficou ligeiramente abaixo em AUC (0.905), porém com F1-Score de 0.9276. A *Regressão Logística*, apesar de ser o modelo mais simples e ocupar apenas 0.74 KB, atingiu surpreendente F1-Score de 0.9176 — demonstrando que, para o horizonte curto, a relação sinal-queda é suficientemente linear para ser capturada por um único hiperplano de decisão.

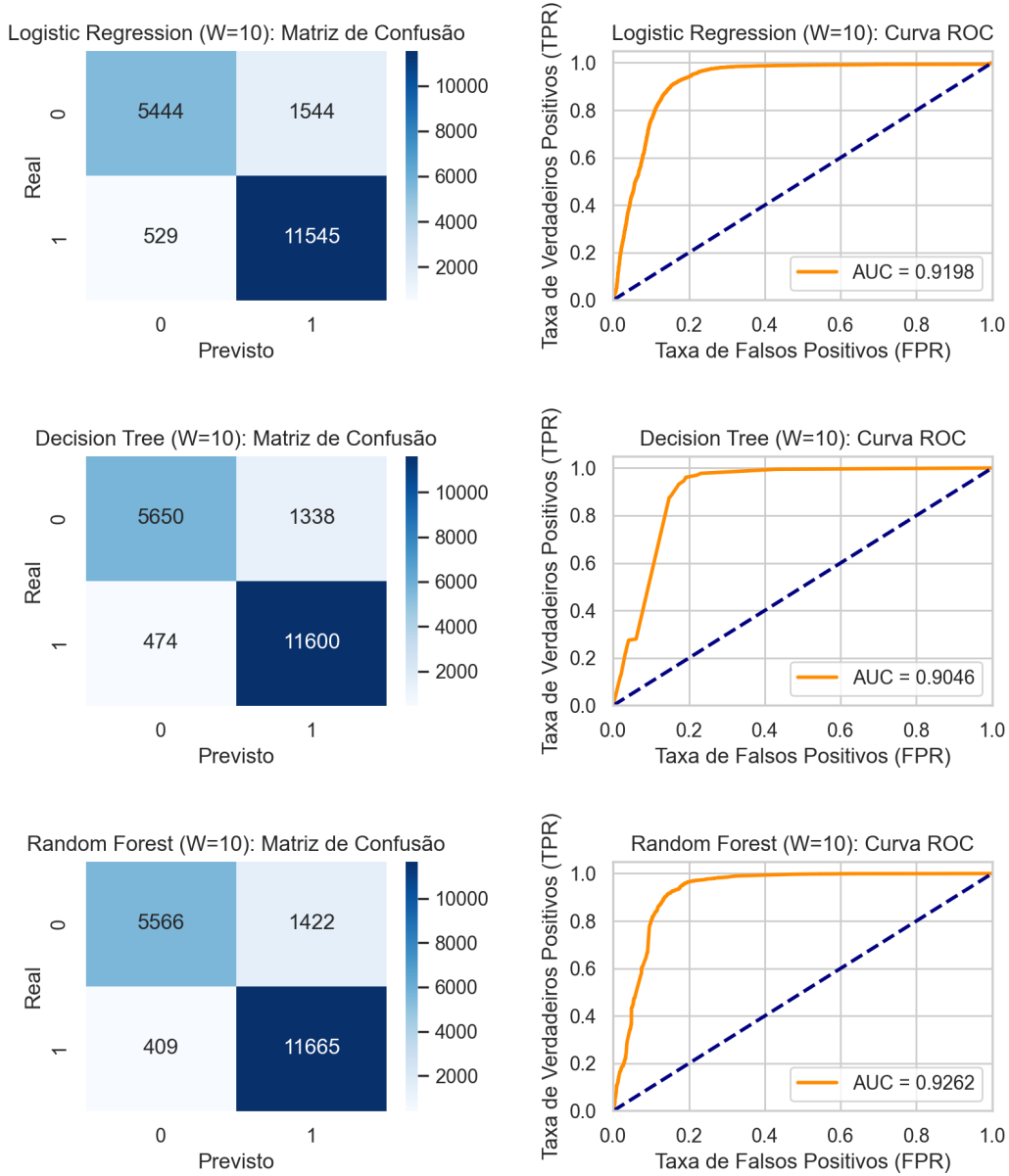


Figura 3: Matrizes de Confusão e Curvas ROC (*Receiver Operating Characteristic*) dos modelos TinyML estatísticos (Regressão, Árvore de Decisão e Random Forest) em $W = 10$.

As arquiteturas de *Deep Learning* (Figura 4) ilustram o principal dilema do paradigma TinyML: precisão preditiva *versus* viabilidade de hardware. A rede GRU atingiu o maior F1-Score absoluto do experimento (0.9307), acompanhada de perto pela LSTM (0.9243). No entanto, o custo imposto ao sistema embarcado é proibitivo: a GRU demanda 16.63 KB de memória e mais de 54 segundos de treinamento; a LSTM requer 22.13 KB e 41 segundos. Considerando dispositivos automotivos típicos, essas redes consomem uma fração

excessiva dos recursos, comprometendo a operação simultânea de outros subsistemas críticos do veículo.

A CNN 1D desponta como a solução de compromisso ideal: demandando apenas 2 KB de memória e com tempo de inferência de ≈ 8 ms, alcançou um F1-Score de 0.9279. Esse desempenho supera marginalmente o *Random Forest* (0.9272), mas com um consumo de RAM 39 vezes menor (o *ensemble* requer 78.76 KB). Isso demonstra que a extração de características via filtros convolucionais (prescindindo de estados recorrentes complexos) é altamente representativa para detectar quedas de sinal no curto prazo, configurando uma excelente abordagem para o TinyML [14].

