

José Raphael Danillo dos Santos

**Soluções baseadas em GMDH e Árvore de Decisão
para o Problema de CQI Desatualizado em Redes
5G**

RECIFE

2023

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
CURSO DE BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

José Raphael Danillo dos Santos

Soluções baseadas em GMDH e Árvore de Decisão
para o Problema de CQI Desatualizado em Redes
5G

Monografia apresentada ao Centro de Informática (CIN) da Universidade Federal de Pernambuco (UFPE), como requisito parcial para conclusão do Curso de Engenharia da Computação, orientado pelo professor Andson Marreiros Balieiro.

RECIFE

2023

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Santos, José Raphael Danillo dos.

Soluções baseadas em GMDH e árvore de decisão para o problema de CQI
desatualizado em redes 5G / José Raphael Danillo dos Santos. - Recife, 2023.
78 p. : il.

Orientador(a): Andson Marreiros Balieiro

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de
Pernambuco, Centro de Informática, Engenharia da Computação - Bacharelado,
2023.

Inclui referências, apêndices.

1. redes 5G. 2. machine learning. 3. CQI. 4. GMDH. 5. árvore de decisão.
I. Marreiros Balieiro, Andson. (Orientação). II. Título.

000 CDD (22.ed.)

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
CURSO DE BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

José Raphael Danillo dos Santos

Soluções baseadas em GMDH e Árvore de Decisão
para o Problema de CQI Desatualizado em Redes
5G

Banca Examinadora:

Professor Andson Marreiros Balieiro
Orientador

Professor Jamilson Ramalho Dantas
Examinador

Dedico este trabalho aos meus pais, familiares
e amigos que sempre me apoiaram nos
momentos mais críticos.

EPÍGRAFE

“A persistência é o caminho do êxito.”
Charles Chaplin

RESUMO

As Redes 5G proporcionam melhor qualidade de conexão e transferência de dados se comparado à geração anterior, o 4G, oferecendo consumo energético de forma eficiente e com maior área de cobertura em regiões desfavoráveis. Mas para manter essa qualidade, a estação base 5G utiliza da Modulação e Codificação Adaptativa (AMC) para definir os esquemas de codificação e modulação a serem empregados na transmissão do canal, sendo que esses esquemas são mapeados pelo Indicador de Qualidade do Canal (CQI), o qual retrata a condição do canal de *downlink* para ajudar na tomada de decisão da estação base sobre qual esquema aplicar. O problema do CQI desatualizado em redes 5G ocorre quando o valor reportado pelo dispositivo do usuário já não expressa mais a condição do canal no momento em que a estação base o utiliza para a tomada de decisão e isso pode causar degradação na comunicação com o usuário. Este trabalho faz a avaliação de desempenho da comunicação do usuário, em termos de métricas como *throughput*, quando GMDH e Árvore de Decisão são empregadas para resolver este problema, estimando a qualidade do canal atualizada e comparando os respectivos resultados desses modelos. Esses resultados indicam uma taxa de acerto do CQI de 55.9% para GMDH e 54% para Árvore de Decisão, sendo o primeiro modelo a apresentar menor frequência de erro e desvio médio do CQI em relação ao segundo.

Palavras-chave: CQI; redes 5G; machine learning; GMDH; árvore de decisão

ABSTRACT

5G Networks provide better quality of connection and data transfer compared to the previous generation, 4G, providing efficient energy consumption and a greater coverage area in unfavorable regions. But to maintain this quality, the 5G base station uses Adaptive Modulation and Coding (AMC) to define the commitment and modulation schemes to be used in channel transmission, and these schemes are mapped by the Channel Quality Indicator (CQI), which depicts the condition of the downlink channel to help the base station make a decision based on which scheme to apply. The outdated Channel Quality Indicator (CQI) problem in 5G networks occurs when the value reported by the user device no longer expresses the condition of the channel at the time the base station uses it for decision making, and this can cause failure to communicate with the user. This work evaluates the performance of user communication, in terms of metrics such as throughput, when GMDH and Decision Tree are used to solve this problem, estimating the quality of the updated channel and comparing the respective results of these models. These results indicate a CQI success rate of 55.9% for GMDH and 54% for Decision Tree, being the first model to present a lower error frequency and CQI average deviation in relation to the second.

Keywords: CQI; 5G network; machine learning; GMDH; decision tree

SUMÁRIO

1. Introdução	14
1.1. Objetivos	15
1.2. Organização do Trabalho	15
2. Referencial Teórico	16
2.1. Aspectos gerais e tipos de serviços das Redes 5G	16
2.1.1. Implantação de 5G no Brasil e no mundo	17
2.1.2. Partes da Rede 5G	20
2.2. Rede de Acesso de Rádio (RAN)	23
2.3. Inteligência Artificial e Regressão	39
2.3.1. Árvore de Decisão	40
2.3.2. GMDH	41
3. Trabalhos Relacionados (Estado da Arte)	44
4. Metodologia	47
4.1. Geração do dataset para treinamento dos modelos	47
4.2. Treinamento dos modelos	52
4.3. Cálculo das métricas comparativas	55
5. Resultados e Análise	57
5.1. Análise para seleção das melhores configurações dos modelos treinados	57
5.2. Análise comparativa dos modelos baseado nas métricas calculadas	61
6. Conclusões e Trabalhos Futuros	68
6.1. Contribuições	68
6.2. Trabalhos Futuros	68
7. Referências	70
Apêndice	76

LISTA DE FIGURAS

Figura 1. Cenários e aplicações do 5G	17
Figura 2. Mercado 5G em Junho 2022	18
Figura 3. Acessos móveis 5G no Brasil (milhares)	19
Figura 4. Arquitetura do núcleo 5G	20
Figura 5. Representação em ponto da arquitetura 5G do núcleo	21
Figura 6. Representação em serviço da arquitetura 5G do núcleo	21
Figura 7. Estrutura de rede para <i>Backhaul</i> 5G	22
Figura 8. Espaçamento das subportadoras de acordo com a numerologia	24
Figura 9. Comprimento do <i>slot</i> de acordo com a numerologia	25
Figura 10. Estrutura do <i>frame</i> para $\mu = 0$ com <i>CP</i> normal	26
Figura 11. Estrutura do <i>frame</i> para $\mu = 1$ com <i>CP</i> normal	26
Figura 12. Estrutura do <i>frame</i> para $\mu = 2$ com <i>CP</i> estendido	27
Figura 13. Constelações para QPSK, 16QAM, 64QAM e 256QAM	35
Figura 14. Fluxo de definição do <i>TBS</i>	37
Figura 15. Representação de uma Árvore de Decisão	40
Figura 16. Representação do <i>GMDH</i> Multicamada	42
Figura 17. Nó do <i>GMDH</i>	42
Figura 18. Estrutura do <i>dataset</i> final	48
Figura 19. Teste de correlação com método Pearson	50
Figura 20. Teste de correlação com método Spearman	50
Figura 21. Teste de correlação com método Kendall	51
Figura 22. Árvore da configuração 42	58
Figura 23. Rede <i>GMDH</i> da configuração 114	60

LISTA DE QUADROS

Quadro 1. Intervalos de frequência das faixas espectrais do 5G NR	27
Quadro 2. Primeiro mapeamento do <i>CQI</i> para <i>BLER</i> $\leq 10\%$	28
Quadro 3. Segundo mapeamento do <i>CQI</i> para <i>BLER</i> $\leq 10\%$	29
Quadro 4. Mapeamento do <i>CQI</i> para <i>BLER</i> $\leq 0,001\%$	29
Quadro 5. Primeiro mapeamento do <i>MCS</i> para <i>PDSCH</i>	30
Quadro 6. Segundo mapeamento do <i>MCS</i> para <i>PDSCH</i>	32
Quadro 7. Terceiro mapeamento do <i>MCS</i> para <i>PDSCH</i>	33
Quadro 8. Número de bits por símbolo de cada modulação	35
Quadro 9. Mapeamento do <i>TBS</i> quando <i>Ninfo</i> ≤ 3824	37
Quadro 10. Valores para <i>OverHead</i> em FR1 e FR2	39
Quadro 11. Trabalhos relacionados sobre o problema de <i>feedback</i> desatualizado do <i>CQI</i>	45
Quadro 12. Configurações dos cenários de simulação	47
Quadro 13. Filtragem de colunas do <i>dataset</i> com correlação ≥ 0.5	51
Quadro 14. Pseudocódigo com etapas de geração do <i>dataset</i>	52
Quadro 15. Critérios de definição dos parâmetros fixos, variados e seus valores	53
Quadro 16. Configurações dos modelos de treinamento	53
Quadro 17. Pseudocódigo com etapas do treinamento dos modelos	54
Quadro 18. Definição dos parâmetros para mapeamento do <i>TBS</i> e cálculo da vazão	55
Quadro 19. Pseudocódigo com etapas do cálculo das métricas	56
Quadro 20. Valores da configuração 42 da Árvore de Decisão	58
Quadro 21. Valores da configuração 114 do <i>GMDH</i>	61

LISTA DE GRÁFICOS

Gráfico 1. Comparação dos valores do MSE para Árvore de Decisão	58
Gráfico 2. $SINR$ (dB) normalizado predito versus real para Árvore de Decisão	58
Gráfico 3. Comparação dos valores do MSE para $GMDH$	59
Gráfico 4. $SINR$ (dB) normalizado predito versus real para $GMDH$	60
Gráfico 5. Erro das camadas da configuração 114	61
Gráfico 6. Comparação dos valores do CQI para Árvore de Decisão	62
Gráfico 7. Comparação dos valores do CQI para $GMDH$	62
Gráfico 8. Taxa de acerto do CQI dos modelos	62
Gráfico 9. Distância média de erro do CQI dos modelos	63
Gráfico 10. Desvio padrão médio do CQI dos modelos	63
Gráfico 11. MSE do CQI dos modelos	63
Gráfico 12. $RMSE$ do CQI dos modelos	64
Gráfico 13. Frequência de erros do CQI	64
Gráfico 14. Comparação dos valores de erro do CQI	64
Gráfico 15. Média total da vazão dos modelos	65
Gráfico 16. Média total da eficiência espectral dos modelos	66
Gráfico 17. Desvio padrão total da vazão dos modelos	66
Gráfico 18. Desvio padrão total da eficiência espectral dos modelos	67

LISTA DE SIGLAS

Sigla	Significado
4G	Quarta Geração
5G	Quinta Geração
AMC	Modulação e Codificação Adaptativa
AP	Ponto de Acesso
BBU	Unidade Banda Base
BLER	Taxa de Erro de Bloco
BS	Estação Base
CP	Prefixo Cíclico
CQI	Indicador de Qualidade do Canal
CRAN	Rede de Acesso de Rádio em Nuvem
DCT	Transformada Discreta de Cosseno
DFT-S-OFDMA	Discrete Fourier Transform Spread OFDMA
DL	Downlink
DMRS	Sinal de Referência de Demodulação
DRNN	Rede Neural Recorrente Profunda
DSS	Compartilhamento Dinâmico de Espectro
eMBB	Banda Larga Móvel Aprimorada
FR1	Intervalo de Frequência 1
FR2	Intervalo de Frequência 2
GMDH	Método de Grupo de Tratamento de Dados
gNodeB	Estação Base 5G
GRU	Unidade Recorrente Fechada
GSA	Associação Global de Fornecedores Móveis
IIS	Interferência Inter-Simbólica
IoT	Internet das Coisas
LSA	Aproximação Estocástica Linear
LSTM	Memória a Longo Prazo
LTE	Long Term Evolution
MAPE	Erro Percentual Absoluto Médio
MCS	Esquema de Modulação e Codificação

MLP	Rede Neural Perceptron Multicamadas
mMIMO	Múltiplas Entradas e Múltiplas Saídas Massivo
mMTC	Comunicação Massiva do Tipo Máquina
mmWave	Ondas milimétricas
M-QAM	Modulação de Amplitude em Quadratura de ordem M
MSE	Erro Médio Quadrático
NR	Novo Rádio
NRMSE	Raiz do Erro Médio Quadrático Normalizado
NS	Funções de Rede
NSA	Não Independente
OFDMA	Acesso Múltiplo por Divisão de Frequências Ortogonais
PDSCH	Canal Físico Compartilhado de Downlink
PRB	Recurso de Bloco Físico
QPSK	Chaveamento de Mudança de Fase em Quadratura
R2 Score	Coefficiente de Determinação
RAN	Rede de Acesso de Rádio
RMSE	Raiz do Erro Médio Quadrático
RRU	Unidade de Rádio Remota
SA	Independente
SBA	Arquitetura Baseada em Serviços
SINR	Relação Sinal-Ruído Interferência
SNR	Relação Sinal-Ruído
TBS	Tamanho do Bloco de Transporte
UAV	Veículos Aéreos Não Tripulados
UE	Equipamento do Usuário
UFRJ	Universidade Federal do Rio de Janeiro
UL	Uplink
URLLC	Comunicação de Baixa Latência Ultra Confiável

1. Introdução

A Quinta Geração (5G) de Comunicações Sem Fio visa suportar uma variedade de serviços, incluindo veículos autônomos, streaming de vídeo de ultra-alta definição e a Internet das Coisas (*IoT*), ao mesmo tempo que fornece alta transferência de dados (algo em torno de dezenas de Gbps), ultra-baixa latência na ordem de alguns milissegundos, alta confiabilidade na ordem de 99,99999%, baixo consumo de energia e alta densidade de conexão [24]. No entanto, a reflexão e difração do sinal, a mobilidade do usuário e a interferência de outras fontes representam desafios para a manutenção da alta qualidade do serviço durante as comunicações sem fio.

Para resolver esse problema, as estações base 5G (*gNodeB*) empregam a técnica Modulação e Codificação Adaptativa (*AMC*) para ajustar os esquemas de modulação e codificação (*MCS*) usados para transmissão com base no indicador de qualidade do canal (*CQI*) relatado pelo equipamento do usuário (*UE*). O *CQI* visa refletir a condição do canal de *downlink* e auxiliar o *gNodeB* na tomada de decisões relativas ao *MCS* e aos recursos de rádio para comunicação. Assim, afeta indiretamente a vazão alcançada, a taxa de erro de bloco (*BLER*) e a eficiência espectral. Portanto, relatar com precisão um *CQI* que denote a condição atual do canal é de suma importância para a adaptação do *link* da rede 5G. Entretanto, o tempo decorrido entre a medição da condição do canal e seu uso efetivo pelo *gNodeB*, que envolve tarefas como cálculo e transmissão do *CQI*, pode tornar o *CQI* obsoleto ou envelhecido, não refletindo a qualidade atual do canal [25][28] e possivelmente degradar a comunicação do *UE*, reduzindo a vazão e aumentando o *BLER* por exemplo.

Soluções para o problema de *feedback* desatualizado do *CQI* foram propostas na literatura [25][26][27][28], que são baseadas na previsão de *CQI* ou Relação Sinal-Ruído (*SNR*) por meio de técnicas como extrapolação linear [27], Redes Neurais de Memória a Longo Prazo (*LSTM*) [27][29], ou estimativa linear com aproximação estocástica [25]. No entanto, estas abordagens negligenciam o impacto da mobilidade do usuário como posição, velocidade e direção do movimento, por exemplo, na condição do canal. Eles consideram apenas *CQI* ou *SNR* para estimar o valor futuro, e alguns deles são implementados no lado *gNodeB* [26][28], o que pode sobrecarregá-lo, especialmente em cenários com alta densidade de usuários. Em contraste, este trabalho propõe uma solução baseada no Método de Grupo de Tratamento de Dados (*GMDH*) e na Árvore de Decisão que leva em conta não apenas *CQI* ou *SINR*, mas também a mobilidade do usuário para estimar o *SINR* atualizado para tradução em

um valor do *CQI*. O *GMDH* é um algoritmo auto-organizado que combina os conceitos de modelagem de caixa preta (*black-box*), conexismo e indução para modelagem matemática computacional de sistemas multivariáveis. Ele automaticamente otimiza sua estrutura com intervenção mínima do analista. A Árvore de Decisão é um algoritmo de tomada de decisão composto por nós e ramos, formando um fluxograma com possíveis caminhos a serem percorridos conforme critérios de escolha, iniciando-se do nó-raiz e encerrando em nós-folha, indicando a resposta final. Essa escolha ocorre em nós de decisão e é dependente da condição presente nesses nós, podendo apenas seguir entre um dos dois caminhos, um positivo e um negativo. Dados de um simulador 5G foram usados para avaliar as soluções propostas em termos de acurácia, eficiência espectral e vazão. Os resultados mostram melhor desempenho para o *GMDH* com taxa de acerto do *CQI* de 55,9%, além de ter menor frequência de erro e desvio médio do *CQI* em relação à Árvore de Decisão.

1.1. Objetivos

Este trabalho visa analisar a performance dos modelos Árvore de Decisão e *GMDH* para identificar o mais adequado para a resolução do problema do *CQI* desatualizado, comparando os respectivos resultados gerados por meio de métricas de avaliação e seus gráficos.

1.2. Organização do Trabalho

Este trabalho está estruturado conforme a seguinte descrição dos sete capítulos:

Capítulo 1 → Apresenta aspectos introdutórios a serem detalhados neste trabalho, além dos objetivos e estrutura do mesmo;

Capítulo 2 → Possui os principais conceitos e contextualização da Rede 5G e sua estrutura, como também os modelos de aprendizado aplicados na solução deste trabalho;

Capítulo 3 → Mostra algumas soluções existentes para o problema abordado e as diferenças entre elas com a solução proposta neste trabalho;

Capítulo 4 → Expõe os métodos aplicados para geração do *dataset*, treinamento dos modelos e cálculo das métricas a serem analisadas;

Capítulo 5 → Contém análise da definição das melhores configurações dos modelos treinados e dos seus resultados;

Capítulo 6 → Apresenta os principais pontos abordados neste trabalho, suas contribuições e direcionamento para trabalhos futuros;

Capítulo 7 → Contém as referências utilizadas neste trabalho;

2. Referencial Teórico

Este capítulo apresenta conceitos envolvidos ao longo deste trabalho, de forma a introduzir aspectos sobre as Redes 5G, suas características, serviços, contextualização no Brasil e mundo, além de detalhes sobre Rede de Acesso de Rádio (*RAN*), área do problema abordado neste trabalho, e modelos de *machine learning* empregados na solução.

2.1. Aspectos gerais e tipos de serviços das Redes 5G

A Quinta Geração de redes móveis (redes 5G) tem como aspectos principais a elevada largura de banda, maior capacidade de dispositivos conectados e menor latência, comparado às gerações anteriores, permitindo variação de velocidade útil na faixa de 1 a 10 Gbps, com latência de até 1 ms, porém também traz pontos considerados mais extrínsecos por não atrelarem apenas ao 5G, como alcance geográfico, que depende das condições e recursos para abranger a região de interesse, e eficiência energética, advindo da durabilidade de dispositivos de baixo consumo energético à medida que utilizam a rede constantemente [1].

Para prover tais características, são adotadas tecnologias de suporte as redes 5G tais como [1][2][3]:

- 1) Múltiplas Entradas e Múltiplas Saídas Massiva (*mMIMO*) aplicadas nas antenas para aumentar a capacidade de intercomunicação entre vários equipamentos de usuário (*UE*) com as estações base (*BS*), ocasionando maior vazão de dados e diminuição da latência.
- 2) Automação em conjunto com inteligência artificial para auxiliar no tráfego das informações e qualidade do canal de transmissão.
- 3) Uso de ondas milimétricas (*mmWave*) que permite utilização de faixas com largura de banda e em setores mais altos do espectro, com foco nas “bandas de 28, 38, 60 GHz e nas *E-Bands* (71~76 GHz e 81~86 GHz)”, segundo pesquisas apontadas pela UFRJ.
- 4) Acesso Múltiplo por Divisão de Frequências Ortogonais (*OFDMA*) para melhor aproveitar o uso da largura de banda do canal, permitindo mais pacotes nas subportadoras e mantendo a ortogonalidade dos sinais (sem sobreposição das frequências, evitando que haja interferências entre as mesmas).

Em relação às aplicações e cenários de uso, há três classes principais, conforme mostrado na Figura 1 e na descrição abaixo [2]:

1) Banda Larga Móvel Aprimorada (*eMBB*): Cenário centrado na velocidade da rede devido a maior tráfego de dados, possibilitando melhor usabilidade em reprodução de vídeos em alta resolução e jogos em nuvem por exemplo.

2) Comunicação Massiva do Tipo Máquina (*mMTC*): Cenário focado em um conjunto de dispositivos em maior quantidade, tal que buscam a melhor gestão energética ao passo que tenham custo-benefício. Sistemas que fazem parte da Internet das Coisas são um exemplo desse caso.

