



Aplicação de machine learning e deep learning para modelagem de uma bacia hidrográfica

Application of machine learning and deep learning for modeling a water basin

Aplicación de aprendizaje máquina y aprendizaje profundo para modelar una cuenca de agua

SILVA, Vinícius de Azevedo¹
SOUZA, Rafael de Feiras²
OLIVEIRA, Mateus Peixoto³
SANTOS, Francisco Lledo dos⁴

¹ Universidade do Estado do Rio de Janeiro, Programa de Pós-Graduação em Rede Nacional em Gestão e Regulação de Recursos Hídricos, Departamento de Tecnologia e Ciências, Barra Mansa, Rio de Janeiro, Brasil.
vinicius.silva@uerj.profagua.br
ORCID ID: 0009-0009-9008-4095

² Universidade de São Paulo, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Departamento de Administração, Economia e Contabilidade, Ribeirão Preto, São Paulo, Brasil.
fsrafael@usp.br
ORCID ID: 0000-0001-6298-5325

³ Universidade Federal Fluminense, Escola de Engenharia Industrial Metalúrgica de Volta Redonda, Programa de Pós-Graduação em Tecnologia Ambiental, Volta Redonda, Rio de Janeiro, Brasil.
mateuspo@id.uff.br
ORCID ID: 0000-0001-5385-5895

⁴ Universidade do Estado de Mato Grosso, Faculdade de Arquitetura e Engenharia, Programa de Pós-graduação em Rede de Gestão e Regulação de Recursos Hídricos Cuiabá, Mato Grosso, Brasil.
franciscolledo@unemat.br
ORCID ID: 0000-0002-7718-8203

Recebido em 07/03/2023. Aceito em 12/07/2023



Resumo

O desenvolvimento econômico associado a aspectos como a desigualdade social são processos que provocam a deterioração de recursos ambientais. São emergenciais o aperfeiçoamento das tecnologias para combater os efeitos dos desastres naturais, como a melhoria dos processos de transformação de chuva em vazão e da modelagem dos fenômenos hidrológicos em bacias hidrográficas. O objetivo desta comunicação científica foi obter, com o auxílio de técnicas de *machine learning*, previsões do comportamento hidrológico na Bacia do Rio Bananal. Essas informações reunidas foram submetidas aos seguintes modelos: redes neurais artificiais de aprendizado profundo (*Deep Learning* - *InceptionTime*, *Resnet*, FCN e ResCNN) e de aprendizado de máquina (*Machine Learning* - regressão linear, árvore de decisão e *support-vector machine*), e *gradient boosting* para levantar séries históricas consistentes e comparar os padrões de precipitação e da vazão (modelo chuva-vazão) do curso hídrico. Foram utilizadas as séries temporais médias diárias de 2002 a 2014 para treinamento a fim de estabelecer um comportamento otimizado para 2015. Foi possível concluir que o coeficiente R^2 com a melhor performance foi da arquitetura *InceptionTime*, com RMSE de 4.62, MAE de 2.36. O resultado com amior desempenho foi a árvore de decisão com R^2 de 0.73.

Palavras-Chave: *Deep Learning*, inundações, chuva-vazão, modelagem, *InceptionTime*.

Abstract

Economic development associated with aspects such as social inequality is a process that leads to the deterioration of environmental resources. It is essential to improve technologies to combat the effects of natural disasters, such as enhancing the processes of rainfall-to-runoff transformation and modeling hydrological phenomena in river basins. The objective of this scientific communication was to obtain hydrological behavior predictions in the Bananal River Basin using machine learning techniques. The gathered information was subjected to the following models: deep learning artificial neural networks (InceptionTime, Resnet, FCN, and ResCNN) and machine learning models (linear regression, decision tree, and support-vector machine), as well as gradient boosting, to analyze consistent historical data and compare precipitation patterns and streamflow (rainfall-runoff model) in the watercourse. Daily average time series data from 2002 to 2014 were used for training to establish optimized behavior for 2015. It was concluded that the architecture with the highest performance, based on the coefficient of determination (R^2), was InceptionTime, with an RMSE of 4.62 and MAE of 2.36. The decision tree model achieved the best overall performance with an R^2 of 0.73.

Key-Words: *Deep Learning*, floods, rain-runoff, modeling, *InceptionTime*.

Resumen

El desarrollo económico asociado con aspectos como la desigualdad social es un proceso que provoca la deterioración de los recursos ambientales. Es urgente mejorar las tecnologías para combatir los efectos de los desastres naturales, como el mejoramiento de los procesos de lluvia en caudal y la modelización de los fenómenos hidrológicos en cuencas hidrográficas. El objetivo fue obtener, mediante técnicas de aprendizaje automático, pronósticos del comportamiento hidrológico en la Cuenca del Río Bananal. Esta información recopilada se sometió a los siguientes modelos: redes neuronales artificiales de aprendizaje profundo (Deep Learning - InceptionTime, Resnet, FCN y ResCNN) y modelos de aprendizaje automático (Machine Learning - regresión lineal, árbol de decisión y máquina de vectores de soporte), así como impulso de gradiente para generar series históricas consistentes y comparar los patrones de precipitación y flujo (modelo de lluvia-caudal) del curso del agua. Se utilizaron las series temporales promedio diarias de 2002 a 2014 para el entrenamiento, con el fin de establecer un comportamiento optimizado para 2015. Se pudo concluir que el mejor rendimiento, según el coeficiente R^2 , fue InceptionTime, con un RMSE de 4,62 y un MAE de 2,36. El modelo de árbol de decisión obtuvo el mejor R^2 de 0,73.

Palabras clave: *Deep Learning*, inundaciones, precipitaciones, modelado, *InceptionTime*.

1. Introdução

Os eventos naturais catastróficos, também designados como desastres naturais, têm se tornado cada vez mais comuns e severos quanto à magnitude dos danos causados. Tais ocorrências impactam a saúde dos indivíduos, sua qualidade de vida, além de toda a esfera ambiental, que por sua vez, recebe não somente consequências instantâneas, como também de médio e longo prazos (FREITAS et al., 2014).

Dentre os desastres naturais em áreas urbanas próximas a cursos d'água, destacam-se as enchentes e inundações, provocadas por uma série de fatores socioambientais, antrópicos e climatológicos. Uma das principais causas do extravasamento das calhas fluviais é a ocupação desordenada das denominadas Áreas de Preservação Permanente (APP), as quais neste caso, correspondem às chamadas Faixas Marginais de Proteção (FMP) dos corpos hídricos, isto é, extensões territoriais que margeiam os ambientes aquáticos naturais estabelecidas e controladas por diversas leis e destinadas à proteção e manutenção da qualidade ambiental e ecológica dos recursos hídricos e dos ecossistemas atrelados a eles. A presença de elementos urbanos nestes espaços, como moradias e empreendimentos, caracteriza o desencadeamento de eventos extremos prejudiciais ao coletivo, tais como o aumento da probabilidade de ocorrência de enchentes e inundações (LIMA; AMORIM, 2015).

Episódios como estes causam inúmeras perdas humanas e econômicas, que se mostraram 23% mais frequentes em 2020 em relação às últimas décadas, gerando prejuízos na ordem de bilhões. Nesta conjuntura, o uso de tecnologias e ferramentas capazes de tratar informações geoespaciais otimizadas e oferecer panoramas altamente precisos e representativos é fundamental no combate à destruição ocasionada pelos fenômenos naturais agravados pela ação humana. As chamadas geotecnologias como o Sistema de Posicionamento Global (GPS), o sensoriamento remoto e o Sistema de Informação Geográfica (SIG) constituem parte dos instrumentos de geoprocessamento cada vez mais difundidos no gerenciamento dos territórios em diversos domínios: econômico, social, de segurança e ambiental (ROSA, 2011). Pode-se verificar os impactos advindos de ondas de cheias e das inundações e alagamentos, nas Figuras 1 e 2.

Figura 1: Entrada que dá acesso aos bairros Colônia Santo Antônio, Vila Ursulino e Santa Maria II, no bairro Bocaininha, no município de Barra Mansa/RJ.



Fonte: G1 Globo, Rio de Janeiro, May 19, 2019. Disponível em: <https://g1.globo.com/rj/sul-do-rio-costa-verde/noticia/2019/05/19/barra-mansa-registra-136-mm-de-chuva-e-deixa-desalojados-apos-rio-transbordar.ghtml>. Acesso em: 5 Set, 2023.

Figura 2: Alagamento no bairro Boa Sorte, no município de Barra Mansa/RJ.



Fonte: G1 Globo, Rio de Janeiro, May 19, 2019. Disponível em: <https://g1.globo.com/rj/sul-do-rio-costa-verde/noticia/2019/05/19/barra-mansa-registra-136-mm-de-chuva-e-deixa-desalojados-apos-rio-transbordar.ghtml>. Acesso em: 5 Set, 2023.