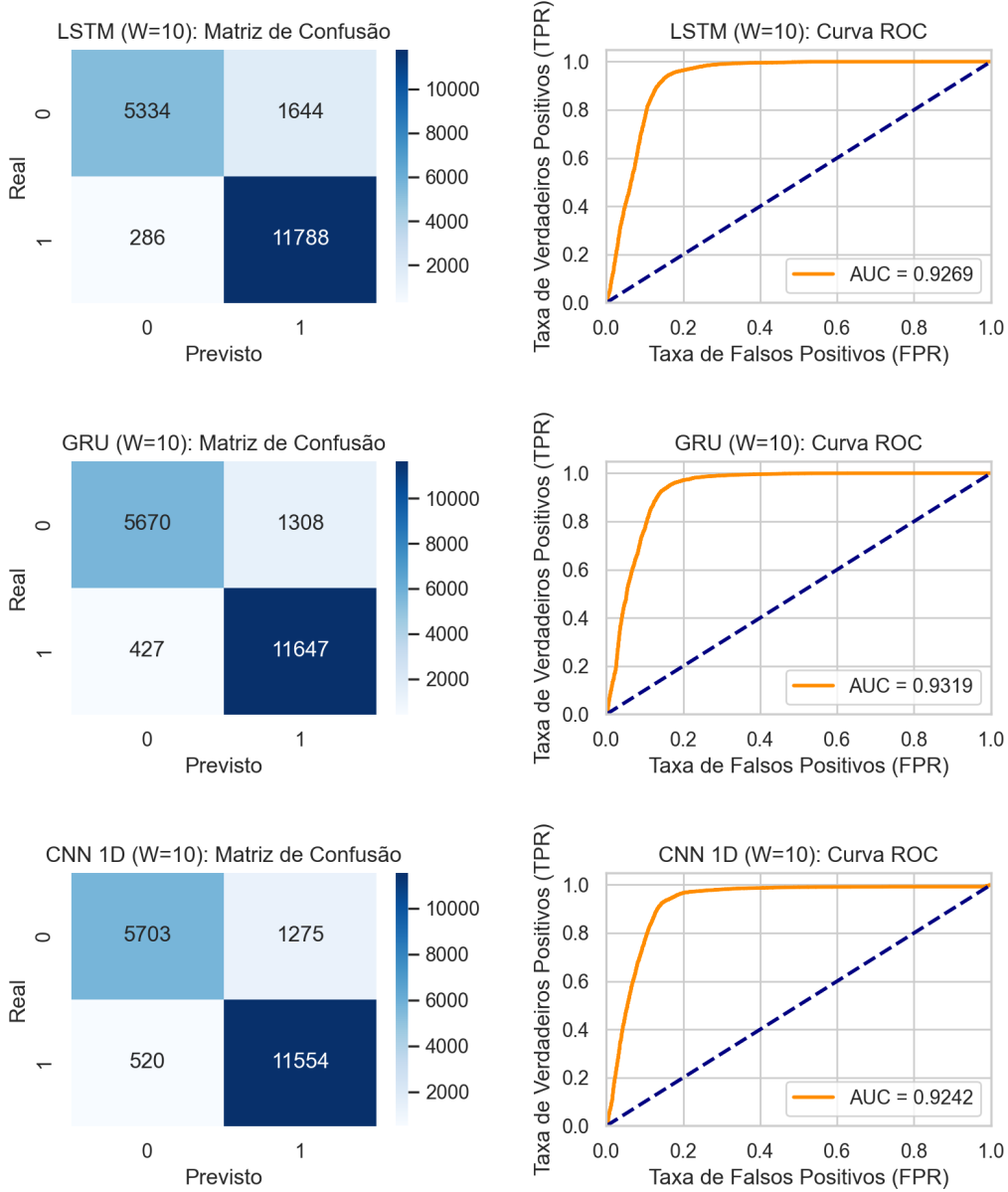


Figura 4: Desempenho das Arquiteturas Neurais Profundas (LSTM, GRU e CNN 1D) no cenário curto ($W = 10$). A topologia CNN 1D apresenta adequação promissora para o uso TinyML.

4.3 Resultados em Médio Prazo ($W = 100$)

Ampliar a janela de predição para 10 segundos ($W = 100$) torna o problema consideravelmente mais difícil. Nesse intervalo, um veículo a 50 km/h percorre cerca de 140 metros, distância suficiente para realizar conversões, sair da cobertura de uma antena ou entrar na

área de sombra de um edifício. Essa variação espacial enfraquece a correlação entre o estado atual do rádio e o estado futuro da conexão, exigindo mais dos algoritmos.

A Tabela 4 e as Figuras 5, 6 e 7 mostram os resultados. A *Árvore de Decisão* é o modelo mais afetado: o F1-Score cai para 0.719 e a AUC para 0.593, evidenciando que suas regras fixas de particionamento não conseguem generalizar bem para eventos mais distantes no tempo. O *Random Forest* se mantém melhor (F1 de 0.809), pois a união das múltiplas árvores tende a absorver parte dos erros individuais.

Chama a atenção o fato de que as redes recorrentes (LSTM e GRU), que em teoria foram concebidas justamente para capturar dependências temporais, sofreram queda expressiva de desempenho (F1 $\approx$ 0.75). O motivo é que o contexto temporal mantido por esses modelos é relativamente curto, e projetar 10 segundos no futuro vai além do que o estado interno acumulado consegue representar de forma útil. A CNN 1D, por outro lado, voltou a se destacar com F1-Score de 0.822. Por trabalhar com uma janela de tempo fixa e extrair padrões locais via filtros convolucionais, sem depender de memória acumulada entre amostras, o modelo conseguiu generalizar melhor para o horizonte médio.

Tabela 4: Valores Analíticos Consolidados - Antecipação de 10 segundos ($W = 100$).

Modelo	Tempo de Treino (s)	Tempo Infer. (s)	Tam. (KB)	Acurácia	Precisão	Recall	F1-Score	AUC
Regressão Logística	0.1099	0.0019	0.74	0.6830	0.6801	0.9451	0.7910	0.7212
Árvore de Decisão	0.1767	0.0014	5.84	0.5734	0.6180	0.8587	0.7187	0.5926
Random Forest	0.1116	0.0377	77.82	0.7468	0.7767	0.8435	0.8087	0.7708
LSTM Neural Network	41.9455	0.0266	22.13	0.6656	0.7075	0.8071	0.7540	0.6309
GRU Neural Network	54.4266	0.0550	16.63	0.6450	0.6660	0.8846	0.7599	0.6248
CNN 1D	18.6718	0.0074	2.00	0.7386	0.7251	0.9474	0.8215	0.7444

