

# Previsão de carga de curto prazo usando um método combinado de redes neurais convolucionais e séries temporais difusas

Hossein Javedani Sadaei<sup>a, b</sup>, Petronio C<sup>^</sup>andido de Lima e Silva<sup>a, c</sup>, Frederico Gadelha Guimaraes<sup>a, d, \*</sup>, Muhammad Hisyam Lee<sup>e</sup>

<sup>a</sup> Laboratório de Inteligência de Máquinas e Ciência de Dados (MINDS), Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil

<sup>b</sup> Machine Learning Practice Lead in Avenue Code, Brazil <sup>c</sup> Programa de Pós-Graduação no Departamento de Engenharia Elétrica, Universidade

Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil <sup>d</sup> Departamento de Engenharia Elétrica, Universidade Federal de Minas Gerais, UFMG,

Belo Horizonte, Brasil <sup>e</sup> Departamento de Ciências Matemáticas, Universidade Teknologi Malásia, Skudai, Johor, Malásia

## artilleinfo

### Histórico do artigo:

Recebido em 26 de julho de 2018

Recebido em forma revisada

12 de março de 2019

Aceito 13 março 2019 Disponível online 19 março 2019

### Keywords:

Séries temporativas multivariadas

Redes neurais convolucionais

Previsão de carga de curto prazo

Previsão da série temporal

Aprendizado profundo

## abstract

Propomos um método combinado baseado nas séries temporais difusas (FTS) e redes neurais convolucionais (CNN) para previsão de carga de curto prazo (STLF). Assim, no método proposto, dados de séries temporais multivariadas que incluem dados de carga horária, séries temporizadas de temperatura hourly e versão difusa de séries temporizadas de carga, foram convertidos em imagens multicanais para serem alimentados a um modelo de CNN de aprendizagem profunda proposto com arquitetura adequada. Utilizando imagens que foram criadas a partir dos valores sequenciados das séries temporais multivariadas, o modelo da CNN proposto poderia determinar e extrair parâmetros importantes relacionados, de forma implícita e automática, sem qualquer necessidade de interação humana e conhecimento especializado, e tudo por si só. Seguindo essa estratégia, foi dito como a contratação do método proposto é mais fácil do que alguns modelos tradicionais de STLF. Portanto, poderia ser visto como uma das grandes diferenças entre o método proposto e algumas metodologias de última geração do STLF. Além disso, o uso da lógica difusa teve grande contribuição para controlar a sobre-ajuste expressando uma dimensão de séries temporais por um espaço confuso, em um espectro, e uma sombra em vez de apresentá-la com números exatos. Vários experimentos em conjuntos de dados de teste suportam a eficiência do método proposto.

© 2019 Elsevier Ltd. Todos os direitos reservados.

## 1. Introdução

### 1.1. Motivação

A previsão de carga é essencial para os recentes sistemas inteligentes de gerenciamento de energia. A cada dia, há cada vez mais usos para previsão de carga. Normalmente, a Previsão de Carga de Curto Prazo (STLF) com prazos de liderança no escopo de meia hora a um dia é necessária para controlar a programação e o agendamento de transferência de energia e o envio de carga. Assim, qualquer melhoria na precisão do STLF pode levar a uma melhoria no desempenho da gestão de energia e uma diminuição nos gastos do sistema de energia. Em um trabalho muito importante, Gross et al. [1] forneceram uma discussão sobre a importância de

STLF e seu papel na formulação de estratégias operacionais ecológicas, confiáveis e seguras para o sistema de energia. Em seu artigo, eles contaram os seguintes impactos do STLF na organização de fornecimento de energia:

O STLF conduz as funções de agendamento que determinam o maior comprometimento econômico das fontes de geração.

É necessário que o STLF seja necessário para que a função de agendamento hidráulico determine as liberações ideais dos reservatórios e os níveis de geração nas usinas.

O STLF é necessário pela função de compromisso da unidade para determinar as estratégias mínimas de custo por hora para a inicialidade e desligamento das unidades para fornecer a carga de previsão, para sistemas puramente térmicos.

Para sistemas hidroelétricos e térmicos mistos, o STLF é exigido pela função de coordenação hidro-térmica para agendar a operação horária dos diversos recursos, de modo a minimizar os custos de produção.

O STLF é útil para uma avaliação preditiva da segurança do sistema de energia.

\* Autor correspondente. Laboratório de Inteligência de Máquinas e Ciência de Dados (MINDS), Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil.

E-mail addresses: [hsadaei@avenuecode.com](mailto:hsadaei@avenuecode.com) (H.J. Sadaei), [petronio.candido@ufmg.edu.br](mailto:petronio.candido@ufmg.edu.br) (P.C. de Lima e Silva), [fredericoguimaraes@ufmg.br](mailto:fredericoguimaraes@ufmg.br) (F.G. Guimaraes). ~

O STLF é um requisito essencial de dados da função de análise de rede off-line para a detecção de condições futuras sob as quais o sistema de energia pode estar vulnerável.

A STLF fornece aos despachantes do sistema informações oportunas, ou seja, a previsão de carga mais recente, com a última previsão meteorológica e behavior aleatório levado em conta.

Como se pode ver, um STLF preciso é a chave para o sucesso das empresas de fornecimento de eletricidade e motivou os pesquisadores do nosso estudo a desenvolver métodos precisos e práticos para esse fim. Para explicar melhor a contribuição do nosso estudo, a próxima seção prevê a revisão bibliográfica de obras relacionadas e também alguns problemas e lacunas são discutidos.

## 1.2. Revisão da literatura

A STLF tornou-se mais interessante nos últimos anos entre os pesquisadores devido à sua crescente importância nas redes e micro-redes de smart, à integração de fontes de energia renovável mais próximas aos consumidores e ao grau de dificuldade que está envolvido nessa linha de pesquisa de previsão. Diferentes métodos clássicos e novos têm sido sugeridos para o STLF, tais como Kalman filtering, Regression, Modelos Auto-regressivos Integrados de Média Móvel (ARIMA), Modelos Sazonais Auto-regressivos Integrados da Média Móvel (SARIMA), Sazonal Auto-regressivo Fracionário Integrado Médio (SARFIMA), etc. Em 2004, Al-Hamadi et al. [2] usaram o modelo de espaço estatal de variação de tempo para modelar a demanda de carga por hora. Kalman filter foi usado recursivamente para estimar os parâmetros ideais de previsão de carga para cada hora do dia [2]. Mais tarde, Song et al. in Ref. [3] propuseram um modelo de regressão linear difuso que foi feito a partir dos dados de carga. Nesse método, utilizando-se três anos anteriores, foram encontrados os coeficientes do modelo, resolvendo o problema de programação linear mista. Em um estudo conduzido por Věhakyland et al. [4], foi utilizada uma análise da série temporal Box-Jenkins usando o modelo ARIMA para STLF, e foi descrito um sistema de previsão desenvolvido na Imatra Power Company. O algoritmo de previsão era simples, rápido e preciso e o modelo de função de transferência foi usado para introduzir efeitos de temperatura, melhorando assim ainda mais a precisão.

Nos últimos anos, outros novos métodos foram aplicados para a previsão de carga, como a Inteligência Artificial (IA) e a lógica difusa. Soluções inteligentes, baseadas em tecnologias de IA, para resolver o STLF estão se tornando mais mais e mais difundidas hoje em dia. Sistemas baseados em IA estão sendo desenvolvidos e implantados em todo o mundo em muitas aplicações, principalmente por causa de seus recursos simbólicos de raciocínio, flexibilidade e explicação. Como exemplo, Ranaweera et al. propuseram uma metodologia que utilizasse regras difusas para incorporar dados históricos de clima e carga. Essas regras difusas foram obtidas a partir dos dados históricos usando um algoritmo do tipo aprendizagem [5]. Yaoyao et al. [6] apresentaram um modelo para quantificar a incerteza ligada à carga de energia, obtendo mais informações sobre carga futura, e então uma rede neural foi empregada para reconstruir o modelo de regressão quântica para a construção do método de previsão probabilística. Em outro trabalho, um método de previsão híbrida da transformação de ondas, network neural e o algoritmo evolutivo foi proposto no Ref. [7] e empregado para STLF. Em uma abordagem semelhante, Chofrani et al. in Ref. [8] introduziram uma combinação de rede neural bayesiana (BNN) e uma transformação de onda para produzir a característica de carga detalhada para o treinamento bnn. Neste modelo, foi aplicada uma soma ponderada das saídas BNN para prever a carga para um dia de especific. Ainda em outro modelo híbrido, um modelo STLF proposto por Fan et al. [9] combina algoritmo de Reconstrução de Espaço de Fase (PSR) com modelo de regressão BSK. Nesse modelo, os dados de carga poderiam ser reproduzidos pelo algoritmo PSR para derivar as tendências evolutivas dos dados históricos de carga e as informações importantes incorporadas para aumentar a confiabilidade da previsão. O modelo BSK, por outro lado, representava as estruturas espaciais entre os pontos de regressão e os pontos vizinhos para obter as regras de rotação e perturbação em cada dimensão. Finalmente, o modelo proposto, incluindo a regressão multidimensional, foi estabelecido com sucesso e utilizado para o STLF [9]. Mamlook et al. in Ref.

[10] empregaram controlador lógico difuso em uma base horária para prever o impacto de vários parâmetros condicionais, por exemplo, clima, tempo, carga de dados históricos e distúrbios aleatórios, em previsão de carga em termos de conjuntos confusos através do processo de geração. Neste estudo, a wavelet transform decompôs a série temporal em seus componentes e, em seguida, cada componente foi previsto por uma combinação de rede neural e um algoritmo evolutivo.

Em um estudo importante, uma lista das principais pesquisas sobre o uso da IA para IST foi fornecida por Metaxiotis et al. [11]. Em comparação com os modelos de IA, as CNNs recentemente mostram um desempenho muito bom na previsão. CNNs foram introduzidos por Fukushima em sua forma muito simples em Ref. [12]. Mais tarde, LeCun et al. in Ref. [13] apresentaram a forma atual de CNNs com conceitos mais avançados. Desde as CNNs propostas por Fukushima, houve muitas melhorias e extensões, como camadas de piscina máxima & normalização de lotes pela LeCun [14]. Na Seção 2, a arquitetura e os componentes das CNNs são explicados com mais detalhes. Embora as CNNs sejam muito bem sucedidas para a previsão [15,16], um dos principais problemas de usá-las que é comum em muitos outros métodos de redes neurais profundas é o excesso de fitting [17]. Como Burnham desafiou em Ref. [18] excesso de fitação é a produção de uma análise que corresponde muito de perto ou exatamente a um determinado conjunto de dados, e pode, portanto, deixar de fit dados adicionais ou prever observações futuras de forma confiável. Existem algumas técnicas para evitar ou reduzir o excesso de fitação em modelos de deep learning, como técnicas de aumento de dados, normalização em lote, uso de termos de regularização na função objetiva, ver Ref. [19] para obter mais detalhes. Algumas técnicas são especificamente projetadas para modelos da CNN, como incluir um leigo de agrupamento na arquitetura [20]. Em resumo, a CNN é uma boa candidata às candidaturas do STLF, desde que o excesso de fitting possa ser controlado. Haverá discussão sobre este assunto na próxima seção.

