



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA
MESTRADO EM ENGENHARIA MECÂNICA

FELIPE PINTO MARINHO

DISPOSITIVO DE PREVISÃO DE IRRADIÂNCIA SOLAR POR MEIO DE
APRENDIZAGEM DE MÁQUINA UTILIZANDO SENSORES DE LUMINOSIDADE
LDR E IMAGENS DO CÉU

FORTALEZA

2020

FELIPE PINTO MARINHO

DISPOSITIVO DE PREVISÃO DE IRRADIÂNCIA SOLAR POR MEIO DE
APRENDIZAGEM DE MÁQUINA UTILIZANDO SENSORES DE LUMINOSIDADE
LDR E IMAGENS DO CÉU

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Mecânica. Área de concentração: Processos, Equipamentos e Sistemas para Energias Renováveis.

Orientador: Prof. Dr. Paulo Alexandre Costa Rocha.

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- M29d Marinho, Felipe Pinto Marinho.
Dispositivo de previsão de irradiância solar por meio de aprendizagem de máquina utilizando sensores de luminosidade LDR e imagens do céu / Felipe Pinto Marinho Marinho. – 2020.
71 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Mecânica, Fortaleza, 2020.
Orientação: Prof. Dr. Paulo Alexandre Costa Rocha.
1. Energia Solar. 2. Aprendizagem de Máquina. 3. Processamento Digital de Imagens. 4. Previsão de Curto Prazo. I. Título.

CDD 620.1

FELIPE PINTO MARINHO

DISPOSITIVO DE PREVISÃO DE IRRADIÂNCIA SOLAR POR MEIO DE
APRENDIZAGEM DE MÁQUINA UTILIZANDO SENSORES DE LUMINOSIDADE
LDR E IMAGENS DO CÉU

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Mecânica, do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial para a obtenção do Título de Mestre em Engenharia Mecânica. Área de Concentração: Processos, Equipamentos e Sistemas para Energias Renováveis.

Aprovada em 28/04/2020

BANCA EXAMINADORA

Prof. Dr. Paulo Alexandre Costa Rocha (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Paulo Cesar Marques de Carvalho
Universidade Federal do Ceará (UFC)

Prof. Dr. Matheus Pereira Porto
Universidade Federal de Minas Gerais (UFMG)

A Deus.

Aos meus pais, Raimunda Pinto Marinho Melo
e Francisco Felix da Silva (In memoriam).

AGRADECIMENTOS

A Deus, por tudo em minha vida.

À minha mãe Raimunda Pinto Marinho Melo por ter nos momentos mais difíceis me motivado a continuar e nunca desistir.

À minha família por todo o suporte fornecido.

À CAPES, pelo apoio financeiro com a manutenção da bolsa de auxílio.

Ao Prof. Dr. Paulo Alexandre Costa Rocha, pela excelente orientação.

Aos professores participantes da banca examinadora Paulo Cesar Marques de Carvalho e Matheus Pereira Porto pelo tempo, pelas valiosas colaborações e sugestões.

Aos colegas da turma de mestrado, em especial aos amigos Gilderlânio Palácio, Victor Soares Gualberto, Lino Wagner Castelo Branco Portela, Alessandra Farias Formiga Queiroga, Kevin Santos Guedes, Fernando Wesley Silva de Oliveira, Anderson dos Santos Dias, Jardel de Queiroz Juvêncio, Nelly Vanessa Pérez Rangel, José Welbson Carneiro de Araújo, Paulo Roberto Nogueira Bastos e Luiz David Ricarte de Souza Custódio.

Aos colegas do Laboratório de Energia Solar e Gás Natural, em especial aos amigos Juarez Pompeu de Amorim Neto, Juliana Silva Brasil, Rahysa Silva de Oliveira e Diego Caitano de Pinho pelas reflexões, críticas e sugestões na realização de diversos trabalhos acadêmicos.

Aos funcionários Valdi Matos e tia Raimunda pelos excelentes trabalhos desempenhados no Departamento de Engenharia Mecânica.

À Leticia Rayanne pelo amor, paciência e apoio em todos os momentos que compartilhamos.

RESUMO

Uma das principais dificuldades para o uso da fonte solar na geração de eletricidade na rede se deve ao seu comportamento intermitente. Assim, previsões de irradiância solar de curto prazo, que são aquelas realizadas para um horizonte de até 6 horas a posteriori, são relevantes para que o operador da rede elétrica possa prever quedas de fornecimento de usinas solares e, de posse destas informações, o mesmo consiga realizar um melhor gerenciamento do fornecimento e da demanda de energia. Neste sentido, no trabalho em questão foram obtidas previsões de irradiância solar global para os horizontes de previsão de 10, 20 e 30 minutos a posteriori, por meio da aplicação de algoritmos de aprendizagem de máquina em conjuntos de dados constituídos por sinais obtidos por sensores de resistência dependente da luz e descritores estatísticos (média aritmética, desvio padrão e entropia de Shannon) extraídos de imagens do céu capturadas por uma câmera, onde a integração de tais sensores foi feita utilizando um *Raspberry Pi 3*. Desta forma, foi possível avaliar se a adição de preditores obtidos de imagens do céu proporcionam uma melhora no desempenho dos modelos de aprendizagem de máquina utilizados. Além disso, também foi avaliado se o uso dos filtros de suavização da mediana e o de aguçamento do laplaciano da gaussiana nas imagens ocasionam aumentos na acurácia de previsão relativo ao caso em que se calcula os descritores estatísticos das imagens sem aplicação de nenhum filtro. Para o horizonte de 30 minutos a posteriori, a média aritmética da Raiz do Erro Quadrático Médio obtida pelos modelos, quando se considerou como atributos apenas os sinais fornecidos pelos sensores de luminosidade, foi de 165,83 W/m², já para o caso em que se adiciona informações de imagens não filtradas obteve-se 154,76 W/m², para o caso de se adicionar informações de imagens filtradas pelo filtro da mediana teve-se um valor de 154,08 W/m². Por fim, para o outro filtro obteve-se 163,26 W/m². Posteriormente, os algoritmos foram aplicados em um novo conjunto de dados que foi construído pela substituição dos preditores relacionados aos sensores de luminosidade por atributos caracterizados por serem funções de recorrência sobre os valores de irradiância de instantes anteriores. Para este caso, no horizonte de 30 minutos, as médias da Raiz do Erro Quadrático Médio foram de 136,43 W/m², 138,82 W/m², 135 W/m² e 140,05 W/m² para os casos sem imagem, com imagem, imagem filtrada com mediana e imagem filtrada com o outro filtro, respectivamente.