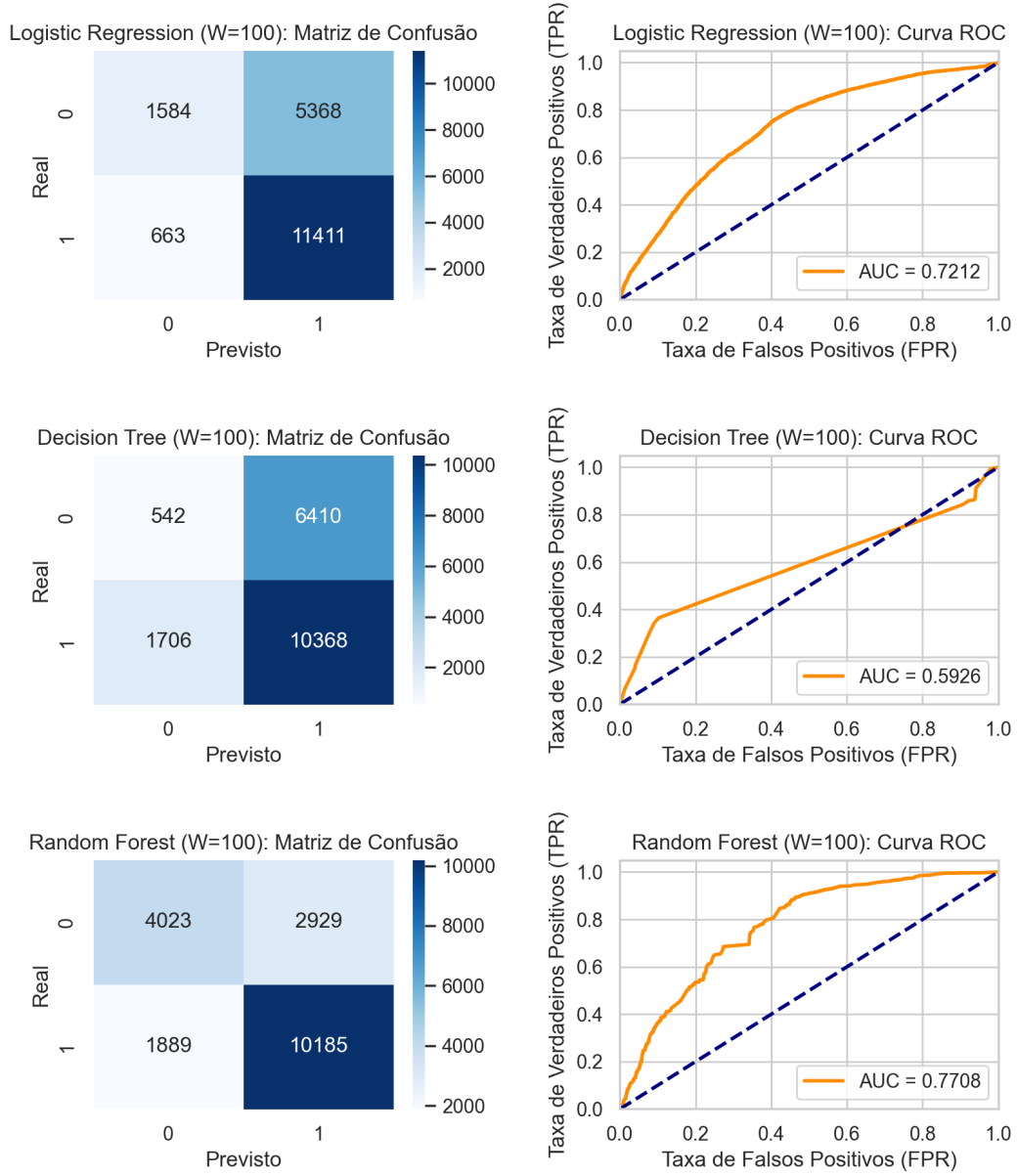


Figura 5: Matrizes de Confusão e Curvas ROC dos modelos TinyML estatísticos em $W = 100$.

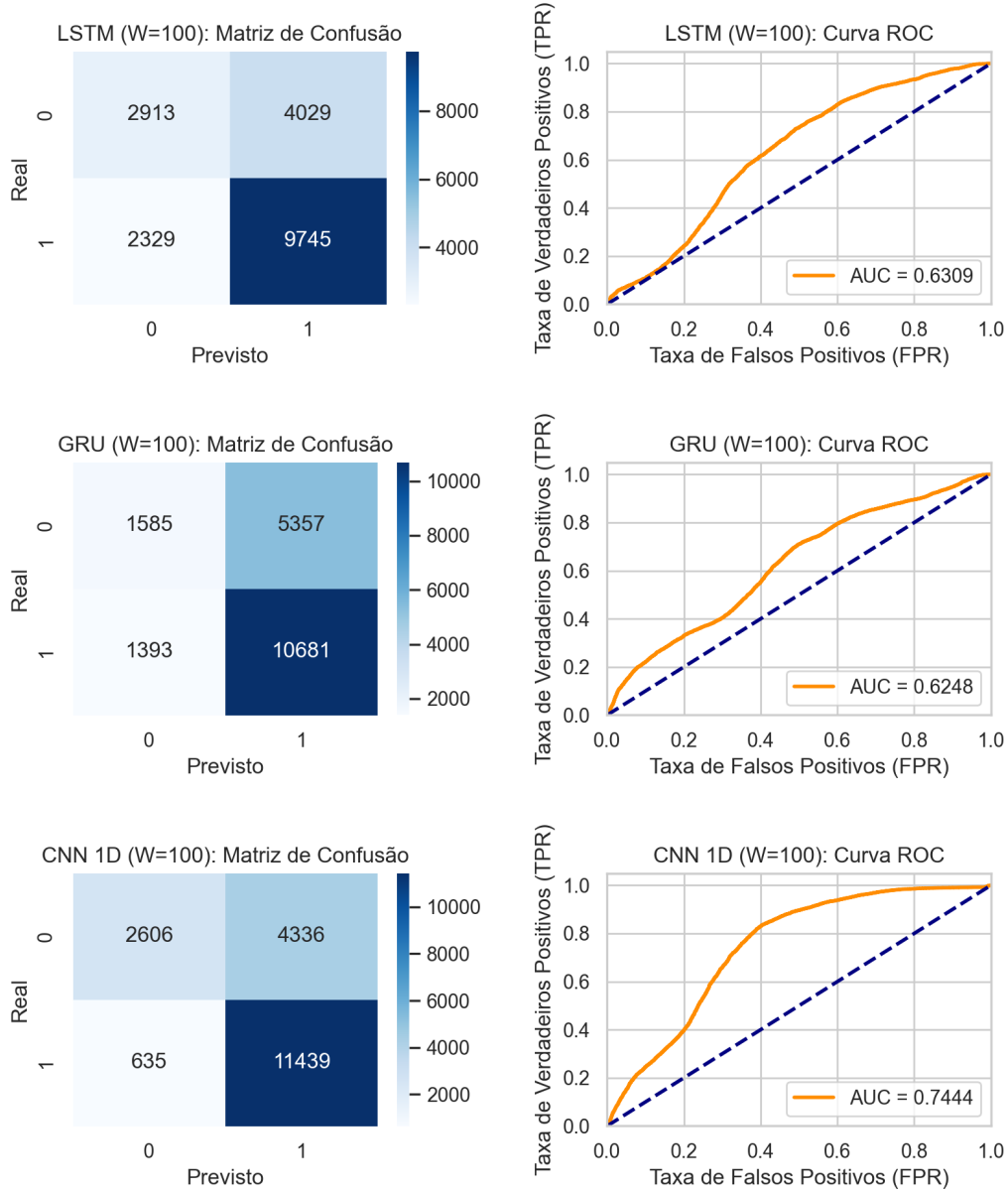


Figura 6: Matrizes de Confusão e Curvas ROC das Arquiteturas Neurais Profundas em $W = 100$.

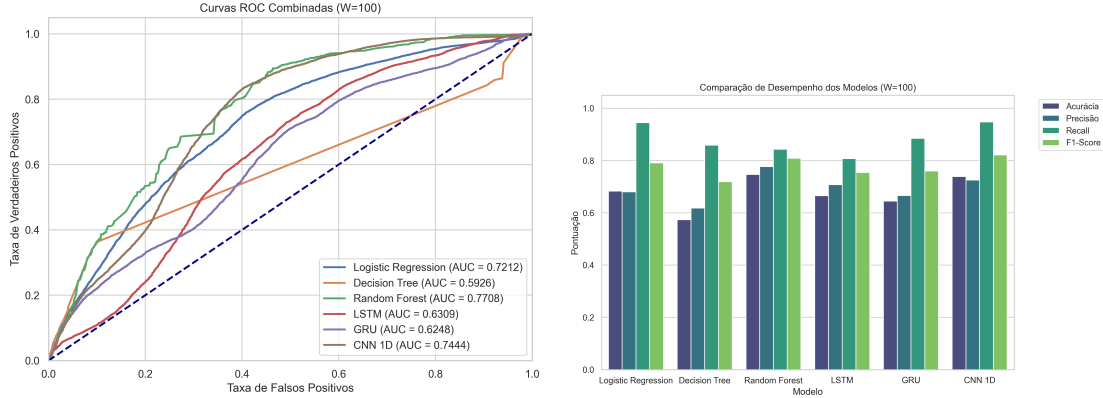


Figura 7: Curvas ROC Combinadas e Distribuição de Desempenho para Janela Média ($W = 100$). A CNN 1D apresenta resiliência superior à degradação probabilística no médio-alcance.