Por outro lado, o Fuzzy Time Series (FTS) tem sido usado como um método baseado em aprendizado pattern em muitas aplicações de séries temporais, incluindo STLF. O método FTS foi introduzido em 1993 por Song [21] e desde então eles foram adotados em muitas aplicações de previsão. Por exemplo, Yu em 2005, ver Ref. [22], desenvolveu uma nova versão de FTS ponderado para previsão do mercado de ações. Em 2009, em outro estudo, conduzido por Wang et al., foi introduzido um novo modelo de FTS para previsão de temperatura e índice de ações [23]. Relacionados ao foco deste artigo, há também alguns studies que usaram modelo FTS por combinação com outra lógica. Por exemplo, Sadaei et al. combinaram um algoritmo evolutivo, ou seja, Improved Harmony Search (IHS) com FTS pesado e usaram dados de carga da França e do Reino Unido para validar o modelo [24]. O modelo That apresentou melhor desempenho em comparação com outros métodos fits e alguns outros métodos de última geração. No entanto, em outro estudo, Sadaei et al. combinaram o modelo sazonal auto-regressivo de média móvel fracionada com FTS para STLF. O model híbrido proposto suprimiu todas as suas contrapartes em termos de precisão e desempenho [25]. Além desses estudos, o FTS tem sido adotado para o STLF em muitas outras pesquisas [26e28].

## 1.3. Os desafios ao construir o modelo STLF

Para fornecer um modelo STLF de alto desempenho, algumas considerações precisam ser levadas em conta. Acredita-se geralmente que uma carga de curto prazo é uma variável que é influnada por muitos fatores, por exemplo, dados históricos de carga, dados meteorológicos, como velocidade do vento, precipitação, pressão atmosférica, temperatura e umidade. É quase impossível fazer uma previsão precisa usando um único modelo. Outro problema relacionado à construção de um modelo ROBUSTO de STLF é a complexidade que existe dentro de seus dados. É mostrado em muitas pesquisas que confirm várias características non-lineares e características intra-sazonalidade complicadas nos dados [29,30] (ver por exemplo

Fig. 1).

Essa complexidade é um obstáculo para o modelo certo para a previsão e segundo para a configuração dos parâmetros necessários desse modelo de forma ideal. Como exemplo, o uso de modelos lineares clássicos para STLTF, por exemplo, SARIMA, ARIMA e SARFIMA envolvem pré-processamento de dados de carga maciça e Funções de Correlação Automática e Correlação Parcial, analisando para figurar os conjuntos adequados de hiper-parâmetros do modelo. Mesmo tendo informações da ACF e PACF, é difícil selecionar as melhores combinações de parâmetros [31,32] (ver, por exemplo, Fig. 2).

A Tabela 1 reúne a revisão da literatura discutida até o momento, resumindo vantagens e desvantagens de cada modelo.

Como se pode ver, os modelos fts híbridos têm mostrado bom desempenho para o STLTF, além disso, os modelos baseados em IA, especialmente CNNs, produziram bons results, se e somente se a fitação excessiva pode ser evitada. Outro bom aspecto do deep learning é que a seleção e extração de recursos podem ser automatizadas, incorporadas na rede. Isso resolverá a complexidade da construção do modelo. A partir dessa discussão, pode-se concluir que a combinação de FTS com CNNs controlando o excesso de fitação pode ser uma boa alternativa para o STLTF. Esta é a principal motivação deste artigo. Na próxima seção, os autores discutem sobre as contribuições deste trabalho.

#### 1.4. Contribuição

Como explicado anteriormente, a principal preocupação deste estudo é propor um modelo combinado com alta precisão e extração automatizada de recursos. Como será mostrado mais tarde na Seção 2, o método proposto, além de sua eficácia no desempenho, é fácil de ser implementado com esforço mínimo para toda configuração necessária. No entanto, como discutimos, um dos problemas ao usar o modelo de aprendizagem profunda, incluindo cnns, é a sobre-fitação. Portanto, uma das principais contribuições em nosso método proposto é lidar com esse problema. Existem algumas abordagens propostas em alguns outros trabalhos para resolver a sobre-fitação em CNNs, por exemplo, usando técnicas de abandono [33]. Além do uso da técnica de abandono, usaremos a lógica difusa que é o núcleo do FTS para aplicar a regularização na camada de entrada. A ideia é usar o FTS juntamente com séries temporais originais e convertê-las para o formato de imagens na camada de entrada. Esta abordagem será discutida na Seção 4. Outra motivação para usar o FTS em nosso método proposto é aproveitar o poder da lógica difusa para lidar com sucesso com incertezas que residem em muitas aplicações de séries temporentais não lineares, incluindo o STLTF. Houve uma discussão sobre a utilidade do uso do FTS como método híbrido para STLTF na seção anterior. Além disso, como é mostrado nos próximos Sections, a combinação de FTS com CNNs melhora o desempenho do método proposto.

#### 1.5. Organização do papel

Este artigo é organizado da seguinte forma: A Próxima Seção revisa os trabalhos relacionados e conceitos preliminares de FTS, CNN e lógica difusa; A Seção 3 descreve os dados que foram utilizados neste estudo e, posteriormente, na Seção 4, a metodologia da pesquisa é apresentada. A Seção 5 discute os resultados, e finally, a Seção 6 conclui o artigo.

## 2. Histórico de estudo

Na primeira parte desta Seção, os fundamentos e desfigurações do FTS são revisados e, em seguida, o conceito cnn é explicado. Finalmente, uma breve introdução à lógica difusa é discutida.

### 2.1. Série temporal difusa

Song e Chissom first apresentaram os conceitos de FTS [21,34], onde os valores em uma série de tempo são representados por conjuntos fuzzy [35]. A partir de agora, baseamos nossa apresentação sobre a literatura convencional sobre FTS, com poucas mudanças.

Que  $U$  seja o universo do discurso, geralmente um domínio contínuo. Um conjunto confuso de  $U$  é desafiado da seguinte forma:

$$\mu_{A_i}(x) = \frac{1}{4} \quad \text{onde } x \in U \quad (1)$$

No

onde  $\mu_{A_i} : U \rightarrow [0, 1]$  é a função de associação do conjunto fuzzy  $A_i$ ,  $x \in U$  é um elemento genérico do conjunto fuzzy  $A_i$ ,  $\mu_{A_i}(x)$  é o grau de adesão de  $x$  a  $A_i$ . A integral não deve ser interpretada algebricamente, esta notação simplesmente indica que a  $IA$  é uma coleção de pares ordenados.

Definição: (Song and Chissom) [21]. Deixe uma série temporal  $y_t : t = 1, 2, \dots$  e deixar a faixa de valores de  $y_t$ , um subconjunto de números reais  $R$ , formar o universo do discurso pelo qual conjuntos confusos  $\mu_{A_i} : U \rightarrow [0, 1]$ ,  $i = 1, \dots, N$ , são desfiéis. Se  $F_t$  é uma coleção de conjuntos confusos no tempo, então  $F_t$  é chamado de Fuzzy Time Series desafiado em  $y_t$ .

### 2.2. Chen uni-variate fuzzy time series algoritmo (1996)

Em 1996, Chen propôs um modelo com operadores aritméticos simplificados, ver Ref. [36], substituindo as complicadas operações de composição max-min apresentadas por Song e Chissom [21]. O algoritmo básico proposto por Chen é revisado aqui.

Passo 1: Desafiar o universo do discurso como um intervalo, cujos limites são desafiados da gama de valores nos dados históricos (in-

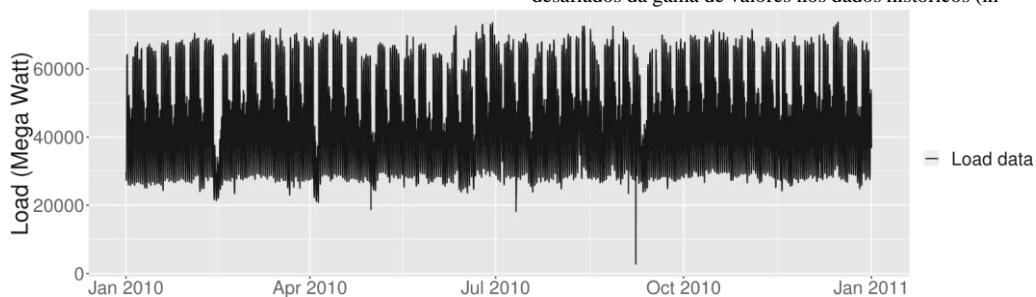


Fig. 1. Dados de carga por hora da Malásia do ano de 2010.

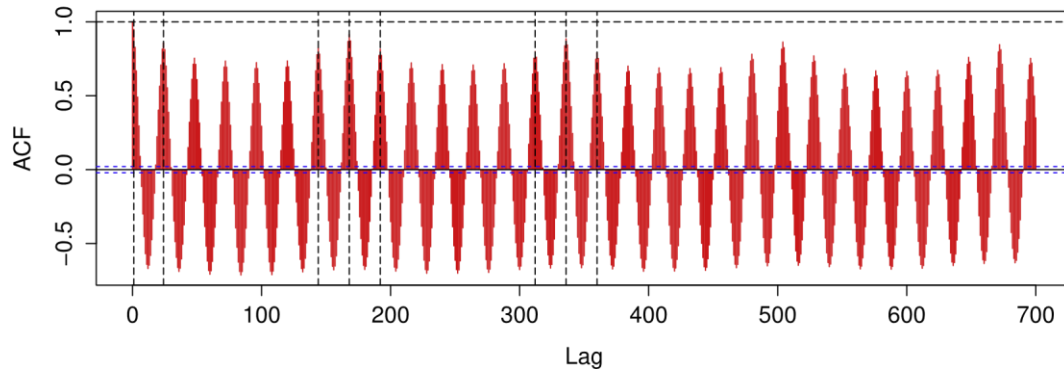


Fig. 2. Correlação automática dos dados de carga de 2010 da Malásia.