Palavras-chave: Energia solar. Aprendizagem de máquina. Processamento digital de imagens. Previsão de curto prazo.

ABSTRACT

One of the main difficulties in using the solar source to generate electricity for the grid is due to its intermittent behavior. Thus, short-term solar irradiance forecasts, which are those made for horizons of until 6 hours ahead, are relevant so that the power grid operator can predict oscillations in the supply of solar plants and, in possession of this information, it can perform a better management of energy supply and demand. In this sense, in the present work, global solar irradiance forecasts were obtained for the forecast horizons of 10, 20 and 30 minutes a posteriori by the application of machine learning algorithms in data sets consisting of signals obtained by light dependent resistors and statistical descriptors (arithmetic mean, standard deviation and Shannon entropy) extracted from images of the sky captured by a camera, where the integration of such sensors was made using a Raspberry Pi 3. In this way, it was possible to evaluate whether the addition of predictors obtained from images of the sky provide an improvement in the performance of the applied machine learning models. In addition, it was also assessed if the use of the median smoothing and the Gaussian Laplacian sharpening filters over the images could increase the forecast accuracy relative to the case in which the statistical descriptors of the images are calculated without any filter. For the horizon of 30 minutes ahead, the arithmetic mean of the Root Mean Square Error obtained by the models which only the signals provided by the luminosity sensors were considered as attributes was 165.83 W / m^2 , while for the case in which it is added information from unfiltered images it was obtained 154.76 W / m^2 . For the case of adding information from images filtered by the median filter, a value of 154.08 W / m^2 was obtained and finally, for the other filter, a RMSE of 163.26 W / m^2 was obtained. Subsequently, the algorithms were applied to a new set of data that was built replacing the predictors related to the luminosity sensors by attributes characterized by being recurrence functions over the irradiance values of previous instants. For this case, for the 30 minute horizon, the mean of root mean square errors were $136,43 \text{ W / m}^2$, 138.82 W / m^2 , 135 W / m^2 and $140,05 \text{ W / m}^2$ for cases without image, with image, image filtered with median and image filtered with the sharpening filter, respectively.

Keywords: Solar energy. Machine learning. Digital image processing. Short-term forecasting

LISTA DE FIGURAS

| | | |
|-----------|--|----|
| Figura 1 | – Representação esquemática dos tipos radiação solar..... | 17 |
| Figura 2 | – Representação dos modelos NWP | 21 |
| Figura 3 | – Neurônio biológico em conjunto com um neurônio artificial com três entradas..... | 25 |
| Figura 4 | – Arquitetura de uma rede MLP com três camadas..... | 26 |
| Figura 5 | – Imagem do céu obtida por uma câmera de segurança..... | 27 |
| Figura 6 | – Imagem obtida por satélite..... | 27 |
| Figura 7 | – Exemplo de imagem obtida neste trabalho..... | 28 |
| Figura 8 | – Sistema de aquisição montado..... | 29 |
| Figura 9 | – Representação da etapa de treinamento..... | 32 |
| Figura 10 | – Representação esquemática do uso de kernels em SVR | 35 |
| Figura 11 | – Representação esquemática de um <i>Bagging</i> aplicado em árvores de decisão..... | 36 |
| Figura 12 | – Interpretação geométrica para a solução do problema de multilateração..... | 39 |
| Figura 13 | – Busca por hiperparâmetro no MLM para outros conjuntos de dados..... | 40 |
| Figura 14 | – Representação do processo de filtragem espacial..... | 41 |
| Figura 15 | – (a) Imagem original, (b) Imagem filtrada com mediana de tamanho 3, (c) Imagem filtrada com mediana de tamanho 5, (d) Imagem filtrada com mediana de tamanho 7, (e) Imagem filtrada com mediana de tamanho 9, (f) Imagem filtrada com mediana de tamanho 11..... | 43 |
| Figura 16 | – (a) Imagem original, (b) Imagem filtrada com LoG de tamanho 3, (c) Imagem filtrada com LoG de tamanho 5, (d) Imagem filtrada com LoG de tamanho 7, (e) Imagem filtrada com LoG de tamanho 9, (f) Imagem filtrada com LoG de tamanho 11..... | 45 |
| Figura 17 | – Busca por hiperparâmetro no MLM..... | 51 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1 – Valores de n para um dado mês do ano..... | 19 |
| Tabela 2 – Resultados obtidos para os ARIMA, DHR, ETS..... | 24 |
| Tabela 3 – Busca por hiperparâmetro SVR..... | 45 |
| Tabela 4 – Busca por hiperparâmetro no algoritmo RF..... | 46 |
| Tabela 5 – Busca por hiperparâmetro no KNN..... | 47 |
| Tabela 6 – Resultados para o conjunto I..... | 50 |
| Tabela 7 – Resultados para o conjunto I..... | 50 |
| Tabela 8 – Resultados para o conjunto I..... | 51 |
| Tabela 9 – Resultados para o conjunto II..... | 52 |
| Tabela 10 – Resultados para o conjunto II..... | 52 |
| Tabela 11 – Resultados para o conjunto II..... | 53 |
| Tabela 12 – Resultados para o conjunto III..... | 54 |
| Tabela 13 – Resultados para o conjunto III..... | 54 |
| Tabela 14 – Resultados para o conjunto III..... | 55 |
| Tabela 15 – Resultados para o conjunto IV..... | 55 |
| Tabela 16 – Resultados para o conjunto IV..... | 56 |
| Tabela 17 – Resultados para o conjunto IV..... | 56 |
| Tabela 18 – Resultados para o conjunto A..... | 60 |
| Tabela 19 – Resultados para o conjunto A..... | 61 |
| Tabela 20 – Resultados para o conjunto A..... | 61 |
| Tabela 21 – Resultados para o conjunto B..... | 61 |
| Tabela 22 – Resultados para o conjunto B..... | 61 |
| Tabela 23 – Resultados para o conjunto B..... | 62 |
| Tabela 24 – Resultados para o conjunto C..... | 62 |