4.4 Resultados em Longo Prazo e Limites Preditivos ($W = 200$)

No cenário mais desafiador ($W = 200$, antecipação de 20 segundos), a relação entre o estado atual do rádio e a condição futura da conexão se torna muito fraca. A 50 km/h, o veículo percorre quase 280 metros nesse intervalo, tempo suficiente para paradas em semáforos, conversões e transições entre zonas de cobertura de diferentes antenas. Com tanta variação, o estado instantâneo do canal oferece pouca informação útil sobre o que acontecerá 20 segundos depois.

Os resultados, apresentados na Tabela 5 e nas Figuras 8, 9 e 10, confirmam essa dificuldade. A GRU foi o modelo mais prejudicado, com F1-Score de 0.661 e AUC de 0.526 — valor muito próximo de uma classificação aleatória. A LSTM também caiu, chegando a F1 de 0.711. O *Random Forest* se mostrou mais resistente (F1 de 0.740), provavelmente porque o conjunto das árvores preserva tendências espaciais mais gerais do trajeto. A CNN 1D, mesmo sofrendo queda de desempenho, manteve a liderança com F1-Score de 0.749 e o maior Recall (0.876), o que indica que ela foi a que menos deixou passar quedas reais sem detectar. No geral, os resultados com $W = 200$ marcam o limite prático da abordagem baseada em *time-shifting* quando se dispõe apenas de telemetria de rádio e GPS.

Tabela 5: Valores Analíticos Consolidados - Antecipação de 20 segundos ($W = 200$).

Modelo	Tempo de Treino (s)	Tempo Infer. (s)	Tam. (KB)	Acurácia	Precisão	Recall	F1-Score	AUC
Regressão Logística	0.0641	0.0017	0.74	0.6084	0.6604	0.7911	0.7199	0.5660
Árvore de Decisão	0.1728	0.0014	6.00	0.6856	0.7613	0.7365	0.7487	0.6523
Random Forest	0.1086	0.0391	78.44	0.6431	0.6891	0.7995	0.7402	0.6622
LSTM Neural Network	41.9666	0.0258	22.13	0.5627	0.6137	0.8437	0.7106	0.4347
GRU Neural Network	54.1676	0.0501	16.63	0.5320	0.6130	0.7171	0.6610	0.5255
CNN 1D	18.6074	0.0078	2.00	0.6263	0.6542	0.8755	0.7488	0.5885

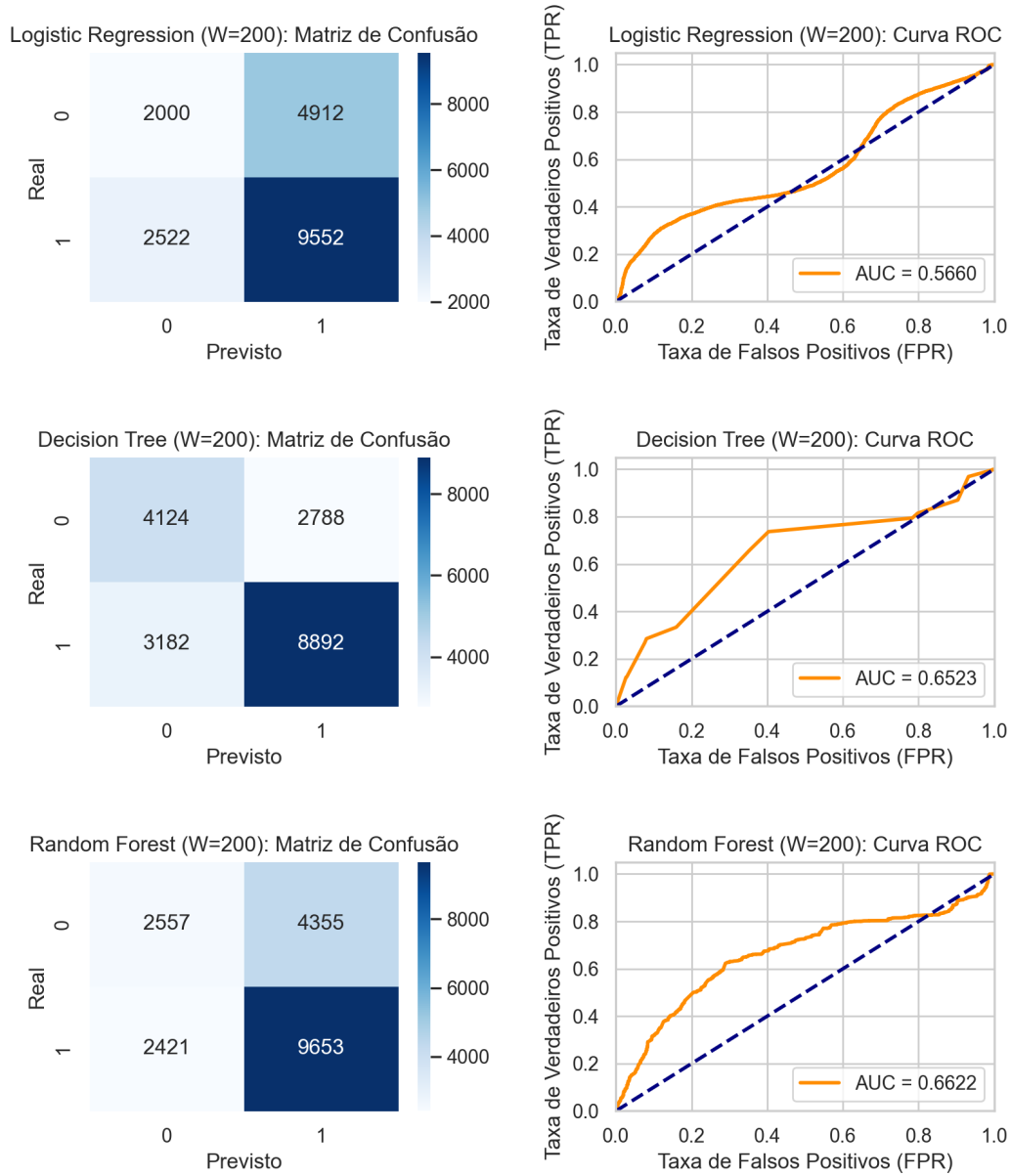


Figura 8: Matrizes de Confusão e Curvas ROC dos modelos TinyML estatísticos em $W = 200$.