Para estudo da hidrologia, bem como seu monitoramento e atualização de informações temporais das modificações ocorridas no território, o sensoriamento remoto se apresenta como uma ferramenta bastante útil e com capacidade temporal e espacial satisfatórias a obtenção de dados. Por meio de imagens com diferentes variações espectrais é possível verificar o comportamento da vegetação que, conforme os autores, influencia diretamente no fluxo da água entre atmosfera e superfície terrestre, afetando sua qualidade e disponibilidade para usos múltiplos em ambientes urbanos (BRITTO, BAPTISTA e LIMA, 2019).

Contudo, muitas das vezes, a utilização pontual das geotecnologias se mostra pouco efetiva na prevenção e mitigação de impactos ambientais futuros e no auxílio ao planejamento estratégico de riscos e à gestão dos recursos naturais. Logo, é essencial a atribuição de ferramentas geoespaciais combinadas a investigações temporais e integradas a técnicas computacionais mais inovadoras como as Redes Neurais Artificiais (RNA), que consistem na avaliação inteligente e versátil de dados capazes de oferecer respostas ainda mais apuradas e aperfeiçoadas (SOUSA et al., 2017; GU et al., 2019).

Segundo MENDONÇA et al a aplicação de técnicas de RNA para previsão hidrológica na região Amazônica, permitiu elaborar modelos chuva-vazão com dados de uma Sub-bacia Hidrográfica do Rio Guamá, no Estado do Pará, Amazônia. A seleção dos vetores de entrada através de coeficientes de correlações linear, cruzada, autocorrelação e autocorrelação parcial mostrou-se ser uma alternativa interessante, possibilitando identificar quais estações pluviométricas e defasagens nas séries temporais são relevantes à simulação das vazões.

Luiz e Schröde utilizaram estruturas baseadas em aprendizagem de máquinas (Machine Learning) para calibrar dois modelos de chuva-vazão em escala diária na Bacia Hidrográfica do Arroio Grande no leste do Rio Grande do Sul. Sendo eles: Redes Neurais Artificiais (RNA) e *Gradient Boosting Machine* (GBM) com a técnica *bootstrap* de reamostragem, onde foi possível observar que a técnica GBM apresentou-se levemente superior à de RNA.

A classificação do uso e cobertura do solo em unidades de planejamento territorial como em bacias hidrográficas é primordial na administração dos elementos integrantes da paisagem, bem como definição

de estratégias para evitar desastres, atenuar danos inerentes a catástrofes e determinar ações emergenciais necessárias (SILVA, 2022).

É necessário ressaltar que os métodos de determinação de vazão máxima de cheia em seções de controle pré-estabelecidas em bacias hidrográficas demandam o tratamento de uma grande quantidade de variáveis, como as características físicas da bacia, a intensidade de precipitação, coeficientes de escoamento superficial, tempo de concentração, declividade de canais, entre outras (MENDONÇA et al., 2021).

Um estudo foi realizado em três pontos de monitoramento de vazão da bacia do rio Piracicaba (Minas Gerais), para a previsão de vazões diárias. Durante o desenvolvimento das RNAs, os dados foram coletados e classificados, para posteriormente definir a arquitetura da rede. Após o aprendizado e validação das redes, a maioria das RNAs desenvolvidas apresentou coeficiente de *Nash-Sutcliffe* maior que 0,80, permitindo que os modelos sejam considerados válidos para estimativa de fluxo (ULIANA et al., 2018).

Em outro estudo, foi avaliada a simulação do comportamento hidrológico da bacia hidrográfica do Alto Canoas, em Santa Catarina, usando redes neurais artificiais *Multi Layer Perceptron* (MLP). Foram testados doze tratamentos com diferentes combinações de variáveis, como precipitação, evapotranspiração e vazão, a fim de determinar as melhores variáveis para modelar a vazão. A MLP foi treinada usando parte das vazões observadas. Além disso, as vazões foram simuladas em *open-loop* e *closed-loop* durante o período de teste. O tratamento com melhor desempenho utilizou a precipitação diária das quatro estações pluviométricas, com um tempo de resposta de -2 dias, juntamente com a vazão simulada do dia anterior. Apesar do baixo valor do erro médio quadrático obtido, a vazão modelada pela MLP tendeu a ser superestimada (DEBASTIANI; SILVA; NETO, 2016).

Sousa & Sousa (2010) modelaram a relação não-linear entre chuva e vazão na bacia hidrográfica do rio Piancó, no semiárido paraibano, utilizando Redes Neurais Artificiais (RNA). A capacidade da RNA em modelar o processo chuva-vazão em base mensal foi avaliada considerando a influência da arquitetura da rede e da inicialização dos pesos durante o treinamento. Após o treinamento, a melhor arquitetura de RNA foi selecionada com base no desempenho do modelo na estimativa de vazões médias mensais. A arquitetura RC315L obteve os melhores resultados, com coeficiente de determinação, eficiência e erro padrão da estimativa de 92,0, 77,0% e 8,29, respectivamente.

Neste contexto, a utilização de metodologias empíricas como os sistemas de chuva-vazão ou através de redes neurais possibilitam simplificar a elaboração de modelos matemáticos capazes de simular a vazão de cursos hídricos.

Dentro do cenário exposto, dada a falta ou divergência nos resultados do emprego de técnicas de aprendizado de máquina para modelos hidrológicos, o objetivo principal dessa comunicação científica foi avaliar quatro tipos de modelos e determinar qual ofereceu maior precisão na previsão do comportamento hidrológico do Rio Bananal junto à seção de controle determinada para análise.

2. Redes Neurais e Inteligência Artificial

A evolução tecnológica não somente no campo das geotecnologias, mas em inúmeras outras categorias, tem caminhado para a fusão de técnicas já existentes com os recursos computacionais atrelados ao conceito de Inteligência Artificial (AI), que engloba artifícios como a programação lógica indutiva e a árvore de aprendizado. Dentro desta concepção, um termo continuamente disseminado no âmbito da Inteligência Artificial é o denominado Machine Learning (ML), isto é, tornar possível o aprendizado de

máquinas com base no tratamento de dados, identificação de padrões lógicos e na organização e processamento de dados fundamentados em algoritmos (PACHECO; PEREIRA, 2018).

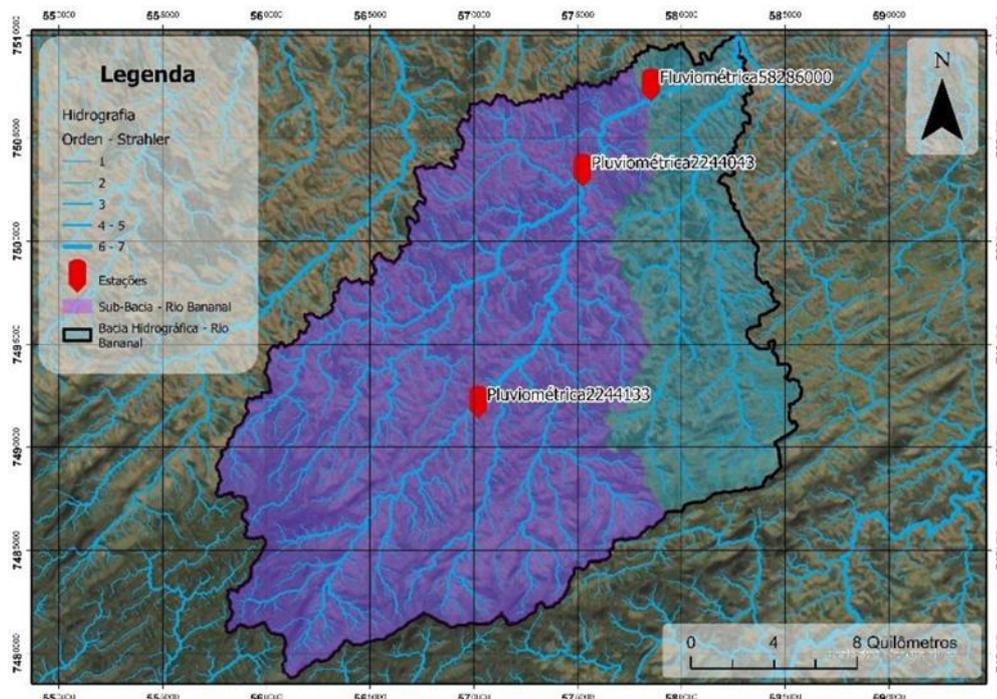
Deste modo, os programas computacionais conseguem aprender a executar ações orientadas no sequenciamento lógico dos algoritmos pressupostos. Ainda mais a fundo, deve-se mencionar a técnica mais aprimorada e complexa de Machine Learning, conhecida como Deep Learning ou aprendizagem profunda, isto é, uma ferramenta do tipo rede neural aplicada em diversos campos do conhecimento, como no processamento de dados não-estruturados e no processamento de imagens destinado ao reconhecimento de objetos/imagens e classificação de áreas (COPELAND, 2016 *apud* PACHECO; PEREIRA, 2018).