3) Comunicação de Baixa Latência Ultra Confiável (*URLLC*): Cenário com serviços mais críticos que necessitam de mais confiabilidade e pouca latência, como por exemplo, sistemas de segurança e carros autônomos.

Figura 1. Cenários e aplicações do 5G

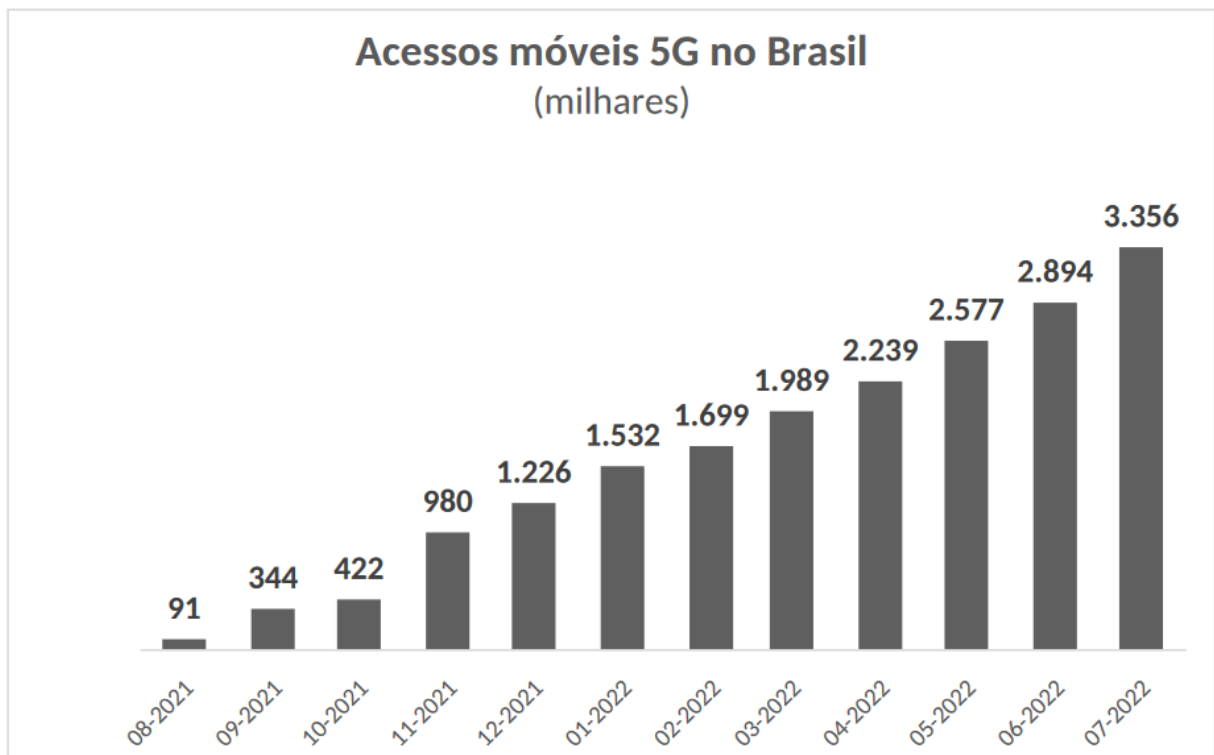


Fonte: Lidiano, Marcelo e Waslon [12].

2.1.1. Implantação de 5G no Brasil e no mundo

O 5G teve seus serviços iniciados na Coreia do Sul em 2019 e conforme dados da *Global Mobile Suppliers Association (GSA)* (apud [4], 2022, p. 8), na primeira metade de 2022 já ocorreu o lançamento dos serviços 5G em 205 operadoras de 80 países, que representa quase o triplo do valor se comparado ao cenário de 2020 com 81 operadoras de 42 países. Atualmente, a implantação do 5G tem maior concentração na Europa, América do

Figura 3. Acessos móveis 5G no Brasil (milhares)



Fonte: Anatel (apud [4], 2022, p. 16).

Algumas cidades que já possuem os serviços do 5G são: Belo Horizonte (BH), Rio de Janeiro (RJ), Porto Alegre (RS), São Paulo (SP), Brasília (DF), Goiânia (GO) e Salvador (BA), ou seja, a implantação iniciou-se pelas capitais com previsão de chegar em todas as cidades do país até final de 2029 [5]. Dentre as implementações disponibilizadas no Brasil, temos [6][7]:

1) 5G SA (*Standalone*): Versão mais pura e sem uso de infraestrutura da geração anterior e com maior probabilidade de troca de chip para usar esse tipo de tecnologia. – *Redes de Núcleo e Acesso 5G*.

2) 5G NSA (*Non Standalone*): Versão “impura” com uso da rede de núcleo da geração anterior e sem necessidade de troca de chip. – *Rede de Núcleo 4G e Rede de Acesso 5G*.

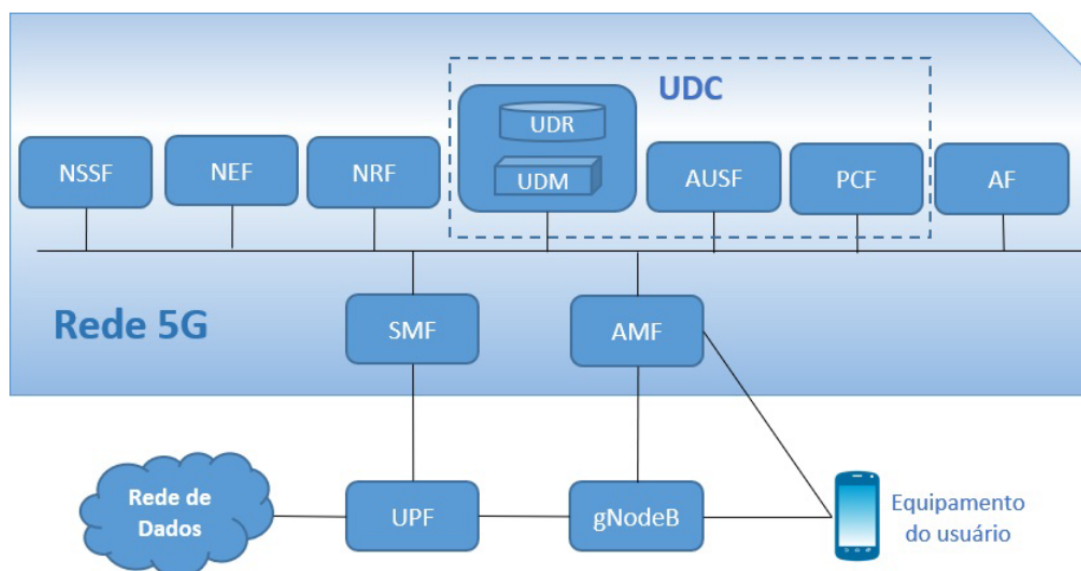
3) 5G DSS (Compartilhamento Dinâmico de Espectro): Versão com uso de toda infraestrutura da geração anterior e possui limite de alcance. – *Voltado a usar faixa espectral existente da rede 4G, com opção de alternar para usar 5G, mas no fundo é um NSA com faixas da rede 4G compartilhadas*.

2.1.2. Partes da Rede 5G

A rede 5G pode ser compreendida a partir das seguintes partes, sendo elas:

1) Rede de Núcleo: O núcleo serve para garantir a segurança e o gerenciamento adequado da rede, permitindo a conexão final até o usuário. Entre outras características, tem sua arquitetura baseada em serviços (*SBA*), utilizando de *edge computing* e *cloud computing* como parte dos componentes da arquitetura, sendo o primeiro buscando trazer mais eficiência no processamento dos dados dos usuários, concentrando na borda da rede, reduzindo latência e o fluxo de tráfego em direção ao núcleo ou rede externa, enquanto o segundo traz operabilidade e acesso aos serviços hospedados na nuvem [10]. Sua arquitetura também é dividida em dois tipos de planos, plano de usuário e plano de controle, para trazer mais flexibilidade e independência entre eles, atentando-se na necessidade de cada serviço [11]. A Figura 4 ilustra funções de rede (*NS*) que formam a arquitetura do núcleo, sendo algumas delas centradas na autenticação, controle de acesso e mobilidade do usuário, outras no gerenciamento de sessão e roteamento dos dados do usuário, por exemplo.

Figura 4. Arquitetura do núcleo 5G

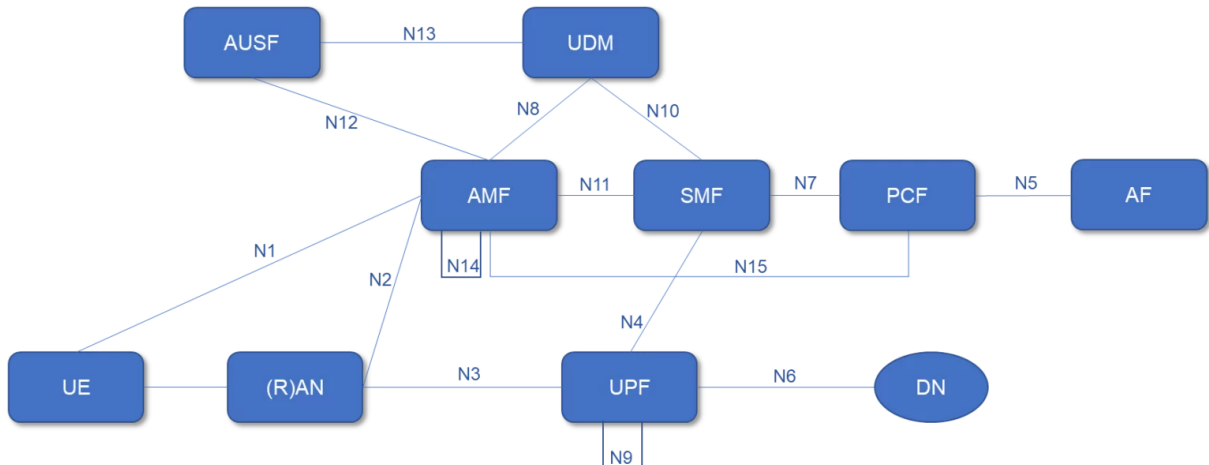


Fonte: Lidiano, Marcelo e Waslon (apud [11], 2020, p. 4).

As funções de rede podem interagir entre si através do consumo e disponibilização dos seus serviços. Além da forma de representação ilustrada na Figura 6, onde há um barramento lógico comum entre as *NSs* de modo que todas as *NS* do plano de controle podem interagir entre si, há também a representação através de pontos de referência, que evidencia quais *NSs*

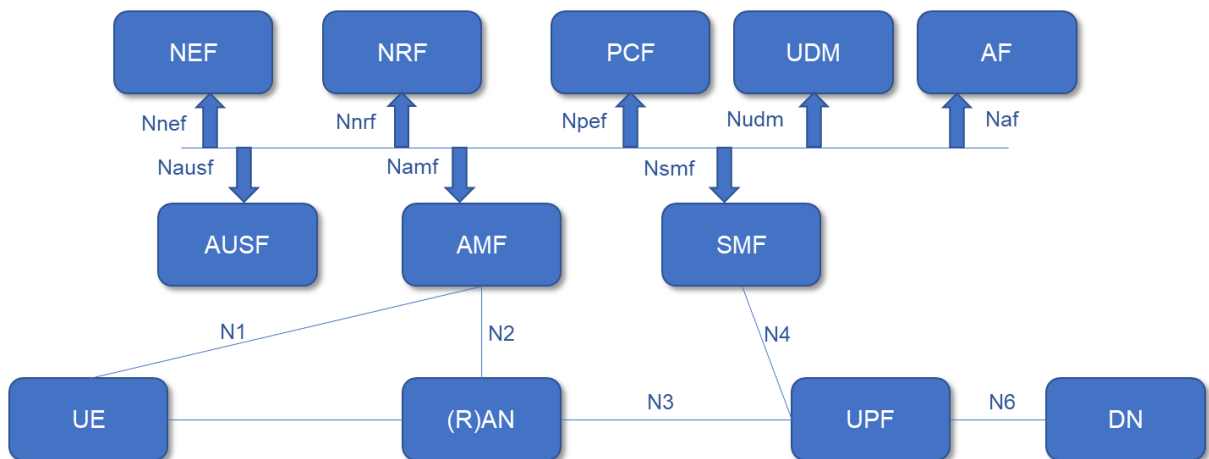
interagem entre si e nomeia a ligação entre elas [11]. As Figuras 5 e 6 apresentam essas duas formas.

Figura 5. Representação em ponto da arquitetura 5G do núcleo



Fonte: 3GPP (apud [11], 2020, p. 4).

Figura 6. Representação em serviço da arquitetura 5G do núcleo

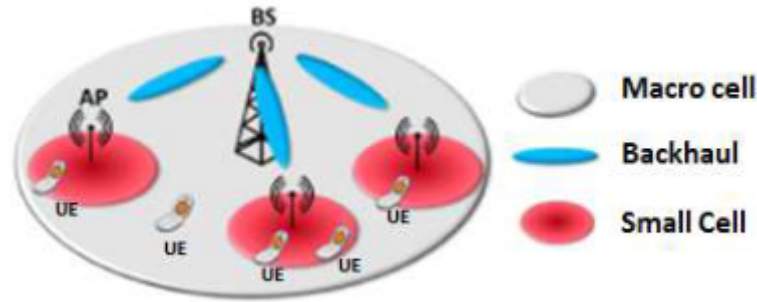


Fonte: 3GPP (apud [11], 2020, p. 4).

2) Rede de *Backhaul*: O *backhaul* serve como uma ponte que une o núcleo às estações base a partir de *links* ou enlaces, podendo ser por meio de um material como cobre e fibra óptica, ou sem fios, como micro-ondas [9]. Uma de suas estruturas para o 5G trata da existência de células (*Macro Cells* e *Small Cells*) para locais de difícil acesso de sinal de rede, os quais são compostos por várias macro células para ampliar e melhorar a qualidade da rede nesses ambientes. No interior da macro célula existe a *BS* que se comunica com o *AP* (Ponto

de Acesso) através do *backhaul* sem fio, e em seguida, o *AP* fornece sinal para os *UEs* na região de cobertura da pequena célula, conforme a Figura 7 [8].

Figura 7. Estrutura de rede para *Backhaul* 5G



Fonte: MiWaveS (2017) [8].

Comumente adotado por fibra óptica, a transmissão sem fio a partir de ondas milimétricas se tornam uma alternativa que mantém a capacidade oferecida pelo 5G, flexibilizando o uso em conjunto ou separado da fibra óptica. Dentre outras técnicas e combinações, têm-se: o uso do *mMIMO* com *mmWave* e a aplicação de múltiplas faixas de espectro, variando entre micro-ondas e *mmWave*.

3) Rede de Acesso: Responsável pelas formas de transmissão final com o *UE*, ou seja, com o último ponto de comunicação. A rede de acesso possui características relacionadas à modulação, codificação e faixas de espectro, conforme apontado por Paulo Florêncio [13].

Dentre as arquiteturas, segundo Vuyo, Oluwaseyi e Vipin [14], existe a *Cloud RAN* (*CRAN*), diferentemente da tradicional Rede de Acesso a Rádio (*RAN*) que possui na mesma torre da *BS* a conexão entre a Unidade de Rádio Remota (*RRU*) com a Unidade Banda Base (*BBU*), a *CRAN* concentra os recursos no sistema de tempo real em nuvem, permitindo a *RRU* se comunicar com a *BBU* de outra *BS*, contribuindo para a existência de menos *BS*'s e consequentemente menor custo e consumo de energia.

Elementos já mencionados, como *mmWave* e *mMIMO*, fazem parte da rede de acesso, no chamado 5G *NR* (*New Radio*), tal que apontado por Editorial Team [15], apresenta flexibilidade na alocação de recursos graças à ortogonalidade das frequências, podendo ser *OFDMA* para *uplink* e *downlink*, enquanto *DFT-S-OFDMA* (*Discrete Fourier Transform Spread OFDMA*) fica como uma opção para *uplink*, porém a *OFDMA* tem mais vantagens pela possibilidade de diversas faixas de espectro. Além disso, outras características

pertencentes ao *OFDMA* são a paralelização dos recursos de diferentes usuários através das subportadoras e impedimento de interferência inter-simbólica (IIS). Entretanto, quando se trata de deslocamento de frequências, traz riscos de interferência entre as subportadoras. Faz parte também nesse cenário, a definição da configuração da *UE* na célula, sendo permanente ou dinâmica, com limite predeterminado de transmissão. O conjunto desses elementos contribuem para maiores transferências de dados e direcionamento da potência para os *UE's*, sendo o aglomerado de portadoras ou *BS's* que auxiliam na ampliação da cobertura de uma região.

2.2. Rede de Acesso de Rádio (RAN)

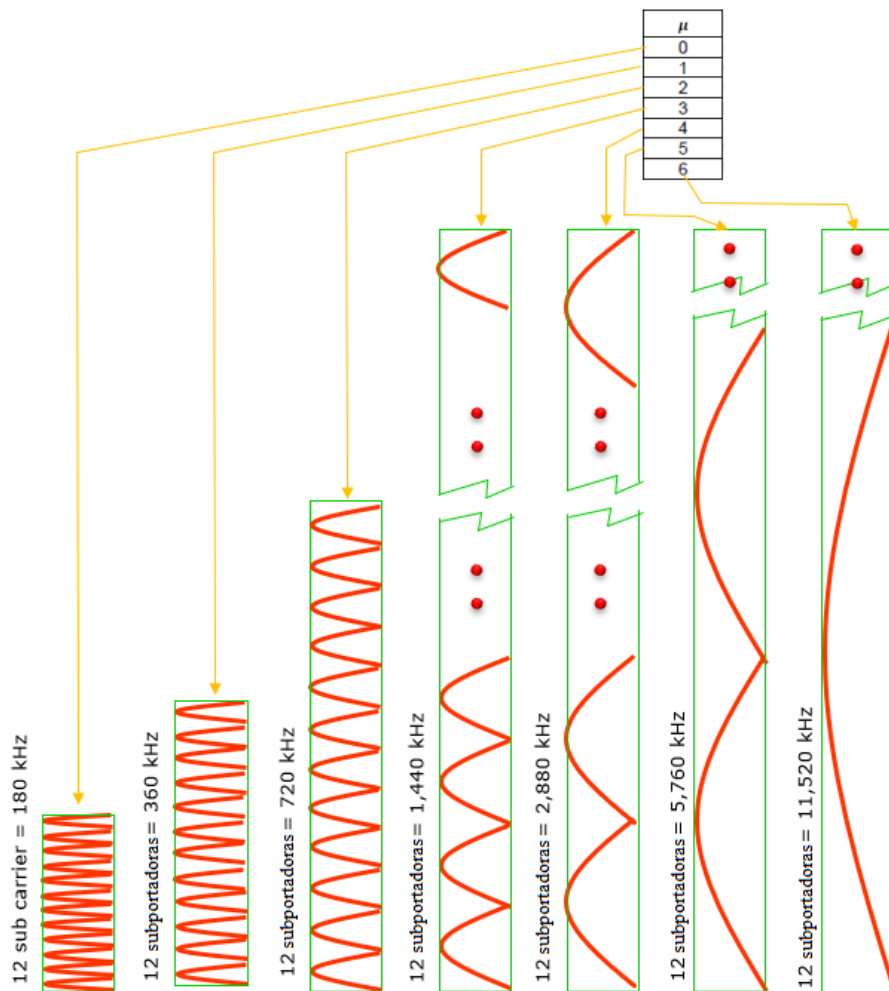
Uma vez que o problema a ser estudado neste trabalho se situa na Rede de Acesso de Rádio, alguns detalhes desse segmento de rede são apresentados nesta seção. Dentre os conceitos mais específicos acerca da *RAN*, tem-se a estrutura do *frame* mais flexível e eficiente de acordo com a aplicação utilizada, novos intervalos de faixas espectrais e o uso de Modulação e Codificação Adaptativa (*AMC*).

Assim como nas redes 4G *LTE*, nas redes 5G *NR*, o *frame* é subdividido em *subframes*, que lida com a transmissão dos dados, flexibilizando a forma de transmitir seja no *uplink (UL)* ou *downlink (DL)* dependendo da aplicação. Cada subframe contém *slots*, que podem ser usados para *UL* e *DL*. O *slot* armazena um determinado número de símbolos e sua duração varia de acordo com o parâmetro de numerologia adotado, que pode ser configurado baseado no tipo de serviço, aplicação a ser suportada. Por exemplo, para sistemas em tempo real os *slots* são menores devido a sua criticidade, enquanto para tráfego massivo de dados os *slots* são maiores, não importando muito o tempo de resposta.