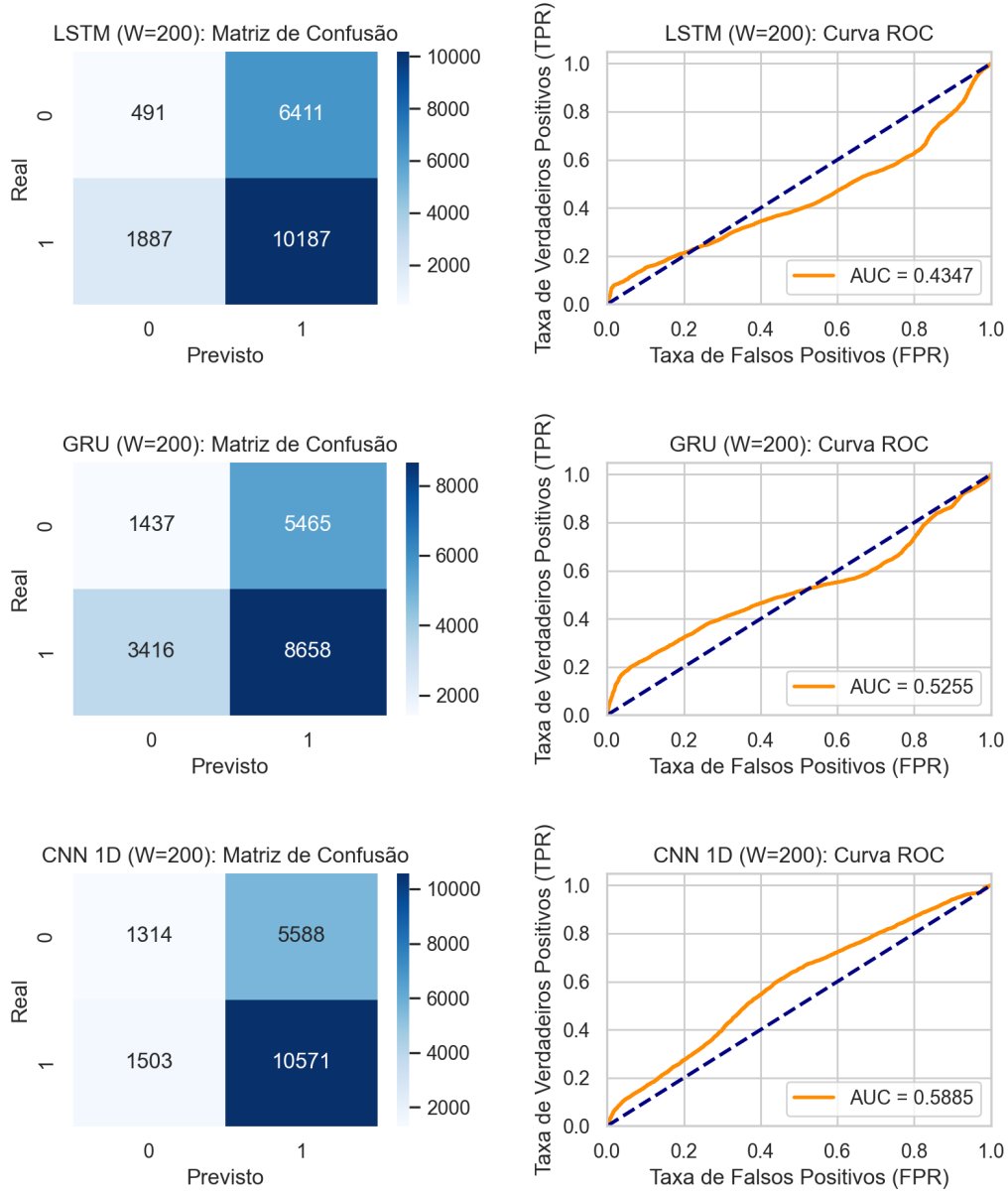


Figura 9: Matrizes de Confusão e Curvas ROC das Arquiteturas Neurais Profundas em $W = 200$.

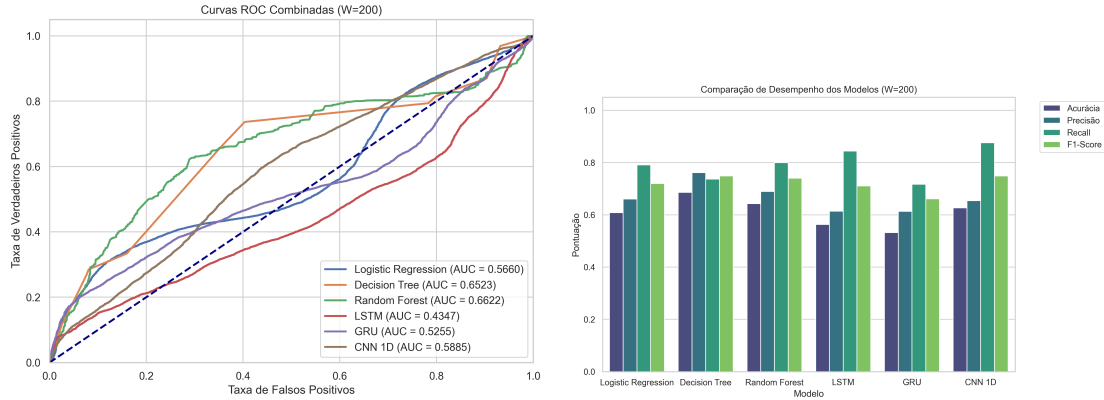


Figura 10: Curvas ROC e Distribuição de Desempenho para Janela Longa ($W = 200$). A degradação do sinal com 20 segundos de antecedência supera a capacidade de generalização de todos os modelos.

A Figura 11 sintetiza visualmente a trajetória de degradação de todos os modelos através dos três horizontes. Dois padrões emergem com clareza: primeiro, a queda é universal — todos os algoritmos perdem desempenho conforme W cresce, confirmando que o problema se torna inerentemente mais difícil. Segundo, a taxa de degradação difere entre os modelos: a CNN 1D exibe a curva mais plana (menor perda relativa de F1-Score), enquanto as RNNs sofrem as quedas mais bruscas. Isso comprova que, para aplicações TinyML em cenários veiculares, a CNN 1D é não apenas a arquitetura mais eficiente em memória, mas também a mais resiliente à expansão do horizonte preditivo.

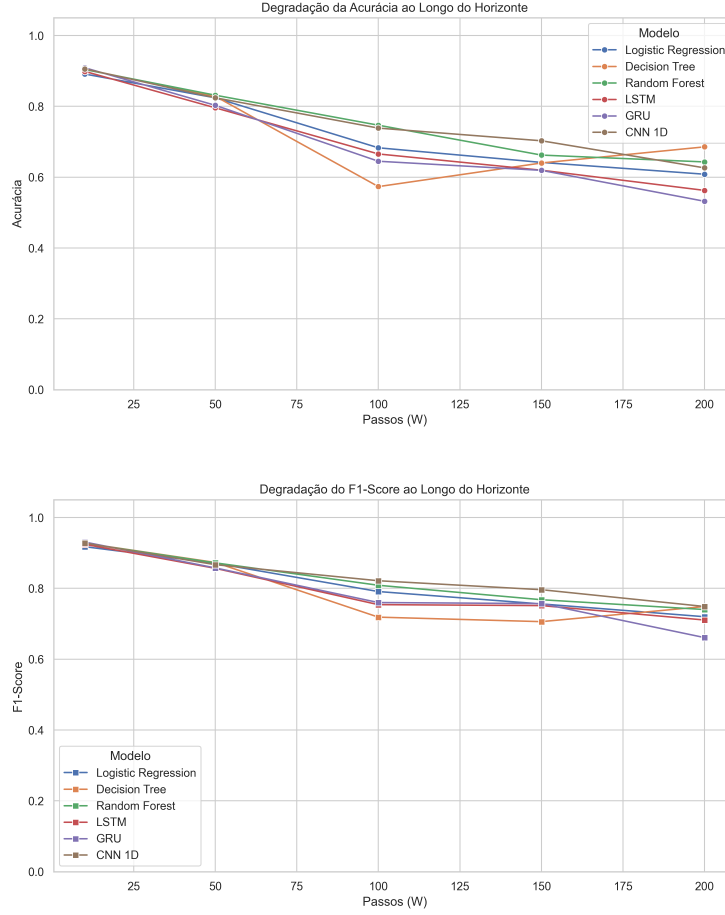


Figura 11: Curvas relativas à degradação da Capacidade Preditiva (Acurácia e F1-Score) acompanhando a variação do distanciamento $W = 10$ a $W = 200$.