O estudo aqui tratado buscou empregar redes neurais típicas para análise de dados em séries temporais hidrometeorológicas de maneira automatizada, aplicando-se arquiteturas relativas aos conceitos de Machine Learning (regressão linear, árvore de decisão e *support-vector machine*) e *Deep Learning* (*InceptionTime*, ResCNN, FCN e *Resnet*), com o intuito de avaliar as previsões de dados ao longo do tempo e determinar suas eficiências.

3. Delimitação da Área de Estudo

O objeto de estudo do presente trabalho é a bacia hidrográfica do Rio Bananal. Respectivamente, as áreas de drenagem da bacia e de sua componente (sub-bacia) são de 521,29 e 389,08 km². A figura 3 exibe as poligonais das bacias estudadas e a localização das respectivas estações pluviométricas e fluviométrica escolhidas para o levantamento de dados.

Figura 3: Bacia/sub-bacia hidrográfica do Rio Bananal e postos de medição de dados de precipitação e vazão fluvial selecionados.



Localizada no município de Bananal/SP, a parte mais alta da bacia do Rio Bananal é caracterizada pela pequena suscetibilidade a extravasamentos na calha do rio. Por outro lado, o lado oposto (parte baixa), situada no município de Barra Mansa/RJ, sofre com recorrência os danos oriundos das inundações,

intensificados em virtude do grau de vulnerabilidade das comunidades nas planícies ribeirinhas.

Figura 4: Mancha de inundação na parte baixa da bacia do rio Bananal.



4. Metodologia

A aquisição dos dados relativos à série temporal de vazões e precipitação foi realizada junto à biblioteca da plataforma HIDROWEB e HydroBR, da Agência Nacional de Águas (ANA). Foram selecionadas duas estações pluviométricas na sub-bacia a montante da seção de controle, as quais foram coletados registros de precipitação média diária ao longo da série histórica. Estes foram utilizados para constituir as variáveis explanatórias. Além disso, selecionou-se uma estação fluviométrica a jusante, na bacia hidrográfica principal, na qual foram levantados os dados de vazão registrados no decorrer dos anos de operação (2002-2015). Foram contabilizados 5113 dados de entrada para cada posto pluviométrico selecionado e 4748 no posto fluviométrico. A série temporal de vazão consiste na variável preditiva/prevista.

A observação das séries temporais pluviométricas mostrou que há uma imensa quantidade de dias secos (sem chuva) ao longo dos anos de operação. Para contornar as lacunas de dados de precipitação, foi aplicada a técnica de *gradient boosting*, a qual promove a filtragem dos valores para articular o processamento dos dados somente nos dias chuvosos. O método proposto, mesmo que robusto, torna simples tal eliminação, mantendo as características da memória da ocorrência dos mesmos.

As figuras 5 e 7 apresentam as localizações dos postos pluviométricos e fluviométricos analisados preliminarmente para empregar o método de *gradient boosting*. Ao mesmo tempo, as figuras 6 e 8 demonstram os gráficos gerados pelo método, nos quais as marcações destacam os dados disponíveis referentes aos dias chuvosos nos respectivos anos. Análise pós processamento ao que se percebe, tendeu a uma clusterização intrínseca.

Figura 5: Localização das estações fluviométricas utilizadas na aquisição de dados temporais.

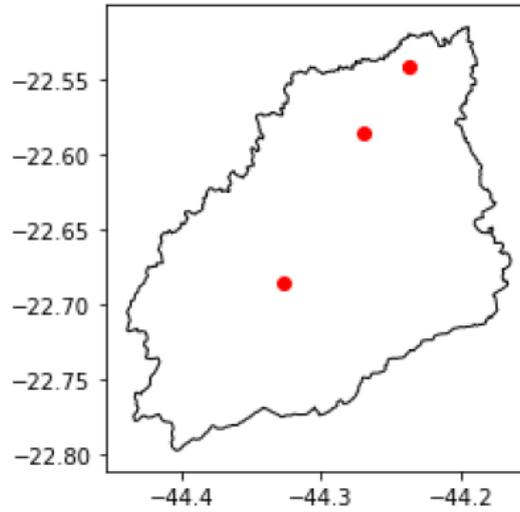


Figura 6: Gráfico da faixa de dados fluviométricos disponíveis após a filtragem *gradient boosting*.

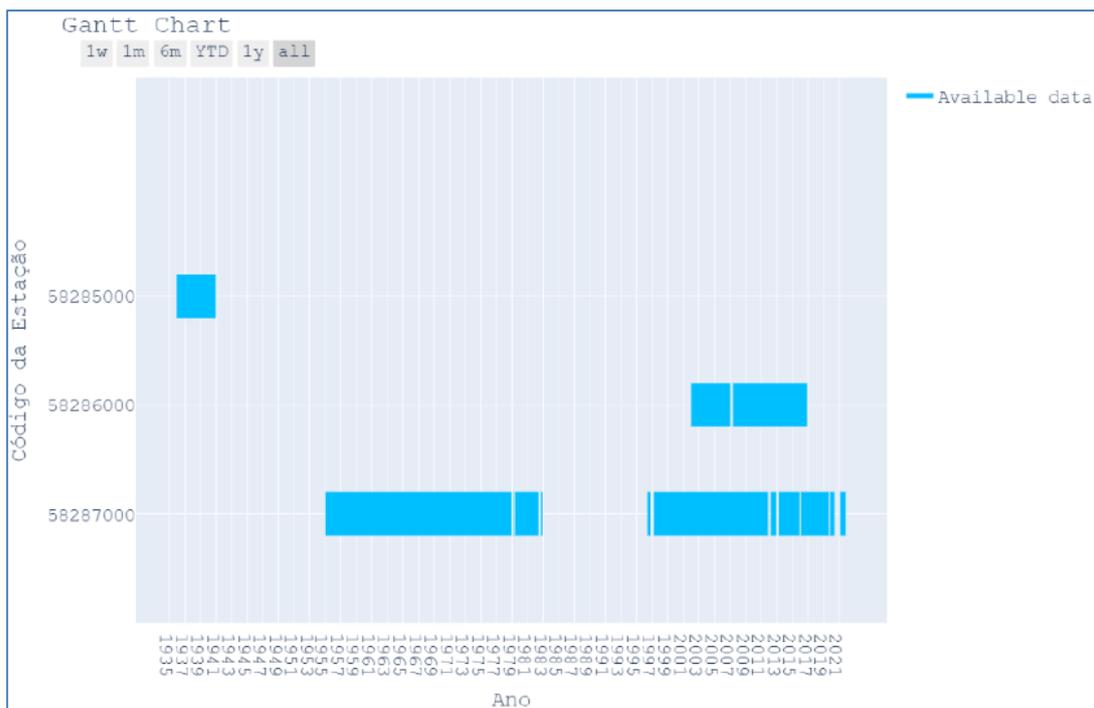
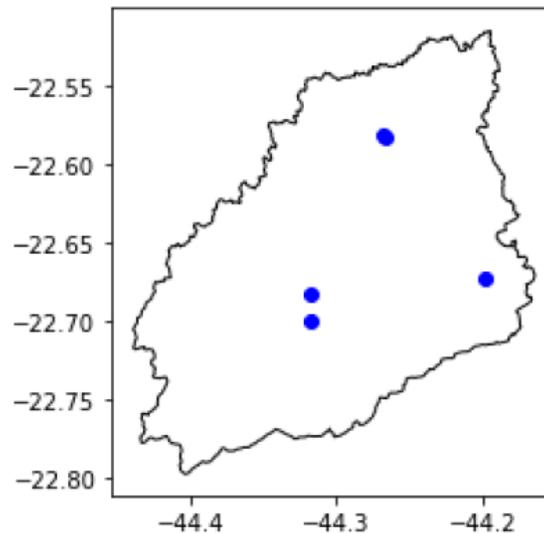
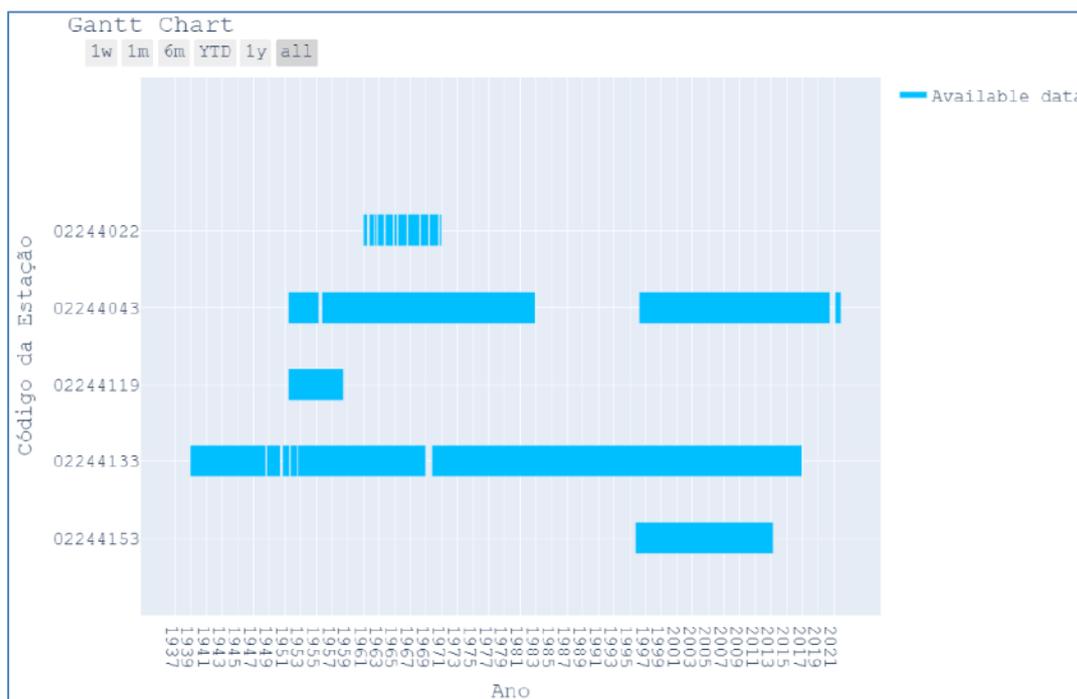