Os símbolos possuem recursos que são alocados no domínio do tempo e da frequência, dividindo-se em blocos, e por conseguinte, subdivididos em elementos de recurso. Por último, tem-se o conceito de numerologia, representado por μ , que refere-se ao tamanho do espaçamento de subportadoras e determina como fica a estrutura do *frame* a partir da duração do *slot* e o comprimento do prefixo cíclico. No 5G *NR*, a numerologia suporta múltiplos espaçamentos de subportadoras, diferentemente do *LTE (Long Term Evolution)* que suporta apenas um tipo de espaçamento, o de 15 kHz. Assim, o índice $\mu = 0$ indica esse espaçamento do *LTE* (garantindo compatibilidade) e à medida que esse índice aumenta, o número de *slot* por *subframe* aumenta, podendo ser representado por um valor de potência com base 2. A

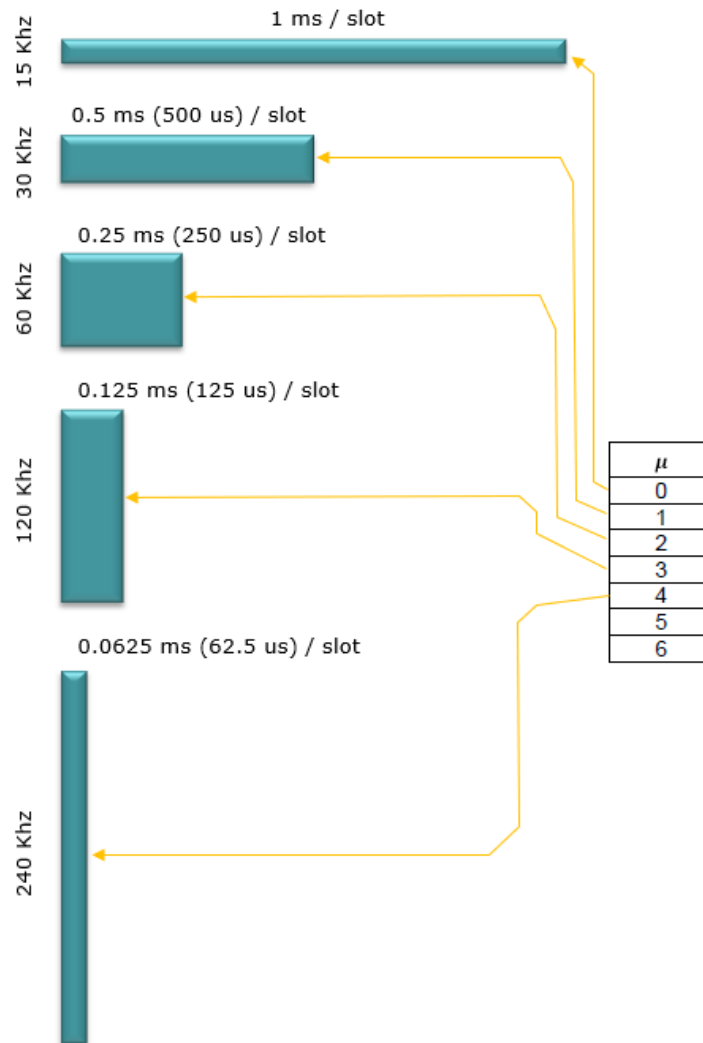
Figura 8 mostra a mudança da largura da subportadora em relação ao índice μ . Por influenciar no espaçamento da subportadora, o comprimento do *slot* consequentemente é afetado, no qual para maiores larguras de subportadoras, o comprimento do *slot* diminui, como pode ser visto na Figura 9 [16].

Figura 8. Espaçamento das subportadoras de acordo com a numerologia



Fonte: Adaptado de [16].

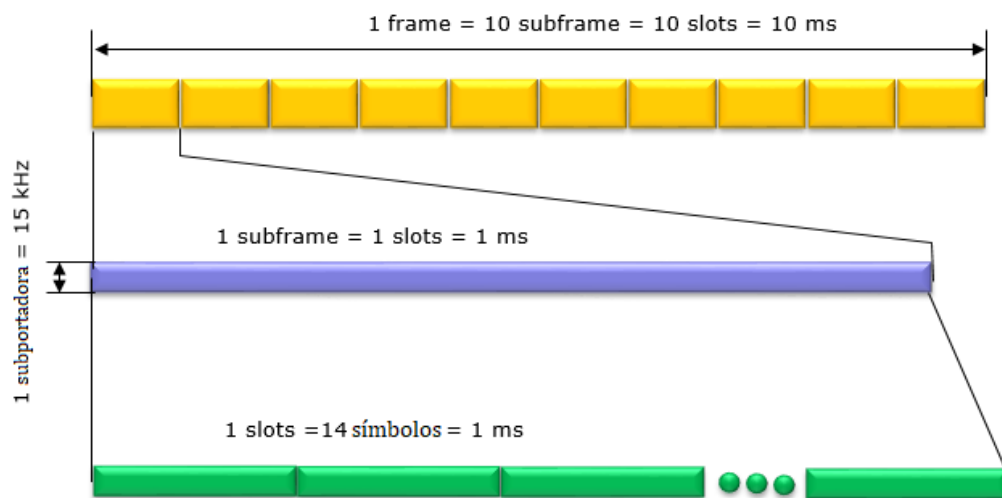
Figura 9. Comprimento do *slot* de acordo com a numerologia



Fonte: Adaptado de [16].

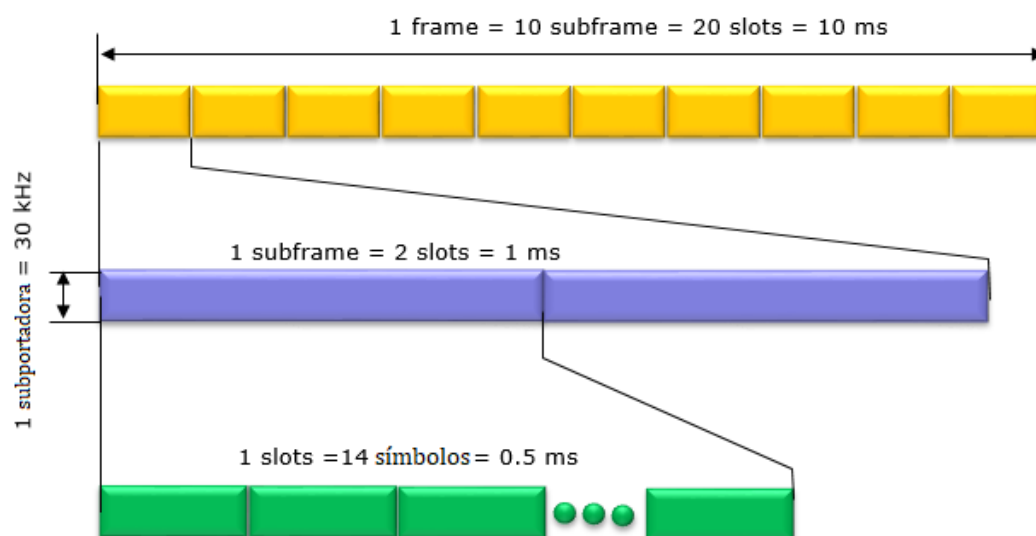
Algumas características do *frame* são predefinidas, como sua duração ser sempre de 10 ms, enquanto a do *subframe* é sempre 1 ms. Apesar disso, a estrutura física é adequada conforme o número de *slots* dentro de um *subframe*. O número de símbolos só varia dependendo do tipo de configuração do *slot*, não interferindo na sua duração. Normalmente, o *slot* apresenta 14 símbolos com o *CP* (Prefixo Cíclico) normal. Já para o caso de *CP* estendido, o número de símbolos passa a ser 12 [16]. As Figuras 10, 11 e 12 exemplificam diferentes estruturas do *frame*.

Figura 10. Estrutura do *frame* para $\mu = 0$ com *CP* normal



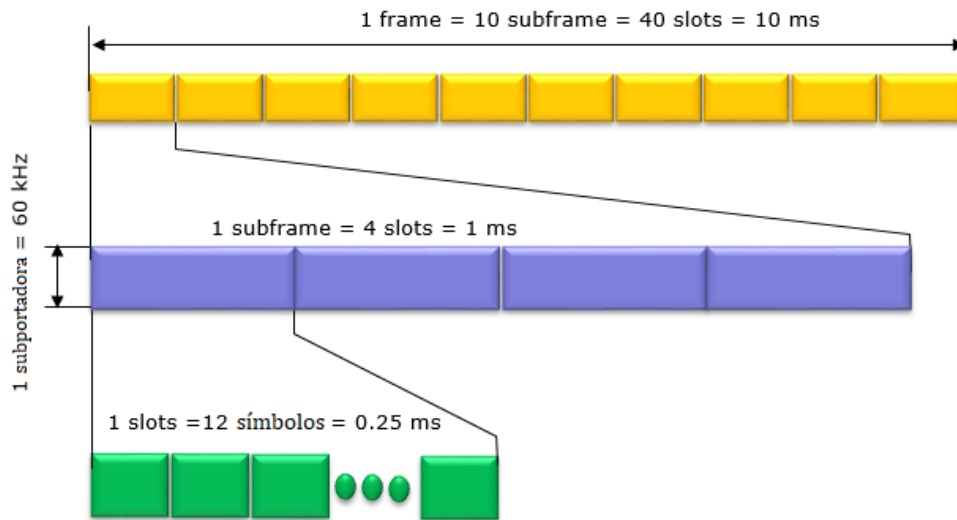
Fonte: Adaptado de [16].

Figura 11. Estrutura do *frame* para $\mu = 1$ com *CP* normal



Fonte: Adaptado de [16].

Figura 12. Estrutura do *frame* para $\mu = 2$ com *CP* estendido



Fonte: Adaptado de [16].

Em relação às faixas espectrais do 5G NR, existe a FR1 (Sub 6 GHz), suportando subportadoras com largura de 15 kHz, 30 kHz e 60 kHz, onde a de 15 kHz e 30 kHz são comumente usadas, e a FR2 (ondas milimétricas), suportando subportadoras com largura de 120 kHz, 240 kHz e 960 kHz, onde a de 120 kHz é comumente usada. A máxima largura de banda para FR1 é 100 MHz e para FR2 é 400 MHz. Há subportadoras que podem ser usadas em apenas uma das faixas e algumas que podem ser usadas em ambas. O intervalo de cada faixa espectral pode ser vista no Quadro 1. Cada uma é representada por um número operador da banda, o qual informa o *range* de *uplink*, *downlink* e o modo de *duplex* [17].

Quadro 1. Intervalos de frequência das faixas espectrais do 5G NR

Designação do intervalo de frequência	Intervalo de frequência correspondente
FR1	450 MHz - 6000 MHz
FR2	24250 MHz - 52600 MHz

Fonte: Adaptado de [17].

Outro elemento presente na rede de acesso é a *AMC*, que busca adaptar a codificação e modulação do sinal de *downlink* baseado no canal percebida pelo *UE* de forma a obter melhor qualidade possível de transmissão (ex. em termos de *throughput* e taxa de erro de blocos), e

consequentemente, trazer melhores resultados de eficiência espectral para o sistema. Para determinar o esquema a ser utilizado, algumas métricas como Relação Sinal-Ruído Interferência (*SINR*) podem ser usadas para tal fim [19]. Logo, para o fluxo interativo entre *BS* e *UE*, tem-se, primeiramente, o envio do sinal de referência da *BS* para o *UE*. Em seguida, o *UE* estima a qualidade do canal baseado nesse sinal recebido. Nesse caso, a partir do *SINR*, por exemplo, a eficiência espectral é calculada e mapeada em um Índice de Qualidade do Canal (*CQI*) de acordo com um intervalo da taxa de erro de bloco (*BLER*) adotada. Os Quadros 2 e 3 apresentam um mapeamento de *CQI* para um *BLER* menor ou igual a 10%, enquanto o Quadro 4 apresenta para um *BLER* menor ou igual a 0,001%. A eficiência espectral corresponde a um quantificador do quão uma banda limitada está sendo eficientemente utilizada, possuindo mais detalhes acerca do seu conceito na p. 35.

Quadro 2. Primeiro mapeamento do *CQI* para *BLER* $\leq$ 10%

Índice do <i>CQI</i>	Eficiência espectral (bit/s/Hz)
0	Fora de alcance
1	0.1523
2	0.2344
3	0.3770
4	0.6016
5	0.8770
6	1.1758
7	1.4766
8	1.9141
9	2.4063
10	2.7305
11	3.3223
12	3.9023
13	4.5234
14	5.1152
15	5.5547

Fonte: Adaptado de [22].

Quadro 3. Segundo mapeamento do *CQI* para *BLER* $\leq 10\%$

Índice do CQI	Eficiência espectral (bit/s/Hz)
0	Fora de alcance
1	0.1523
2	0.3770
3	0.8770
4	1.4766
5	1.9141
6	2.4063
7	2.7305
8	3.3223
9	3.9023
10	4.5234
11	5.1152
12	5.5547
13	6.2266
14	6.9141
15	7.4063

Fonte: Adaptado de [22].

Quadro 4. Mapeamento do *CQI* para *BLER* $\leq 0,001\%$

Índice do CQI	Eficiência espectral (bit/s/Hz)
0	Fora de alcance
1	0.0586
2	0.0977
3	0.1523
4	0.2344

5	0.3770
6	0.6016
7	0.8770
8	1.1758
9	1.4766
10	1.9141
11	2.4063
12	2.7305
13	3.3223
14	3.9023
15	4.5234

Fonte: Adaptado de [22].

Quando a *BS* recebe o *CQI* enviado pelo *UE*, o índice irá definir o Esquema de Modulação e Codificação (*MCS*), o qual influencia na alocação dos recursos de bloco (*PRB*) para a transmissão. Três tabelas de mapeamento do *MCS* podem ser vistas nos Quadros 5, 6 e 7. As modulações suportadas pelo 5G no Canal Físico Compartilhado de *Downlink* (*PDSCH*) são QPSK, 16QAM, 64QAM e 256QAM, cujo número de bits por símbolo dependem da ordem de modulação, informadas no Quadro 8 [21]. Quanto maior a ordem de modulação, mais dados são transmitidos, porém isso acarreta em perdas de potência, possibilitando o aumento do *BLER*. Assim, o uso em canais com muito ruído (interferência) promove esquemas de ordem mais baixa, o que implica em menor vazão dos dados. A Figura 13 exibe as constelações das modulações mencionadas.

Quadro 5. Primeiro mapeamento do *MCS* para *PDSCH*

Índice do MCS I_{MCS}	Ordem de Modulação Q_m	Taxa de Codificação Alvo x [1024] R	Eficiência espectral
0	2	120	0.2344
1	2	157	0.3066

2	2	193	0.3770
3	2	251	0.4902
4	2	308	0.6016
5	2	379	0.7402
6	2	449	0.8770
7	2	526	1.0273
8	2	602	1.1758
9	2	679	1.3262
10	4	340	1.3281
11	4	378	1.4766
12	4	434	1.6953
13	4	490	1.9141
14	4	553	2.1602
15	4	616	2.4063
16	4	658	2.5703
17	6	438	2.5664
18	6	466	2.7305
19	6	517	3.0293
20	6	567	3.3223
21	6	616	3.6094
22	6	666	3.9023
23	6	719	4.2129
24	6	772	4.5234
25	6	822	4.8164

26	6	873	5.1152
27	6	910	5.3320
28	6	948	5.5547
29	2	reservado	
30	4	reservado	
31	6	reservado	

Fonte: Adaptado de [18].

Quadro 6. Segundo mapeamento do MCS para PDSCH

Índice do MCS I_{MCS}	Ordem de Modulação Q_m	Taxa de Codificação Alvo x [1024] R	Eficiência espectral
0	2	120	0.2344
1	2	193	0.377
2	2	308	0.6016
3	2	449	0.877
4	2	602	1.1758
5	4	378	1.4766
6	4	434	1.6953
7	4	490	1.9141
8	4	553	2.1602
9	4	616	2.4063
10	4	658	2.5703
11	6	466	2.7305
12	6	517	3.0293
13	6	567	3.3223
14	6	616	3.6094

15	6	666	3.9023
16	6	719	4.2129
17	6	772	4.5234
18	6	822	4.8164
19	6	873	5.1152
20	8	682.5	5.332
21	8	711	5.5547
22	8	754	5.8906
23	8	797	6.2266
24	8	841	6.5703
25	8	885	6.9141
26	8	916.5	7.1602
27	8	948	7.4063
28	2	reservado	
29	4	reservado	
30	6	reservado	
31	8	reservado	

Fonte: Adaptado de [18].

Quadro 7. Terceiro mapeamento do MCS para PDSCH

Índice do MCS I_{MCS}	Ordem de Modulação Q_m	Taxa de Codificação Alvo x [1024] R	Eficiência espectral
0	2	30	0.0586
1	2	40	0.0781
2	2	50	0.0977
3	2	64	0.1250

4	2	78	0.1523
5	2	99	0.1934
6	2	120	0.2344
7	2	157	0.3066
8	2	193	0.3770
9	2	251	0.4902
10	2	308	0.6016
11	2	379	0.7402
12	2	449	0.8770
13	2	526	1.0273
14	2	602	1.1758
15	4	340	1.3281
16	4	378	1.4766
17	4	434	1.6953
18	4	490	1.9141
19	4	553	2.1602
20	4	616	2.4063
21	6	438	2.5664
22	6	466	2.7305
23	6	517	3.0293
24	6	567	3.3223
25	6	616	3.6094
26	6	666	3.9023
27	6	719	4.2129
28	6	772	4.5234
29	2	reservado	
30	4	reservado	

31	6	reservado
----	---	-----------

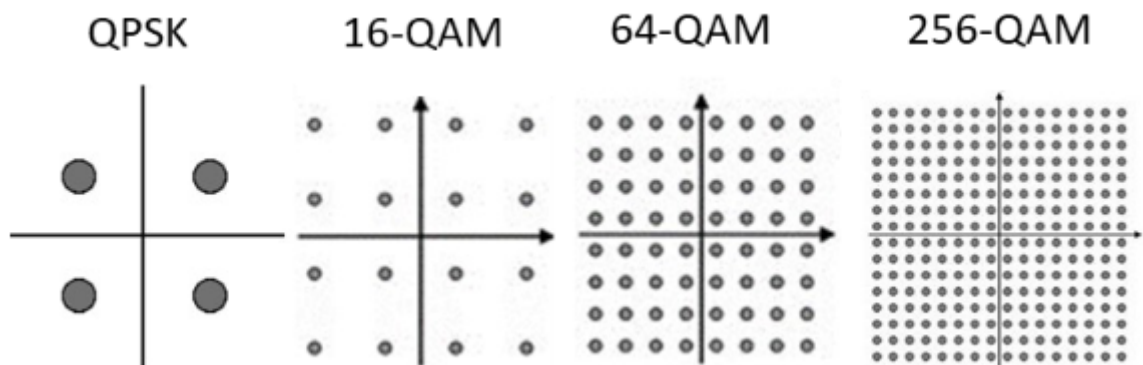
Fonte: Adaptado de [18].

Quadro 8. Número de bits por símbolo de cada modulação

Modulação	Número de bits por símbolo
QPSK	2
16QAM	4
64QAM	6
256QAM	8

Fonte: Elaborado pelo autor.

Figura 13. Constelações para QPSK, 16QAM, 64QAM e 256QAM



Fonte: Samuel Davi de Lima [21].