4.5 Importância das Variáveis

Para entender melhor o comportamento dos modelos, analisou-se a importância atribuída a cada variável (*Feature Importances*). Essa análise ajuda a verificar se o modelo aprendeu padrões físicos reais de degradação do sinal (como queda de RSSI e aumento de perda de pacotes) ou se apenas memorizou características geográficas específicas do dataset de Aveiro — o que caracterizaria *overfitting* espacial e comprometeria a generalização do modelo para outros cenários urbanos.

Comparando os resultados nos horizontes de $W = 10$ e $W = 100$ (Figuras 12 e 13), observam-se mudanças relevantes:

- **Regressão Logística:** Em $W = 10$, o modelo atribui peso elevado à Latitude, deixando métricas de rede em segundo plano. Ao ampliar para $W = 100$, a distância até a antena (*dist_rsu*) passa a dominar. O problema é que tentar estabelecer uma

relação linear entre posição geográfica e falha de sinal não funciona bem em ambientes urbanos densos, onde a propagação do sinal é altamente irregular.

- **Árvore de Decisão:** Em $W = 10$, o modelo foca em métricas de rede relevantes, como perda de pacotes (*loss*) e *bitrate*. Ao passar para $W = 100$, ele abandona essas variáveis e concentra quase todo o peso na Latitude absoluta. Isso é um sinal claro de *overfitting* geográfico: em vez de aprender o comportamento geral do sinal, o modelo memorizou trechos específicos do percurso de treinamento.
- **Random Forest:** Apresentou o comportamento mais equilibrado. Em $W = 10$, distribui bem o peso entre qualidade de rede e posição. Em $W = 100$, embora também passe a depender mais da Latitude, mantém contribuição relevante de outras variáveis como perda de pacotes e direção do veículo (*heading*). Esse equilíbrio explica sua maior robustez em relação às demais abordagens clássicas.
- **Redes Neurais Profundas:** Como essas arquiteturas não produzem diretamente um ranking de importância de variáveis, não é possível comparar seus pesos da mesma forma. No entanto, o desempenho superior da CNN 1D em $W = 100$ sugere que ela conseguiu extrair padrões mais genéricos da janela de dados, combinando informações de rádio e posição sem depender de uma única variável dominante.

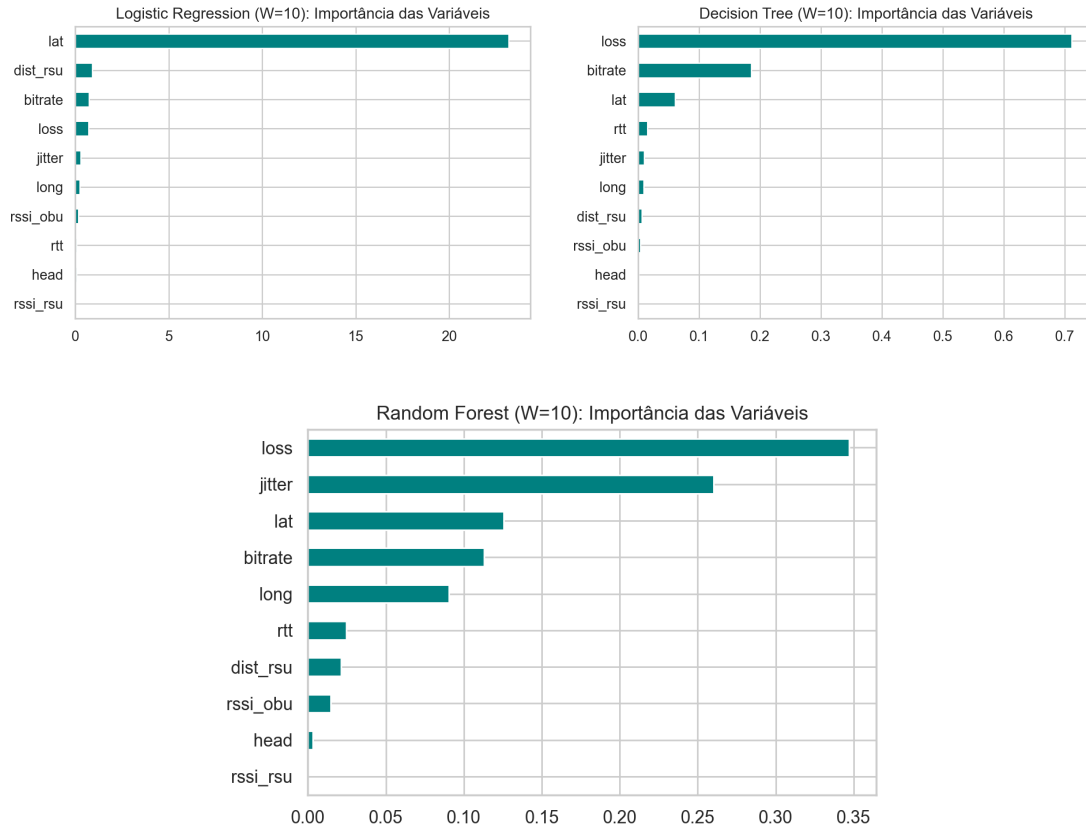


Figura 12: Importância das Variáveis no Horizonte Curto ($W = 10$). Nota-se o foco da Árvore na perda de pacotes (*loss*), enquanto o Random Forest exibe grande equilíbrio preditivo.