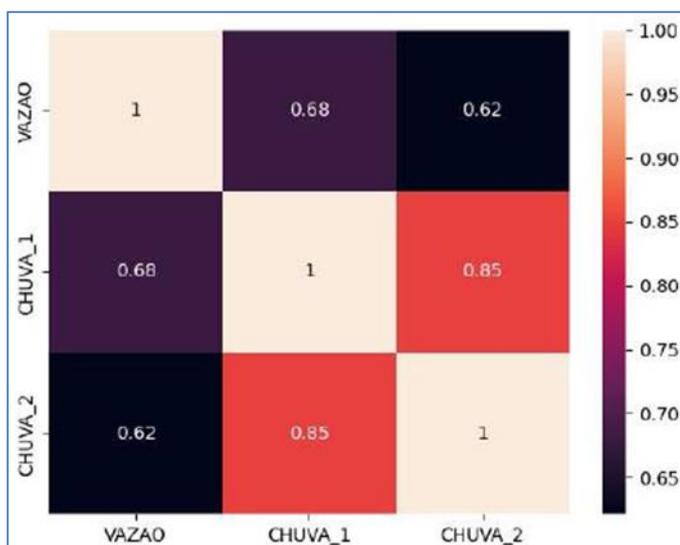
Figura 7: Localização das estações pluviométricas utilizadas na aquisição de dados temporais.**Figura 8:** Gráfico da faixa de dados pluviométricos disponíveis após a filtragem *gradient boosting*.

Anteriormente ao tratamento dos dados levantados, foram determinados os coeficientes de correlação dos mesmos, ou seja, valores que expressam o quão proporcionais ou padronizados entre si as variáveis avaliadas se apresentaram. A figura 9 retrata os valores de correlação matemática para chuva e vazão.

Para permitir a construção do modelo encarregado de processar os dados de entrada, foi aplicada a linguagem de programação Python na plataforma ArcGIS Pro, junto à interface interativa *Jupyter Notebook*. A utilização desta interface no ArcGIS Pro combina a funcionalidade do ambiente de desenvolvimento *Jupyter* com a capacidade de acessar dados e ferramentas geoespaciais do ArcGIS. Foi utilizado o *ArcGIS API for Python: Deep learning on time series*, que já possui os pacotes FCN, CNN

baseado na INceptionTime, ResCNN e ResNet pré-otimizado. Foi utilizada a biblioteca *pandas*, *numpy*, *sklearn.metrics*, *tensorflow* entre outras. Foi utilizado o pacote principal *TimeSeriesModel* que se integra às redes acima citadas. Foi optada a utilização dos parâmetros *default set* para evitar a criação de inúmeras arquiteturas. Como a estratégia de agrupamento de incorporação de pixels (*pooling: mean*), *epochs=none* – número de épocas otimizado (quando estabilizado), *early_stopping* = é necessária uma diferença mínima de 0.001 para que seja considerada uma melhoria. Em síntese, a utilização do ambiente Jupyter, aliado ao API (*Application Programming Interface*) do ArcGIS para Python, disponibiliza os modelos sejam de RNA's ou outras técnicas utilizadas nessa pesquisa, em ambiente colaborativo, sem a necessidade de aprimoramento das técnicas utilizadas, sendo as mesmas já em modo default (<https://developers.arcgis.com/python/guide/dl-on-time-series-data/>).

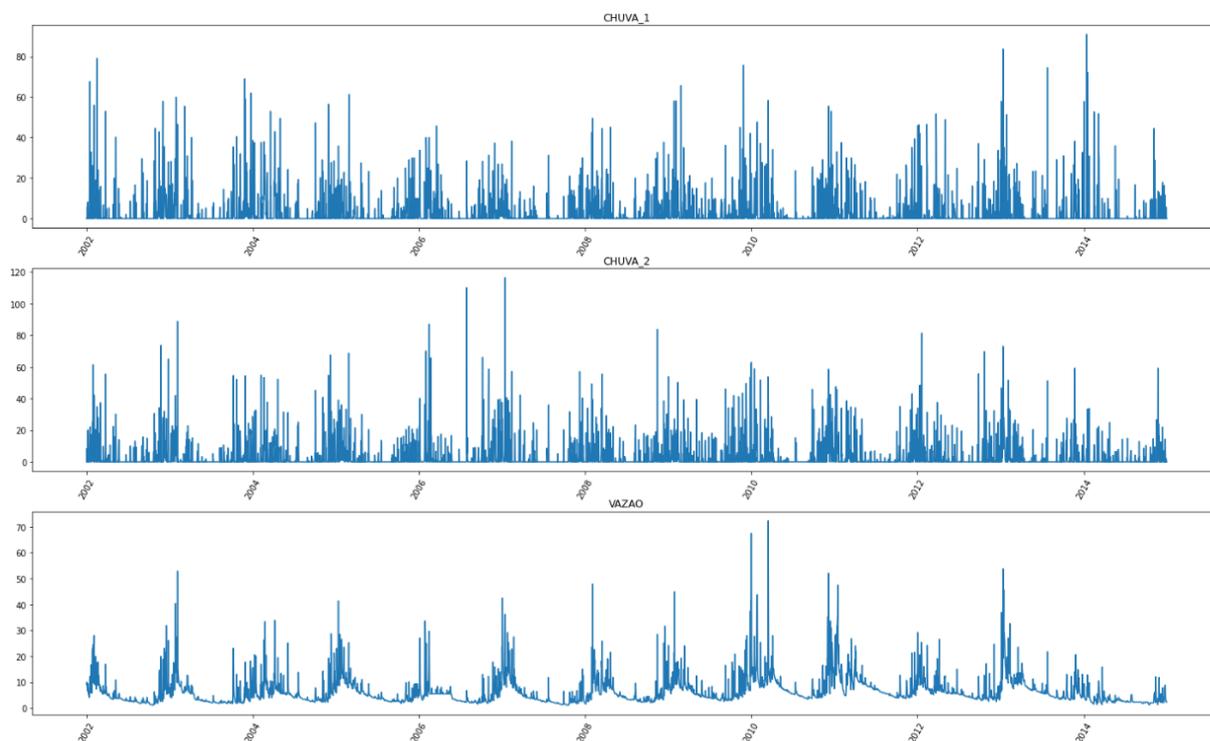
Figura 9: Valores correlacionados entre os parâmetros de análise utilizados.



A reprodução dos modelos de aprendizagem de máquina e aprendizagem profunda foi executada em computador com processador Intel Core i7 e placa de vídeo NVIDIA Quadro com memória dedicada. Dessa forma, com a posse dos dados necessários, foi elaborado o programa responsável por executar a análise comparativa entre os comportamentos da chuva e vazão ao longo da série temporal, com base na inserção das variáveis explanatórias (chuva e vazão) e prevista (vazão).

Foram utilizados dados de treinamento levantados entre os anos de 2002 e 2014, devido ao fato de os postos pluviométricos e fluviométricos se encontrarem consistentes no mesmo período das séries temporais. Em relação ao ano de 2015, no qual foi avaliada a variável prevista (vazão), foi testada a acurácia de previsão. Os dados de treinamento do programa consistiram em 4748 dados de chuva coletados de ambos os postos pluviométricos e 4748 dados de vazão levantados do posto fluviométrico. O programa foi capaz de reunir os valores de precipitação e vazão médias diárias (figura 10), perceber como se comportavam, aprender estes padrões e estabelecer uma correlação matemática a ser extrapolada comparativamente para o ano de 2015, com base nas arquiteturas de *deep learning* e *machine learning*. Assim, foram construídos gráficos comparativos entre a vazão real e a prevista, com a finalidade de verificar se a previsão dos padrões de eventos hidrológicos foi bem-sucedida.