Tabela 1  
Modelos STLFL.

Categoria Modelo	Exatidão	Excesso de fitação	Fácil de usar
Método clássico	baixo	Não	Não
Modelos baseados em IA	justo	Sim	Não
Série tempo difusa	justo	Não	Não
FTS híbrido	Muito bom	Não	Não

dados de amostra ou treinamento).

Passo 2: Partição U em intervalos de comprimento igual como

$U = \{u_1, u_2, \dots, u_n\}$

Passo 3: Define fuzzy define  $A_i$  onde  $1 \leq i \leq n$  em U e, em seguida, fuzzify valores nos dados da amostra. Se o valor pertence ao intervalo  $u_i$ , ele é fuzzified para  $A_i$ .

Passo 4: Estabelecer flr de primeira ordem, por exemplo,  $F_{\delta t} = 1/F_{\delta t}$ . Considerando os dados históricos fuzzified, identifique os padrões  $A_i/A_j$ , onde  $1 \leq i, j \leq n$ , nos dados.

Passo 5: Estabelecer flr groups (FLRGs) usando o FLR reconhecido. Esses grupos são estabelecidos pelo agrupamento daqueles FLR which têm o mesmo lado esquerdo. Por exemplo, dada a seguinte FLR,

$A_i/A_j$

$A_i/A_k$

$A_i/A_m$

$A_j/A_n$

$A_j/A_m$

$A_m/A_i$  onde  $1 \leq i, j, m \leq n$ , produza estes FLRG:

$A_i/A_j; A_k; \text{Sou}$

$A_j/A_n; A_m A_m/A_i$

Passo 6: Previsão e desfiação. Dado o valor atual, calcule o valor de previsão com base nos conjuntos fuzzy no lado direito do FLRG. Se em tempo t,  $y_{\delta t}$  é fuzzified como  $A_i$  e o RHS contém um conjunto fuzzy na sequência i.e.  $A_i/A_j$ , onde  $1 \leq i, j \leq n$ , então

$1^{\circ} 1/4 M_j$

se na hora t, o RHS contém mais de um conjunto confuso, ou seja.

$A_i/A_{j1}; A_{j2}; \dots; A_{jm}$ , então.

$1/4 M_{j1} \text{ p } M_{j2} \text{ p } M_{j3} \text{ p } \dots \text{ p } M_{jm} \text{ f t l}$

m

se no momento t, o RHS não contém conjuntos fuzzy na sequência ou seja,  $A_i/\#$  em seguida

$1^{\circ} 1/4 M_i$

onde  $f_{\delta t}$  é o valor defuzzificado do FTS  $F_{\delta t}$  e  $M_i$  corresponde ao valor desfiação de  $A_i$ , geralmente o ponto médio do intervalo associado a este conjunto confuso.

No modelo proposto, ele requer apenas o conceito fuzzification gerando a sequência fts e não haverá necessidade de estabelecer FLRs, FLRGs e etapa de desfuzzificação. Assim, ao following algoritmo Chen até o Passo 3, o FTS pode ser gerado, fulfilling nossos requisitos para o método proposto. Mais tarde, na Seção 4, será discutido como o FTS pode ser usado.

### 2.3. Redes neurais convolucionais

Redes Neurais Convolucionais (CNNs, ou ConvNets) é uma classe de redes neurais artificiais profundas e alimentares que empregam uma variedade de perceptrons multicamadas projetados para exigir o mínimo de pré-processamento. Eles foram incentivados por processos biológicos em que o padrão de conectividade entre neurônios é inspirado na organização do córtex visual animal. Neurônios corticais individuais respondem a estímulos apenas em uma região limitada do field visual identificado como o field receptivo. Os fields receptivos de diferentes neurônios se sobrepõem parcialmente, de modo que cobrem toda a field visual [37].

#### 2.3.1. Contorno de arquitetura

Redes Neurais regulares não escalam bem para dados dimensionais elevados.

Por exemplo, uma imagem de tamanho, digamos 200 200 3, levaria

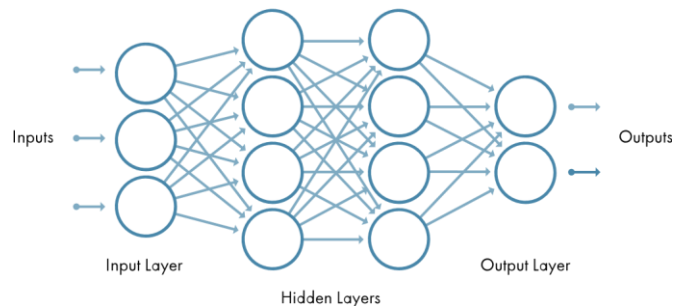


Fig. 3. Um exemplo de rede neural profunda.

aos neurônios que possuem 200 200 3 1/4 120; 000 pesos. Além disso, quase certamente queremos ter algumas camadas desses neurônios, portanto os parâmetros se somariam muito rápido (ver Fig. 3). Claramente, essa conectividade completa é ineficiente e o grande número de parâmetros rapidamente levaria a excesso de fitação.

As CNNs se aproveitam do fato de que a entrada consiste em dados e restringem a arquitetura de forma mais prática. Em particular, as CNNs incluem quatro camadas diferentes da seguinte forma: A camada de Convolução: AS CNNs herdaram seu nome do operador de "convolução". O objetivo principal da Convolução no caso de uma CNN é extrair recursos dos dados de entrada. A convolução preserva a conexão espacial entre os itens de dados, aprendendo recursos de dados utilizando pequenos quadrados de dados de entrada. Cada entrada de dados pode ser contada como uma matriz de alguns valores. No vocabulário da CNN, um 'filter' é a matriz que se executou em dados de entrada, e a matriz formada por deslizar o filter sobre a imagem e computação do produto ponto é nomeado o

'Recurso complicado'. Uma operação adicional chamada Unidade Linear Rectified (ReLU) é aplicada após cada operação de Convolução. ReLU é uma operação não linear. Sua saída é desafiada e mostrada na Fig. 4.

$$1/4 \text{ no máximo } 00; xTH \quad (2)$$

ReLU é um processo de elemento-wise (implementado por pixel) e substitui todos os valores negativos no mapa de recursos por zero. A ideia da ReLU é introduzir a não linearidade nas CNNs. Recentemente, a ReLU tornou-se muito popular, pois foi provado que ela teve muita melhoria na convergência da função Tangente Hiperbólica (ver Experimento 1). Isso se deve principalmente à sua forma simples (ver eq. (2)). Assim, evita e recita o problema do gradiente de desaparecimento. Quase todos os modelos de aprendizagem de deep usam ReLU hoje em dia. Em contraste, duas outras funções de ativação que eram populares no passado, a saber, Sigmoid e Tangente Hiperbólica, estão sofrendo de lenta convergência e problemas de gradiente de desaparecimento, que causam muitos problemas à train, degradam a precisão e o desempenho de um modelo de rede neural profunda. A outra vantagem da ReLU é a esparsidade. A esparsidade ocorre quando  $x < 0$ . Quanto mais unidades existirem em uma camada, mais esparsas é a representação resultante. Sigmoids e Tangente Hiperbólica, por outro lado, são constantemente esperados para produzir algum valor não-zero resultando em representações densas. Representações esparsas parecem ser mais vantajosas do que representações densas.

A camada de agrupamento: O pooling espacial (também chamado de sub-sampling ou down-sampling) reduz a dimensão de cada mapa de características, mas de alguma forma preserva as informações mais valiosas. O Pooling Espacial pode ser de vários tipos: Max, Average, Sum etc. No caso do Max Pooling, uma área espacial é determinada e o maior elemento do mapa de características rectified dentro dessa janela é tomado. A ideia do Pooling é reduzir frequentemente o tamanho espacial da representação de entrada [38].

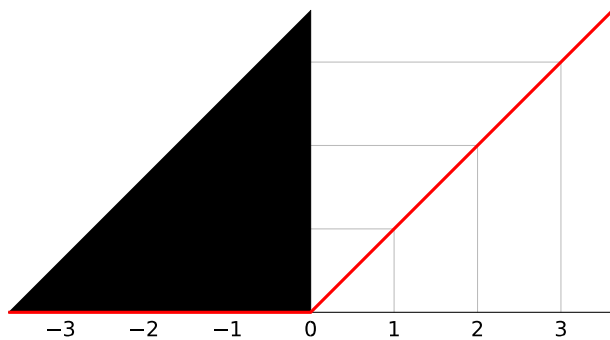


Fig. 4. Unidade linear rectificada.

O pooling torna as representações de entrada (dimensão do recurso) menores e mais controláveis

O pooling reduz o número de parâmetros e cálculos na rede, portanto, controlando o excesso de fitação

O pooling torna a rede invariante a pequenas transformações, distorções e traduções nos dados de entrada (uma pequena distorção na entrada não mudará a saída do Pooling porque tomamos o valor máximo/médio em um bairro local).

Camada totalmente conectada: A camada totalmente conectada é um perceptron multi-camada tradicional que emprega uma função de ativação softmax na camada de saída. O termo The "Totalmente Conectado" significa que cada neurônio na camada anterior está conectado a todos os neurônios na camada seguinte. A saída das camadas convolucionais e de agrupamento tornam características de alto nível dos dados de entrada. A ideia da camada Totalmente Conectada é usar essas proezas para classificar os dados de entrada em várias classes em problemas de classificação de classe. Para o problema de regressão, no entanto, a última camada Totalmente Conectada está apenas ligada a uma saída linear [39].

Camada de abandono: O abandono é um método aplicado para controlar o excesso de sfitting em redes neurais, restringindo co-adaptações complexas em dados de treinamento. Nesta técnica de regularização, algumas unidades (escondidas e visíveis), aleatoriamente, por uma taxa de know-out específico serão retiradas, durante a fase de treinamento [33].

### 2.3.2. arquitetura de CNNs para problema de regressão

Como é conhecido, na topologia da CNN, a Convolução juntamente com as camadas de pooling atuam como Extratores de Recursos a partir dos dados de entrada (por exemplo, imagens). Como estamos usando CNNs para resolver um problema de regressão, em vez de uma classificação, em nosso método proposto, a camada totalmente conectada será usada como classifier, exceto a camada final totalmente conectada que será usada para a previsão de valor em vez de previsão de classe de rótulo.