| | |
|---|----|
| Tabela 25 – Resultados para o conjunto C..... | 63 |
| Tabela 26 – Resultados para o conjunto C..... | 63 |
| Tabela 27 – Resultados para o conjunto D..... | 64 |
| Tabela 28 – Resultados para o conjunto D..... | 64 |
| Tabela 29 – Resultados para o conjunto D..... | 65 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|-------|--|
| NWP | Previsão Numérica do Clima (<i>Numerical Weather Prediction</i>) |
| LDR | Resistência Dependente da Luz (<i>Light Dependent Resistor</i>) |
| MLM | Máquina de Aprendizagem Mínima (<i>Minimal Learning Machine</i>) |
| KNN | K-Vizinhos mais Próximos (<i>K-Nearest Neighbors</i>) |
| SVR | Regressão por Vetor Suporte (<i>Support Vector Regression</i>) |
| RF | Florestas Aleatórias (<i>Random Forests</i>) |
| MBE | Erro Médio por Viés (<i>Mean Bias Error</i>) |
| MAE | Erro Médio Absoluto (<i>Mean Absolute Error</i>) |
| RMSE | Raiz do Erro Quadrático Médio (<i>Root Mean Squared Error</i>) |
| rRMSE | Raiz do Erro Quadrático Médio Relativo (<i>Relative Root Mean Squared Error</i>) |
| s | Habilidade de Previsão (<i>Forecast Skill</i>) |
| DNI | Irradiância Normal Direta (<i>Direct Normal Irradiance</i>) |
| GHI | Irradiância Horizontal Global (<i>Global Horizontal Irradiance</i>) |
| CSP | Usinas de Potência Solar Térmica Concentrada (<i>Concentred Solar Power</i>) |
| FV | Usinas Fotovoltaicas (<i>Photovoltaics Farm</i>) |
| ARIMA | Modelos auto-regressivos integrados de médias móveis (<i>Autoregressive Integrated Moving Average Model</i>) |
| DHR | Regressão Harmônica Dinâmica (<i>Dynamic Harmonic Regression</i>) |
| ETS | Suavização Exponencial no Espaço de Estados (<i>Esponential Smoothing State Space</i>) |
| RNA | Redes Neurais Artificiais (<i>Artificial Neural Networks</i>) |
| MLP | Percéptrons de Múltiplas Camadas (<i>Multilayer Perceptron</i>) |
| CNN | Redes Neurais Convolucionais (<i>Convolutional Neural Networks</i>) |
| LSTM | Redes Neurais Recorrentes de Memória de Longo Prazo (<i>Long Short-Term Memory Networks</i>) |
| CMV | Vetores de Movimento de Nuvens (<i>Cloud Moving Vectors</i>) |
| LESGN | Laboratório de Energia Solar e Gás Natural |
| QP | Problema Quadrático Convexo (<i>Quadratic Programming</i>) |
| LoG | Filtro do Laplaciano da Gaussiana |

SUMÁRIO

| | | |
|-------|---|----|
| 1 | INTRODUÇÃO | 14 |
| 2 | OBJETIVOS | 16 |
| 2.1 | Objetivo geral | 16 |
| 2.2 | Objetivos específicos | 16 |
| 3 | REVISÃO BIBLIOGRÁFICA | 17 |
| 3.1 | Irradiância solar, irradiância solar direta, irradiância solar difusa e irradiância solar global | 17 |
| 3.2 | Previsões para as componentes GHI, DNI e G_D | 17 |
| 3.3 | Índices de clareza e de céu claro | 18 |
| 3.4 | Principais metodologias para a previsão de irradiância | 20 |
| 3.4.1 | <i>Numerical Weather Prediction</i> | 20 |
| 3.4.2 | <i>Métodos baseados em análise de séries temporais</i> | 22 |
| 3.4.3 | <i>Métodos baseados em aprendizagem de máquina</i> | 24 |
| 3.4.4 | <i>Métodos baseados em processamento de imagens</i> | 26 |
| 4 | METODOLOGIA | 29 |
| 4.1 | Dados | 29 |
| 4.2 | Modelos utilizados para realizar as previsões | 32 |
| 4.2.1 | <i>Modelo de baseline</i> | 32 |
| 4.2.2 | <i>K-Vizinhos mais Próximos</i> | 33 |
| 4.2.3 | <i>Regressão por Vetor Suporte</i> | 33 |
| 4.2.4 | <i>Árvore de decisão com a aplicação de um procedimento de Bagging</i> | 35 |
| 4.2.5 | <i>Florestas Aleatórias</i> | 36 |
| 4.2.6 | <i>Máquina de Aprendizagem Mínima</i> | 37 |
| 4.3 | Filtragem espacial de imagens digitais | 40 |
| 4.3.1 | <i>Convolução e Correlação espacial</i> | 40 |
| 4.3.2 | <i>Filtro de suavização da mediana</i> | 42 |
| 4.3.3 | <i>Filtro de aguçamento do laplaciano da gaussiana</i> | 43 |
| 4.3.4 | <i>Descritores estatísticos</i> | 45 |
| 4.4 | Métricas de erro | 46 |
| 4.4.1 | <i>Erro Médio Absoluto</i> | 46 |
| 4.4.2 | <i>Erro Médio por Viés</i> | 46 |

| | | |
|-------|---|----|
| 4.4.3 | <i>Raiz do Erro Quadrático Médio</i> | 47 |
| 4.4.4 | <i>Raiz do Erro Quadrático Médio Relativo</i> | 47 |
| 4.4.5 | <i>Habilidade de Previsão</i> | 47 |
| 5 | RESULTADOS E DISCUSSÃO | 48 |
| 5.1 | Busca por hiperparâmetros | 48 |
| 5.1.1 | <i>SVR</i> | 48 |
| 5.1.2 | <i>Florestas Aleatórias</i> | 49 |
| 5.1.3 | <i>Árvore de regressão com a aplicação de um procedimento de Bagging</i> | 50 |
| 5.1.4 | <i>KNN</i> | 50 |
| 5.1.5 | <i>MLM</i> | 52 |
| 5.2 | Resultados obtidos pela aplicação dos modelos ajustados no conjunto de validação | 52 |
| 6 | CONCLUSÃO | 66 |