A eficiência espectral é uma das métricas de avaliação expressa em unidades de bit/s/Hz e é definida como a razão entre a taxa de dados da comunicação e a banda do canal [21]. Ela pode ser calculado a partir do $SINR$ (em unidade linear) e do $BLER$, conforme a seguinte equação:

$$Eficiência\ espectral = \log_2 \left(1 + \frac{SINR}{-\ln(5*BLER)/1.5} \right) \quad (1)$$

Outro elemento importante que é influenciado pelo MCS escolhido pela BS é o Tamanho do Bloco de Transporte (TBS). Nas redes 5G NR, o TBS é definido de acordo com o algoritmo ilustrado na Figura 14. Nele se observa que, dependendo do resultado da

quantidade de informações (N_{info}), o TBS será mapeado via tabela ou calculado por equações. A partir da codificação e modulação selecionados por tabela, tem-se o cálculo do recurso de um elemento (r_{cs_elem}), e posteriormente, o total de recursos ($total_{rcs}$) [18], dados pelas Equações 2 e 3, respectivamente.

$$r_{cs_elem} = Num_{subportadora\ por\ RB} * Num_{símbolos\ por\ slot} - Num_{DMRS\ por\ PRB} - OH_{PD SCH} \quad (2)$$

Onde:

$Num_{subportadora\ por\ RB}$ é o número de subportadoras por recurso de bloco, sendo fixo igual a 12;

$Num_{símbolos\ por\ slot}$ é o número de símbolos por $slot$, podendo ser 12 ou 14 dependendo do CP ;

$Num_{DMRS\ por\ PRB}$ é o número de recursos para $DMRS$ (Sinal de Referência de Demodulação) por recurso de bloco;

$OH_{PD SCH}$ é o *overhead* de $PD SCH$ configurado. Quando não está configurado, seu valor é zero;

$$total_{rcs} = \min(156, r_{cs_elem}) * Num_{PRB} \quad (3)$$

Onde:

Num_{PRB} é o número de recursos de bloco alocados para o usuário DL ;

Ao final, N_{info} é expressa como:

$$N_{info} = total_{rcs} * R * Q_m * v \quad (4)$$

Onde:

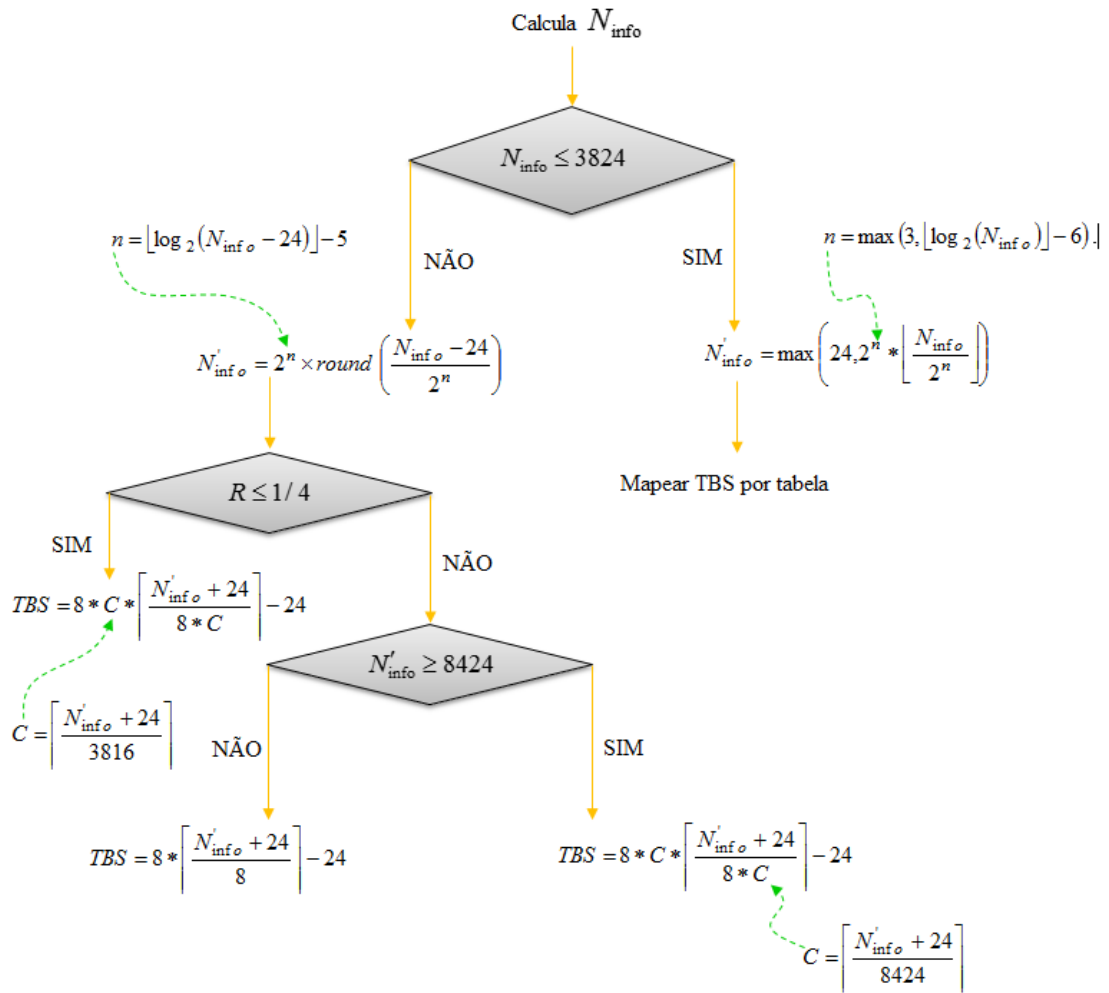
R é a taxa de codificação mapeada pelo índice do MCS ;

Q_m é a ordem de modulação mapeada pelo índice do MCS ;

v é o número de camadas do canal;

Baseado no N_{info} , determina-se o fluxo de definição do TBS conforme Figura 14. No Quadro 9 é apresentado o mapeamento do TBS através do valor de N_{info} [18].

Figura 14. Fluxo de definição do TBS



Fonte: Adaptado de [18].

Quadro 9. Mapeamento do TBS quando $N_{info} \leq 3824$

Índice	TBS	Índice	TBS	Índice	TBS	Índice	TBS
1	24	31	336	61	1288	91	3624
2	32	32	352	62	1320	92	3752
3	40	33	368	63	1352	93	3824
4	48	34	384	64	1416		
5	56	35	408	65	1480		

6	64	36	432	66	1544		
7	72	37	456	67	1608		
8	80	38	480	68	1672		
9	88	39	504	69	1736		
10	96	40	528	70	1800		
11	104	41	552	71	1864		
12	112	42	576	72	1928		
13	120	43	608	73	2024		
14	128	44	640	74	2088		
15	136	45	672	75	2152		
16	144	46	704	76	2216		
17	152	47	736	77	2280		
18	160	48	768	78	2408		
19	168	49	808	79	2472		
20	176	50	848	80	2536		
21	184	51	888	81	2600		
22	192	52	928	82	2664		
23	208	53	984	83	2728		
24	224	54	1032	84	2792		
25	240	55	1064	85	2856		
26	256	56	1128	86	2976		
27	272	57	1160	87	3104		
28	288	58	1192	88	3240		
29	304	59	1224	89	3368		
30	320	60	1256	90	3496		

Fonte: Adaptado de [18].

A vazão, de forma geral, mede o quanto de uma informação é processada por um meio durante um determinado período de tempo, sendo no contexto das Redes 5G, uma métrica em bps (bits por segundo) que mede o fluxo de dados por um canal. Nesse cenário, seu cálculo é baseado na duração do *slot* (determinado pela numerologia) e pelo número de bits por *slot*, que é calculado por meio do *overhead* do canal *downlink* (OH_{dw}) e do *TBS*. No Quadro 10 é informado os valores de *overhead* de *DL* e *UL* nas faixas FR1 e FR2. As Equações 5, 6 e 7 denotam a duração do *slot* ($slot_{time}$), número de bits por *slot* ($bits_{slot}$) e vazão, respectivamente.

Quadro 10. Valores para *OverHead* em FR1 e FR2

Direção	Faixa de frequência	<i>Overhead</i>
DL	FR1	14%
DL	FR2	18%
UL	FR1	8%
UL	FR2	10%

Fonte: Adaptado de J. G. Proakis e M. Slehi (apud [21], 2023, p. 43).

$$slot_{time} = \frac{1}{2^{\mu}} * 10^{-3} \quad (5)$$

$$bits_{slot} = (1 - OH_{dw}) * TBS \quad (6)$$

$$Vazão = \frac{bits_{slot}}{slot_{time}} \quad (7)$$

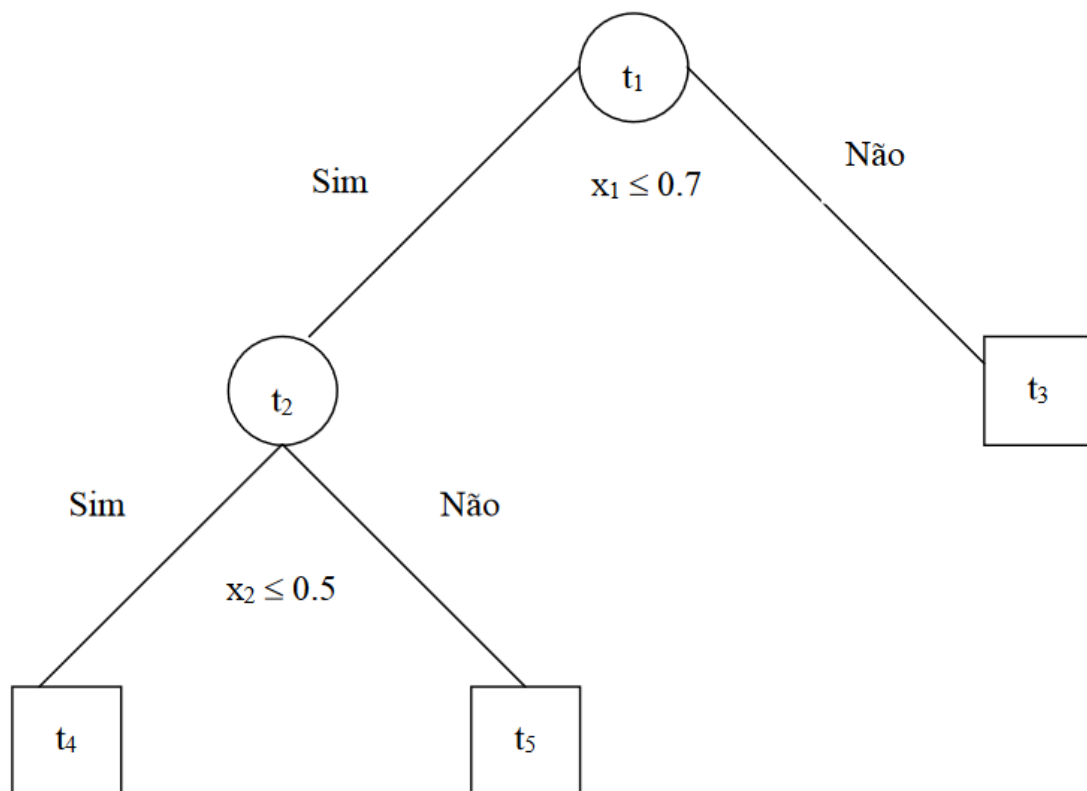
2.3. Inteligência Artificial e Regressão

Os modelos de aprendizado supervisionado empregados na solução deste trabalho envolvem a regressão, ou seja, verificam a influência entre variáveis numéricas, nesse caso entre as variáveis de entrada e a variável de saída a ser predita, o *SINR*. Na regressão, busca-se minimizar o erro calculado pela diferença entre o resultado real e o predito, podendo utilizar, por exemplo, como métricas de avaliação o Coeficiente de Determinação (*R2 Score*), *MSE* ou *RMSE*.

2.3.1. Árvore de Decisão

A Árvore de Decisão é um algoritmo de aprendizado que realiza tomadas de decisão, formado por nós e ramos (ligações entre os nós), conectando-se entre si e caracterizando uma hierarquia em formato de árvore com prováveis caminhos a serem percorridos do nó inicial (raiz da árvore, sempre será o ponto de partida único) até os nós terminais (as folhas da árvore, contendo o valor final do resultado de um atributo). Os nós intermediários (nós de decisão) possuem uma condição de algum atributo (*feature*), no qual dependendo da resposta, só é possível se decidir entre dois caminhos, um para o caso afirmativo e outro o negativo. Assim, seu funcionamento caracteriza uma recursividade na qual conforme ocorrem as divisões dos nós (*splits*), o nível de profundidade da árvore aumenta até chegar nos pontos de parada (as folhas), de tal maneira que quando calcula o caminho esquerdo, por exemplo, não leva em consideração o resultado do caminho direito, caracterizando-o como um algoritmo guloso, além de buscar minimizar o custo dessas divisões [30]. A Figura 15 ilustra um exemplo de árvore de decisão.

Figura 15. Representação de uma Árvore de Decisão



Fonte: PUC-Rio [33].

Nessas divisões, critérios de seleção das *features* são utilizados para determinar quais atributos serão levados em consideração nos nós intermediários (por exemplo, a partir da raiz quadrada do total de *features*), como também qual critério de parada será considerado para determinar um nó terminal. Para a Árvore de Decisão de Regressão, os atributos são numéricos e seus intervalos que caracterizam a forma de divisão da árvore, tal que o *MSE* é considerado para o custo do *split* por padrão, sendo considerado também o grau de pureza para divisão [31][32]. Dentre as formas de realizar uma parada, tem-se a definição de um número mínimo de amostras para um nó folha, altura máxima da árvore e o valor mínimo de decréscimo da impureza, por exemplo [32].

Para melhorar o desempenho de uma árvore, existe a podagem, que remove ramos de menor importância da árvore, visando evitar o sobreajuste dos dados (*overfitting*) e reduzir a complexidade da árvore. Uma das formas de podagem é pelo Custo Mínimo de Complexidade da Poda, partindo de um parâmetro de complexidade (*alpha*) no qual mede o custo de complexidade de nós e ramos, podendo removê-los com base no custo das subárvores [31][32][34].

2.3.2. GMDH

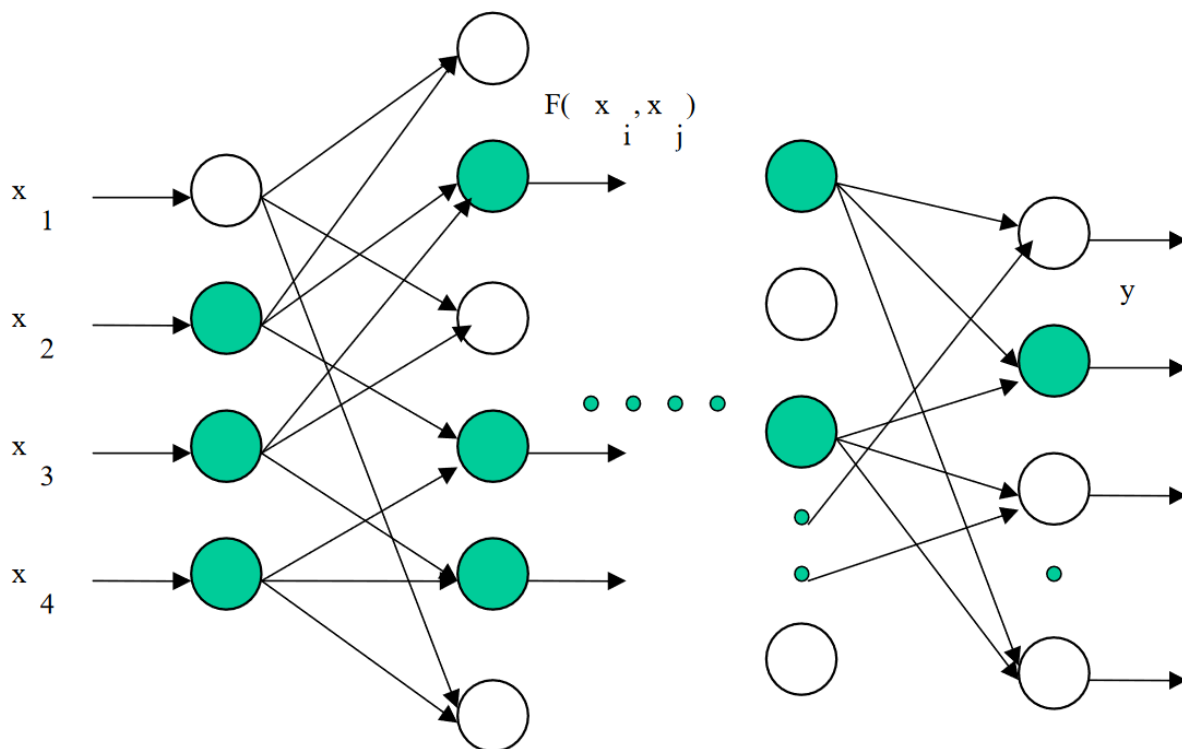
O *GMDH* é um algoritmo auto-organizado formando uma espécie de rede polinomial a partir de funções de ativação, definindo a arquitetura do modelo no momento de aprendizado e de forma otimizada, minimizando a intervenção de um analista [35][36]. O algoritmo envolve algumas heurísticas como [37]:

- 1) Ter um conjunto relevante de observações para o objeto.
- 2) Divisão dessas observações em dois grupos, um para estimar os coeficientes e outro para separar informação útil dos dados.
- 3) Ter um conjunto de funções que aumentarão a complexidade iterativamente, permitindo a produção de diferentes modelos.
- 4) Determinação do modelo ótimo a partir de um critério externo.

A estrutura do *GMDH* com multicamadas é composta por três partes: as variáveis de entrada, as camadas ocultas e a variável de saída, conforme Figura 16, onde os pontos verdes são os nós selecionados de acordo com o critério externo, dispensando os outros nós que não são mais necessários [37]. À medida que avança nos nós, as estimativas são propagadas de acordo com os resultados da função de ativação, melhorando a cada nível passado. A Figura 17 apresenta um nó do *GMDH*, composto pelas entradas (x_i e x_j), coeficientes polinomiais (os

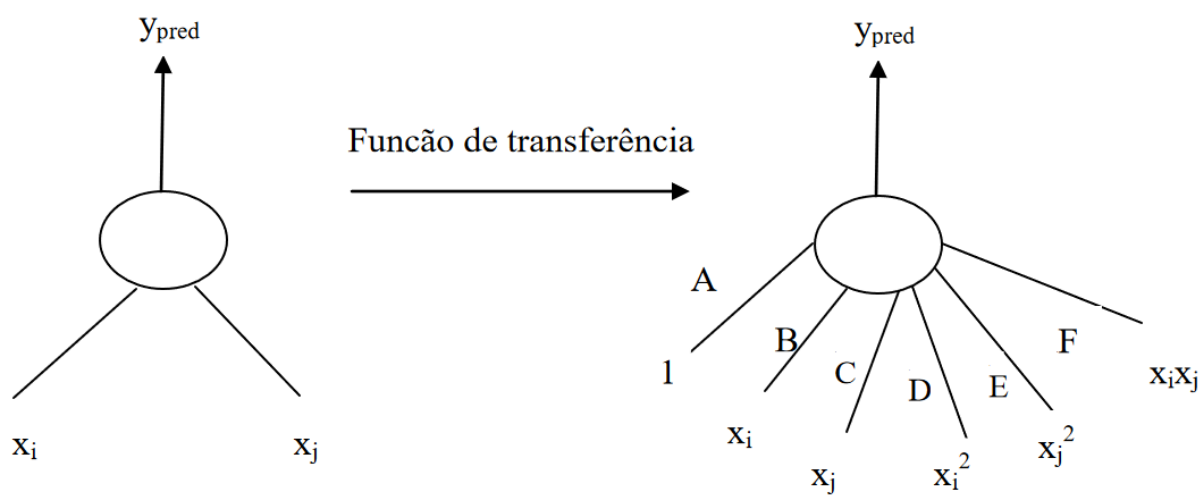
pesos de A a F) e a saída (y_{pred}), estabelecendo ao final, a relação entre as entradas e a saída através de uma função com os pesos ajustados [35].

Figura 16. Representação do *GMDH* Multicamada



Fonte: Anastasakis e Mort [37].

Figura 17. Nó do *GMDH*



Fonte: Elaine Inacio Bueno [35].

Alguns meios de melhoria do desempenho do *GMDH* são: definição do número máximo de camadas, limite de erro estabelecido e alteração do α para aumentar ou diminuir a influência dos pesos, por exemplo [38]. Dentre alguns exemplos de se estabelecer o critério de seleção dos nós tem-se: o Critério de Regularidade, Critério de Viés Mínimo e *Cross-Validation*.

3. Trabalhos Relacionados (Estado da Arte)

Soluções baseadas em previsão foram propostas para o problema de *feedback* desatualizado do *CQI* na literatura. Por exemplo, os autores em [25] analisam diferentes técnicas, como Linear, Aproximação Estocástica Linear (*LSA*), Filtros de Kalman e sequências de Transformada Discreta de Cosseno (*DCT*) para prever o *CQI*, com *LSA* superando o outro em termos de compensação de complexidade-desempenho. Em [27], os autores usam valores anteriores de relação sinal-ruído (*SNR*) e extrapolação linear para prever a *SNR* futura. Porém, a proposta não funciona adequadamente em cenários com usuários de velocidade moderada ou alta.

Por outro lado, [26] adota uma Rede Neural de Memória a Longo Prazo (*LSTM*) para prever o *CQI* e usa retreinamento online para obter alta acurácia mesmo em cenários dinâmicos. Para prever o *CQI*, [28] propõe uma abordagem de redes neurais recorrentes profundas (*DRNNs*) baseada na série temporal de valores de *CQI* anteriores. Os autores defendem que a solução é projetada para informações de controle de Veículos Aéreos Não Tripulados (*UAV*) com base em comunicação Ultra-confiável e de Baixa Latência (*URLLC*), mas não consideram os aspectos de mobilidade do dispositivo, como posição do dispositivo, velocidade de movimento e direção para estimar o *CQI* futuro. Essas três abordagens (ou seja, [26], [27] e [28]) são previsores de entrada de tipo único, mas diferem entre si em termos da técnica de previsão, onde [28] considera a camada de Memória a Longo Prazo (*LSTM*) e a Unidade Recorrente Fechada (*GRU*). Além disso, [26] e [28] prevêm o *CQI* e são executados na estação base enquanto [27] foca no *SNR* e é incorporado ao dispositivo.