### 2.4. Lógica difusa

A lógica difusa é uma espécie de lógica de muito valor em que os valores da verdade das variáveis podem ser qualquer número real entre 0 e 1. Está correndo para lidar com a teoria da verdade parcial, onde o valor da verdade pode variar dentro completamente verdadeiro e falso. Em contraste, na lógica booleana, os valores da verdade das variáveis podem ser apenas os valores inteiros 0 ou 1 [40]. Além disso, se forem aplicadas variáveis linguísticas, esses graus poderão ser regulados por funções específicas (adesão). O termo lógica difusa foi introduzido com a proposta de 1965 da teoria dos conjuntos confusos por Lotfi Zadeh [35]. A lógica difusa ainda era estudada desde a década de 1920, como lógica valorizada por infinita notavelmente por Lukasiewicz e Tarski. A lógica clássica só permite inferências que são verdadeiras ou falsas [41]. Em nossa metodologia proposta, no entanto, a lógica difusa já foi incorporada no FTS. Assim, o método proposto implicitamente se isola de suas características.



### 3. Dados e configurações

Os dados de carga de curto prazo são um dos tipos mais complicados de séries temporais para previsão. Esses dados são considerados não lineares com muitas intra-sazonalidade que dificultam ainda mais a resolução deste problema (ver Fig. 2). Devido à dificuldade que existe no uso de dados de carga de curto prazo, e para mostrar como o método proposto era capaz de resolver problemas complexos de previsão, os autores selecionaram para validar o modelo proposto. Além disso, como discutido na seção de introdução, ter UM STLTF preciso é vital para a maioria das empresas de fornecimento de energia, ad foi a segunda razão para a escolha deste estudo de caso. Portanto, os dados de carga horária da empresa de fornecimento de energia da cidade de Johor, na Malásia, gerados em 2009 e 2010 (ver Fig. 1) foram implantados para cumprir os objetivos deste estudo.

Aqui, o  $\hat{y}_i$  real é o valor real de carga no tempo  $i$ , e o  $\hat{y}_i$  previsto é o valor final previsto no tempo  $i$ .

#### 3.2. Configuração do sistema

Para verificar e validar o método proposto, é utilizada uma plataforma Keras versão 2.1.5 com o back-end da versão 1.8.0 do Tensorflow em um PC com cpU Intel Core i7, CPU de 2,3 GHz, memória de 8 GB e disco rígido de 1 TB com sistema operacional Linux.

#### 4. Proposta de metodologia

A metodologia proposta deste estudo inclui três fases. A primeira fase é sobre o processamento da série temporal para deixá-las prontas para converter séries temporais multivariadas em imagens multicanais. A segunda fase explica a forma como as imagens de entrada são construídas a partir de saídas da fase I e finally, na fase III, explica-se a topologia da CNN proposta.

##### 4.1. Fase I: processamento de dados

Nesta seção, o termo  $t$  refere-se ao estado atual da época para séries temporais de carga, temperatura e índices de conjuntos difusos. Assim, pode ser qualquer índice de tempo desde o início do conjunto de treinamento mais tamanho de largura de imagem até o comprimento do conjunto de treinamento. Para o treinamento do modelo proposto, como com o treinamento de todos os outros modelos, os dados até o tempo podem ser usados para prever o valor de carga para o tempo  $t + 1$ , que se supõe ser desconhecido (em nossa metodologia apenas séries temporais de  $t + 1$  a  $t$  serão usadas para calcular a previsão para cada  $t + 1$ , onde  $n$  será desafiado mais tarde nesta parte). Agora, por esta explicação, assumir o estado atual da série temporal é  $t$  e o estado-alvo de previsão é  $t + 1$ . Para cada série de tempo em mãos, por exemplo, dados de carga originais, temperatura e índices de conjuntos difusos em séries temporestáticas geradas, atribuem  $j = 1, 2$  e  $3$ , respectivamente e depois, realize o seguinte:

conjunto  $j = 1$  e até  $j$  número de variáveis em séries temporentas multivariada (em nosso caso de uso é  $3$ ), fazer:

e selecionar o valor da série original de tempo de carga no momento  $t + 1$  como uma previsão (valor-alvo) que corresponde para aquele conjunto de imagens que serão criados em próximos passos.

e decidir sobre o tamanho das imagens, digamos  $n \times n$  (em nosso caso de uso selecionamos  $n = 32$ ). e definir  $i = 0$ .

e enquanto  $t - 1 \geq 0$  do

estudo. A série de tempo de temperatura foi acoplado com dados de carga por hora, para melhorar a precisão do modelo. Estas séries de tempo de temperatura foram plotadas em Figs. 5 e 6, respectivamente.

Além dessas parcelas informativas, uma lista curta de dados de carga juntamente com dados de temperatura e fuzzified é apresentada na Tabela 2. Ressalta-se que, para a produção de FTS, neste estudo, implementamos o método apresentado no Ref. [42] para particionar o universo do discurso. Os dados são divididos em conjuntos de trem e teste conforme detalhado na Tabela 3.

#### 3.1. Critérios de comparação

O Erro percentual absoluto (APE), erro percentual absoluto médio (MAPE) e erro quadrado de média raiz (RMSE), introduzido no Eqs. (3)e(5), são os fatores mais importantes a serem utilizados quando se comparam aos resultados dos modelos em problemas de STLTF. Neste estudo, a comparação se baseia nessas métricas, que são descritas da seguinte forma:  $APE_i = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100$  (3)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

Neste estudo também foram utilizadas medidas baseadas em erros relativos. Para calcular mdRAE first  $n$  deve ser obtido da seguinte forma:

$$rt = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}$$

onde  $f_i$  é o valor previsto obtido por meio de uma previsão de modelo de referência (modelo de benchmark). A prática main é usar um modelo ingênuo como modelo de referência, ou seja,  $f_i = y_i$ . Em seguida, mdRAE é calculado como de acordo com

$$MdRAE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (6)$$

\* a partir de  $t_i$ , selecione  $n$  valores anteriores na sequência de séries temporais e crie uma lista de sequência deles, ou seja,  $list_i$ . Assim,  $list_i$  contém itens de  $itemti \setminus P$  para  $itemti$ .

\* cada vez mais classificar  $list_i$  para criar outra lista classificada, ou seja,  $slist_i$ .

Um exemplo simplificado com  $n = 1/4 \cdot 4$  pode ser:  $list_i = 1/4 \cdot 1/230361 \cdot 29155$

$29155 \cdot 28031$

e sua lista classificada correspondente pode ser:

$slist_i = 1/4 \cdot 1/228031291552915530361$

\* usar  $list_i$  e  $slist_i$  para estabelecer um  $matrix_i$ ,  $j$  para  $j$ th canal de imagem correspondente ao índice  $i$  da seguinte forma:

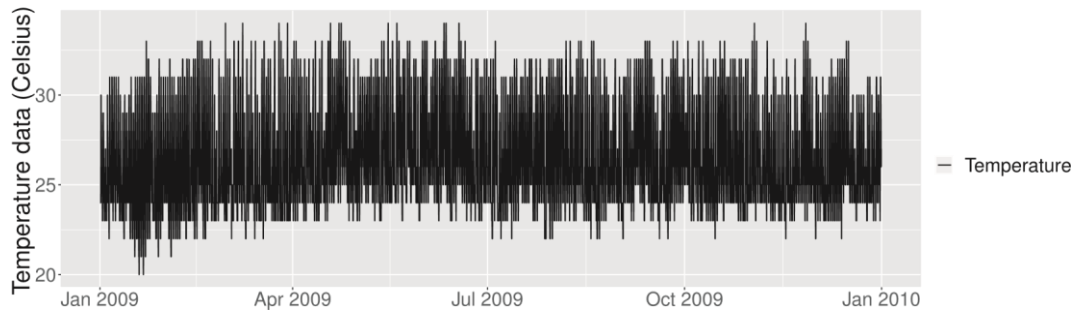


Fig. 5. Dados de temperatura do ano de 2009.

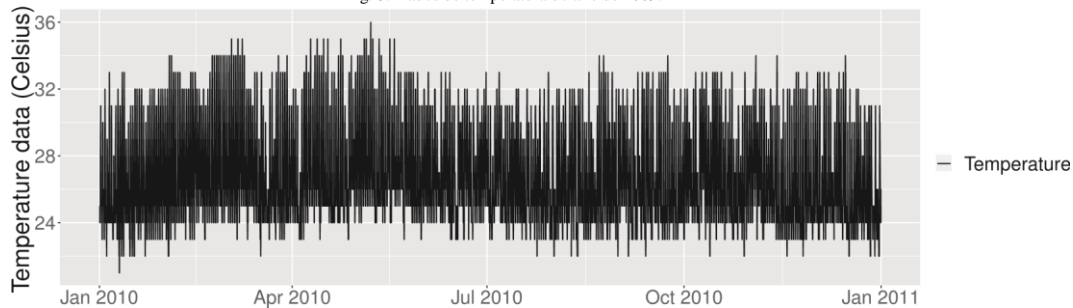


Fig. 6. Dados de temperatura do ano de 2010.

Tabela 2

Dados de temperatura, carga e carga fuzzified.

Índice de data e hora	Temperatura (Celsius)	Carga (MW)	Fuzzified
01/01/09 01:00 .m.	24	30,360	A2
01/01/09 02:00 a.m.	24	29,155	A2
01/01/09 03:00 a.m.	24	28,086	A2
01/01/09 04:00 a.m.	24	28,031	A2
01/01/09 05:00 a.m.	24	27,730	A1
01/01/09 06:00 a.m.	24	30,490	A2
01/01/09 07:00 a.m.	24	37,081	A4
01/01/09 08:00 a.m.	24	46,975	A6
01/01/09 09:00 a.m.	25	57,261	A8
01/01/09 10:00 a.m.	28	63,679	A10
01/01/09 11:00.m.	30	65,123	A10
01/01/09 12:00.m.	30	65,142	A10
01/01/09 01:00 p.m.	30	64,745	A10
01/01/09 02:00 p.m.	28	64,783	A10
«	«	«	«
31/12/09 03:00 p.m.	30	65,003	A10
31/12/09 04:00 p.m.	30	60,655	A10
31/12/09 05:00 p.m.	29	51,507	A9
31/12/09 06:00 p.m.	28	50,743	A7
31/12/09 07:00 p.m.	27	51,317	A7
31/12/09 08:00 p.m.	27	48,414	A7
31/12/09 09:00 p.m.	26	39,262	A4
31/12/09 10:00 p.m.	26	35,809	A3
31/12/09 11:00.m.	26	33,426	A3

atribua um  $vector_k$  com tamanho  $n$  de tal forma que o elemento  $pth$  de

$vector_k$  será 1 e outros elementos de  $vector_k$  serão zero.

verticalmente empilhar todos os  $vector_k$  obtidos:  $1 \times k \times n$  para estabelecer

$matrix_i$ ;  $J$ .