1 INTRODUÇÃO

O potencial de inserção de energias renováveis na matriz energética global, em especial da energia solar, torna necessária uma adequada gestão do fornecimento de energia pelo órgão de controle, uma vez que se tratam de fontes com comportamento estocástico e intermitente. Ter uma previsão confiável da irradiância solar futura na região de um parque solar permite ao operador da rede elétrica avaliar em que momentos serão necessárias complementações do abastecimento com outras fontes.

O operador da rede precisa de previsões confiáveis em diferentes horizontes temporais, onde aquelas feitas em intervalos de horas e minutos são adequadas para prever quedas de fornecimento, enquanto as realizadas em intervalos de dias se destacam na programação de manutenções preventivas. Problemas relacionados à flutuação de tensão, qualidade de energia e estabilidade da rede elétrica podem ser evitados por meio da realização de previsões de irradiância solar.

Neste sentido, muitas metodologias têm sido utilizadas para que previsões de irradiância em diversos horizontes de tempo sejam efetuadas. Para tanto, é feito o uso de modelos *Numerical Weather Prediction* (NWP), análise de séries temporais, aplicações envolvendo o processamento de imagens, utilização de algoritmos de aprendizagem de máquina, etc. (INMAN et al., 2013)

Neste estudo, foram realizadas previsões de irradiância solar global para os horizontes de 10, 20 e 30 minutos a posteriori por meio da aplicação de modelos de aprendizagem de máquina em conjuntos de dados constituídos por sinais coletados por sensores de resistência dependente da luz e descritores estatísticos (média, desvio padrão e entropia de Shannon) extraídos de imagens do céu capturadas por uma câmera.

Posteriormente, os algoritmos foram aplicados nos mesmos conjuntos de dados, diferindo apenas pela substituição dos atributos correspondentes aos sinais obtidos pelos sensores de luminosidade por funções de recorrência sobre os valores de irradiância de instantes anteriores.

Assim é possível avaliar se a adição de atributos relacionados à imagens do céu, uso de informações dos valores de irradiância de instantes passados, bem como se o uso de determinados filtros digitais (mediana e laplaciano da gaussiana) nas imagens proporcionam uma melhora na acurácia dos algoritmos de aprendizagem de máquina utilizados, a saber: Máquina de Aprendizagem Mínima (MLM), Regressão por Vetor Suporte (SVR), K-Vizinhos mais Próximos (KNN), Florestas Aleatórias (RF) e uma árvore de regressão com a aplicação de

um procedimento de *Bagging*. A avaliação do desempenho dos modelos foi feita por meio do cálculo das métricas Erro Médio por Viés (MBE), Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE), Raiz do Erro Quadrático Médio Relativo (rRMSE) e Habilidade de Previsão (s).

2 OBJETIVOS

2.1 Objetivo geral

O objetivo desta dissertação foi o de desenvolver um dispositivo de previsão de irradiância solar por meio de aprendizagem de máquina utilizando LDR's e imagens do céu e por meio deste avaliar se a adição de atributos extraídos de imagens do céu e o uso de informações relacionadas aos valores de irradiância solar de instantes passados, contribuem para uma melhor generalização dos algoritmos de aprendizagem de máquina utilizados quando os mesmos são aplicados para propósitos de previsão de irradiância solar de curto prazo, bem como se o uso de determinados filtros digitais (mediana e laplaciano da gaussiana) nas imagens proporcionam melhorias na acurácia dos métodos.

2.2 Objetivos específicos

1. Aplicar um sistema *Raspberry Pi 3* para realizar a integração dos sensores LDR's com a câmera, desenvolvendo assim um sistema de aquisição de baixo custo;
2. Implementar código em Python com o intuito de se extrair os descritores estatísticos das imagens, realizar a filtragem das mesmas e a aquisição dos sinais obtidos pelos LDR's, posteriormente alocando todos estes dados em uma planilha;
3. Implementar código em Scilab para o cálculo das funções de recorrência sobre os valores de irradiância solar de instantes anteriores, posteriormente alocando estes em uma planilha;
4. Construir os bancos de dados analisados pela realização de coletas dos dias 03/06/2019 à 07/06/2019 em Fortaleza, CE, onde as mesmas englobaram valores de irradiância solar global na faixa das 8:00 às 17:30;
5. Implementar os algoritmos de aprendizagem de máquina utilizados no software R;
6. Realizar as previsões para os horizontes de 10, 20 e 30 minutos a posteriori pela aplicação dos modelos implementados nos conjuntos de dados;
7. Avaliar o desempenho dos métodos na realização das previsões pelo cálculo das métricas de erro MBE, MAE, RMSE, rRMSE e s.