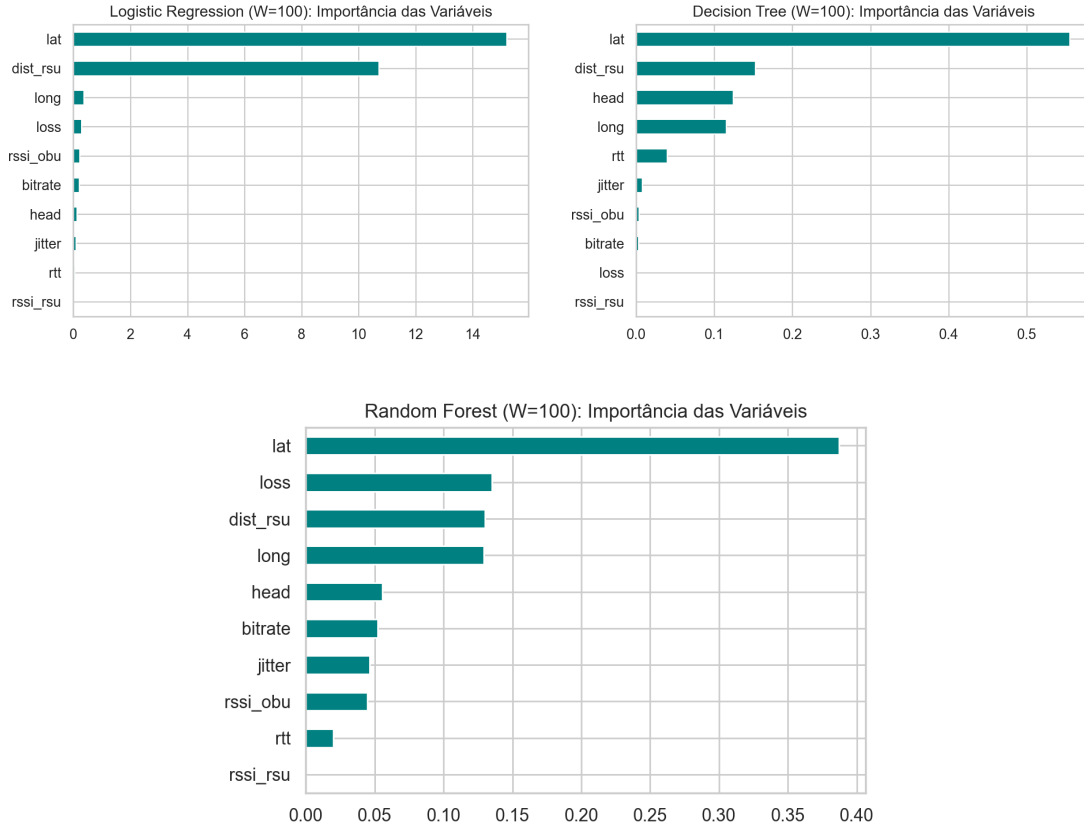


Figura 13: Importância das Variáveis no Horizonte Médio ($W = 100$). Aumentar a janela preditiva força os modelos a priorizarem massivamente as métricas espaciais (Latitude e *dist_rsu*).

5 Conclusão

Este trabalho investigou o uso de *TinyML* para viabilizar *handover* preditivo diretamente na Unidade de Bordo do veículo. A proposta partiu da ideia de antecipar a degradação do enlace com modelos leves, capazes de operar sob restrições reais de memória e processamento. Para isso, foi empregada a técnica de *time-shifting* no treinamento, permitindo que classificadores relativamente simples aprendessem a prever falhas futuras a partir do estado atual da rede.

Os resultados mostraram que a predição é bastante eficaz em horizontes curtos, nos quais as medições instantâneas de rádio ainda mantêm forte relação com a falha iminente. À medida que o horizonte de previsão aumenta para 10 e 20 segundos, o problema se torna mais difícil, pois a mobilidade do veículo, a mudança de trajetória e as variações do ambiente urbano reduzem a correlação entre o estado atual do canal e a condição futura da conexão. Como consequência, todos os modelos sofrem degradação de desempenho, ainda que em

ritmos diferentes.

Entre os métodos clássicos, o *Random Forest* apresentou boa robustez. No entanto, a melhor relação entre desempenho preditivo e custo computacional foi obtida pela CNN 1D. Esse modelo manteve resultados competitivos ao mesmo tempo em que exigiu apenas cerca de 2 KB de memória e poucos milissegundos por inferência. Em contraste, LSTM e GRU alcançaram desempenho semelhante ou inferior em vários cenários, com custo de memória muito maior, o que reduz sua adequação ao contexto embarcado considerado neste estudo.

De forma geral, os resultados indicam que a CNN 1D é a alternativa mais equilibrada para a proposta investigada, sobretudo quando se considera não apenas a qualidade da predição, mas também a viabilidade prática de execução na OBU. Assim, o trabalho reforça que a discussão sobre *handover* inteligente em redes veiculares não deve se limitar à acurácia dos modelos, mas incluir também os limites reais do hardware onde a inferência será executada.

Como continuidade natural, trabalhos futuros podem avançar na implementação nativa do modelo em microcontroladores, na avaliação com outras combinações de tecnologias de acesso e na integração do mecanismo preditivo a políticas reais de decisão de *handover* em veículos conectados.

Declaração de Uso de Inteligência Artificial

Durante a elaboração deste documento, ferramentas de Inteligência Artificial foram empregadas como recurso de apoio. O seu uso se concentrou na otimização de pesquisas e levantamentos bibliográficos, na reformulação de frases para garantir maior clareza e adequação à linguagem acadêmica, e na formatação de elementos estruturais (como a criação de tabelas em LaTeX). Ressalta-se que a concepção da pesquisa, as análises de resultados, o embasamento teórico e as conclusões apresentadas são de autoria e responsabilidade exclusiva do autor.

Referências

- [1] ALJERI, N.; BOUKERCHE, A. A two-tier machine learning-based handover management scheme for intelligent vehicular networks. **Ad Hoc Networks**, [S. l.], v. 94, 101930, 2019. DOI: 10.1016/j.adhoc.2019.101930. Disponível em: <https://doi.org/10.1016/j.adhoc.2019.101930>. Acesso em: 1 jun. 2026.
- [2] BREIMAN, L. Random Forests. **Machine Learning**, [S. l.], v. 45, n. 1, p. 5-32, 2001. DOI: 10.1023/A:1010933404324. Disponível em: <https://doi.org/10.1023/A:1010933404324>. Acesso em: 4 maio 2026.
- [3] CHANG, B.-J.; CHEN, J.-F. Cross-Layer-Based Adaptive Vertical Handoff With Predictive RSS in Heterogeneous Wireless Networks. **IEEE Transactions on Vehicular Technology**, [S. l.], v. 57, n. 6, p. 3679-3692, 2008. DOI: 10.1109/TVT.2008.921619. Disponível em: <https://doi.org/10.1109/TVT.2008.921619>. Acesso em: 1 jun. 2026.

- [4] CHO, K. *et al.* Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In: **CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP)**, 2014, Doha. **Proceedings [...]**. Stroudsburg: Association for Computational Linguistics, 2014. p. 1724-1734. DOI: 10.3115/v1/D14-1179. Disponível em: <https://doi.org/10.3115/v1/D14-1179>. Acesso em: 4 maio 2026.
- [5] DIAS, D.; ROSMANINHO, R.; FIGUEIREDO, A. *et al.* A dataset of ITS-G5 and cellular vehicular connectivity in urban environment. **Data Brief**, [S. l.], v. 52, 109846, 2023. DOI: 10.1016/j.dib.2023.109846. Disponível em: <https://doi.org/10.1016/j.dib.2023.109846>. Acesso em: 4 maio 2026.
- [6] EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE (ETSI). **ETSI EN 302 663**: Intelligent Transport Systems (ITS); Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency band. Sophia Antipolis: ETSI, 2013. Disponível em: https://www.etsi.org/deliver/etsi_en/302600_302699/302663/01_02_01_60/en_302663v010201p.pdf. Acesso em: 9 maio 2026.
- [7] GILLANI, M. *et al.* Data collection protocols for VANETs: a survey. **Complex & Intelligent Systems**, [S. l.], v. 8, p. 2351-2376, 2022. DOI: 10.1007/s40747-021-00629-x. Disponível em: <https://doi.org/10.1007/s40747-021-00629-x>. Acesso em: 4 maio 2026.
- [8] HATİPOĞLU, A. *et al.* Machine Learning-Based Handover Performance Prediction for Beyond 5G Communication Networks. In: **INTERNATIONAL CONFERENCE ON COMMUNICATIONS, SIGNAL PROCESSING, AND THEIR APPLICATIONS (ICCSPA)**, 6., 2024, Sharjah. **Proceedings [...]**. Piscataway: IEEE, 2024. p. 1-6. DOI: 10.1109/ICCSPA61559.2024.10794376. Disponível em: <https://doi.org/10.1109/ICCSPA61559.2024.10794376>. Acesso em: 4 maio 2026.
- [9] HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, [S. l.], v. 9, n. 8, p. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735. Disponível em: <https://doi.org/10.1162/neco.1997.9.8.1735>. Acesso em: 4 maio 2026.
- [10] HOSMER JR., D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied Logistic Regression**. 3. ed. Hoboken: John Wiley & Sons, 2013. DOI: 10.1002/9781118548387. Disponível em: <https://doi.org/10.1002/9781118548387>. Acesso em: 4 maio 2026.
- [11] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS (IEEE). **IEEE Std 802.11p-2010**: IEEE Standard for Information technology—Local and metropolitan area networks—Specific requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments. New York: IEEE, 2010. DOI: 10.1109/IEEESTD.2010.5514475. Disponível em: <https://ieeexplore.ieee.org/document/5514475>. Acesso em: 12 jul. 2026.