Segundo nossos exemplos numéricos de  $list_i$  e  $slist_i$ ,  $vector1$ ;  $vector2$ ;  $vector3$  e  $vector4$  serão:

$vector1 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$   
 $vector2 = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$   
 $vector3 = \begin{bmatrix} 1 \\ 4 \\ 6 \end{bmatrix}$   
 $vector4 = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$

Tabela 3

$selectfind$  a posição  $ofk$   $klistinislist$ : 1 inamelyk  $n$ , i.e.,  $plk$





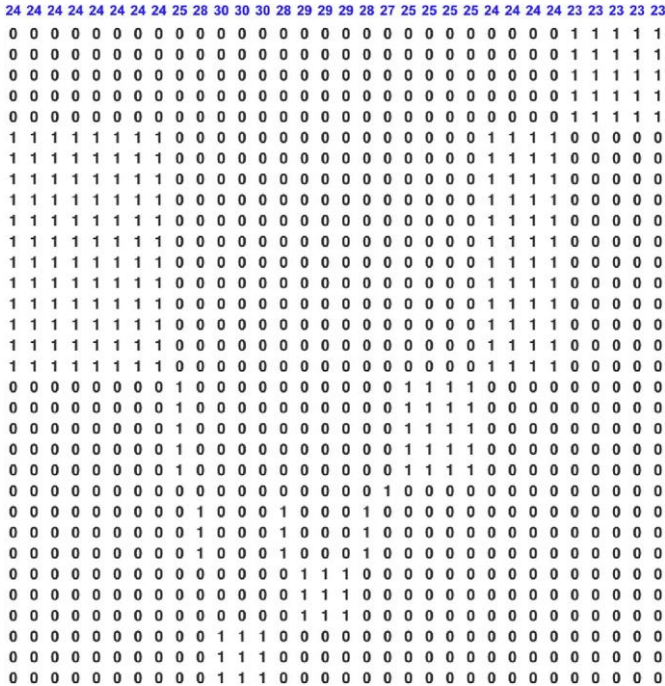


Fig. 8. Matriz de dados de temperatura por hora do ano de 2010 para i 1/4 8728 e n 1/4 32.

série de imagens. Para este objetivo, selecionamos um modelo CNN com uma arquitetura da seguinte forma;

tamanho da imagem 1/4 32

tamanho do lote 1/4 100

número de épocas 1/4 20 taxa de

aprendizagem 1/4 0,001

Fig. 9. Matriz de dados de carga por hora embacados do ano de 2010 para i 1/4 8728 e n 1/4 32.

a forma de entrada será 1/232; 32; número de canal, em nosso exemplo número do canal 1/4 3.

A rede utilizada tem 2 camadas convolucionais e 5 totalmente conectadas. Há uma camada de abandono com a taxa de 40%, após a segunda totalmente conectada, durante o treinamento, para evitar excesso de fitação. Todas as camadas convolucionais são seguidas por Unidades Lineares Rectified (ReLU), e as duas mais antigas são seguidas por uma piscina máxima de 2 por 2 com um passo de 2. Por fim, o treinamento para regressão é realizado com uma função de resumição de 1 via (ver Fig. 11).

## 5. Resultados e discussões

Para avaliar a eficácia do método proposto, nesta Seção projetamos quatro Experimentos para testar o modelo a partir do ponto de vista diferente como follows:

### 5.1. Experimento 1: tempo de convergência para várias funções de ativação

Como explicado na Seção 2, o ReLU é uma das melhores funções de ativação que está sendo usada em muitos modelos de deep learning. Neste estudo, a ReLU tem sido utilizada, porém projetamos essa experiment para

confirmar outras findings sobre a superioridade da ReLU em relação a outras funções de ativação em nossos estudos de caso.

Como é evidente na Tabela 4, o tempo de convergência ao usar o ReLU é muito menor do que outros (menos da metade do tempo em comparação com a Tangente Hiperbólica). Esses resultados estão em consonância com os de estudos anteriores.

### 5.2. Experimento 2: determinar o tamanho das imagens

Como pode ser visto a partir dos resultados da Tabela 5, o método proposto não é muito sensível aos tamanhos de imagem, se o tamanho da imagem for selecionado na faixa lógica. Como exemplo, com um tamanho de imagem de 224, o MAPE no conjunto de testes é de 3,77 para o ano de 2009 que tem

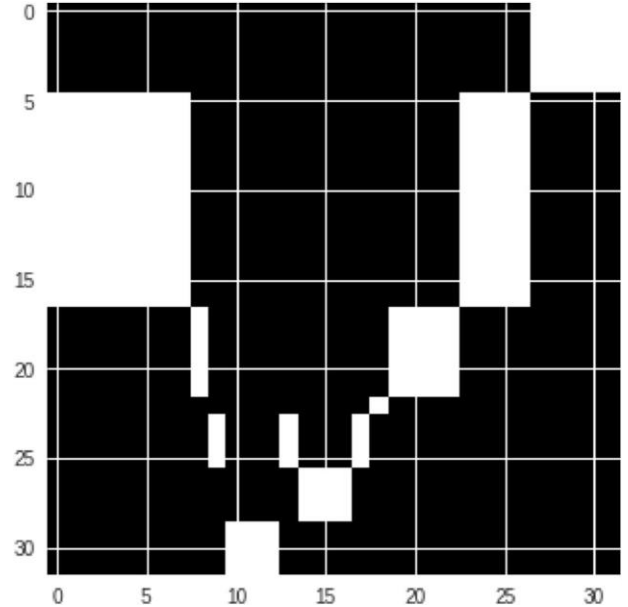


Fig. 10. Imagem de dados de carga por hora embacados do ano de 2010 para i 1/4 8728 e n 1/4 32. Nota: nos Figs. 8 e 9, a primeira linha está lá apenas para ser como uma diretriz, não pertence à matriz.

o pequeno desvio (apenas 0,75) dos resultados quando a imagem é 32. Isso mostra a robustez da forma como os dados são explicados pelas imagens e também o poder das CNNs no mapeamento de uma relação complexa entre imagens de entrada e saída, com pequenos erros.

### 5.3. Experimento 3: Sinalização das camadas de entrada Neste experimento,

é mostrado que, para o nosso caso de uso, usando todos

Tabela 5

MAPE do método proposto com base em vários tamanhos de imagem.

Tamanho da imagem	32 32	48 48	64 64	128 128	224 224
Malásia 2009	3.02	3.26	3.45	3.52	3.77
Malásia 2010	2.89	3.43	3.49	3.34	3.71

três canais de entrada juntos produzem melhores resultados. Para facilitar a apresentação dos resultados, estamos usando as seguintes abreviaturas:

Série original do tempo 1/4 OTS

Fuzzy Time Series 1/4 FTS

Série de tempo de temperatura 1/4 TTS

Como é evidente na Tabela 6, o melhor cenário possível é usar todas as camadas juntas. Os segundos melhores resultados aconteceram quando OTS e FTS são usados juntos. O método é tão forte que mesmo utilizando apenas imagens com o canal único, ou seja, apenas OTS, ele é capaz de produzir resultados consideráveis.

#### 5.4. Experimento 4: comparação da precisão dos modelos de estado da arte no conjunto de dados de teste

Repetimos 30 vezes os experimentos com os métodos propostos para o conjunto de dados de teste e, em seguida, calculamos mape médio, RMSE e MdRAE. Os resultados obtidos com outros métodos STLF são reunidos e apresentados nas Tabelas 7e9. Para desenvolver a Média Móvel Integrada Sazonal Autoregressiva (SARIMA) em dados que

Tabela 6  
MAPE do método proposto com base no uso de diferentes canais de imagem.

Tipo de imagem	OTS	OTS & FTS	OTS & FTS & TTS	FTS & TTS
Malásia 2009	3.34	3.16	3.02	3.20
Malásia 2010	3.19	3.15	2.98	3.18

empregamos uma função famosa no pacote de previsão de R, ou seja, `auto.arima`, que realiza uma pesquisa sobre o melhor modelo possível dentro da mesma faixa lógica de parâmetros [43]. Acreditamos que o modelo linear selecionado, SARIMA(1,0,3)(2,1,2) [24] para a Malásia 2009 e SARIMA (1,0,1)(1,1,2) [24] para a Malásia 2010, por essa abordagem poderia produzir os melhores resultados entre todos os modelos lineares. Outro modelo selecionado, neste experimento, é o Probabilistic Weighted Fuzzy Time Series (PWFTS) introduzido em `pyFTS` [44]. `pyFTS` é uma nova biblioteca FTS que abrange uma ampla gama de modelos neste domínio, que recentemente é equipado com otimização de hiperparâmetro também. Portanto, os resultados relatados do PWFTS podem ser vistos como as melhores saídas possíveis de FTS. Além disso, outros modelos de FTS foram selecionados da biblioteca `pyFTS` no experimento, ou seja, Séries Tempo Difusas Ponderadas (WFTS) e Séries Tempo Difusas Ponderadas Integradas (IWFTS) e os resultados obtidos são coletados nas Tabelas 7e9. Além dos modelos lineares e FTS de última geração para STLF, testamos um modelo muito importante de aprendizado profundo para previsão de sequência, ou seja, redes LSTM (Long Short Term Memory, memória de longo prazo). LSTM é um tipo especial de RNNs (Redes Neurais Recorrentes), capazes de aprender dependências de longo prazo. Eles foram apresentados em Ref. [45] e foram refiados e popularizados por muitos pesquisadores e praticantes em uma grande variedade de problemas, especialmente, previsão de séries temporizadas. Neste estudo foram selecionadas

várias configurações para LSTM e os melhores resultados são produzidos quando, foram selecionados os 168 (one week) antigos atrasos de séries temporais para todos os dados originais, temperatura e índices de séries temporais difusos. Para os modelos LSTM 1, 2 e 3, foram selecionados como entrada as 24, 48 e 72 antigas lags. Para todos os casos de LSTM, os dados originais no momento  $t$  e  $t-1$  foram targeted para ser previsto.