3 REVISÃO BIBLIOGRÁFICA

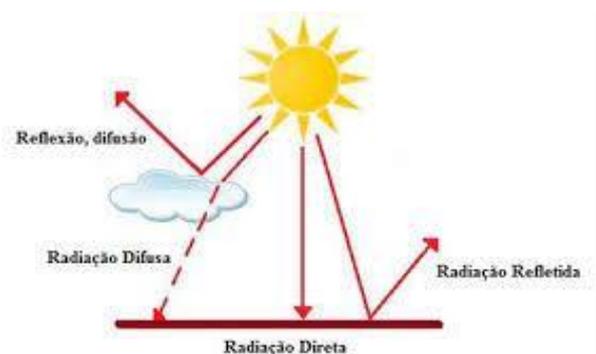
3.1 Irradiância solar, irradiância solar direta, irradiância solar difusa e irradiância solar global

Nesta seção é estabelecido alguns conceitos fundamentais para um melhor entendimento do significado físico da irradiância solar, bem como seus tipos. A **Irradiância solar (G)** é a taxa na qual a energia radiante incide sobre uma superfície por unidade de área, geralmente dada em W/m^2 . A **Irradiância normal direta (DNI)** é a irradiância solar gerada pela energia radiante proveniente do Sol sem que haja um espalhamento da mesma na atmosfera, já a **irradiância solar difusa (G_D)** é a irradiância solar gerada pela energia radiante proveniente do Sol após sua direção ter sido alterada devido ao espalhamento na atmosfera (DUFFIE, BECKMAN, 2013). No trabalho em questão, o interesse maior está sobre a **irradiância global horizontal (GHI)**, que é a soma entre a irradiância solar direta multiplicada pelo cosseno do ângulo de zênite com a irradiância solar difusa, como representado pela Equação 1.

$$GHI = DNI \cos \theta_z + G_D \quad (1)$$

A Figura 1 serve para ilustrar os tipos de radiação solar que acabam por desenvolver as irradiâncias anteriormente descritas.

Figura 1 – Representação esquemática dos tipos de radiação solar



Fonte: Disponível em <<https://aulasdegeografiaodivelas.wordpress.com/2015/06/12/nocoes-radiacao-solar/>>

Acesso em Julho 2019.

3.2 Previsões para as componentes GHI, DNI e G_D

Os dois principais meios de geração de eletricidade proveniente do recurso solar são por usinas de potência solar térmica concentrada (CSP) e por usinas fotovoltaicas (PV). Sistemas de potência solar térmica concentrada convertem a irradiância solar direta em calor que

posteriormente é utilizado em um ciclo de potência, como o ciclo Rankine, para a geração de eletricidade. Entretanto, dispositivos fotovoltaicos geram eletricidade pelo aproveitamento tanto da componente direta quanto da difusa. Desta forma, previsões de irradiância solar direta são de maior relevância no contexto de sistemas de potência solar térmica, enquanto previsões de irradiância solar global apresentam maior importância em aplicações que envolvam o uso de dispositivos fotovoltaicos (BENALI *et al.*, 2019).

Segundo os trabalhos de Marquez e Coimbra (2011) e Pedro e Coimbra (2015) previsões de irradiância solar direta apresentam uma menor acurácia do que as realizadas para irradiância solar global. No primeiro trabalho, para um horizonte de previsão de um dia a posteriori foram encontrados para a previsão de GHI valores de rRMSE variando na faixa de 15 % até 22% e de 28% até 35% para o caso de DNI. Já no segundo trabalho, onde se realizaram previsões em um horizonte de até 30 minutos a frente, foram encontrados valores da métrica RMSE na faixa de 30 W/m² até 36,5 W/m² para GHI e de 56,9 W/m² até 77,5 W/m² para o caso DNI.

Além disso, outros autores (TRAPERO *et al.*, 2015) utilizaram os modelos baseados em análise de séries temporais: suavização exponencial no espaço de estados (DONG *et al.*, 2013), auto-regressivos integrados de médias móveis (ARIMA) (BOX, 1994) e uma regressão harmônica dinâmica (YOUNG *et al.*, 1999). Conseguiram como melhores resultados um rRMSE de 29,66% para previsões de GHI e 46,79% para DNI em horizontes de 1 hora até 1 dia a posteriori. Novamente, pela avaliação destes resultados, tem-se uma indicação de que previsões DNI são menos confiáveis do que aquelas realizadas para GHI.

Benalli *et al.* (2019) realizou previsões para as componentes GHI, DNI e difusa em intervalos de 1 hora até 6 horas à frente onde foram encontrados valores de rRMSE na faixa de 19,65% até 27,78%, 34,11% até 49,08% e 35,08% até 49,14%, para as previsões das componentes GHI, DNI e difusa, respectivamente, ilustrando assim que previsões de irradiância solar difusa são ainda menos confiáveis do que para DNI. O trabalho em questão focou na realização de previsões GHI.

3.3 Índices de clareza e de céu claro

A maioria dos trabalhos que realizam previsões de irradiância solar pela aplicação de métodos de aprendizagem de máquina utilizam como entradas para tais algoritmos variáveis endógenas (PEDRO *et al.*, 2015), (PEDRO *et al.*, 2018), (BENALI *et al.*, 2018), que são relacionadas aos valores de irradiância solar para instantes anteriores, bem como o atual. Estes modelos de aprendizagem são mais eficientes na realização de previsões quando os mesmos são aplicados em séries históricas com um comportamento estacionário. Desta forma, para

tornar os dados de irradiância solar estacionários e separar efeitos climáticos e sazonais, costuma-se utilizar um índice adimensional chamado de índice de clareza, K_t , que é definido como a razão entre a irradiância solar na superfície da terra pela irradiância extraterrestre, que não sofreu a influência da atmosfera, como indicado pela Equação 2.

$$K_t = \frac{G}{G_0} \quad (2)$$

Onde, G é a irradiância global na superfície da terra em uma dada localização, enquanto G_0 é a irradiância extraterrestre. O valor de G_0 é calculado por meio de considerações geométricas, como ilustrado pela Equação (3).

$$G_0 = G_{SC}(1 + 0,033 \cos (360n/365)) (\cos \phi \cos \delta \cos \omega + \sin \phi \sin \delta) \quad (3)$$

Onde, G_{SC} é a constante solar que assume o valor de 1367W/m^2 , ϕ é a latitude, δ é a declinação, ω é o ângulo horário e n é o dia do ano que pode ser determinado pelo uso da Tabela 1. Para um melhor detalhamento destes parâmetros recomenda-se Duffie & Beckman (2013).