Figura 10: Dados de precipitação e vazão médias diárias levantados nos postos selecionados entre os anos de 2002 e 2014.



Dentre os modelos de aprendizado de máquina existentes, cita-se as três utilizadas nesta pesquisa: regressão linear, árvore de decisão e *support-vector machine*. Os melhores resultados em termos de previsibilidade de dados foram fornecidos pelo modelo árvore de decisão (*random tree* ou *decision tree*), que é definido como uma técnica semelhante a um fluxograma, que designa um algoritmo a ser seguido a partir de uma abordagem na qual o programa é processado iniciando-se pela raiz, que prossegue para uma sequência de separação de dados e operações booleanas (nós) de decisão lógica, hierarquicamente estruturadas, até alcançar um resultado final representado pelas folhas (JIJO & ABDULAZEEZ, 2021).

Além do mais, as arquiteturas selecionadas para integrar os testes de aprendizado profundo são geralmente utilizadas para detecção de objetos e tratamento de imagens, porém elas mostraram resultados significativos sendo submetidas a processamento de séries temporais, por serem classificadas como redes neurais convolucionais (CNN). Esta categoria é amplamente aplicada no campo da classificação de séries temporais, muito provavelmente em razão da sua robustez e baixo tempo de gasto computacional de treinamento em relação a outras arquiteturas mais complexas (ISMAIL et al., 2019).

A arquitetura Resnet (*Residual Networks*) é uma das variantes das redes neurais convolucionais, capaz de elevar substancialmente a acurácia do modelo, por meio da adição de vínculos lineares às camadas convolucionais (HE et al., 2016).

Ademais, o modelo conhecido como FCN (*Fully Convolutional Neural Networks*) consiste em uma rede que mantém a extensão de uma série temporal no decorrer da sequência de convoluções, isto é, não possuem camadas de agrupamento locais para segregar os períodos de dados. Uma de suas vantagens é a redução acentuada do número de parâmetros de uma rede neural de maneira simultânea à utilização dos chamados Mapas de Ativação de Classe (CAM), uma vez que substitui a camada final tradicional

por uma de agrupamento médio global (GAP) (ZHOU et al., 2016).

Outra arquitetura testada é a designada como ResCNN (*Residual Convolutional Neural Networks*), uma técnica de aprendizado profundo que descarta conjuntos de camadas convolucionais através de atalhos de conexão. Deste modo, tem potencial para superar o problema do gradiente de fuga/explosão, isto é, uma complicação recorrente no treinamento de redes neurais na qual os gradientes de atualização de dados aumentam de maneira exponencial, causando instabilidade no aprendizado do programa (LONG; YAN; LIANG, 2019).

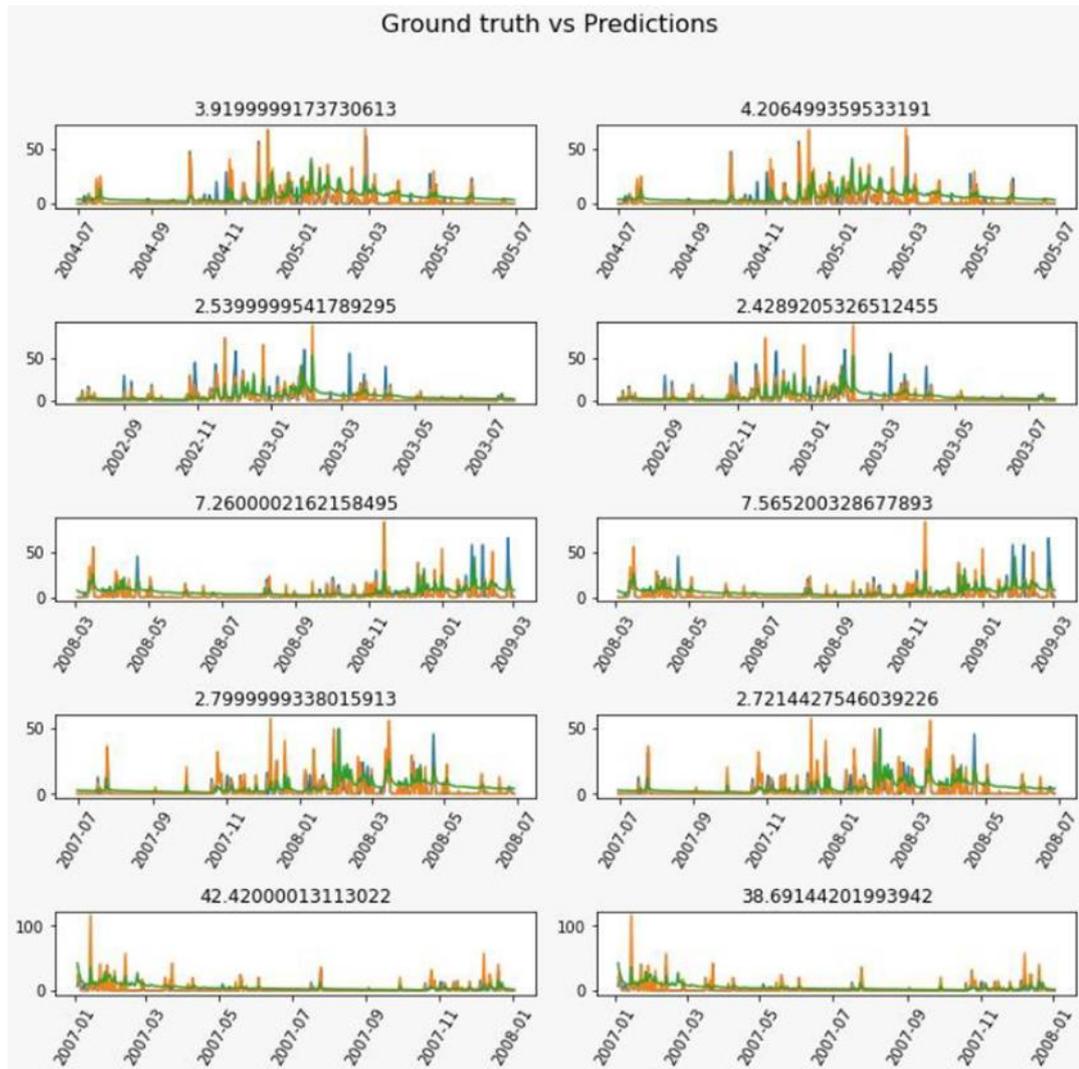
Finalmente, o modelo que melhor descreveu o comportamento da série histórica deste trabalho é o chamado *InceptionTime*, caracterizado como um conjunto de cinco modelos de *Deep Learning* apropriados para análises de séries temporais, os quais apresentam a mesma construção, porém são iniciados com pesos diferentes. Além disso, a arquitetura submete o volume dados inúmeros filtros com comprimentos variados. Em suma, isso permite que o modelo colete um apanhado de recursos relevantes das séries históricas (SZEGEDY et al., 2015).

O modelo de *Deep Learning Inception*, também conhecido como *InceptionNet*, é uma arquitetura de rede neural convolucional (CNN) desenvolvida pela *Google*. Foi projetado para lidar com o desafio de equilibrar o compromisso entre a profundidade da rede e o poder computacional necessário para treiná-la. Uma das características distintivas do modelo *Inception* é o uso de módulos *Inception*, que são blocos de convoluções paralelas de diferentes tamanhos de filtro. Esses módulos permitem que a rede capture informações em diferentes escalas e resoluções, o que ajuda a melhorar a capacidade de aprendizado e a eficiência computacional (SZEGEDY et al., 2015). O modelo *Inception* é uma rede neural convolucional própria para o processamento de dados de séries temporais, isto é, o algoritmo da arquitetura é capaz de realizar todos os procedimentos necessários de transformação de dados e treinamento.

Quando as arquiteturas são aplicadas a uma série temporal, é necessário validá-las com as chamadas métricas de validação do modelo. Uma vez que o modelo é treinado, é essencial visualizar os resultados e medir seu desempenho, utilizando as métricas explicadas a seguir.