Conforme os resultados obtidos sugeridos, o melhor desempenho pertencia ao método proposto, seguido pelo LSTM e PWFTS. Embora a LSTM estivesse na segunda posição em termos de precisão neste experimento, o método proposto poderia alcançar 22,84% de precisão melhor do que o LSTM, para o ano de 2009 como exemplo. É cerca de 27,81% melhor que o PWFTS em termos de comparação MAPE. Figs. 12 e 13 fornecem uma melhor visão para os três principais modelos STLF usados neste experimento. Além disso, a RMSE para o método proposto é inferior a 460 unidades a menos que o PWFTS para o ano de 2010. Para o ano de 2010, o método proposto produziu resultados com resultados 20% melhores que o LSTM utilizando métrica MdRAE.

#### 5.5. Experimento 5: use redes residuais e visuais do Grupo de Geometria

Como é mostrado na antiga seita, a CNN com topologia proposta já superou alguns modelos avançados de séries temporais para STLF com melhores resultados. No entanto, a próxima pergunta válida e verdadeira é se os resultados podem ser melhorados usando derivação mais avançada e nova de modelos da CNN. Esta seção visa abordar esta questão. Para isso, empregamos alguns desses modelos, como ResNet 50, VGG 16, etc., nos mesmos dados. Na verdade, ResNet 50 é a CNN que venceu a competição ILSRVC 2015 e

Tabela 7  
Avaliação do método médio MAPE do método the proposto e alguns outros modelos em STLF.

Métodos	Malásia 2009	Malásia 2010
FTS-CNN (método proposto)	3.02	2.89
SARIMA	4.68	4.23
PWFTS	3.86	4.00
WFTS com diferenças	5.33	9.09
IWFTS com diferenças	5.32	7.69
PWFTS com diferenças	4.59	5.83
LSTM	3.71	3.45
LSTM modelo 1	4.55	4.21
LSTM modelo 2	4.11	4.23
LSTM modelo 3	3.93	3.88

Tabela 8

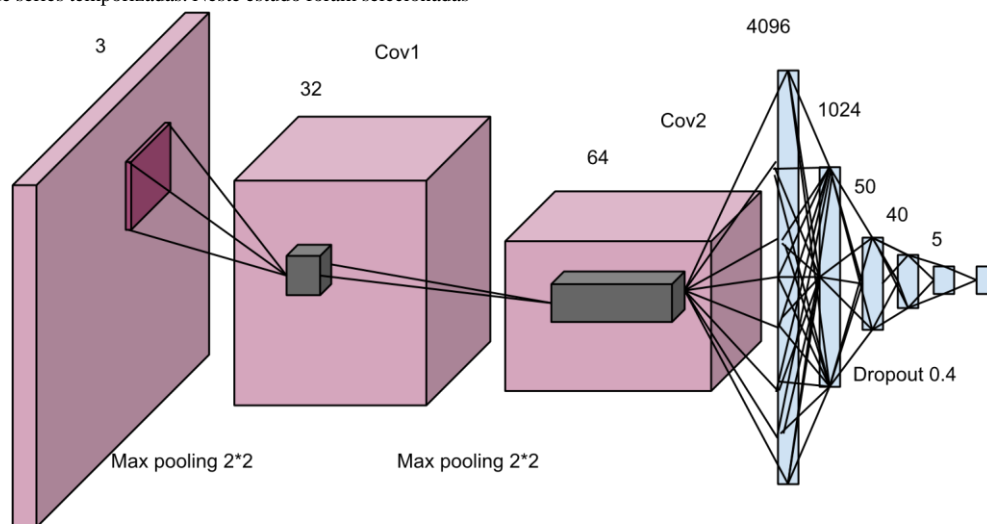


Fig. 11. CNN architecture used for STLF.

Tabela 4  
Tempo de convergência (em minutos) para várias funções de ativação.

Função de ativação	Sigmóide	Tangente hiperbólico	relu
Tempo de convergência para o conjunto de trens da Malásia 2009	52.12	45.32	17.23
Tempo de convergência para o conjunto de trens da Malásia 2010	54.51	47.43	18.01

Avaliação da RMSE média do método proposto e de alguns outros modelos em STLf.

Métodos	Malásia 2009	Malásia 2010
FTS-CNN (método proposto)	1777.99	1702.70
SARIMA	2763.66	2501.25
PWFTS	2230.91	2162.57
WFTS com diferenças	2930.36	4419.11
IWFTS com diferenças	2961.17	4663.17
PWFTS com diferenças	2987.40	3797.35
LSTM	2194.19	2037.49
LSTM modelo 1	2689.42	2044.68
LSTM modelo 2	2413.60	2483.71
LSTM modelo 3	2317.88	2279.23

VGG16, mude a parte superior para usar a entrada 3 camada VGG16, mude a parte superior para usar a entrada 3 camadas, com apenas 5 blocos

Embora se esperasse que esses modelos avançados de deep learning pudessem produzir bons resultados, tanto ResNet 50 quanto

Tabela 9  
Avaliação do MdRAE médio do profissional

O STLf.	método posed e alguns	outros modelos em
Métodos	Malásia 2009	Malásia 2010
FTS-CNN (método proposto)	0.4857	0.4537
SARIMA	0.7504	0.6655
PWFTS	0.6107	0.6319
WFTS com diferenças	1.0838	2.1105
IWFTS com diferenças	1.0780	1.2185
PWFTS com diferenças	0.5284	0.7638
LSTM	0.5981	0.5435
LSTM modelo 1	0.7313	0.6622
LSTM modelo 2	0.6610	0.6673
LSTM modelo 3	0.6273	0.5991

superou o desempenho humano no conjunto de dados ImageNet [46]. ResNet 50 é um dos métodos empregados em experimentos. Neste experimento, os seguintes cenários de uso do ResNet 50 foram testados em dados:

ResNet 50, treine totalmente: Neste cenário, ajustamos os parâmetros necessários, durante a fase de treinamento, e testamos o modelo em dados de teste.

ResNet 50, trem de última camada: Neste caso, usamos a técnica de aprendizagem de transferência, pela qual usamos apenas todos os pesos pré-ajustados, exceto a última camada. Neste caso, a última camada totalmente conectada foi treinada e o modelo então testado.

ResNet 50, mude top para usar imagens de entrada com o tamanho de 32 32 3: A arquitetura do ResNet 50 não permite usar imagens com altura e peso inferiores a 224. No entanto, desenvolvemos uma arquitetura another do ResNet alterando a camada superior para aceitar imagens com o tamanho menor. Neste caso, apenas 5 blocos de ResNet foram empregados.

Outra metodologia principal e nova derivada das CNNs é o VGG 16, (também chamado oxfordnet), que é uma arqui-abertura nomeada em homenagem ao Visual Geometry Group de Oxford, que a desenvolveu. Foi usado para ganhar a competição ILSVR (ImageNet) em 2014. Até hoje ainda é considerado um excelente modelo de visão, embora tenha sido um pouco superado por modelos more avançados como Inception e ResNet. Ele foi empregado neste experimento seguindo os mesmos cenários que foram usados na ResNet na parte anterior [47].

VGG16, totalmente trem

VGG16, último trem de camada

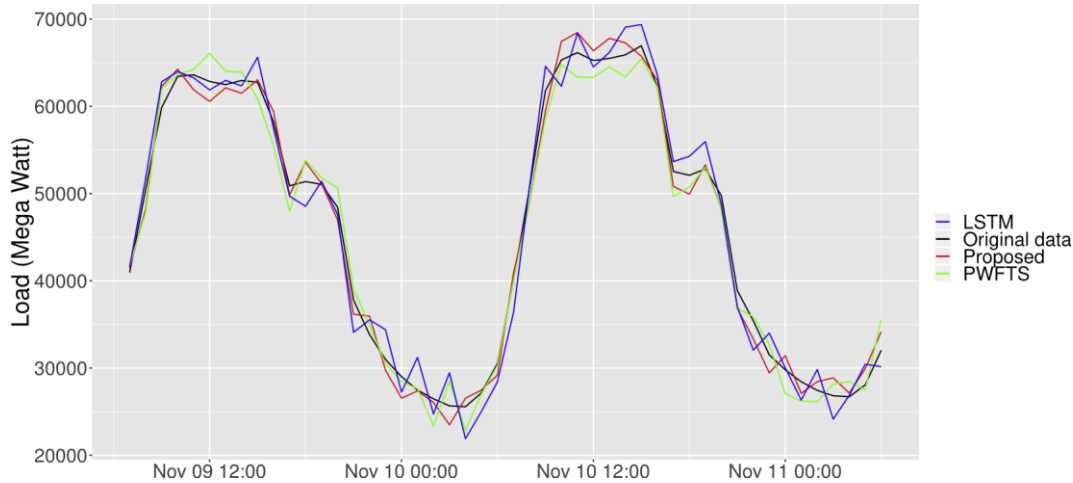


Fig. 12. Comparação prevista e série temporal real entre os três principais métodos para o conjunto de dados de teste do ano de 2010.

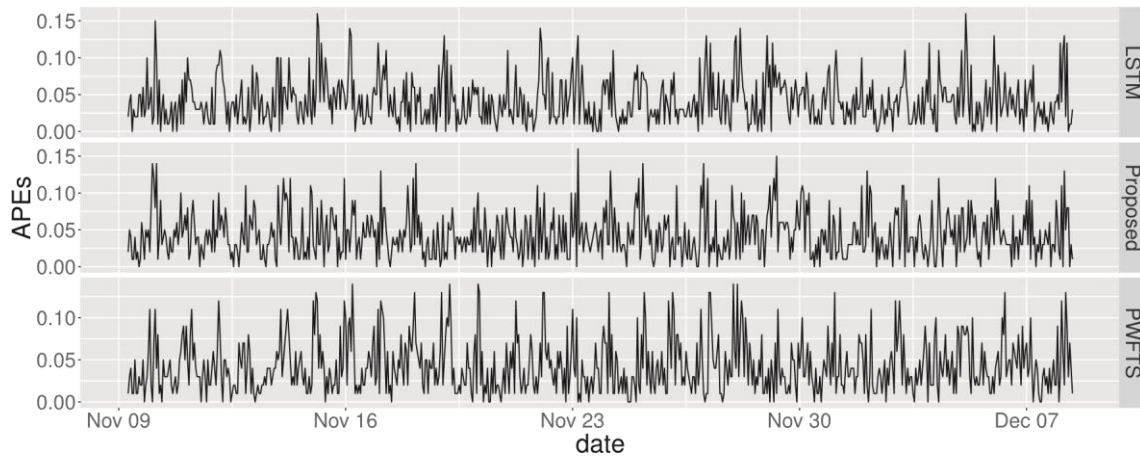


Fig. 13. Comparação de APES entre os três principais métodos para conjunto de dados de teste.