Em [23], os autores abordam o problema de *feedback* desatualizado do *CQI* propondo uma solução baseada em uma Rede Neural Perceptron Multicamadas (*MLP*) que leva em conta o contexto de mobilidade do *UE* (por exemplo, posição, velocidade e direção do movimento), comprimento do atraso, relação sinal-interferência mais ruído (*SINR*) atual para prever o *SINR* atualizado que será mapeado em um valor do *CQI*. Este trabalho também considera essas variáveis de entrada (com exceção do comprimento de atraso), mas adota o Método de Grupo de Tratamento de Dados e a Árvore de Decisão para prever o *SINR* atualizado a ser mapeado em um valor do *CQI*. Ao contrário do *MLP*, que requer informações significativas sobre a topologia da rede neural (por exemplo, número de camadas e neurônios ocultos, bem como sua função de ativação) que podem ser definidas subjetivamente pelo analista ou determinadas através de testes como em [23], o *GMDH* otimiza automaticamente

sua estrutura. Além disso, este trabalho não apenas avalia a precisão das soluções *GMDH* e *Árvore de Decisão*, mas também considera seu impacto na eficiência espectral e vazão. O Quadro 11 sumariza os principais pontos dos trabalhos mencionados, inclusive este trabalho, como forma de apresentar a diferença deste trabalho em relação aos outros.

Quadro 11. Trabalhos relacionados sobre o problema de *feedback* desatualizado do *CQI*

Proposta	Objetivo(s)	Técnica(s)	Lado de atuação	Entrada(s)	Saída
Uso de <i>Machine Learning</i> para o problema de <i>CQI</i> desatualizado [23]	Estimar o <i>SINR</i> conforme o <i>feedback</i> de atraso do <i>CQI</i>	Rede Neural Perceptron Multicamadas (<i>MLP</i>)	<i>UE</i>	Mobilidade do <i>UE</i> (posição, velocidade e direção do movimento), comprimento do atraso e <i>SINR</i> atual	<i>SINR</i>
Uso de preditores de <i>CQI</i> para compensar o problema de envelhecimento do <i>CQI</i> no canal <i>downlink</i> do <i>LTE</i> [25]	Melhorar capacidade da célula por meio da previsão do <i>CQI</i>	Linear, Aproximação Estocástica Linear (<i>LSA</i>), Filtros de Kalman e sequências de Transformada Discreta de Cosseno (<i>DCT</i>)	<i>UE</i>	Ganho do canal baseado no atraso do <i>CQI</i> aplicado no conjunto de observações	<i>CQI</i>
Uso da aprendizagem profunda <i>LSTM</i> e retreinamento online para predição do <i>CQI</i> [26]	Aprimorar a predição do <i>CQI</i> de forma a torná-la mais precisa	Rede Neural de Memória a Longo Prazo (<i>LSTM</i>)	<i>BS</i>	<i>CQI</i>	<i>CQI</i>
Uso de extrapolação linear para predição de <i>SNR</i> e mapeamento do <i>CQI</i> [27]	Aprimorar precisão de <i>feedback</i> do canal por meio do <i>CQI</i> mapeado	Extrapolação Linear	<i>UE</i>	<i>SNR</i>	<i>SNR</i> e <i>CQI</i> mapeado
Uso de <i>DRNNs</i> para predição do <i>CQI</i> para informações de controle de <i>UAV</i> com base em comunicação <i>URLLC</i> [28]	Aprimorar a predição do <i>CQI</i> de forma a torná-la mais precisa	Redes Neurais Recorrentes Profundas (<i>DRNNs</i>)	<i>BS</i>	Série temporal de valores de <i>CQI</i> anteriores	<i>CQI</i>
Uso de <i>Árvore de Decisão</i> e <i>GMDH</i> para o problema de <i>CQI</i> desatualizado (Este trabalho)	Estimar o <i>SINR</i> conforme o <i>feedback</i> de atraso do <i>CQI</i> , além de verificar seu	<i>Árvore de Decisão</i> e <i>GMDH</i>	<i>UE</i>	Mobilidade do <i>UE</i> e <i>SINR</i> atual	<i>SINR</i>

	impacto na eficiência espectral e vazão				
--	--	--	--	--	--

Fonte: Elaborado pelo autor.

4. Metodologia

Este capítulo trata da geração do *dataset*, treinamento dos modelos e cálculo das métricas de análise dos resultados de treinamento. Os modelos estimam o valor do *SINR*, o qual é mapeado em *CQI*, visando estabelecer que seu valor não fique desatualizado para uma determinada janela de atraso. Dessa forma os recursos podem ser alocados de forma devida e expressando a resposta da qualidade do canal esperada para a *BS*.

4.1. Geração do *dataset* para treinamento dos modelos

A base usada para o tratamento dos dados foram os cenários gerados a partir do simulador *mmWave* [29], no qual pode-se informar como entrada o tempo de execução da simulação, número de *UE* e *BS*, área de cobertura (coordenadas *x*, *y* e *z*) e velocidade direcional (representadas por *xVel*, *yVel* e *zVel*). Algumas das informações geradas como saída do simulador são *SINR(dB)*, momento de execução (*time*), velocidade absoluta (*AbsVel*), ângulo (*Angle*), distância da *BS* mais próxima (*CalcClosestEnbDist*), posição corrente (*CurXPos*, *CurYPos* e *CurZPos*), velocidade corrente (*CurXVel*, *CurYVel* e *CurZVel*) e direção (*Direction*). Estas informações são usadas para a construção final do *dataset*. Com o intuito de obter dados diversificados, realizaram-se oito execuções, onde as configurações são mostradas no Quadro 12.

Quadro 12. Configurações dos cenários de simulação

Cenário	Parâmetros de Entrada								
	Posição x	Posição y	Posição z	xVel	yVel	zVel	Tempo de Execução	Número de UE	Número de BS
A1	100	100	100	100	100	100	30s	1	1
A2	50	30	20	100	100	100	30s	1	1
A3	30	50	100	100	100	100	30s	1	1
A4	20	30	50	100	100	100	30s	1	1
A5	10	100	50	100	100	100	30s	1	1
A6	100	80	20	100	100	100	30s	1	1
A7	40	50	60	100	100	100	30s	1	1
A8	80	80	40	100	100	100	30s	1	1

Fonte: Elaborado pelo autor.

Com os resultados da simulação, calculou-se uma coluna nova (*Delay*) relacionada à diferença dos tempos da coluna *time* (momento de execução) de acordo com a janela de *delay* predefinida ($\tau = 1$), correspondente à quantos “saltos” os dados usados como entrada estão em relação ao dado alvo. Com $\tau = 1$, o atraso é calculado utilizando cada amostra de momento de execução e a amostra seguinte, computando a diferença absoluta entre elas. O conjunto formado no *dataset* final é composto com os dados de entrada (*Delay*, *SINR(dB)*, *AbsVel*, *Angle*, *CalcClosestEnbDist*, *CurXPos*, *CurYPos*, *CurZPos*, *CurXVel*, *CurYVel*, *CurZVel* e *Direction*) e o dado alvo ou de saída (*SINR(dB)* + τ , onde $\tau = 1$), possuindo a junção do processamento de todos os cenários do Quadro 12. A Figura 18 apresenta o estado final do *dataset*.

Figura 18. Estrutura do *dataset* final

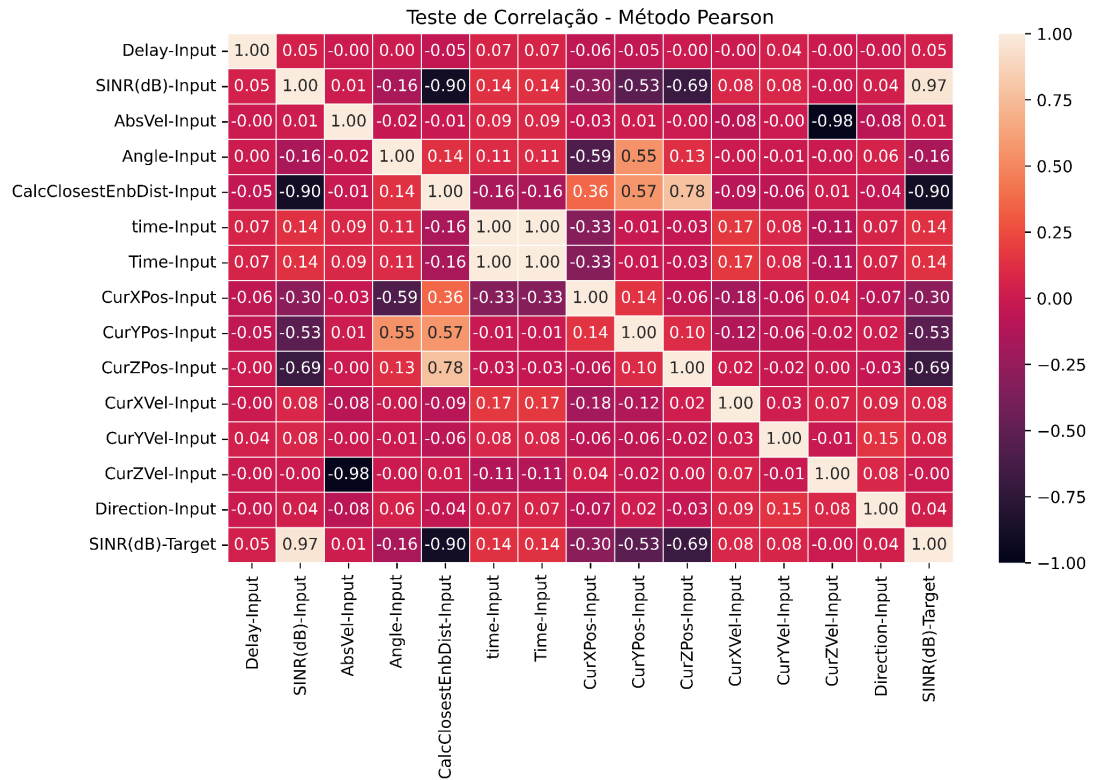
Generating Final Dataset - After [A8]:					
	Delay-Input	SINR(dB)-Input	AbsVel-Input	Angle-Input	\
0	0.019383	-6.38835	19.6441	0.784444	
1	0.011000	-5.58315	19.6441	0.782529	
2	0.001104	-4.92377	19.6441	0.781568	
3	0.000200	-5.34894	19.6441	0.780606	
4	0.000200	-5.49176	19.6441	0.779642	
...	...	...	...	...	
23971	0.010000	10.63250	21.4752	1.503560	
23972	0.010000	13.01300	21.4752	1.500170	
23973	0.010000	10.19090	21.4752	1.496740	
23974	0.010000	12.61790	21.4752	1.493280	
23975	0.010000	9.84467	21.4752	1.489780	
	CalcClosestEnbDist-Input	time-Input	Time-Input	CurXPos-Input	\
0	173.0880	0.004008	0.01	99.99440	
1	172.8540	0.023391	0.03	99.98310	
2	172.7380	0.034391	0.04	99.97750	
3	172.6210	0.035496	0.05	99.97180	
4	172.5050	0.035696	0.06	99.96620	
...	...	...	...	...	
23971	50.1322	29.841300	29.94	2.26125	
23972	50.0094	29.851300	29.95	2.36222	
23973	49.8871	29.861300	29.96	2.46319	
23974	49.7655	29.871300	29.97	2.56417	
23975	49.6445	29.881300	29.98	2.66514	

	CurYPos-Input	CurZPos-Input	CurXVel-Input	CurYVel-Input	\
0	99.8036	99.9990	-0.563485	-19.6358	
1	99.4109	99.9971	-0.563485	-19.6358	
2	99.2146	99.9961	-0.563485	-19.6358	
3	99.0182	99.9952	-0.563485	-19.6358	
4	98.8219	99.9942	-0.563485	-19.6358	
...	...	...	...	...	
23971	33.5808	37.1545	10.097300	-18.9530	
23972	33.3913	37.1534	10.097300	-18.9530	
23973	33.2018	37.1524	10.097300	-18.9530	
23974	33.0123	37.1513	10.097300	-18.9530	
23975	32.8227	37.1503	10.097300	-18.9530	
	CurZVel-Input	Direction-Input	SINR(dB)-Target		
0	-0.096559	1.54211	-5.58315		
1	-0.096559	1.54211	-4.92377		
2	-0.096559	1.54211	-5.34894		
3	-0.096559	1.54211	-5.49176		
4	-0.096559	1.54211	-5.64696		
...	...	...	...		
23971	-0.104172	5.20190	13.01300		
23972	-0.104172	5.20190	10.19090		
23973	-0.104172	5.20190	12.61790		
23974	-0.104172	5.20190	9.84467		
23975	-0.104172	5.20190	11.38500		
[23976 rows x 15 columns]					

Fonte: Elaborado pelo autor.

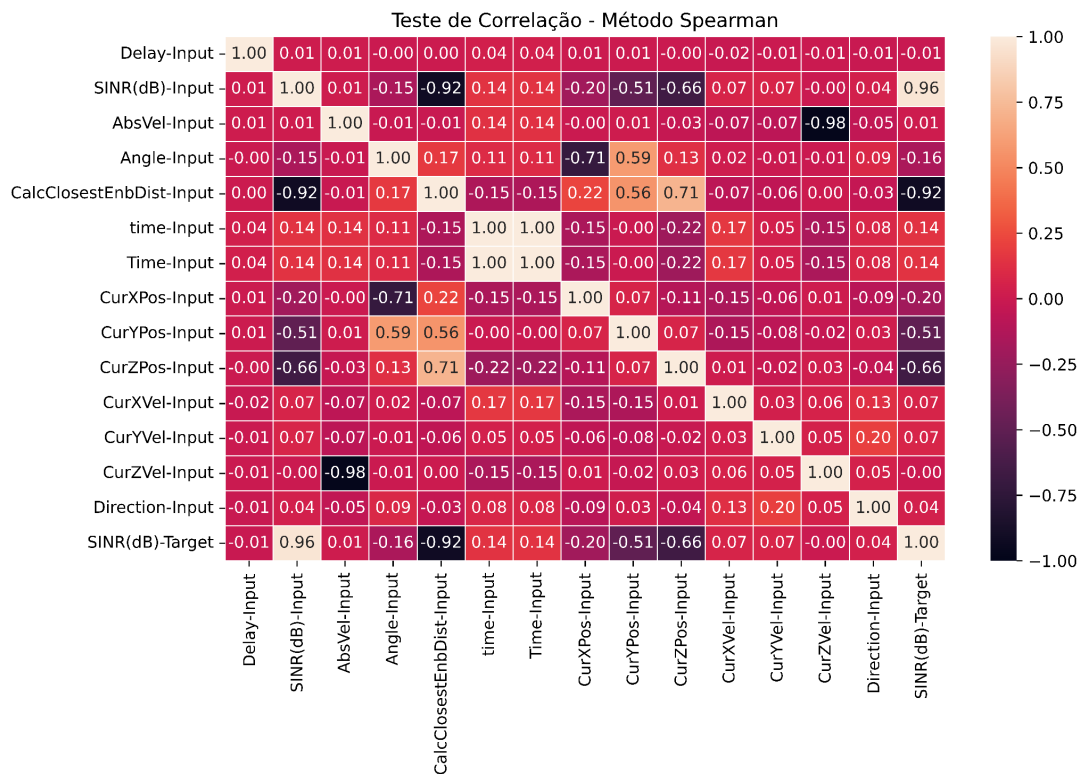
Para verificar a influência dos dados de entrada com o dado alvo, aplicou-se testes de correlação com os métodos Pearson, Spearman e Kendall, representados nas Figuras 19, 20 e 21 respectivamente. Posteriormente, restringiu-se com correlação maior ou igual a 0.5 (Padrão de Correlação de Cohen), identificando as que tiveram maiores valores de correlações para cada tipo. O Quadro 13 mostra a filtragem do *dataset* após a aplicação desta condição de correlação. Com isso, além do *dataset* final, há mais três *datasets*, uma para cada teste de correlação com o limiar aplicado. Por último, há a normalização (entre 0 e 1) desses *datasets* utilizando a escala *MinMax*, onde o menor valor de um atributo é representado por zero e o maior valor é representado por um. Assim, ficam dispostos no total oito *datasets* para uso. O Quadro 14 mostra um pseudocódigo com as etapas de geração do *dataset*.

Figura 19. Teste de correlação com método Pearson



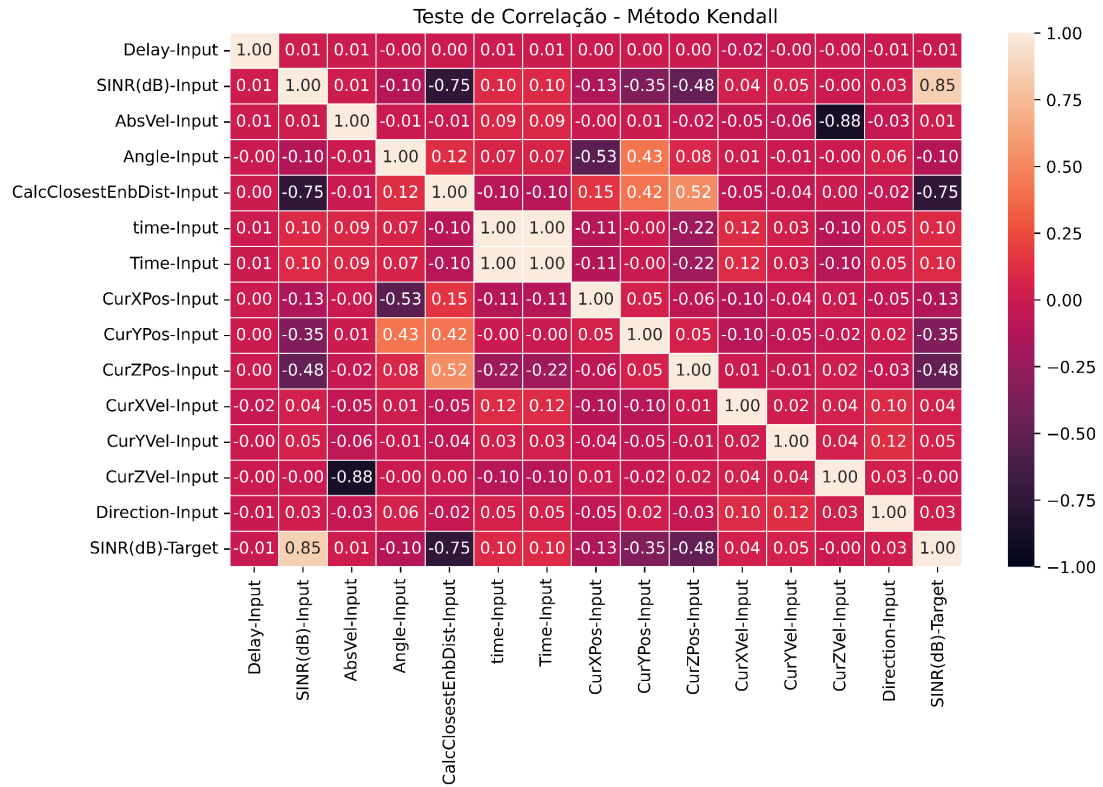
Fonte: Elaborado pelo autor.

Figura 20. Teste de correlação com método Spearman



Fonte: Elaborado pelo autor.

Figura 21. Teste de correlação com método Kendall



Fonte: Elaborado pelo autor.

Quadro 13. Filtragem de colunas do *dataset* com correlação ≥ 0.5

Teste de Correlação	Colunas selecionadas
Pearson	SINR(dB)-Input, CalcClosestEnbDist-Input, CurYPos-Input, CurZPos-Input e SINR(dB)-Target
Spearman	SINR(dB)-Input, CalcClosestEnbDist-Input, CurYPos-Input, CurZPos-Input e SINR(dB)-Target
Kendall	SINR(dB)-Input, CalcClosestEnbDist-Input e SINR(dB)-Target

Fonte: Elaborado pelo autor.

Quadro 14. Pseudocódigo com etapas de geração do *dataset*

Pré-condições:

Cenários gerados por dados de simulador.

Definir identificador dos cenários, número de cenários, tamanho da janela de atraso e limiar para filtragem da correlação.

Algoritmo:

Para cada identificador do cenário:

Para cada número do cenário:

Captura as colunas de interesse dos dados gerados pelo simulador.

Gera nova coluna de *Delay* baseado na diferença absoluta do momento de execução da amostra atual e da amostra seguinte de acordo com janela de atraso.

Gera novas colunas de *input* com base nas colunas de interesse e de acordo com janela de atraso.