Tabela 1 – Valores de n para um dado mês do ano

| Mês | n para o i -ésimo dia do mês |
|-----------|----------------------------------|
| Janeiro | i |
| Fevereiro | $31 + i$ |
| Março | $59 + i$ |
| Abril | $90 + i$ |
| Mai | $120 + i$ |
| Junho | $151 + i$ |
| Julho | $181 + i$ |
| Agosto | $212 + i$ |
| Setembro | $243 + i$ |
| Outubro | $273 + i$ |
| Novembro | $304 + i$ |
| Dezembro | $334 + i$ |

Muitos estudos (LINARES-RODRIGUEZ *et al.*, 2011) mostraram que a irradiância G_0 pode ser substituída pela irradiância solar em condições de céu claro. Nestas condições, tem-se o índice de céu claro dado pela Equação 3.

$$K_{t,cs} = \frac{G}{G_{0,CS}} \quad (3)$$

Onde $G_{0,CS}$ é a irradiância solar de céu claro.

Vários modelos de irradiância de céu claro estão disponíveis na literatura (BADESCU *et al.*, 2012), onde as principais diferenças entre os mesmos está nas entradas de cada modelo.

Para os conjuntos com sinais obtidos pelos sensores de luminosidade, não se utilizou tais índices, ao invés disso, para que se obtivesse dados estacionários, realizou-se um procedimento de normalização estatística para cada conjunto de dados, onde a normalização estatística de uma variável x é dada como na Equação (4).

$$x^* = \frac{x - \bar{x}}{\sigma_x} \quad (4)$$

Onde \bar{x} , σ_x são, respectivamente, a média e o desvio padrão amostral para o atributo x .

3.4 Principais metodologias para a previsão de irradiância

3.4.1 Numerical Weather Prediction

Os modelos NWP realizam previsões das condições atmosféricas, incluindo valores de irradiância solar, por meio do conhecimento do estado atmosférico para um dado instante inicial e pelo uso de leis físicas (conservação de massa, conservação de energia, etc.) que governam a transição de um estado atmosférico para outro.

Primeiro, um domínio físico é selecionado, então o mesmo é discretizado até uma resolução desejada. Posteriormente, as leis do movimento e/ou termodinâmicas são resolvidas numericamente. É o passo temporal nas soluções numéricas das equações que providenciam uma base para a realização das previsões NWP (INMAN *et al.*, 2013).

Nos primeiros trabalhos com modelos NWP, observaram-se dificuldades relacionadas à obtenção das soluções numéricas, uma vez que o modelo matemático resultante era um sistema de equações diferenciais parciais não lineares (BJERKNES, 1904). Simplificações nas

equações governantes foram possíveis graças ao uso de modelos filtrados e pelo uso da hipótese quase-geostrófica (CHARNEY, 1948). As primeiras previsões por modelos NWP assistidos por computador a terem sucesso ocorreram no ano de 1950 (CHARNEY *et al.*, 1950). Estas foram baseadas na equação de vorticidade barotrópica, representada pela Equação (5)

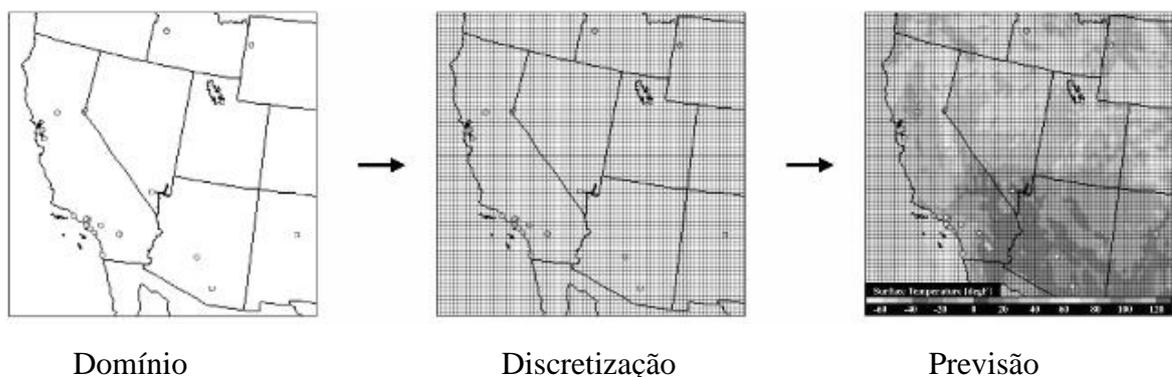
$$\frac{\partial \zeta_b}{\partial t} + \nabla_h \cdot \{(\zeta_b + f)\mathbf{v}_\psi\} = 0 \quad (5)$$

$$\zeta_b = \nabla_h^2 \psi_b \quad \mathbf{v}_\psi = k \times \nabla_h \psi_b$$

Onde, ζ_b é a vorticidade barotrópica, ∇_h é o gradiente bidimensional horizontal, ∇_h^2 é o laplaciano bidimensional horizontal, \mathbf{v}_ψ é a velocidade do vento bidimensional puramente rotacional, ψ_b é a função de corrente barotrópica, k é o vetor unitário normal a superfície geopotencial e $f = 2\Omega \sin \phi$ é o parâmetro de Coriolis (EHRENFORFER, 2012).

Desde então, muitas estratégias têm sido desenvolvidas, proporcionando melhorias nestes modelos. Dentre estas abordagens pode-se citar: Refinamento das equações primitivas (SMAGORINSKY, 1958), modelagem da circulação geral (PHILLIPS, 1956), extensivo uso de modelos espectrais e métodos de transformadas (SILBERMAN, 1954), (PLATZMAN, 1960), (ORSZAG, 1970), (ELIASSEN *et al.*, 1970), além do uso de diferenciação semi-implícita no tempo (ROBERT *et al.*, 1985). De uma maneira geral, estes modelos funcionam com a estrutura já discutida anteriormente e ilustrada na Figura 2.

Figura 2 – Representação dos modelos NWP



Fonte: Inman *et al.* (2013)

Diversos trabalhos (ARYAPUTERA *et al.*, 2015; MEJIA *et al.*, 2018; NONNEMACHER *et al.*, 2016) indicaram que os modelos NWP apresentaram um bom desempenho para realização

de previsões de irradiância solar de longo prazo, que são aquelas com horizontes acima de 6 horas.