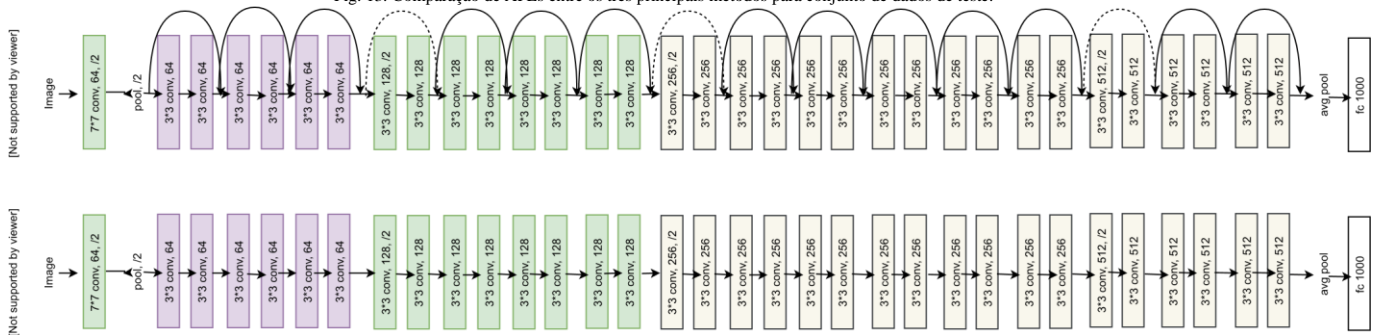


Fig. 14. A arquitetura resnet evita o problema de dados de desaparecimento (imagem do ResNet).

O VGG 16 falhou nesse sentido e o resultado é contra-intuitivo. A próxima seção forneceu alguma explicação sobre possíveis razões.

#### 5.5.1. Interpretação do resultado da ResNet 50 e VGG 16 sobre dados

Os achados da seção anterior sugerem que a principal razão para as falhas foi o pequeno tamanho do conjunto de dados de treinamento. O pequeno tamanho do conjunto de dados significava que não era possível ajustar o grande número de parâmetros que esses modelos avançados exigiam. Nosso conjunto de dados para cada ano de dados de carga contém imagens about 8000 que não são suficientes para treinar CNNs muito profundas, como resNet (por exemplo, a complexidade ResNet 34 pode ser vista na Fig. 14).

Mais coleta de dados é necessária para determinar exatamente se há alguma possibilidade de melhorar os resultados usando ResNet e VGG.

Outra grande fonte de incerteza está na necessidade de empregar muitos parâmetros para nosso simples problema de manipulação de imagens. O caso de uso de STLF proposto, em comparação com muitos aplicativos complicados de visão computacional, como detecção de objetos e reconhecimento de face, problemas de classificação de imagem, que envolvem o manuseio de várias imagens com diferentes tamanhos, cores, objetos que ResNet e VGG podem resolver, é muito simples. A variação das imagens em nosso caso de uso não é muito, por isso não há necessidade de tune muitos parâmetros (por exemplo, para ResNet 50 há 25.610.216 parâmetros a serem ajustados). Em nosso último experimento, embora os resultados não tenham sido satisfatórios a serem relatados, a versão simplificada do ResNet e do VGG com menos camadas, ou seja, 5 blocos, produziu melhores resultados do que a ResNet 50 completa. Essa



findagem também está de acordo com a regra da parcimônia, que diz que a solução mais simples tende a ser a correta.

## 6. Conclusões

Uma nova abordagem foi sugerida neste artigo para previsão de carga de curto prazo (STLF) com base em um modelo híbrido da CNN e FTS. Foi demonstrado que a ideia de converter séries temporais multivariáveis em imagens e, em seguida, usar CNNs para STLF pode ser considerada útil e promissora. Além disso, a lógica difusa que é usada implicitamente durante a construção do FTS a partir dos dados originais de carga ajudou a diminuir o efeito negativo da sobrefitagem, que foi reflectada nos resultados dos testes. Este estudo tentou avançar a nova ideia de usar imagens em vez de séries temporais. Os resultados mostraram que sua ideia realmente teve sucesso no STLF. Outra findagem importante foi a importância dos dados de temperatura que poderiam ser acoplados com informações difusas na melhoria do desempenho do modelo. As findings desta pesquisa podem ser usadas para referenciar when realizando estudos adicionais sobre a utilização de métodos revisados propostos usando cns de vários tipos com diferentes tipologias, bem como outros tipos de FTSs, como FTSa ponderado para outras aplicações de séries temporal, como carga de médio e longo prazo fo reformulação, índice de estoque e previsão de temperatura, número de previsão turística, etc. É evidente que alguns outros estudos futuros em potencial podem seguir aplicando uma configuração semelhante à introduzida neste artigo.

## Reconhecimento

Esta pesquisa foi apoiada pelo Avenue Code e pelas seguintes agências brasileiras:

Conselho Nacional de Cientific e Desenvolvimento Tecnológico (CNPq). Bolsas n° 405840/2017-9 e n° 306850/2016-8.

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Os autores também gostariam de expressar sua mais profunda gratidão a UniversidastecnologiaMalaysiathroughgrantnum

R.J130000.7301.4B359 para o financiamento desta pesquisa.