Gera novas colunas de *target* com base nas colunas de interesse e de acordo com janela de atraso.

Seleciona as colunas de *input* e *target* desejadas e concatena no *dataset* final a ser formado.

Realiza os testes de correlação de cada método (Pearson, Spearman e Kendall) no *dataset* final gerado.

Gera *dataset* filtrado para cada método de acordo com limiar de correlação (correlação da variável de *input* com a variável *target* que seja maior ou igual ao limiar de correlação).

Normaliza cada *dataset* gerado com a escala *MinMax*.

Pós-condições:

Datasets gerados (4 normalizados e 4 não normalizados).

Fonte: Elaborado pelo autor.

4.2. Treinamento dos modelos

Para os modelos utilizados, foi escolhido o *dataset* normalizado com o método Pearson devido possuir a mesma seleção de colunas realizada pelo método Spearman. Este por sua vez abrange variáveis quantitativas ou não, ao passo que o método Kendall é voltado para variáveis ordinais apenas. Como os dados de entrada possuem diferentes escalas, usou-se a normalização para um dado não se sobressair sobre outro.

Dentre os parâmetros de configuração de cada modelo, estabeleceram-se os valores fixos e variados de acordo com critérios dispostos no Quadro 15. A divisão do *dataset* efetuou-se com 70% para treino e 30% para teste, sendo o mesmo conjunto de divisão utilizado para os dois modelos. Outras configurações pré estabelecidas foram o limite de trinta iterações para cada configuração com o propósito de acarretar o efeito randômico no treinamento, além do limiar da métrica de avaliação *MSE* (Erro Médio Quadrático) ser de 0,01 (1%). Outras métricas geradas para nível de análise das predições foram *RMSE* (Raiz do Erro Médio Quadrático), *MAPE* (Erro Percentual Absoluto Médio), *NRMSE* (Raiz do Erro Médio Quadrático Normalizado) e *R2 Score*. Obteve-se também a média desses erros das

trinta execuções de cada configuração. O Quadro 16 apresenta a sumarização dos valores providos para os parâmetros de cada modelo e o número total de configurações geradas para cada um. O Quadro 17 mostra um pseudocódigo com as etapas do treinamento dos modelos.

Quadro 15. Critérios de definição dos parâmetros fixos, variados e seus valores

Natureza do <i>dataset</i>
Característica da definição do parâmetro de configuração
Valores prováveis dos parâmetros contribuintes para efeito randômico nas trinta iterações
Número de configurações próxima de 200

Fonte: Elaborado pelo autor.

Quadro 16. Configurações dos modelos de treinamento

Modelo	Total de configurações	Parâmetro	Valor(es)
Árvore de Decisão	216	Critério de medida da qualidade do <i>split</i>	squared_error
		Número mínimo de amostras por <i>split</i>	2
		Fração mínima ponderada das folhas	0.0
		Aleatoriedade do estimador	None
		Diminuição mínima da impureza	0.0
		Estratégia de <i>split</i> de cada nó	['best', 'random']
		Número máximo de <i>features</i> a considerar durante o <i>split</i>	['sqrt', 'log2']
		Profundidade máxima da árvore	[5, 10]
		Número mínimo de amostras para uma folha	[20, 30, 40]
		Número máximo de folhas	[30, 40, 50]
		Parâmetro de complexidade do Custo Mínimo de Complexidade da Poda	[0.0, 0.01, 0.02]
		Método de <i>split</i> dos dados em treino e validação	random

GMDH	192	Nome original das <i>features</i>	True
		Número máximo de camadas	10
		Critério de largura mínima	5
		Limite de parada do treinamento	0.001
		Normalização das <i>features</i>	False
		Número de <i>threads</i> para treinamento do modelo	1
		Seleção manual dos melhores nós	False
		Critério de seleção dos melhores nós	['validate', 'bias', 'validate_bias', 'bias_retrain']
		Função de referência	['linear_cov', 'quadratic', 'cubic', 'linear']
		Critério do cálculo do erro de camada	['top', 'avg']
		Parâmetro de penalidade	[0.001, 0.01, 0.1, 0.5, 1.0, 2.0]

Fonte: Elaborado pelo autor.

Quadro 17. Pseudocódigo com etapas do treinamento dos modelos

Pré-condições:

Dataset a ser utilizado ter sido gerado.

Definir número de iterações de treinamento dos modelos e limiar do *MSE* de referência.

Algoritmo:

Leitura do *dataset* a ser utilizado.

Divisão do *dataset* em conjunto de treino e de teste.

Para cada modelo de aprendizagem:

 Definir os parâmetros fixos e variados.

 Para cada combinação dos parâmetros variados (configuração):

 Atribui parâmetros do modelo de acordo com a configuração.

 Para cada iteração de treinamento até atingir o limite determinado:

 Treina o modelo.

 Realiza a predição.

 Calcula *MSE*, *RMSE*, *MAPE*, *NRMSE* e *R2 Score* do conjunto de treino e teste.

 Armazena resultados da configuração.

Pós-condições:

Número de configurações geradas próxima de 200 para cada modelo treinado.

Valores dos parâmetros variados terem contribuído para efeito randômico.

4.3. Cálculo das métricas comparativas

A partir das melhores configurações de cada modelo já selecionadas, realiza-se a leitura dos dados preditos normalizados (y_{pred}) e dos utilizados como teste (y_{test}), necessitando o auxílio do dataset não normalizado para o devido cálculo da transformação inversa dos mesmos, retornando os dados à escala original.

O mapeamento das métricas ocorre para o resultado de cada iteração ($SINR$ predito) como também para o resultado a ser usado para comparação ($SINR$ real). Inicialmente, alguns valores como *downlink overhead*, número de recursos de bloco, índice μ da numerologia e taxa de bloco de erro alvo são predefinidos segundo o Quadro 18. Outras informações como CQI , MCS e TBS são mapeadas por tabelas. Com o $SINR$ (dB), faz-se uma conversão de decibel para linear em razão do cálculo da eficiência espectral, e em seguida, exerce-se o mapeamento do CQI . Com o valor do CQI , é definido o MCS , que por sua vez, mapeia o esquema de codificação e a ordem de modulação para o cálculo do recurso de um elemento e do total de recursos. Nesses cálculos, há também valores predeterminados conforme indicado no Quadro 18. Por fim, calculando o número de bits de informação, segue o mapeamento do TBS , prosseguindo com o cálculo da vazão em Mbps a partir do tempo de *slot* e número de *bits* por *slot*. Os cálculos e mapeamentos aplicados dispõem-se nos Quadros 2, 6, 9, Figura 14 e Equações de 1 a 7 do Capítulo 2.4. O Quadro 19 mostra um pseudocódigo com as etapas do cálculo das métricas.

Quadro 18. Definição dos parâmetros para mapeamento do TBS e cálculo da vazão

Parâmetro	Valor
<i>Downlink overhead</i>	0.18
Número de recursos de bloco alocados	1
Índice μ	3
Taxa de bloco de erro alvo	0.00005
Número de camadas	1
Número de subportadoras por bloco de recursos	12
Número de símbolos por <i>slot</i>	14

Número de recursos para demodulação por bloco	0
<i>PDCH overhead</i>	0

Fonte: Elaborado pelo autor.

Quadro 19. Pseudocódigo com etapas do cálculo das métricas

Pré-condições:

Ter determinado as configurações a serem calculadas as métricas para cada modelo.
 Determinar as tabelas de mapeamento do *CQI*, codificação e modulação a serem usadas.
 Definir valores dos parâmetros do Quadro 18.

Algoritmo:

Leitura de *dataset* final não normalizado.
 Caso os dados preditos estejam normalizados, retornam à escala original.
 Para cada amostra predita e real da configuração escolhida dos modelos treinados:
 Realiza cálculo da eficiência espectral e mapeamento do *CQI*.
 Realiza mapeamento do *MCS*.
 Realiza mapeamento do *TBS*.
 Calcula a vazão.
 Armazena resultados gerados.

Pós-condições:

Métricas calculadas para todas as amostras da configuração escolhida de cada modelo treinado.

Fonte: Elaborado pelo autor.

5. Resultados e Análise

Este capítulo apresenta as análises para escolha das melhores configurações dos modelos *GMDH* e Árvore de Decisão, e baseado nessa seleção, analisa comparativamente os resultados gerados por ambos através das métricas calculadas como vazão, eficiência espectral e *CQI* das predições em relação aos valores reais.

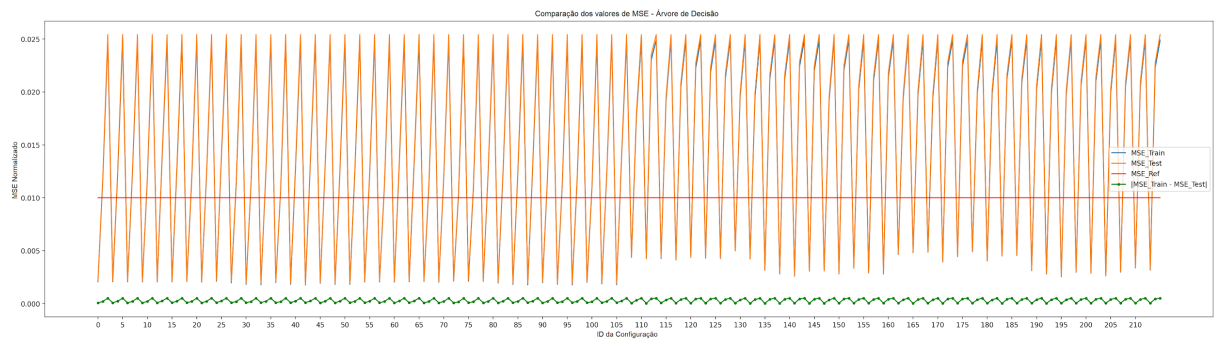
5.1. Análise para seleção das melhores configurações dos modelos treinados

Com o total de 216 configurações para Árvore de Decisão e 192 configurações para *GMDH*, a seleção prosseguiu-se por algumas etapas gerais, especificando-se as particularidades mais adiante. Os passos gerais são:

- 1) Restringir configurações com menor *MSE* abaixo do *MSE* referencial (1%).
- 2) Separar a de menor erro absoluto da diferença dos *MSE*'s (treino e teste).
- 3) Verificar as métricas de erro e considerar a menor média *MSE* com maior média *R2 Score*.
- 4) Verificar os resultados das configurações candidatas.

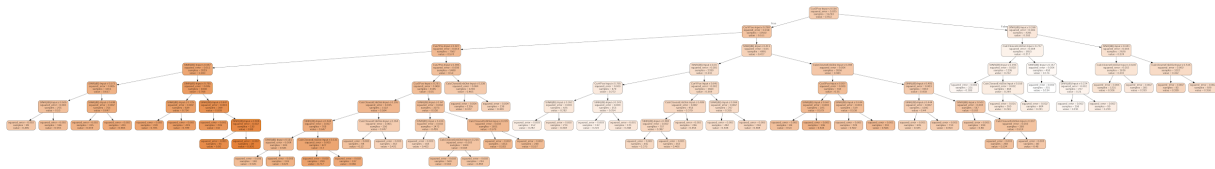
Para a Árvore de Decisão, as configurações com menor *MSE* abaixo de 0,01 são 33, 42, 87, 96 e 105 conforme o Gráfico 1. A que apresenta menor erro absoluto da diferença dos *MSE*'s é a 42. Baseado nas métricas de erros das configurações selecionadas, os valores encontram-se próximos entre si, sendo a configuração 42 a apresentar a menor média *MSE* com maior *R2 Score*. Esta configuração de árvore de decisão possui altura nove com variabilidade na distribuição das folhas ilustrada na Figura 22 com visualização ampliada no Apêndice. Comparando o *SINR*(dB) normalizado predito por ela com o real, tem-se que a maioria das entradas com valores próximos e alguns registros parcialmente distantes, conforme mostrada no Gráfico 2. Para esta configuração, os erros das iterações tem valores menores que 0,1 exceto o *MSE* que se encontra na primeira condição, sendo o *R2 Score* em sua maioria com valores maiores que 0,92. Assim, a configuração 42 foi a escolhida e seus parâmetros são expostos no Quadro 20.

Gráfico 1. Comparação dos valores do *MSE* para Árvore de Decisão



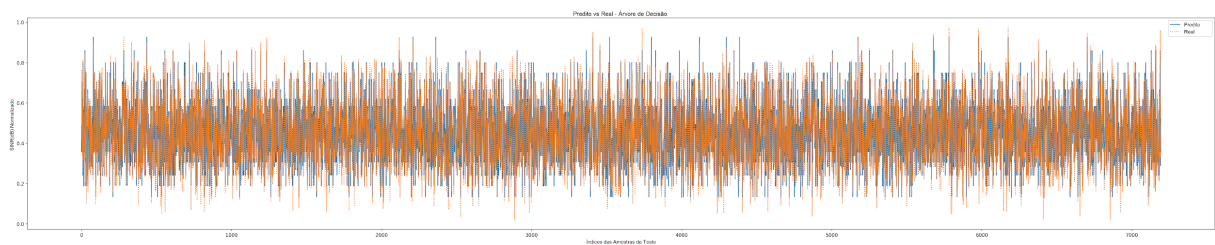
Fonte: Elaborado pelo autor.

Figura 22. Árvore da configuração 42



Fonte: Elaborado pelo autor.

Gráfico 2. *SINR*(dB) normalizado predito versus real para Árvore de Decisão



Fonte: Elaborado pelo autor.

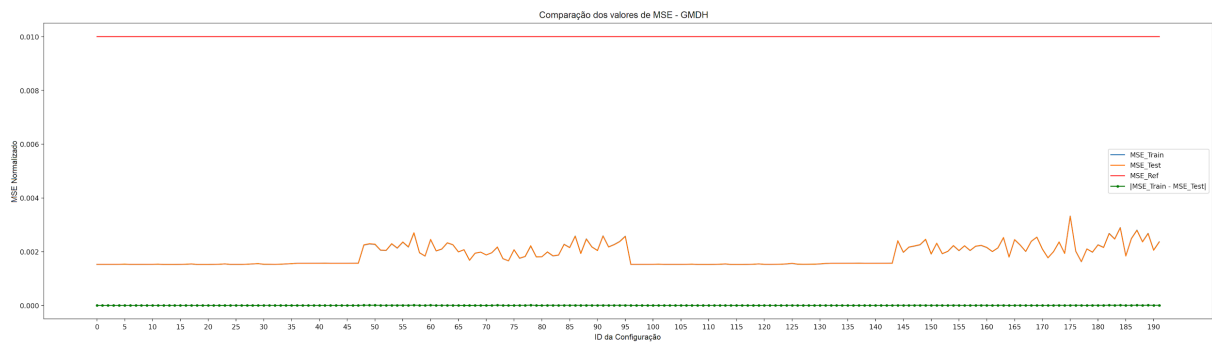
Quadro 20. Valores da configuração 42 da Árvore de Decisão

Parâmetro	Valor
Estratégia de <i>split</i> de cada nó	best
Número máximo de <i>features</i> a considerar durante o <i>split</i>	sqrt
Profundidade máxima da árvore	10
Número mínimo de amostras para uma folha	30
Número máximo de folhas	50
Parâmetro de complexidade do Custo Mínimo de Complexidade da Poda	0

Fonte: Elaborado pelo autor.

Para o *GMDH*, apesar de todas as configurações possuírem *MSE* abaixo de 0,01, é percebido que as configurações entre 96 e 143 (incluindo-las) possuem *MSE* ainda menor conforme Gráfico 3, sendo que a 119 é a que apresenta menor erro absoluto da diferença entre os *MSE*'s. Baseado nas métricas de erros das configurações selecionadas, os valores encontram-se próximos entre si, sendo as configurações 108, 109, 110, 114, 115 e 116 a estarem no critério de menor média *MSE* com maior *R2 Score*. Dessas, a 114 tem menor erro absoluto da diferença entre os *MSEs* (treino e teste). Comparando os resultados da configuração 114 com os de outras dessa seleção, através das plotagens dos erros das camadas e da comparação do valor real da *SINR*(dB) versus o predito, tem-se as configurações 114 e 108 como as melhores configurações. Como critério de desempate, determinou-se a que possui menor média de *MSE* e maior média de *R2 Score*. Assim, a configuração escolhida é a 114 e seus parâmetros são exibidos no Quadro 21. Na Figura 23, Gráficos 4 e 5 são apresentados a rede do modelo, comparação do valor real com o predito do *SINR*(dB) e plotagem dos erros das camadas dessa configuração respectivamente.

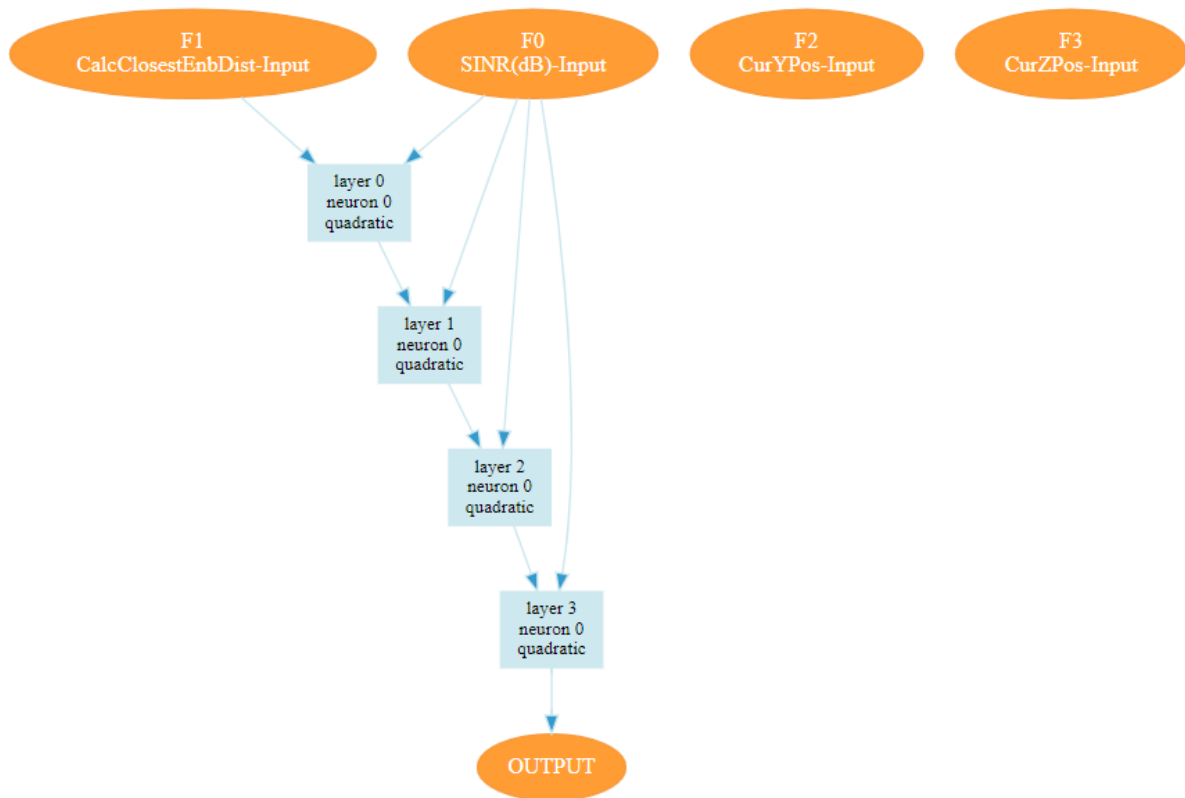
Gráfico 3. Comparação dos valores do *MSE* para *GMDH*



Fonte: Elaborado pelo autor.