- [12] KASSAR, M.; KERVELLA, B.; PUJOLLE, G. An overview of vertical handover decision strategies in heterogeneous wireless networks. **Computer Communications**, [S. l.], v. 31, n. 10, p. 2607-2620, 2008. DOI: 10.1016/j.comcom.2008.01.044. Disponível em: <https://doi.org/10.1016/j.comcom.2008.01.044>. Acesso em: 1 jun. 2026.
- [13] KIM, W.-I.; LEE, B.-J.; SONG, J.-S. *et al.* Ping-Pong Avoidance Algorithm for Vertical Handover in Wireless Overlay Networks. In: **IEEE 66TH VEHICULAR TECHNOLOGY CONFERENCE (VTC-FALL)**, 2007, Baltimore. **Proceedings [...]**. Piscataway: IEEE, 2007. p. 1509-1512. DOI: 10.1109/VETECF.2007.321. Disponível em: <https://doi.org/10.1109/VETECF.2007.321>. Acesso em: 1 jun. 2026.
- [14] KIRANYAZ, S. *et al.* 1D convolutional neural networks and applications: A survey. **Mechanical Systems and Signal Processing**, [S. l.], v. 151, 107398, 2021. DOI: 10.1016/j.ymssp.2020.107398. Disponível em: <https://doi.org/10.1016/j.ymssp.2020.107398>. Acesso em: 4 maio 2026.
- [15] LÓPEZ-PÉREZ, D.; GÜVENÇ, I.; CHU, X. Theoretical analysis of handover failure and ping-pong rates for heterogeneous networks. In: **IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (ICC)**, 2012, Ottawa. **Proceedings [...]**. Piscataway: IEEE, 2012. p. 6774-6779. DOI: 10.1109/ICC.2012.6364722. Disponível em: <https://doi.org/10.1109/ICC.2012.6364722>. Acesso em: 1 jun. 2026.
- [16] LIMA, J. P. S. H. *et al.* Deep Learning-Based Handover Prediction for 5G and Beyond Networks. In: **IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (ICC)**, 2023, Rome. **Proceedings [...]**. Piscataway: IEEE, 2023. p. 3468-3473. DOI: 10.1109/ICC45041.2023.10279195. Disponível em: <https://doi.org/10.1109/ICC45041.2023.10279195>. Acesso em: 4 maio 2026.
- [17] LÊ, M. T.; WOLINSKI, P.; ARBEL, J. Efficient Neural Networks for Tiny Machine Learning: A Comprehensive Review. **ACM Transactions on Intelligent Systems and Technology**, [S. l.], 2026. DOI: 10.1145/3798276. Disponível em: <https://doi.org/10.1145/3798276>. Acesso em: 4 maio 2026.
- [18] PARK, H.-S.; KIM, H.; LEE, C.; LEE, H. Mobility Management Paradigm Shift: From Reactive to Proactive Handover Using AI/ML. **IEEE Network**, [S. l.], v. 38, n. 2, p. 18-25, 2024. DOI: 10.1109/MNET.2024.3357108. Disponível em: <https://doi.org/10.1109/MNET.2024.3357108>. Acesso em: 1 jun. 2026.
- [19] PEIXOTO, F. *et al.* Coexistence of ITS-G5 and 5G NR-U in the 5.9 GHz Band. In: **IEEE VEHICULAR NETWORKING CONFERENCE (VNC)**, 2025. **Proceedings [...]**. Piscataway: IEEE, 2025. p. 1-8. DOI: 10.1109/VNC64509.2025.11054093. Disponível em: <https://doi.org/10.1109/VNC64509.2025.11054093>. Acesso em: 17 jun. 2026.
- [20] QI, K.; LIU, T.; YANG, C. Federated Learning Based Proactive Handover in Millimeter-wave Vehicular Networks. In: **IEEE INTERNATIONAL CONFE-**

- RENCE ON SIGNAL PROCESSING (ICSP)**, 15., 2020, Beijing. **Proceedings** [...]. Piscataway: IEEE, 2020. p. 317-322. DOI: 10.1109/ICSP48669.2020.9320974. Disponível em: <https://doi.org/10.1109/ICSP48669.2020.9320974>. Acesso em: 4 maio 2026.
- [21] QUINLAN, J. R. Induction of decision trees. **Machine Learning**, [S. l.], v. 1, n. 1, p. 81-106, 1986. DOI: 10.1007/BF00116251. Disponível em: <https://doi.org/10.1007/BF00116251>. Acesso em: 4 maio 2026.
- [22] RITO, P. *et al.* Aveiro Tech City Living Lab: A Communication, Sensing, and Computing Platform for City Environments. **IEEE Internet of Things Journal**, [S. l.], v. 10, n. 18, p. 15894-15911, 2023. DOI: 10.1109/JIOT.2023.3262627. Disponível em: <https://doi.org/10.1109/JIOT.2023.3262627>. Acesso em: 4 maio 2026.
- [23] SOMVANSI, S. *et al.* From Tiny Machine Learning to Tiny Deep Learning: A Survey. **ACM Computing Surveys**, [S. l.], v. 58, n. 7, p. 1-33, 2025. DOI: 10.1145/3776588. Disponível em: <https://doi.org/10.1145/3776588>. Acesso em: 4 maio 2026.
- [24] WARDEN, P.; SITUNAYAKE, D. **TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers**. Sebastopol: O'Reilly Media, 2019.
- [25] YAN, L.; DING, H.; ZHANG, L. *et al.* Machine Learning-Based Handovers for Sub-6 GHz and mmWave Integrated Vehicular Networks. **IEEE Transactions on Wireless Communications**, [S. l.], v. 18, n. 10, p. 4873-4885, 2019. DOI: 10.1109/TWC.2019.2930193. Disponível em: <https://doi.org/10.1109/TWC.2019.2930193>. Acesso em: 1 jun. 2026.
- [26] YOO, S. J.; CYPHER, D. E.; GOLMIE, N. T. Predictive link trigger mechanism for seamless handovers in heterogeneous wireless networks. **Wireless Communications and Mobile Computing**, [S. l.], v. 9, n. 5, p. 685-703, 2009. DOI: 10.1002/wcm.620. Disponível em: <https://doi.org/10.1002/wcm.620>. Acesso em: 1 jun. 2026.
- [27] ZHANG, Z.; HOU, Y.; SUN, C. Deep Learning Based Handover for High-Speed Connected Vehicles in Ultra-Dense Networks. In: **IEEE VEHICULAR NETWORKING CONFERENCE (VNC)**, 2024, Kobe. **Proceedings** [...]. Piscataway: IEEE, 2024. p. 1-8. DOI: 10.1109/VNC61989.2024.10575996. Disponível em: <https://doi.org/10.1109/VNC61989.2024.10575996>. Acesso em: 4 maio 2026.