## Referências

- [1] Gross G, Galiana F. Previsão de carga de curto prazo. *Proc IEEE* 1987;75(12): 1558e73. <https://doi.org/10.1109/proc.1987.13927>. <https://doi.org/10.1109/proc.1987.13927>.
- [2] Al-Hamadi H, Soliman S. Previsão de carga elétrica de curto prazo com base no algoritmo kalman filtering com tempo de janela em movimento e model de carga. *Electr Power Syst Res* 2004;68(1):47e59. [https://doi.org/10.1016/s0378-7796\(03\)00150-0](https://doi.org/10.1016/s0378-7796(03)00150-0).
- [3] Song K-B, Baek Y-S, Hong D, Jang G. Short-term Load forecasting for the holidays using fuzzy linear regression method. *IEEE Trans Power Syst* 2005;20(1): 96e101. <https://doi.org/10.1109/tpwrs.2004.835632>. <https://doi.org/10.1109/tpwrs.2004.835632>.
- [4] Veahakyla P, Hakonen E, Le eman P. Previsão de curto prazo da carga de grade usando técnicas box-jenkins. *Int J Electr Power Energy Syst* 1980;2(1):29e34. [https://doi.org/10.1016/0142-0615\(80\)90004-6](https://doi.org/10.1016/0142-0615(80)90004-6). [https://doi.org/10.1016/0142-0615\(80\)90004-6](https://doi.org/10.1016/0142-0615(80)90004-6).
- [5] Ranaweera D, Hubele N, Karady G. Fuzzy logic para previsão de carga de curto prazo. *Int J Electr Power Energy Syst* 1996;18(4):215e22. [https://doi.org/10.1016/0142-0615\(95\)00060-7](https://doi.org/10.1016/0142-0615(95)00060-7). [https://doi.org/10.1016/0142-0615\(95\)00060-7](https://doi.org/10.1016/0142-0615(95)00060-7).
- [6] He Y, Xu Q, Wan J, Yang S. Previsão de densidade de densidade de carga de energia de curto prazo com base na rede neural de regressão quântica e função do núcleo do triângulo. *Energia* 2016;114:498e512. <https://doi.org/10.1016/j.en.ergia.2016.08.023>. <https://doi.org/10.1016/j.en.ergia.2016.08.023>.
- [7] AMJADY N, KEYNIA F. Previsão de carga de curto prazo de sistemas de energia por combinação de transformação de ondas e algoritmo neurovoluntário. *Energia* 2009;34(1):46e57. <https://doi.org/10.1016/j.en.ergia.2008.09.020>. <https://doi.org/10.1016/j.en.ergia.2008.09.020>.
- [8] Ghofrani M, Ghayekhloo M, Arabali A, Ghayekhloo A. Uma previsão híbrida de carga de curto prazo com uma nova estrutura de seleção de entrada. *Energia* 2015;81:777e86. <https://doi.org/10.1016/j.en.ergia.2015.01.028>. <https://doi.org/10.1016/j.en.ergia.2015.01.028>.
- [9] Fan G-F, Peng L-L, Hong W-C. Previsão de carga de curto prazo com base no algoritmo de reconstrução de espaço de fase e modelo de regressão do kernel bi-quadrado. *Appl Energy* 2018;224:13e33. <https://doi.org/10.1016/j.apenergy.2018.04.075>. <https://doi.org/10.1016/j.apenergy.2018.04.075>.
- [10] Mamlook R, Badran O, Abdulhadi E. Um modelo de inferência difusa para previsão de carga de curto prazo. *Política Energética* 2009;37(4):1239e48. <https://doi.org/10.1016/j.enpol.2008.10.051>. <https://doi.org/10.1016/j.enpol.2008.10.051>.
- [11] Metaxiotis K, Kagiannas A, Askounis D, Psarras J. Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher. *Energia Convers Manag* 2003;44(9):1525e34. [https://doi.org/10.1016/s0196-8904\(02\)00148-6](https://doi.org/10.1016/s0196-8904(02)00148-6). [https://doi.org/10.1016/s0196-8904\(02\)00148-6](https://doi.org/10.1016/s0196-8904(02)00148-6).
- [12] Fukushima K, Miyake S. Neocognitron: Um modelo de rede neural auto-organizadora para um mecanismo de reconhecimento de padrões visuais. In: *Competição e cooperação em redes neurais*. Springer Berlin Heidelberg; 1982. p. 267e85. [https://doi.org/10.1007/978-3-642-46466-9\\_18](https://doi.org/10.1007/978-3-642-46466-9_18). [https://doi.org/10.1007/978-3-642-46466-9\\_18](https://doi.org/10.1007/978-3-642-46466-9_18).
- [13] LeCun Y, Haffner P, Bottou L, Bengio Y. Reconhecimento de objetos com aprendizado baseado em gradiente. In: *Forma, contorno e agrupamento na visão computacional*. Springer Berlin Heidelberg; 1999. p. 319e45. [https://doi.org/10.1007/3-540-46805-6\\_19](https://doi.org/10.1007/3-540-46805-6_19). [https://doi.org/10.1007/3-540-46805-6\\_19](https://doi.org/10.1007/3-540-46805-6_19).
- [14] Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. Overfeat: Reconhecimento integrado, localização e detecção usando redes convolucionais. 2013. arXiv:1312.6229.
- [15] Abdulnabi AH, Wang G, Lu J, Jia K. Modelo multitarefa da CNN para previsão de atributos. *IEEE Trans Multimed* 2015;17(11):1949e59. <https://doi.org/10.1109/tmm.2015.2477680>. <https://doi.org/10.1109/tmm.2015.2477680>.
- [16] Levi G, Hassner T. Idade e classe de gênero usando redes neurais convolucionais. In: 2015 Conferência IEEE sobre visão computacional e oficinas de reconhecimento de padrões (CVPRW). IEEE; 2015. <https://doi.org/10.1109/cvprw.2015.7301352>. <https://doi.org/10.1109/cvprw.2015.7301352>.
- [17] Xie L, Wang J, Wei Z, Wang M, Tian Q. DisturbLabel: Regularizando a CNN na camada de perda. In: 2016 Conferência IEEE sobre visão computacional e reconhecimento de padrões (CVPR). IEEE; <https://doi.org/10.1109/cvpr.2016.514>. <https://doi.org/10.1109/cvpr.2016.514>.
- [18] Burnham KP, Anderson DR, editores. Seleção de modelos e inferência multimodel. Nova York: Springer; <https://doi.org/10.1007/b97636> de 2004. <https://doi.org/10.1007/b97636>.
- [19] Bom companheiro I, Bengio Y, Courville A. Deep learning. MIT Press; <http://www.deeplearningbook.org>.
- [20] Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. *Neurocomunicação* 2016;187:27e48. <https://doi.org/10.1016/j.neucom.2015.09.116>. <https://doi.org/10.1016/j.neucom.2015.09.116>.
- [21] Canção Q, Chissom BS. Previsão de inscrições com série d fuzzy série d parte i. *Fuzzy Sets Syst* 1993;54(1):1e9. [https://doi.org/10.1016/0165-0114\(93\)90355-1](https://doi.org/10.1016/0165-0114(93)90355-1). [https://doi.org/10.1016/0165-0114\(93\)90355-1](https://doi.org/10.1016/0165-0114(93)90355-1).
- [22] Yu H-K. Modelos de séries temporais difusas ponderadas para previsão de TAIEX. *Phys Stat MechAppl*2005;349(3e4):609e24. <https://doi.org/10.1016/j.physa.2004.11.006>. <https://doi.org/10.1016/j.physa.2004.11.006>.
- [23] Wang N-Y, Chen S-M. Previsão de temperatura e previsão de TAIEX com base em técnicas automáticas de agrupamento e séries temporais de alta ordem de alta ordem. *Especialista Syst Appl* 2009;36(2):2143e54. <https://doi.org/10.1016/j.eswa.2007.12.013>. <https://doi.org/10.1016/j.eswa.2007.12.013>.
- [24] Sadaei HJ, Enayatifar R, Abdullah AH, Gani A. Previsão de carga de curto prazo usando um model híbrido com uma série de tempo embaçada exponencialmente ponderada e uma melhor busca de harmonia. *Int J Electr Power Energy Syst* 2014;62: 118e29. <https://doi.org/10.1016/j.ijepes.2014.04.026>. <https://doi.org/10.1016/j.ijepes.2014.04.026>.
- [25] Sadaei HJ, Guimaraes FG, da Silva CJ, Lee MH, Eslami T. Método de previsão de carga de curto prazo~ baseado em séries temporais difusas, sazonalidade e longo processo de memória. *Int J Aprox Reason* 2017;83:196e217. <https://doi.org/10.1016/j.ijar.2017.01.006>. <https://doi.org/10.1016/j.ijar.2017.01.006>.
- [26] Efendi R, Ismail Z, Deris MM. Uma nova abordagem linguística de séries temporeis temporísticas difusas para previsão diária da demanda de carga elétrica da Malásia. *Appl Soft Comput* 2015;28:422e30. <https://doi.org/10.1016/j.asoc.2014.11.043>. <https://doi.org/10.1016/j.asoc.2014.11.043>.
- [27] Enayatifar R, Sadaei HJ, Abdullah AH, Gani A. Imperialista algoritmo competitivo combinado com refined high-order weighted série tempo fuzzy (RHWFTSeICA) para previsão de carga de curto prazo. *Energia Convers Manag* 2013;76:1104e16. <https://doi.org/10.1016/j.enconman.2013.08.039>. <https://doi.org/10.1016/j.enconman.2013.08.039>.
- [28] Lee W-J. Um modelo híbrido dinâmico e embaçada da série temporal para previsão de carga de energia a médio prazo. *Int J Electr Power Energy Syst* 2015;64:1057e62. <https://doi.org/10.1016/j.ijepes.2014.08.006>. <https://doi.org/10.1016/j.ijepes.2014.08.006>.
- [29] Taylor JW, McSharry PE. Métodos de previsão de carga de curto prazo: uma avaliação baseada em dados europeus. *IEEE Trans Power Syst* 2007;22(4):2213e9. <https://doi.org/10.1109/tpwrs.2007.907583>. <https://doi.org/10.1109/tpwrs.2007.907583>.
- [30] Taylor JW. Previsão de demanda de eletricidade de curto prazo usando dupla suavização exponencial sazonal. *J Oper Res Soc* 2003;54(8):799e805. <http://www.jstor.org/stable/4101650>.
- [31] Huang S-J, Shih K-R. Previsão de carga de curto prazo via identificação do modelo ARMA, incluindo considerações de processo não gaussiana. *IEEE Trans Power Syst* 2003;18(2):673e9. <https://doi.org/10.1109/tpwrs.2003.811010>. <https://doi.org/10.1109/tpwrs.2003.811010>.



- [32] Gooijer JGD, Hyndman RJ. 25 sims de previsão de séries temporas. *Previsão Int J* 2006;22(3):443e73. <https://doi.org/10.1016/j.ijforecast.2006.01.001>. <https://doi.org/10.1016/j.ijforecast.2006.01.001>.
- [33] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: uma maneira simples de evitar que as redes neurais se sobreapilham. *J Mach Learn Res* 2014;15:1929e58. <http://jmlr.org/papers/v15/srivastava14a.html>.
- [34] Canção Q, Chissom BS. Previsão de matrículas com série d d fuzzy. *Fuzzy Sets Syst* 1994;62(1):1e8. [https://doi.org/10.1016/0165-0114\(94\)90067-1](https://doi.org/10.1016/0165-0114(94)90067-1). [https://doi.org/10.1016/0165-0114\(94\)90067-1](https://doi.org/10.1016/0165-0114(94)90067-1).
- [35] Zadeh L. Fuzzy Sets, *Inf Control* 1965;8. 3):338e53. [https://doi.org/10.1016/s0019-9958\(65\)90241-x](https://doi.org/10.1016/s0019-9958(65)90241-x). [https://doi.org/10.1016/s0019-9958\(65\)90241-x](https://doi.org/10.1016/s0019-9958(65)90241-x).
- [36] Chen S-M. Previsão de matrículas com base em séries tempores difusas. *Fuzzy Sets Syst* 1996;81(3):311e9. [https://doi.org/10.1016/0165-0114\(95\)00220-0](https://doi.org/10.1016/0165-0114(95)00220-0). [https://doi.org/10.1016/0165-0114\(95\)00220-0](https://doi.org/10.1016/0165-0114(95)00220-0).
- [37] Sermanet P, Chintala S, LeCun Y. Redes neurais convolucionais aplicadas à classificação de números de dígitos domésticos. 2012. arXiv:1204.3968.
- [38] Zeiler MD, Fergus R. Stochastic pooling para regularização de redes neurais convolucionais profundas. 2013. arXiv:1301.3557.
- [39] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classificação com redes neurais convolucionais profundas. In: *Proceedings of the 25th conference on neural information processing systems*, vol. 1. EUA: Curran Associates Inc.; 2012. p. 1097e105. NIPS'12, <http://dl.acm.org/citation.cfm?id=142999134.2999257>.
- [40] Cornelis C, Deschrijver G, Kerre EE. Implicação na teoria de conjuntos difusos intuitivos e valorizados no intervalo: construção, classificação, aplicação. *Int J Razão aproximada* 2004;35(1):55e95. [https://doi.org/10.1016/s0888-613x\(03\)00072-0](https://doi.org/10.1016/s0888-613x(03)00072-0). [https://doi.org/10.1016/s0888-613x\(03\)00072-0](https://doi.org/10.1016/s0888-613x(03)00072-0).
- [41] Hajek P. *Metamathematics de lógica difusa*. Springer Holanda; 1998. <https://doi.org/10.1007/978-94-011-5300-3>. <https://doi.org/10.1007/978-94-011-5300-3>.
- [42] Lee MH, Sadaei HJ, Suhartono. Melhorando a previsão do TAIEX usando séries temporáticas difusas com a transformação de energia do BoxeCox. *J Appl Stat* 2013;40(11):2407e22. <https://doi.org/10.1080/02664763.2013.817548>. <https://doi.org/10.1080/02664763.2013.817548>.
- [43] R. J. Hyndman, Y. Khandakar, Previsão automática da série temporal: o pacote de previsão para R, *J Stat Softw* 27 (3). doi:10.18637/jss.v027.i03. URL <https://doi.org/10.18637/jss.v027.i03>
- [44] De Lima E Silva Petronio C^ andido. Petroniocandido/pyfts: Export for zenodo^ doi generation (ref. pkg1.2.3). 2018. <https://doi.org/10.5281/zenodo.1194859>. <https://zenodo.org/record/1194859>.
- [45] Hochreiter S, Schmidhuber J. Longa memória de curto prazo. *Neural Comput* 1997;9(8):1735e80. <https://doi.org/10.1162/neco.1997.9.8.1735>. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [46] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 Conferência IEEE sobre visão computacional e reconhecimento de padrões (CVPR). IEEE; 2016. <https://doi.org/10.1109/cvpr.2016.90>. <https://doi.org/10.1109/cvpr.2016.90>.
- [47] Simonyan K, Zisserman A. Redes convolucionais muito profundas para reconhecimento de imagem em larga escala. 2014. arXiv:1409.1556.