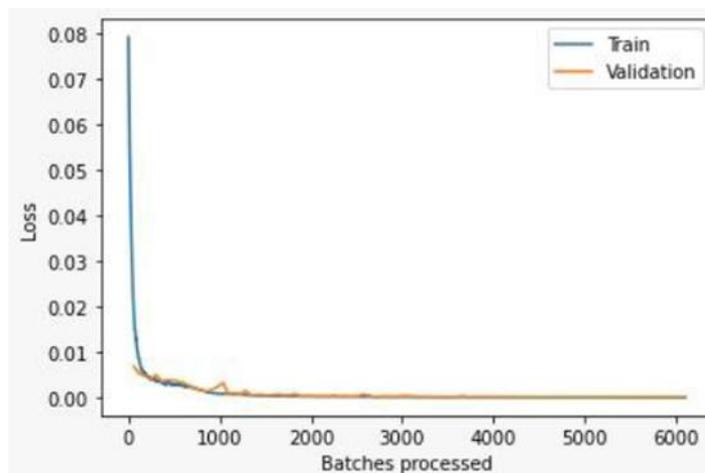
Os resultados do treinamento podem ser impressos via `show_results()`, ou seja, o método para avaliar a qualidade do modelo treinado e calcular as métricas do modelo. Para modelos de séries temporais, o `show_results` (método) normalmente imprimirá dois gráficos adjacentes com os valores de verdade do terreno plotados no gráfico do lado esquerdo e as previsões no lado direito. As previsões seriam os valores previstos pelo modelo após serem aplicados ao conjunto de dados de validação. A comparação dos valores previstos com a variável de destino real ou a verdade de base fornecerá alguma indicação do desempenho do modelo treinado. Assim, o comando `model.show_results(rows=5)` retornaria à comparação de cinco valores de destino versus valores de previsão, conforme mostra a figura 11.

Figura 11: Valores de destino x valores de previsão, para os modelos de *Deep Learning*.



Outro método utilizado para avaliar o desempenho das arquiteturas aplicadas pode ser observado no gráfico contido na figura 12, que compara a perda de treinamento vs. a validação.

Figura 12: Bateladas processadas em treinamento e validação.



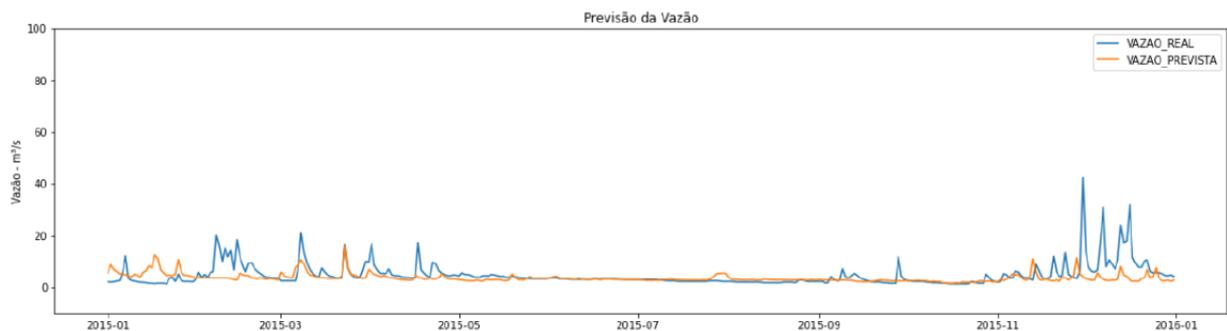
5. Resultados e Discussão

No que tange ao tratamento da série relativo aos dias secos, foi aplicada a técnica de *gradient boosting*, a qual promoveu a filtragem dos valores para articular o processamento dos dados somente nos dias chuvosos. O método proposto, mesmo que robusto, torna simples tal eliminação, mantendo as características da memória da ocorrência dos mesmos. No entanto, as RNA's ou os outros modelos de ML já incorporam tal memória, não necessitando dessa aplicação, mas, por já ter o modelo construído, optou-se por sua utilização, não impactando o tratamento da série. Não foi verificado avanços no sentido da melhoria do tratamento. Trabalhos futuros são pretendidos utilizando-se modelos de zeros inflacionários para dias secos, mas não objeto dessa comunicação científica.

A partir da operação do programa elaborado, foram experimentadas quatro arquiteturas de aprendizado profundo (*InceptionTime*, Resnet, FCN e ResCNN) e três de aprendizado de máquina (regressão linear, árvore de decisão e support-vector machine) diferentes e próprias para séries históricas de dados, com o objetivo de determinar a correlação dos mesmos e de elaborar as curvas comparativas. Os principais resultados obtidos estão descritos nas figuras 13, 14, 15 e 16 para *Deep Learning* e 17 para *Machine Learning*.

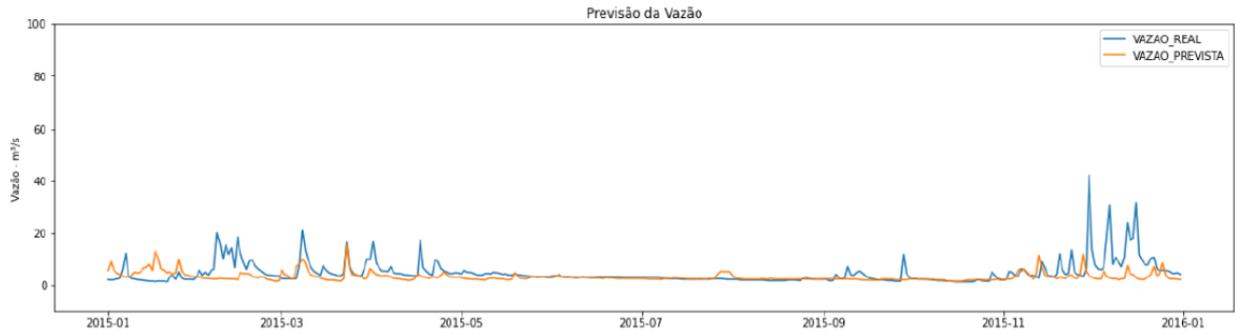
É importante adiantar que as arquiteturas de aprendizado de máquina utilizadas manifestaram resultados mais representativos em se tratando de picos de vazão, ou seja, a previsibilidade de altos valores de vazão é mais eficiente. No entanto, para vazões baixas e mais homogêneas no tempo, o aprendizado profundo exibiu resultados mais satisfatórios entre as vazões reais e previstas. Outro resultado está relacionado ao tempo de processamento de ambas as modalidades de aprendizagem, no qual o aprendizado de máquina manifestou cerca de 5 minutos para a geração dos resultados e o aprendizado profundo, aproximadamente 5 segundos.

Figura 13: Distribuição de vazão real e prevista no ano de 2015 (*InceptionTime*).



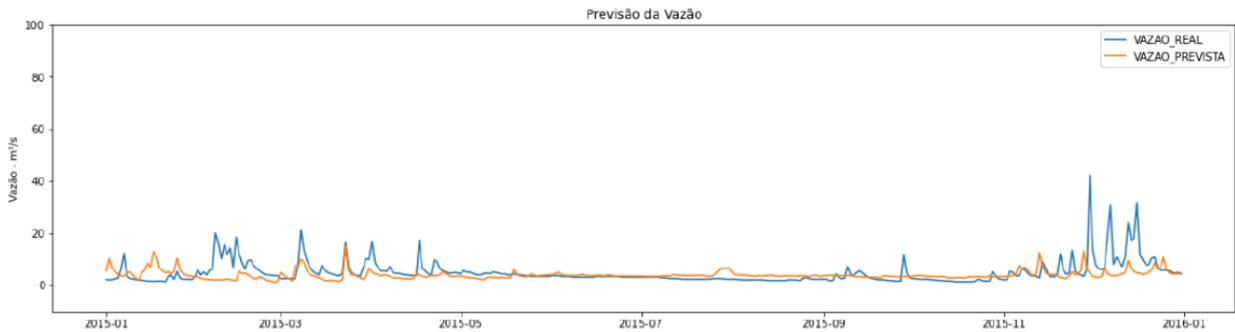
- RMSE: 4.6183;
- MAE: 2.3629.

Figura 14: Distribuição de vazão real e prevista no ano de 2015 (Resnet).



- RMSE: 4.7722;
- MAE: 2.4897.

Figura 15: Distribuição de vazão real e prevista no ano de 2015 (FCN).

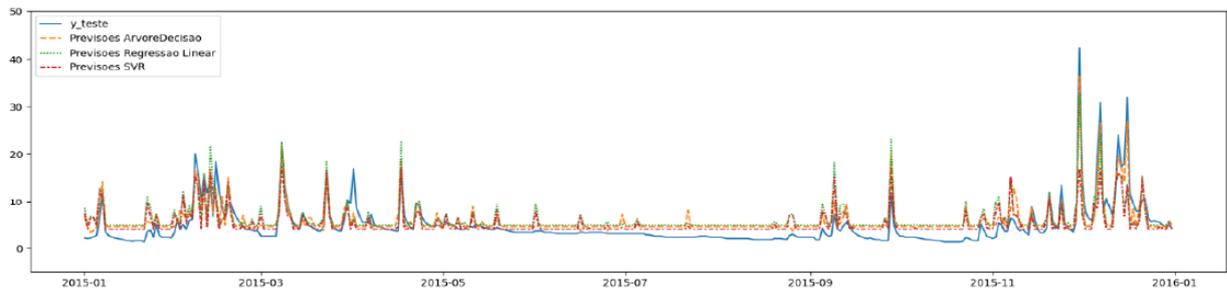


- RMSE: 4.6518;
- MAE: 2.5989.