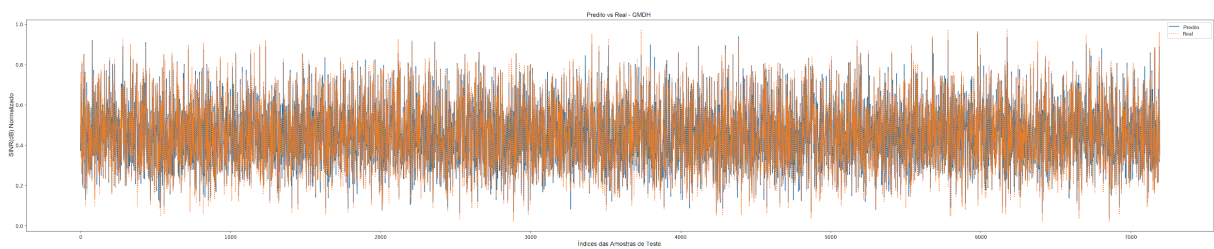
Figura 23. Rede *GMDH* da configuração 114

Rede neural polinomial de aprendizagem profunda auto-organizada



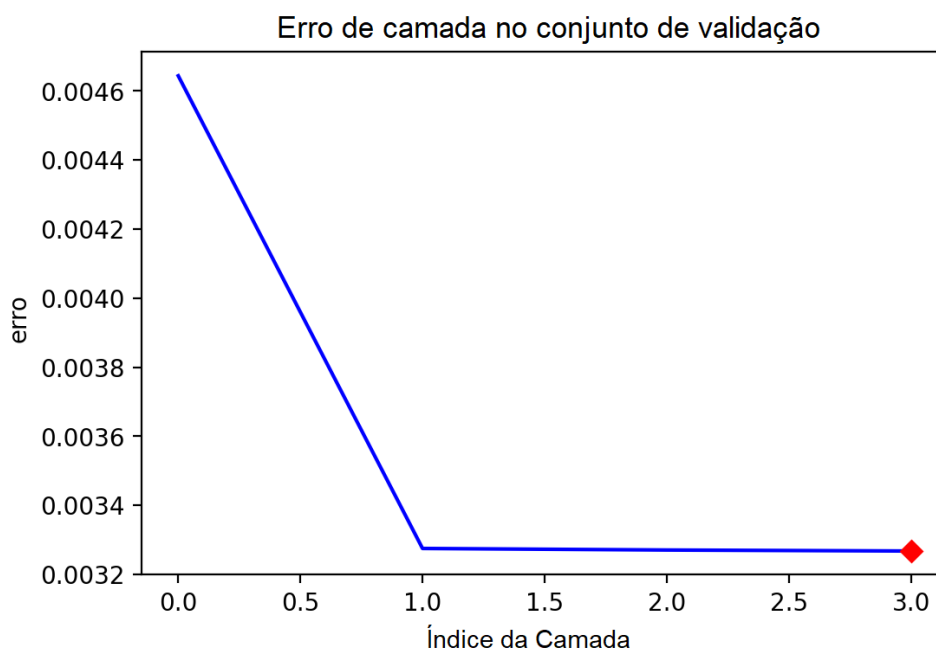
Fonte: Gerado por *software* (adaptado).

Gráfico 4. *SINR*(dB) normalizado predito versus real para *GMDH*



Fonte: Elaborado pelo autor.

Gráfico 5. Erro das camadas da configuração 114



Fonte: Gerado por *software* (adaptado).

Quadro 21. Valores da configuração 114 do *GMDH*

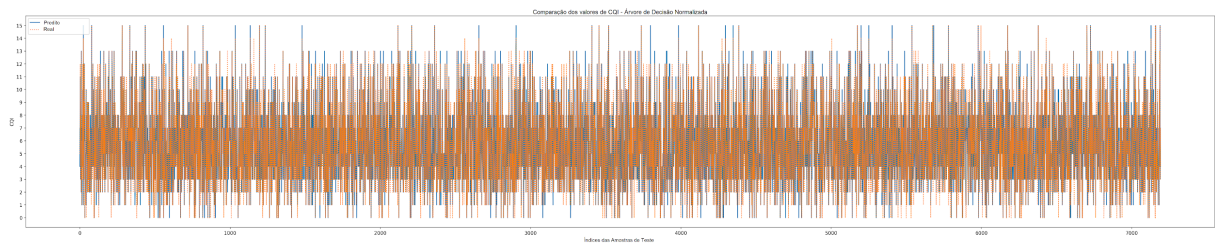
Parâmetro	Valor
Critério de seleção dos melhores nós	validate_bias
Função de referência	quadratic
Critério do cálculo do erro de camada	avg
Parâmetro de penalidade	0.001

Fonte: Elaborado pelo autor.

5.2. Análise comparativa dos modelos baseado nas métricas calculadas

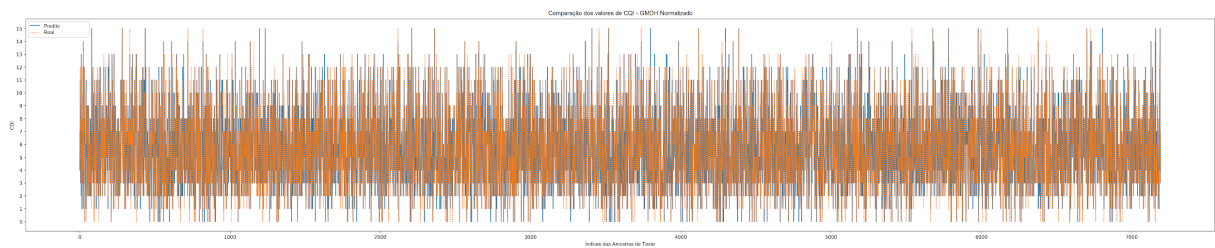
Cada configuração escolhida tem o mapeamento do *CQI* segundo os Gráficos 6 e 7, demonstrando semelhanças entre suas plotagens, de tal forma que o *GMDH* tem as linhas preditas mais próximas do real em relação à Árvore de Decisão. Isso também pode ser visto pela taxa de acerto dos modelos no Gráfico 8, com o *GMDH* apresentando maior taxa em comparação com a Árvore de Decisão, ou seja, 55,9% contra 54,01%.

Gráfico 6. Comparação dos valores do CQI para Árvore de Decisão



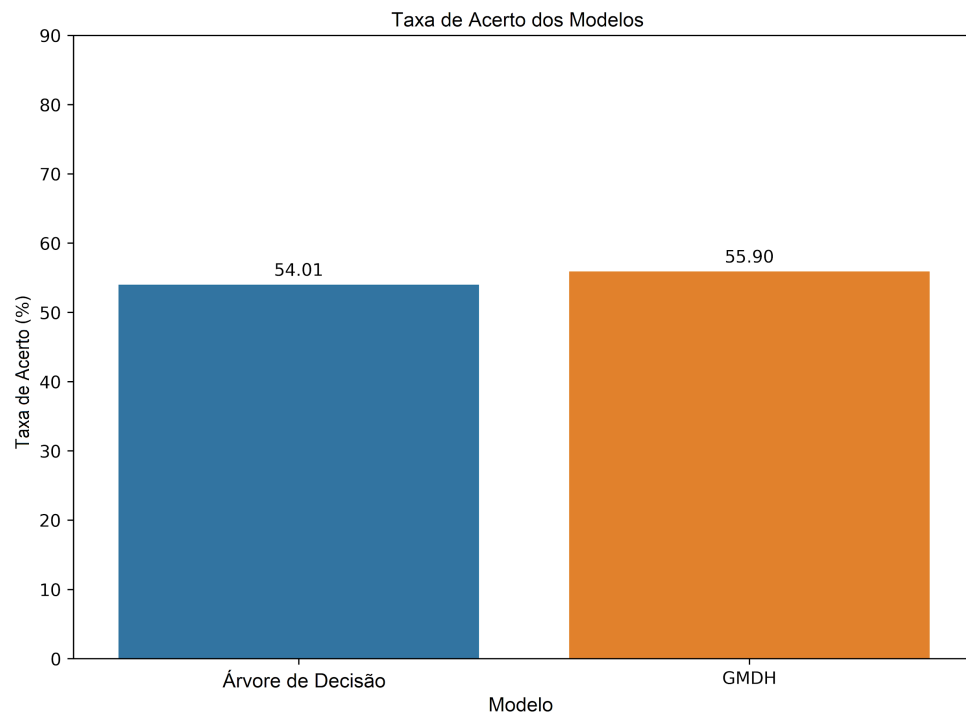
Fonte: Elaborado pelo autor.

Gráfico 7. Comparação dos valores do CQI para GMDH



Fonte: Elaborado pelo autor.

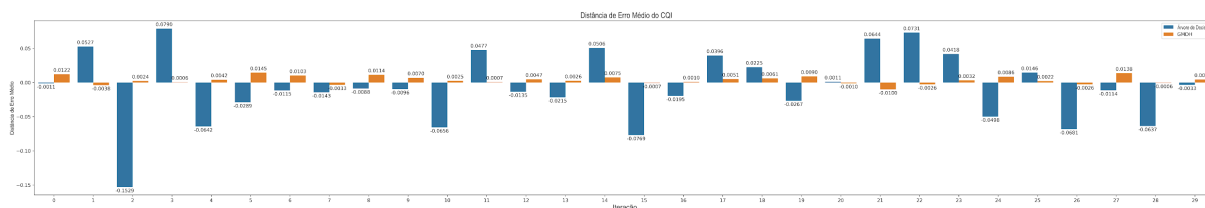
Gráfico 8. Taxa de acerto do CQI dos modelos



Fonte: Elaborado pelo autor.

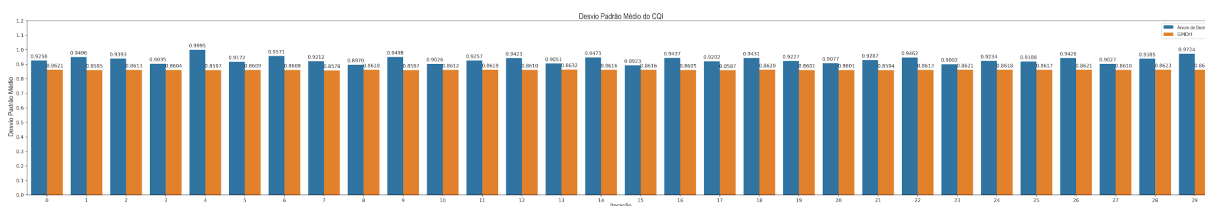
Na medição da distância média do erro de *CQI* segundo Gráfico 9, a Árvore de Decisão contém erro maior que *GMDH* em 26 das 30 iterações, tal qual o segundo tem erros menores que 0,02 absoluto para todas as iterações, enquanto o primeiro tem a mesma característica em apenas 11 iterações. Na medição do desvio padrão médio do *CQI*, o *GMDH* tem resultados sem muita variação por volta de aproximadamente 0,85 ao passo que a Árvore de Decisão tem erros maiores para todas as iterações com valores na faixa de aproximadamente 0,9 a 0,95 de acordo com o Gráfico 10. Na medição do *MSE* do *CQI* no Gráfico 11, o *GMDH* também exibe pouca variabilidade em cada iteração com resultados próximos de 0,75, em contrapartida com a Árvore de Decisão variando os resultados entre 0,8 e 1. Comportamento semelhante é percebido no Gráfico 12 com a medição do *RMSE* do *CQI*, exibindo valores para o *GMDH* em torno de 0,74 e para a Árvore de Decisão entre 0,8 e 1.

Gráfico 9. Distância média de erro do *CQI* dos modelos



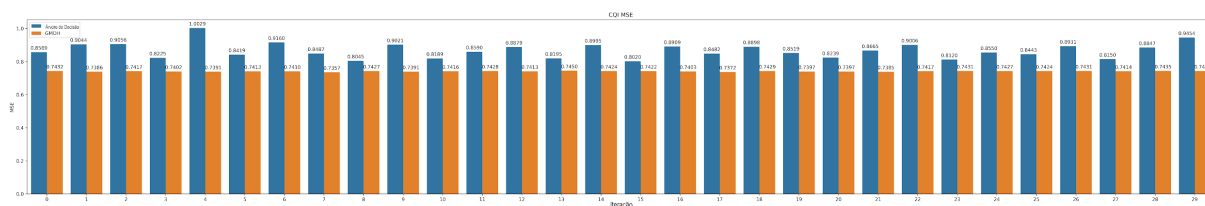
Fonte: Elaborado pelo autor.

Gráfico 10. Desvio padrão médio do *CQI* dos modelos



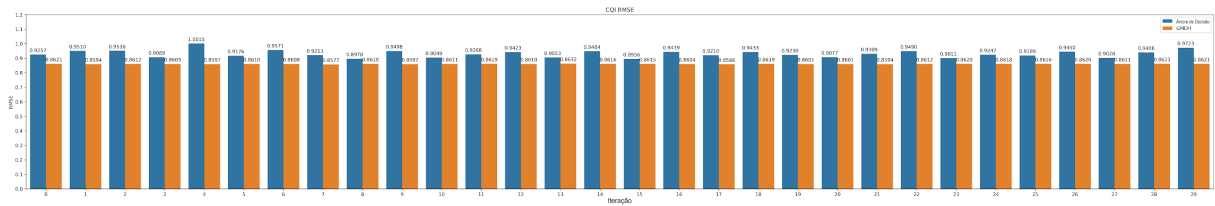
Fonte: Elaborado pelo autor.

Gráfico 11. *MSE* do *CQI* dos modelos



Fonte: Elaborado pelo autor.

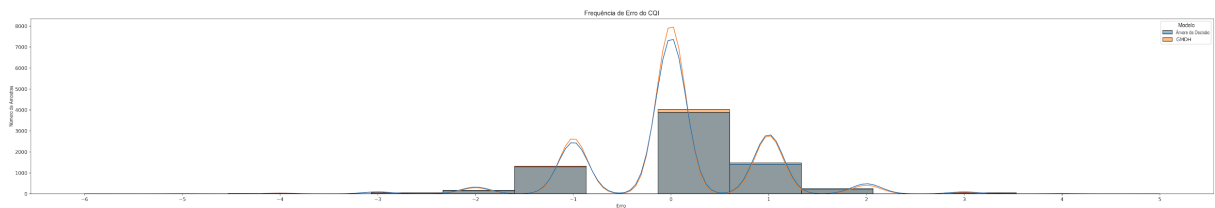
Gráfico 12. *RMSE* do *CQI* dos modelos



Fonte: Elaborado pelo autor.

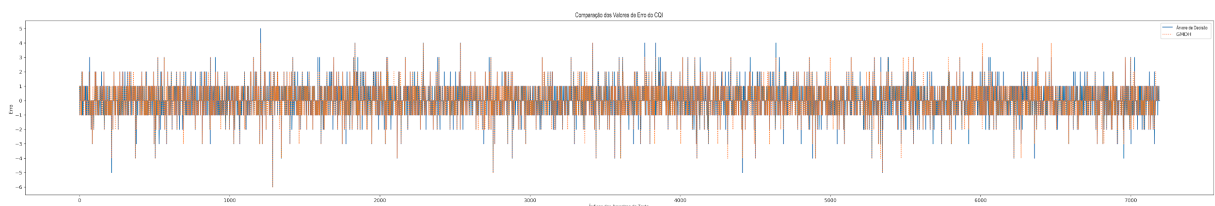
Na frequência dos erros do *CQI* pelo Gráfico 13, tem-se o número de amostras para *GMDH* por volta de 4400 e para a *Árvore de Decisão* em cerca de 4200 no ponto central de erro 0. No ponto do erro -1, o *GMDH* também se sobressai, enquanto para os outros pontos a *Árvore de Decisão* possui mais amostras. É válido ressaltar também essa característica pelo formato da distribuição nos dois modelos, tal que o *GMDH* tem maior densidade no ponto 0 e -1 em relação à *Árvore de Decisão*, embora nos outros pontos haja proximidade nas curvaturas de distribuição dos modelos. Na comparação dos valores de erro do *CQI* conforme Gráfico 14, há maior concentração dos resultados na faixa de -1 a 1 como também de maneira parcial para os outros intervalos, ocorrendo similarmente essa distribuição nos dois modelos. Outra característica é o ponto máximo atingido por ambos com valor de erro -6.

Gráfico 13. Frequência de erros do *CQI*



Fonte: Elaborado pelo autor.

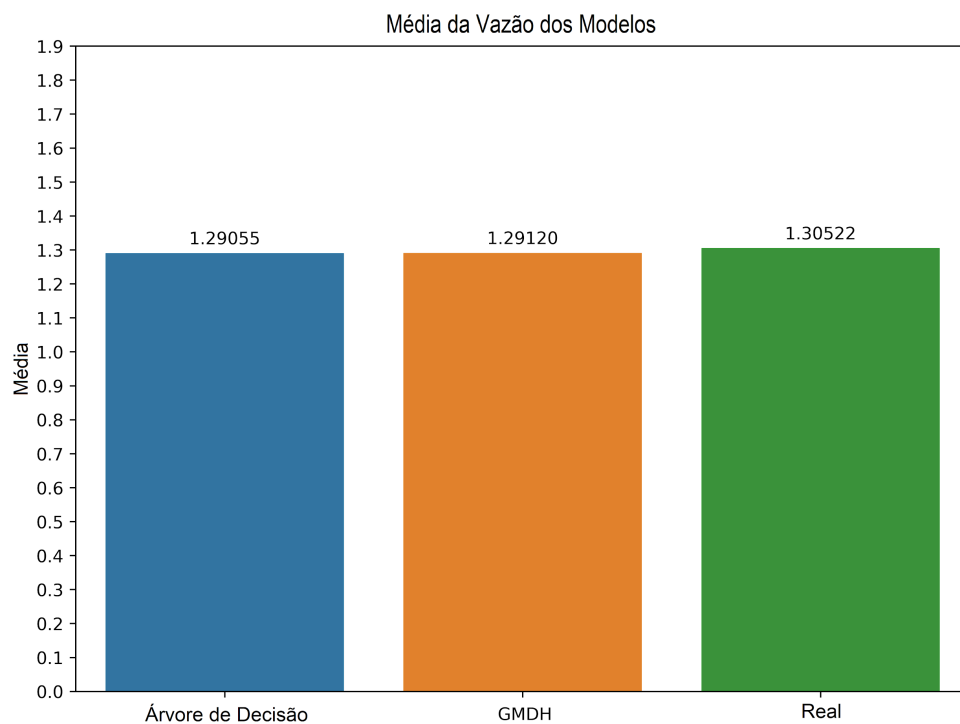
Gráfico 14. Comparação dos valores de erro do *CQI*



Fonte: Elaborado pelo autor.

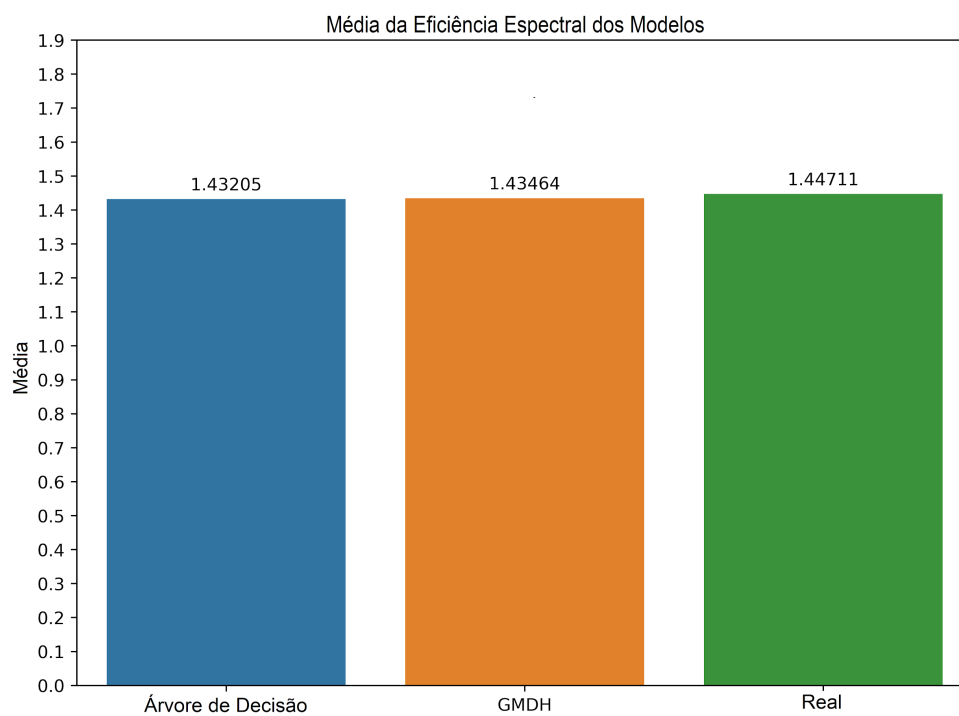
Na média total das métricas de vazão e eficiência espectral pelos Gráficos 15 e 16 respectivamente, o *GMDH* apresenta valor mais próximo comparado ao Real com diferença em cerca de 0,01 tanto para vazão quanto eficiência espectral, enquanto no desvio padrão total dessas métricas mostradas nos Gráficos 17 e 18 respectivamente, o *GMDH* também possui valor mais próximo ao Real mas com diferença em 0,03 aproximadamente tanto para vazão quanto eficiência espectral.