3.4.2 Métodos baseados em análise de séries temporais

Estes modelos apresentam como principal característica a utilização de valores de irradiância de instantes passados para realizar a estimação de informações futuras, por isso muitas vezes são denominados modelos de regressão. Entretanto, em seus primeiros usos tais modelos fundamentavam-se em considerações que não condiziam com a natureza dos dados de irradiância como, por exemplo, consideravam que as observações individuais apresentavam uma variação independente, quando na verdade sabe-se que as condições climáticas locais favorecem ao desenvolvimento de séries temporais de irradiância que são fortemente dependentes.

Portanto, a natureza correlacionada dos dados de irradiância é uma característica essencial que deve ser inserida nos modelos para que os mesmos possam obter sucesso nas previsões (PEDRO *et al.*, 2017). Existem diversos métodos de análise de séries temporais, mas as abordagens que utilizam modelos auto-regressivos integrados de médias móveis (ARIMA) (ALSHARIF *et al.*, 2019; COLAK *et al.*, 2015) regressão harmônica dinâmica (TRAPERO *et al.*, 2015) e suavização exponencial no espaço de estados (DONG *et al.*, 2013) destacam-se pelo bons resultados apresentados para previsões de irradiância de curto prazo.

- ARIMA

Este modelo pode ser considerado uma combinação entre os métodos de médias móveis e o auto-regressivo, utilizado para análise de séries temporais não estacionárias, aquelas que flutuam em torno de uma média variável. Tal modelo é representado pela Equação (6) para uma série temporal de uma variável X .

$$\Phi_p(B)W_t = \theta_q(B)e_t \quad (6)$$

Onde, $\Phi_p(B)$ é um operador auto-regressivo de ordem p , $\theta_q(B)$ é um operador de média móvel de ordem q e $W_t = \Delta^d X_t$. Com $\Delta^d X_t = (1 - B)^d X_t$ representando a equação geral de diferenças sucessivas para a d -ésima diferença. Geralmente, a ordem da diferença d é 1 ou 2, além disso, B representa o operador atraso unitário, como indicado na Equação (7).

$$\Delta^1 X_t = (1 - B)X_t = X_t - X_{t-1} \quad (7)$$

Para um melhor detalhamento sobre o método ARIMA, recomenda-se os trabalhos de Box (1994) e Morettin (2006).

- Regressão Harmônica Dinâmica (DHR)

Este modelo de componentes não observadas (UC) caracteriza-se por ser uma abordagem flexível para análise de séries temporais não estacionárias. O mesmo é particularmente útil para ajustes sazonais adaptativos, extração de sinais e interpolação, bem como para a previsão (YOUNG *et al.*, 1999). O método consiste em considerar a decomposição de uma série temporal, por exemplo de irradiância solar, y_t , em uma soma das componentes de tendência, T_t , uma componente sazonal, S_t e uma componente irregular, e_t que muitas vezes é modelada como um ruído gaussiano de média zero e variância constante, σ . Em seguida, considera-se tal soma como sendo equivalente a uma regressão harmônica com coeficientes que variam no tempo, como ilustra a Equação (8) (TRAPERO *et al.*, 2015).

$$y_t = T_t + S_t + e_t = \sum_{k=0}^{P/2} [a_{k,t} \cos(\omega_k t) + b_{k,t} \sin(\omega_k t)] + e_t \quad (8)$$

O modelo é completado quando o comportamento dinâmico da tendência estocástica e das subcomponentes sazonais são especificados. Existe uma larga faixa de opções disponíveis na literatura para se fazer essa especificação (WEST e HARRISON, 2006), (HARVEY, 1990), (PEDREGAL e YOUNG, 2002). Os algoritmos do filtro de Kalman (WELCH e BISHOP, 1995) e da suavização dos intervalos fixados (WALL *et al.*, 1981) são muitas vezes utilizados para realizar a determinação dos parâmetros $a_{k,t}$ e ω_k .

- Suavização exponencial no espaço de estados (ETS)

O modelo de suavização no espaço de estados é uma abordagem estocástica sofisticada que fornece uma estrutura para a obtenção de previsões de irradiância, bem como também para determinação de intervalos de previsão correspondentes. Todos os métodos de suavização exponencial correspondem a dois modelos no espaço de estados, um com erros aditivos e outro com erros multiplicativos. O modelo de suavização exponencial utilizado nos trabalhos de Dong *et al.* (2013), que por meio deste realizaram previsões de irradiância solar de curto prazo, provém de uma expansão da estrutura geral encontrada em Ord *et al.* (1997), como indicado pelas Equações (9) e (10)

$$y_t = h(x_{t-1}) + k(x_{t-1})e_t \quad (9)$$

$$x_t = f(x_{t-1}) + g(x_{t-1})e_t \quad (10)$$

Onde a Equação (9) é chamada de equação de observação e a Equação (10) de equação de estado. Para um melhor detalhamento deste método recomenda-se o trabalho de Dong *et al.* (2013).

Em Trapero *et al.* (2015), os autores realizaram previsões GHI de curto prazo utilizando os três métodos abordados anteriormente. Os resultados são reportados na Tabela 2.

Tabela 2 – Resultados obtidos para os ARIMA, DHR, ETS

| Modelos | rRMSE (%) |
|---------|-----------|
| ETS | 34,18 |
| ARIMA | 37,34 |
| DHR | 29,66 |

Fonte: Trapero *et al.* (2015)

3.4.3 Métodos baseados em aprendizagem de máquina

Estes métodos fazem parte do ramo da inteligência artificial e caracterizam-se pela sua capacidade de aprender e generalizar por meio da realização de um treinamento adequado. Os dois principais paradigmas de aprendizagem destes modelos são: Aprendizagem supervisionada e aprendizagem não supervisionada (JAMES *et al.*, 2013). Na aprendizagem supervisionada, a generalização ou ajuste ocorre por meio da aplicação dos algoritmos em uma amostra de treinamento rotulada, e cada observação consiste de um vetor de entradas único e de uma resposta desejada (HAYKIN, 2009).