Figura 16: Distribuição de vazão real e prevista no ano de 2015 (ResCNN).



- RMSE: 4.8556;
- MAE: 2.5498.

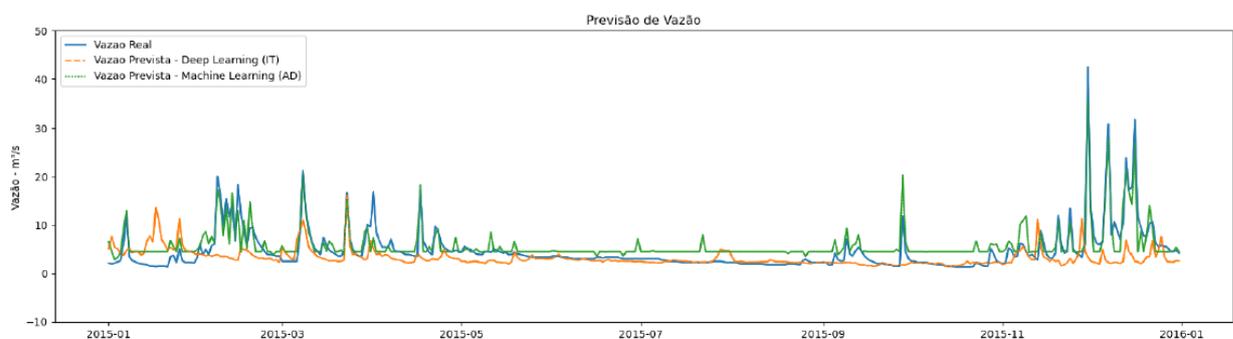
Figura 17: Distribuição de vazão real e prevista no ano de 2015 (Machine Learning).

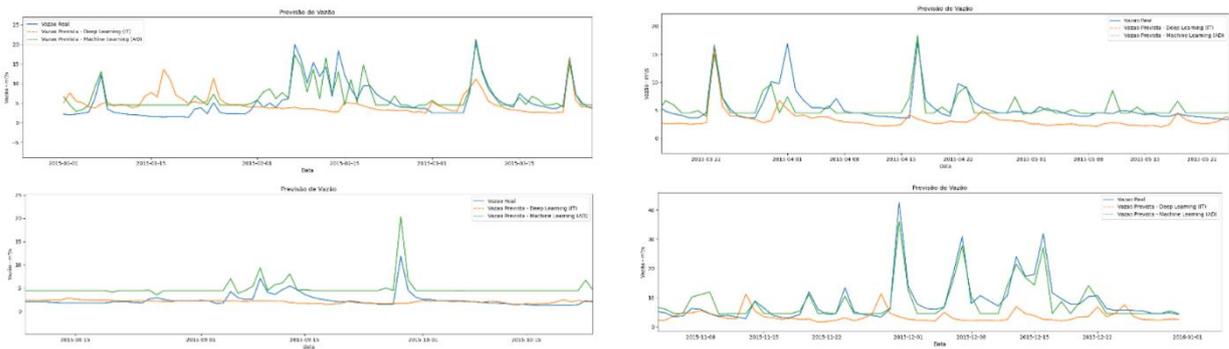
- R-Square: 0.4725 (Regressão Linear);
- R-Square: 0.7261 (Árvore de Decisão);
- R-Square: 0.5056 (SVR).

Os parâmetros de avaliação do desempenho das correlações nas arquiteturas testadas, listados abaixo dos respectivos gráficos comparativos, foram: RMSE, MAE. Os coeficientes RMSE e MAE indicam o grau de erros provenientes da comparação entre as variáveis real e prevista, isto é, quanto menores estes valores, mais eficaz é a explicação do modelo em relação à realidade. O parâmetro RMSE se diferencia do MAE por ser mais útil nos casos em que erros de grande escala são indesejáveis. Como pode-se observar, os menores valores de ambos os coeficientes de erro obtidos na análise dos dados de chuva e vazão foram gerados pela arquitetura *InceptionTime*.

No cenário estudado, a série temporal de vazão possui um comportamento randômico, como pode ser visto nos gráficos elaborados, isto é, os valores distribuídos nos gráficos apresentam padrões altamente irregulares e aleatórios, diferenciando-se intensamente de retas. Portanto, todos os valores de R^2 são negativos. Devido a estas condições, quanto mais próximos os coeficientes estiverem de 0, mais êxito a arquitetura obteve na previsão dos dados reais. Assim, a arquitetura *InceptionTime* também demonstrou a melhor performance de previsão, por manifestar um R^2 de -0,05.

No tocante aos testes de aprendizado de máquina, a arquitetura árvore de decisão forneceu resultados suficientemente consideráveis no que se refere à simulação do comportamento anual das vazões do Rio Bananal. A figura 18 exibe um gráfico comparativo entre a vazão real e as modelagens de aprendizado profundo e aprendizado de máquina. Como já mencionado, é possível observar que a técnica de *machine learning* prevê com mais precisão os picos de vazão existentes ao longo do ano. Em compensação, o modelo de aprendizagem profunda responde a previsão dos eventos contínuos de baixas vazões de maneira mais efetiva.

Figura 18: Previsão de vazão por *Machine Learning* (Árvore de Decisão) e *Deep Learning* (*InceptionTime*) em comparação com vazão real.



6. Conclusão

Apesar do estágio inicial destes estudos quanto à bacia hidrográfica analisada, ainda pretende-se demonstrar ao longo do tempo, que as técnicas de *Deep Learning* e associadas à previsão do comportamento de séries históricas e consequentemente de variáveis hidrometeorológicas responsáveis por provocar desastres naturais, possui capacidade para apresentar resultados mais coerentes e satisfatórios nos modelos de chuva-vazão.

Salienta-se que o trabalho em tela se trata de uma comparação entre modelos pré-processados em ambiente colaborativo com interface no ArcGIS, em sua forma otimizada e apresentação de seus comparativos resultados. Não há aprofundamento de todas as técnicas, o que proporcionaria outros trabalhos com essa essência. Tratou-se de avaliar a utilização desses modelos com precisão aceitável em ambientes de fácil utilização, mantendo o rigor das técnicas, e preparando para novas propostas.

Verifica-se que a literatura, apesar de divergir nas técnicas utilizadas, tem como base que a técnica de RNA apresenta resultados promissores na previsão de vazão na bacia hidrográfica do rio Bananal. No entanto, cada autor apresenta informações diferentes sobre os resultados obtidos, com destaque para as diferentes arquiteturas de RNAs utilizadas e as análises estatísticas realizadas para avaliar o desempenho das RNAs.

Foram utilizadas diferentes modelos, tanto de Machine Learning quanto de Deep Learning, para avaliar as previsões de dados ao longo do tempo e determinar suas eficiências. As arquiteturas de *Deep Learning* utilizadas foram: *InceptionTime*, LSTM, GRU e *WaveNet*. Já as técnicas de Machine Learning utilizadas foram: *Árvore de Decisão*, *Random Forest*, Regressão Linear e Regressão Logística.

Foram aplicadas na análise de dados em séries temporais hidrometeorológicas da bacia hidrográfica do rio Bananal, nos períodos de 2002 a 2014. Foram utilizados três coeficientes de avaliação do desempenho das correlações nas arquiteturas testadas: RMSE, MAE e R^2 .

Em virtude da melhor performance de tempo de processamento e previsão dos dados, a aprendizagem profunda na arquitetura *InceptionTime* apresentou os menores coeficientes de erro MAE e RMSE. Além do mais, foi possível observar o baixo valor de R-Quadrado, que indica um bom desempenho na previsão dos dados reais do ano de 2015, os quais poderão ser utilizados para construir um sistema de alertas a eventos extremos, além de mais otimizado no tempo. Deste modo, seriam gerados incontáveis benefícios para a gestão dos recursos hídricos e manutenção da integridade do meio ambiente e da saúde e segurança públicas.



Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal Nível Superior - Brasil (CAPES) - Código de Financiamento 001, agradeço também ao Programa de Mestrado Profissional em Rede Nacional em Gestão e Regulação de Recursos Hídricos - ProfÁgua, Projeto CAPES/ANA AUXPE Nº 2717/2015, pelo apoio técnico científico aportado até o momento.

Referências

BRITTO, M.; BAPTISTA, G. M. de M.; LIMA, E. A. de. O estudo dos componentes do ciclo hidrológico desde métodos tradicionais até o uso de sensoriamento remoto: uma revisão. **Revista Paranoá**, N.23, p. 127–146, 2019. Disponível em: <https://doi.org/10.18830/issn.1679-0944.n23.2019.11>.