Gráfico 15. Média total da vazão dos modelos



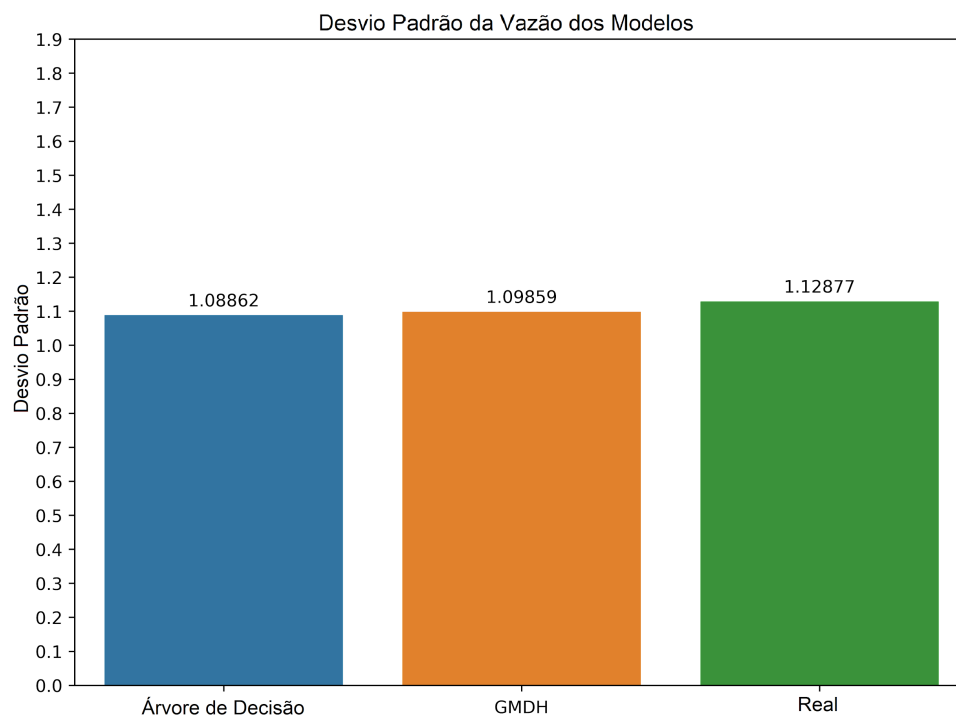
Fonte: Elaborado pelo autor.

Gráfico 16. Média total da eficiência espectral dos modelos



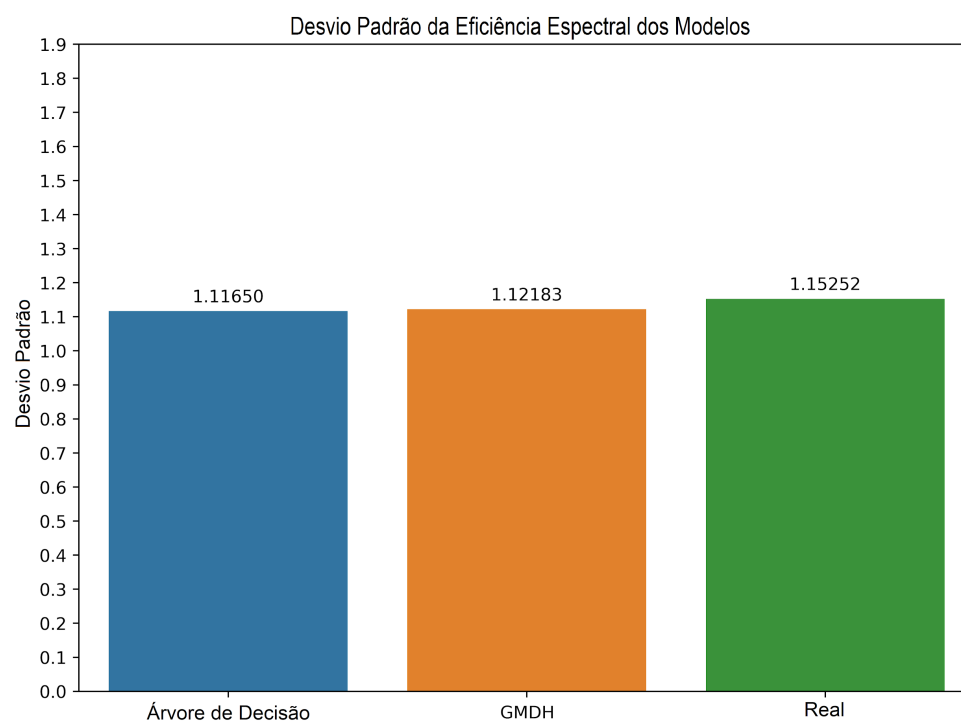
Fonte: Elaborado pelo autor.

Gráfico 17. Desvio padrão total da vazão dos modelos



Fonte: Elaborado pelo autor.

Gráfico 18. Desvio padrão total da eficiência espectral dos modelos



Fonte: Elaborado pelo autor.

6. Conclusões e Trabalhos Futuros

Este capítulo resume os principais pontos deste trabalho como também suas contribuições e considerações finais a respeito dos resultados obtidos e análises realizadas. Aponta-se, inclusive, direções sugestivas para trabalhos futuros envolvendo o tópico estudado.

6.1. Contribuições

No âmbito do problema do *CQI* desatualizado para redes 5G, foi proposto soluções baseadas em outros modelos de inteligência artificial de forma a serem comparados em busca de analisar suas performances e estabelecer qual apresentou melhores resultados a partir das métricas calculadas, como por exemplo, a vazão. Para isso, foi tratado a geração do dataset a ser utilizado no treinamento dos modelos, estabelecendo a escolha de colunas de interesse a partir de testes de correlação, sendo o método Spearman a ser considerado, o qual selecionou as colunas *SINR(dB)-Input*, *CalcClosestEnbDist-Input*, *CurYPos-Input*, *CurZPos-Input* e *SINR(dB)-Target*. Com isso, após treinamento dos modelos e seleção das melhores configurações de cada um, foi calculado a eficiência espectral a partir do *SINR(dB)* predito, e assim, mapeado o *CQI*, *MCS*, *TBS* e calculado a vazão, seguindo o mesmo fluxo para o dado alvo (*SINR(dB)-Target*) e permitindo a comparação dos resultados obtidos por cada modelo.

Por meio dos resultados adquiridos, percebe-se que apesar da semelhança do mapeamento do *CQI* ocasionado pelos modelos e da diferença da taxa de acerto entre eles ser de 1,89%, a Árvore de Decisão obteve distância média e frequência de erros do *CQI* maiores em relação ao *GMDH*, o qual apresentou erro da distância média aproximadamente entre +0,02 e -0,02 na maioria de suas iterações, caracterizando também, valores médios de vazão e eficiência espectral mais próximos do referencial alvo. Logo, o *GMDH* demonstrou melhor desempenho comparado à Árvore de Decisão, contribuindo, assim, para que em abordagens específicas sobre esse modelo sejam mais exploradas na resolução do problema discutido.

6.2. Trabalhos Futuros

Como indicação de possíveis propostas a serem trabalhos futuros baseados nos resultados gerados neste trabalho e métodos utilizados, tem-se:

- 1) Comparar resultados gerados a partir da aplicação de diferentes janelas de atraso (*delay*) na geração do *dataset*.
- 2) Usar os resultados preditos de $SINR(\text{dB})$ para realizar mapeamento do CQI e cálculo das métricas por meio de um simulador, além de comparar com os resultados deste trabalho.
- 3) Aplicar os resultados dos modelos treinados para outros esquemas de mapeamento do CQI , e conseqüentemente de modulação e codificação do 5G.
- 4) Aprimorar a taxa de acerto do modelo que apresentou melhor performance, ou seja, o *GMDH*.

7. Referências

- [1] QUEIROZ, A. *et. al.* Desafios no desenvolvimento de redes 5G. **GTA/UFRJ**. Disponível em: https://www.gta.ufrj.br/ensino/eel879/trabalhos_vf_2015_2/5G/desafios.html. Acesso em: 18 dez. 2022.
- [2] 5G: O que é 5G, suas características e aplicações. **Embarcados**. Disponível em: <https://embarcados.com.br/o-que-e-5g-caracteristicas-aplicacoes/>. Acesso em: 20 dez. 2022.
- [3] COUTINHO, Pedro Smith. Rádios Cognitivos. **GTA/UFRJ**. Disponível em: [https://www.gta.ufrj.br/ensino/eel879/trabalhos_vf_2008_2/coutinho/OFDM.OFDMA\(OrthogonalFrequencyDivisionMultiplexing\).html](https://www.gta.ufrj.br/ensino/eel879/trabalhos_vf_2008_2/coutinho/OFDM.OFDMA(OrthogonalFrequencyDivisionMultiplexing).html). Acesso em: 20 dez. 2022.
- [4] SENADO FEDERAL. Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática. **RELATÓRIO DE AVALIAÇÃO DAS POLÍTICAS PÚBLICAS RELATIVAS À IMPLANTAÇÃO DAS REDES MÓVEIS DE QUINTA GERAÇÃO (5G)**. Disponível em: https://www12.senado.leg.br/noticias/arquivos/2022/09/21/relatorio_redes.pdf. Acesso em: 20 dez. 2022.
- [5] ALGAR TELECOM. Rede 5G: o que é, quanto custa e como vai funcionar a nova tecnologia?. **Blog Conexão Algar Telecom**. Disponível em: <https://blog.algartelecom.com.br/inovacao/o-que-e-rede-5g/>. Acesso em: 3 mar. 2023.
- [6] MOREIRA, Rômulo Diego. Entenda a diferença entre 5G SA, 5G NSA e 5G DSS”. **TechTudo**. Disponível em: <https://www.techtudo.com.br/noticias/2022/08/entenda-a-diferenca-entre-5g-sa-5g-nsa-e-5g-dss.ghml>. Acesso em: 4 mai. 2023.
- [7] BRAGA, Lucas. Afinal, o 5G DSS que temos no Brasil. **TecnoBlog**. Disponível em: <https://tecnoblog.net/responde/afinal-o-5g-dss-que-temos-no-brasil-e-5g-de-verdade/>. Acesso em: 4 mai. 2023.
- [8] MUNIZ, C. M.; LOPES, E. M. **Backhaul móvel para 5G**. INSTITUTO NACIONAL DE TELECOMUNICAÇÕES – INATEL, 2020. Disponível em: <https://biblioteca.inatel.br/cict/acervo%20publico/Sumarios/Artigos%20de%20TCC/TCC>

_Pos%20Gradua%C3%A7%C3%A3o/SRST-%20Engenharia%20de%20Redes%20e%20Sistemas%20de%20Telecomunica%C3%A7%C3%B5es/2020/TCCRedesParticipantesdoSemin%C3%A1rio/TCC_Backhaul_m%C3%B3vel_para_5G.pdf. Acesso em: 6 mar. 2023.

- [9] CARVALHO, M. S.; ABBADE, A. L. R. **Descrição do backhaul para as redes 5G**. INSTITUTO NACIONAL DE TELECOMUNICAÇÕES – INATEL, 2020. Disponível em:
https://biblioteca.inatel.br/cict/acervo%20publico/Sumarios/Artigos%20de%20TCC/TCC_Pos%20Gradua%C3%A7%C3%A3o/SRST-%20Engenharia%20de%20Redes%20e%20Sistemas%20de%20Telecomunica%C3%A7%C3%B5es/2020/TCCRedesParticipantesdoSemin%C3%A1rio/TCC_Descri%C3%A7%C3%A3o_do_backhaul_para_as_redes_5G.pdf. Acesso em: 6 mar. 2023.
- [10] Introdução ao 5G Core. **LABCOM**. Disponível em:
<https://labcom.ufsc.br/2021/04/24/introducao-ao-5g-core/>. Acesso em: 6 mar. 2023.
- [11] ALMEIDA, V. M.; OLIVEIRA, F. C. **Características de uma rede 5G**. INSTITUTO NACIONAL DE TELECOMUNICAÇÕES – INATEL, 2020. Disponível em:
https://biblioteca.inatel.br/cict/acervo%20publico/Sumarios/Artigos%20de%20TCC/TCC_Pos%20Gradua%C3%A7%C3%A3o/SRST-%20Engenharia%20de%20Redes%20e%20Sistemas%20de%20Telecomunica%C3%A7%C3%B5es/2020/TCCRedesParticipantesdoSemin%C3%A1rio/TCC_Caracter%C3%ADsticas_de_uma_rede_5G.pdf. Acesso em: 6 mar. 2023.
- [12] OLIVEIRA, L. A. N.; ALENCAR, M. S.; LOPES, W. T. A.; Evolução da Arquitetura de Redes Móveis Rumo ao 5g. **RTIC - Revista de Tecnologia da Informação e Comunicação**, v. 8, n. 2, 2018. Disponível em:
https://d1wqtxts1xzle7.cloudfront.net/81241311/102-libre.pdf?1645547356=&response-content-disposition=inline%3B+filename%3DEvolucao_da_Arquitetura_de_Redes_Moveis.pdf&Expires=1677808736&Signature=eBUCiZbeTcL3WTxO1nd6tClaJpT8gdU6fqNDRFWl7ZKvtSMcI0ywOyxkFA73KAvVeX1dGnRIp3zKF4NSlfVIYnIcg603OAavqBb3TxKNkfBBSg6GYU4HMJEzzi3kZ4CoaaKr~pv7VgZjfYNRKH86NR-juzsvQi9JSSFnMT26srSTe-s2Q1e9Rb5ck3iXUJvJbiG2Ippyq3MBRpxYmzVPe6EMUTwJ6kVIwi4bcfHJNlz hXyKVzkpegVkJxftwLLaOnubE51N7cjpFRdEao7yruDxGSyKc2DI8cy6ehCm3s4jj1lqR40n6uMkK1Y2B~pH5fC7jEn6G1VwhL9Q1Bz0wQ__&Key-Pair-Id=APKAJLOHF5GG

SLRBV4ZA. Acesso em: 3 mar. 2023.

- [13] FLORENCIO, Paulo. Conceitos básicos de Redes de Telecomunicações. **Target**. Disponível em:
<https://www.targetso.com/2017/03/09/redes-de-telecomunicacoes-conceitos-basicos/>. Acesso em: 2 mar. 2023.
- [14] PANA, V. S.; BABALOLA, O. P.; BALYAN, V. 5G radio access networks: A survey. **Array**, v. 14, 2022. Disponível em:
https://www.sciencedirect.com/science/article/pii/S2590005622000315?ref=pdf_download&fr=RR-2&rr=80bf783a7c47264c. Acesso em: 2 jul. 2023.
- [15] 5G OFDM Technology. **5GHub**. Disponível em:
<https://5ghub.us/5g-ofdm-technology/>. Acesso em: 30 jul. 2023.
- [16] 5G/NR - Frame Structure. **ShareTechnote**. Disponível em:
https://www.sharetechnote.com/html/5G/5G_FrameStructure.html. Acesso em: 12 mar. 2023.
- [17] 5G/NR - FR/Operating Bandwidth. **ShareTechnote**. Disponível em:
https://www.sharetechnote.com/html/5G/5G_FR_Bandwidth.html. Acesso em: 12 mar. 2023.
- [18] 5G/NR - MCS/TBS/Code Rate. **ShareTechnote**. Disponível em:
https://www.sharetechnote.com/html/5G/5G_MCS_TBS_CodeRate.html. Acesso em: 12 mar. 2023.
- [19] BUNN, Marcus Vinicius. **Controle de Modulação e Codificação Adaptativo Utilizando GNU Radio**. Instituto Federal de Santa Catarina – IFSC, 2018. Disponível em: https://wiki.sj.ifsc.edu.br/images/b/b3/TCC290_Marcus_Vinicius_Bunn.pdf. Acesso em: 12 ago. 2023.
- [20] TECHNICAL SPECIFICATION: 5G; NR; Physical channels and modulation (3GPP TS 38.211 version 16.2.0 Release 16). **ETSI**, 2020. Disponível em:
https://www.etsi.org/deliver/etsi_ts/138200_138299/138211/16.02.00_60/ts_138211v1602

00p.pdf. Acesso em: 13 mar. 2023.

- [21] LIMA, Samuel Davi de. **ANÁLISE E SIMULAÇÃO DOS ASPECTOS DE TRANSMISSÃO DE SINAIS 5G EM MACRO CÉLULAS EM AMBIENTES URBANOS**. UNIVERSIDADE FEDERAL DE UBERLÂNDIA, 2023. Disponível em: <https://repositorio.ufu.br/bitstream/123456789/38645/1/An%C3%A1liseSimula%C3%A7%C3%A3oAspectos.pdf>. Acesso em: 17 set. 2023.
- [22] TECHNICAL SPECIFICATION: 5G; NR; Physical layer procedures for data (3GPP TS 38.214 version 16.2.0 Release 16). ETSI, 2020. Disponível em: https://www.etsi.org/deliver/etsi_ts/138200_138299/138214/16.02.00_60/ts_138214v160200p.pdf. Acesso em: 18 set. 2023.
- [23] A. Balieiro, K. Dias and P. Guarda, “A Machine Learning Approach for CQI Feedback Delay in 5G and Beyond 5G Networks,” 30th Wireless and Optical Communications Conference (WOCC), 2021, pp. 26-30, doi: 10.1109/WOCC53213.2021.9603019.
- [24] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. DaiJ, “A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions,” IEEE Communications Surveys & Tutorials, vol. 20, 2018, pp.3098--3130.
- [25] R. A. Akl, S. Valentin, G. Wunder and S. Stánczak, “Compensating for CQI aging by channel prediction: The LTE downlink,” IEEE Global Communications Conference (GLOBECOM), 2012, pp. 4821-4827, doi:10.1109/GLOCOM.2012.6503882.
- [26] H. Yin, X.Guo, P. Liu, X. Hei and Y. Gao, “Predicting Channel Quality Indicators for 5G Downlink Scheduling in a Deep Learning Approach,” arXiv 2008.01000, 2020.
- [27] Mingjian Ni, Xiang Xu and R. Mathar, “A channel feedback model with robust SINR prediction for LTE systems,” 7th European Conference on Antennas and Propagation (EuCAP) pp. 1866-1870, 2013.
- [28] G. Bartoli and D. Marabissi, “CQI prediction through recurrent neural network for UAV control information exchange under URLLC regime,” IEEE Transactions on

Vehicular Technology, 2022, doi: 10.1109/TVT.2022.3152408.

- [29] MARCO, M. *et. al.* End-to-End Simulation of 5G mmWave Networks. **IEEE**. Disponível em: <https://ieeexplore.ieee.org/document/8344116>. Acesso em: 23 set. 2023.
- [30] SACRAMENTO, Gabriel. ÁRVORE DE DECISÃO: ENTENDA ESSE ALGORITMO DE MACHINE LEARNING. **TeraBlog**. Disponível em: <https://blog.somostera.com/data-science/arvores-de-decisao>. Acesso em: 23 set. 2023.
- [31] ABREU, Leonardo Evangelista de. **PEOPLE ANALYTICS: USO DE ÁRVORES DE DECISÃO NA RETENÇÃO DE TALENTOS**. UNESP, 2022. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/217419/abreu_le_tcc_prud.pdf?sequence=6&isAllowed=y. Acesso em: 23 set. 2023.
- [32] GUPTA, Prashant. Decision Trees in Machine Learning. **Towards Data Science**, 2017. Disponível em: <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>. Acesso em: 14 mar. 2023.
- [33] Árvore de decisão. **PUC-Rio**. Disponível em: https://www.maxwell.vrac.puc-rio.br/3710/3710_4.PDF. Acesso em: 23 set. 2023.
- [34] Decision Trees. **Scikit Learn**. Disponível em: <https://scikit-learn.org/stable/modules/tree.html#minimal-cost-complexity-pruning>. Acesso em: 23 set. 2023.
- [35] BUENO, Elaine Inacio. Group Method of Data Handling (GMDH) e Redes Neurais na Monitoração e Detecção de Falhas em sensores de centrais nucleares. **IPEN**, 2011. Disponível em: <https://teses.usp.br/teses/disponiveis/85/85133/tde-02092011-140535/publico/2011BuenoGroup.pdf>. Acesso em: 20 set. 2023.
- [36] SHEN, C. *et. al.* Group Method of Data Handling (GMDH) Lithology Identification Based on Wavelet Analysis and Dimensionality Reduction as Well Log Data

Pre-Processing Techniques. **Energies**, 2019. Disponível em:
<https://www.mdpi.com/1996-1073/12/8/1509>. Acesso em: 20 set. 2023.

[37] ANASTASAKIS, L.; MORT, N. **THE DEVELOPMENT OF SELF-ORGANIZATION TECHNIQUES IN MODELLING: A REVIEW OF THE GROUP METHOD OF DATA HANDLING (GMDH)**. Department of Automatic Control & Systems Engineering, The University of Sheffield, 2001. Disponível em:
https://gmdhsoftware.com/GMDH_%20Anastasakis_and_Mort_2001.pdf. Acesso em: 14 mar. 2023.

[38] Multilayered iterational algorithm of the Group Method of Data Handling. **Github**. Disponível em: <https://github.com/kvoyager/GmdhPy>. Acesso em: 14 mar. 2023.

Apêndice

Plotagem da Árvore de Decisão da configuração 42.