DEBASTIANI, A. B.; SILVA, R. D.; RAFAELI NETO, S. L. Eficácia da arquitetura MLP em modo closed-loop para simulação de um Sistema Hidrológico. **RBRH**, v. 21, n. 4, p. 821–831, out. 2016.

FREITAS, C. M. de et al. Desastres naturais e saúde: uma análise da situação do Brasil. **Ciência & Saúde Coletiva** [online], v. 19, n. 9, p. 3645-3656, 2014. Disponível em: <https://doi.org/10.1590/1413-81232014199.00732014>. ISSN 1678-4561. Acesso em: 30 Jul. 2022.

GU, Q. et al. Characterizing the spatial variations of the relationship between land use and surface water quality using self-organizing map approach. **Ecological Indicators**, v. 102, p. 633-643. mar. 2019.

HE, K. et al. **Deep residual learning for image recognition**. In: IEEE conference on computer vision and pattern recognition, p. 770–778, 2016.

ISMAIL, F. H. et al. Deep learning for time series classification: a review. **Data Min Knowl Disc**, 33, p. 917–963, 2019. Disponível em: <https://doi-org.ez24.periodicos.capes.gov.br/10.1007/s10618-019-00619-1>.

JJO, B. T.; ABDULAZEEZ, A. M. Classification based on decision tree algorithm for machine learning. **Journal of Applied Science and Technology Trends**, 2(01), p. 20-28, 2021.

LIMA, A. P.; AMORIM, M. C. de C. T. Análise de episódios de alagamentos e inundações urbanas na cidade de são carlos a partir de notícias de jornal. **Revista Brasileira de Climatologia**, [S.l.], v. 15, abr. 2015. ISSN 2237-8642. Disponível em: <https://revistas.ufpr.br/revistaabclima/article/view/33406>. Acesso em: 30 jul. 2022.

LUIZ, T.B.P; SCHRODER, T. Modelos chuva-Vazão: Uso de Técnicas de Aprendizagem de Máquinas para Calibração de Modelos em uma Pequena Bacia Hidrográfica. **Geoambiente on-line**, N. 37, p.304-321, 2020.

MENDONÇA, L. M. et al. Modelagem chuva-vazão via redes neurais artificiais para simulação de vazões de uma bacia hidrográfica da Amazônia. **Revista de Gestão de Água da América Latina**, v. 18, 2021. Disponível em: <https://doi.org/10.21168/rega.v18e2>.

PACHECO, C.; PEREIRA, N. Deep Learning Conceitos e Utilização nas Diversas Áreas do Conhecimento. **Revista Ada Lovelace**, v. 2. P. 34-49, 2018.

ROSA, R. Análise Espacial em Geografia. **Revista da ANPEGE**, v. 7, n. 1, p. 275-289, 2011. Disponível em: <https://ojs.ufgd.edu.br/index.php/anpege/article/view/6571/0>. Acesso em: 30 jul. 2022.

SILVA, C. V. Uso de redes neurais artificiais para análise multitemporal da dinâmica do uso e cobertura da terra em bacias hidrográficas. 2022.

SOUSA, L. M. et al. Avaliação do Uso e Cobertura da Terra em Paragominas e Ulianópolis PA, Utilizando Dados do Projeto TERRACLASS. **Revista Brasileira de Cartografia**, v. 3, n. 69. p. 421-431, 2017. Disponível em: <http://www.seer.ufu.br/index.php/revistabrasileiracartografia/article/view/44339>. Acesso em: 30 jul. 2022.



SZEGEDY, C. et al. **Going deeper with convolutions**. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 1–9, 2015.

ULIANA, E. M. et al. Desenvolvimento de redes neurais artificiais para estimativa das vazões diárias na bacia do rio Piracicaba. *IRRIGA*, v. 23, n. 4, p. 756–772, 2018. DOI: 10.15809/irriga.2018v23n4p756-772. Disponível em: <https://actaarborea.fca.unesp.br/index.php/irriga/article/view/2740>. Acesso em: 10 jun. 2023.

WEN, Long; DONG, Yan; GAO, Liang. A new ensemble residual convolutional neural network for remaining useful life estimation. **Mathematical Biosciences and Engineering**, v. 16(2), p. 862-880, 2019. Disponível em: <https://doi:10.3934/mbe.2019040>.

ZHOU, B. et al. **Learning deep features for discriminative localization**. In: IEEE conference on computer vision and pattern recognition, pp 2921–2929, 2016.



Vinicius de Azevedo Silva

Atual Secretário de Meio Ambiente do município de Barra Mansa/RJ pela Secretaria Municipal de Meio Ambiente e Desenvolvimento Sustentável, Diretor do CBH Médio Paraíba onde o Diretório Colegiado dirige administrativamente o Comitê, Membro do Conselho Estadual de Recursos Hídricos do Rio de Janeiro onde exerce diretrizes para a formação, a organização e funcionamento dos Comitês de Bacia Hidrográfica (CBHs) e Agências de Água, exercer a arbitragem dos conflitos entre os Comitês, estabelecer os critérios gerais sobre a outorga de direito de uso de recursos hídricos e a sua cobrança, deliberar sobre os projetos de aproveitamento de recursos hídricos dentro do Estado, além de analisar as propostas de alteração da legislação de recursos hídricos e à Política Estadual de Recursos Hídricos; Engenheiro Ambiental pela Fundação Oswaldo Aranha. Tem experiência nas áreas de Termodinâmica e Energia, Poluição Atmosférica, Poluição Hídrica, Hidráulica, Climatologia, Hidrologia e bom conhecimento de Informática.

Contribuição de coautoria: Concepção; Curadoria de dados; Análise; Coleta de dados; Metodologia; Software; Supervisão; Validação.

Rafael de Feiras Souza

Doutor em Administração pela Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo (FEA/USP). Servidor público há dezessete anos, tendo ocupado cargos de chefia, direção, coordenação e assessoria; além do magistério no Ensino Superior em cursos de graduação e MBAs. Atualmente, é Professor da Área de Ciência de Dados da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto (FEARP/USP). Possui experiência com a programação direcionada à coleta, ao tratamento, à análise e à apresentação de dados, atuando com técnicas de Machine Learning e de Deep Learning, incluindo as Análises Espaciais. Também é um dos autores da library `overdisp` for R, módulo computacional registrado pelo CRAN (The Comprehensive R Archive Network).

Contribuição de coautoria: Metodologia e Software.

Mateus Peixoto Oliveira

Mestre do Programa de Pós-Graduação em Tecnologia Ambiental pela Universidade Federal Fluminense - UFF (2020-2023). Membro do grupo de pesquisa GEUFF (Grupo de Eletroquímica e Energia da UFF). Possui graduação em Engenharia Ambiental pelo Centro Universitário de Volta Redonda - UniFoa (2015-2019). Formado pela Escola Técnica Pandiá Calógeras - ETPC - no curso técnico de Eletrônica (2012-2014). Trabalhou como Aprendiz Administrativo no setor de Meio Ambiente da Saint-Gobain Canalização (Barra Mansa/RJ). Trabalhou como estagiário e Analista Ambiental na empresa Azevedo Soluções Ambientais LTDA (Barra Mansa/RJ). Tem experiência na área de Consultoria Ambiental.

Contribuição de coautoria: Metodologia; Validação; Visualização; Redação.

Francisco Lledo dos Santos

Professor Adjunto da Faculdade Arquitetura e Engenharia do Campus de Barra do Bugres da Universidade do Estado de Mato Grosso (UNEMAT). Tem experiência na área de Hidráulica de Superfície Livre e Recursos Hídricos, atuando principalmente nos seguintes temas: mecânica dos fluidos, reologia, canais abertos, hidrologia, redes neurais artificiais e *machine learning*. Foi o 1o Chefe de Departamento de Engenharia Civil - Unemat/Sinop (2006-2008), Diretor da Faculdade de Ciências Exatas da Universidade do Estado do Mato Grosso (2008-2010). Foi 1o Diretor do Núcleo de Inovação Tecnológica. Foi o 1o Pró Reitor de Assuntos Estudantis da Unemat (2011) e Pró Reitor de Planejamento e Desenvolvimento Institucional (2011). Pró-reitor de Planejamento e Tecnologia da Informação da UNEMAT de 2013 a 2018. Hoje atua como Assessor de Assuntos Externos da Reitoria e Coordenador do Programa em Rede de Mestrado Profissional Gestão e Regulação de Recursos Hídricos - ProfÁgua-UNEMAT.



Contribuição de coautoria: Metodologia; Supervisão; Validação; Redação - revisão e edição.

Como citar: SILVA et al. Aplicação de machine learning e deep learning para modelagem de uma bacia hidrográfica. Revista Paranoá. N.34, jan/jun 2023. DOI 10.18830/issn.1679-0944.n34.2023.20

Editores responsáveis: Daniel Sant'Ana, Livia Santana, Ronaldo Rodrigues Lopes Mendes, Sílvio Roberto Magalhães Orrico e Thiago Alberto da Silva Pereira.