Desta forma, é possível construir métricas de erro que funcionam como funções de custo e a seleção dos hiperparâmetros presentes nos modelos é feita com base na minimização de tais funções. A maioria dos trabalhos que utilizam aprendizagem de máquina para realizar previsões de irradiância solar aplicam algoritmos de aprendizagem supervisionada (PEDRO e COIMBRA, 2015; PEDRO *et al.*, 2018; YAGLI *et al.*, 2019; BENALI *et al.*, 2019). Modelos de aprendizagem supervisionada podem ser utilizados para a resolução de problemas de regressão e classificação.

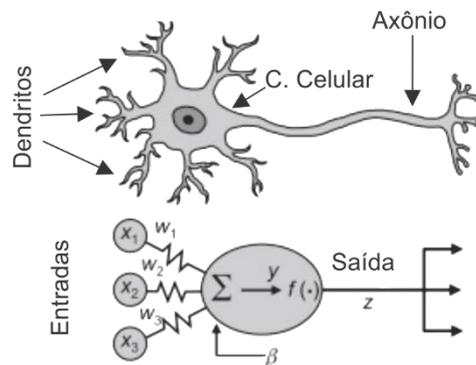
Dentre os mais populares pode-se citar: KNN (ROUSSOPOULOS *et al.*, 1995), SVR (CARRASCO *et al.*, 2019; DRUCKER *et al.*, 1997), regressão por mínimos quadrados (MYERS, 1990), métodos baseados em Splines (MARSH e CORMIER, 2001), árvores de

decisão (DIETTERICH, 2000), lasso (TIBSHIRANI, 1996), ridge (HOERL e KENNARD, 1970), entre outros.

Na aprendizagem não supervisionada, nenhum tipo de informação sobre a resposta desejada é repassada ao algoritmo que tenta generalizar de forma independente. Os seus principais usos estão relacionados a análise exploratória de dados e para execução de procedimentos de *clustering*. Seus principais representantes são os algoritmos de *clustering* hierárquico (JOHNSON, 1967) e o *K-means* (JAIN, 2010).

Vale destacar também os percéptrons de múltiplas camadas (MLP) que são o tipo mais clássico de redes neurais artificiais (RNA) e são considerados modelos de *benchmarking* devido ao seu bom desempenho em muitas aplicações. Tal modelo tem inspiração biológica, especificadamente nos neurônios biológicos, como indicado pela Figura 3, que ilustra uma representação de um neurônio biológico e a arquitetura de um neurônio artificial com três entradas.

Figura 3 – Neurônio biológico em conjunto com um neurônio artificial com de três entradas



Fonte: Pedro e Coimbra (2017)

Pela análise da Figura 3, tem-se na terminologia de RNA que o vetor $\mathbf{x} = [x_1, x_2, x_3]$ é o vetor de entradas, $\mathbf{w} = [w_1, w_2, w_3]$ é o vetor de pesos sinápticos, $f(\cdot)$ é a função de ativação e y é muitas vezes chamado de campo local induzido (HAYKIN, 2009), que é definido pela Equação (11).

$$y = \beta + \sum_{i=1}^3 w_i x_i \quad (11)$$

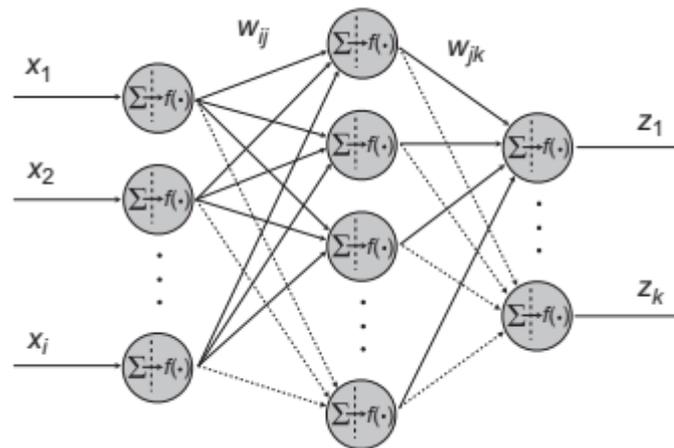
Onde β representa um viés externo. Então, a saída z é determinada quando se calcula o valor da função de ativação no campo local induzido, assim $z = f(y)$. Existem diversos tipos de funções de ativação que podem ser selecionadas (HAYKIN, 2009), mas a função sigmóide ou

logística é geralmente a mais aplicada devido às suas atrativas propriedades matemáticas, tais como, monotonicidade, continuidade e diferenciabilidade. Assim, a saída z é calculada pela Equação (12)

$$z = \frac{1}{1 + e^{-(\beta + \sum_{i=1}^3 w_i x_i)}} \quad (12)$$

A arquitetura mais comumente utilizada de redes MLP é aquela com uma camada oculta e as camadas de entrada e saída, como representado na Figura 4.

Figura 4 – Arquitetura de uma rede MLP com três camadas



Fonte: Pedro e Coimbra (2017)

Os trabalhos mais recentes que utilizam RNA para a previsão de irradiância solar utilizam redes *deep learning*, tais como as redes neurais convolucionais (CNN) (ZHAO *et al.*, 2019) e as redes neurais recorrentes de memória de longo prazo (LSTM) (QING e NIU, 2018).

3.4.4 Métodos baseados em processamento de imagens

Previsão de irradiância solar por meio de técnicas de processamento de imagens podem ser divididas em duas categorias: as que usam algoritmos de processamento de imagens para detecção de vetores de movimento de nuvens (CMV), que em alguns casos utilizam o método da correlação cruzada máxima para realizar essa determinação (ALONSO-MONTESINOS *et al.*, 2015) ou o uso de algoritmos de fluxos ópticos (NONNENMACHER e COIMBRA, 2014) e aquelas que utilizam métodos de processamento de imagens para extrair atributos de imagens do céu para a posterior adição dos mesmos em algoritmos de aprendizagem de máquina para a

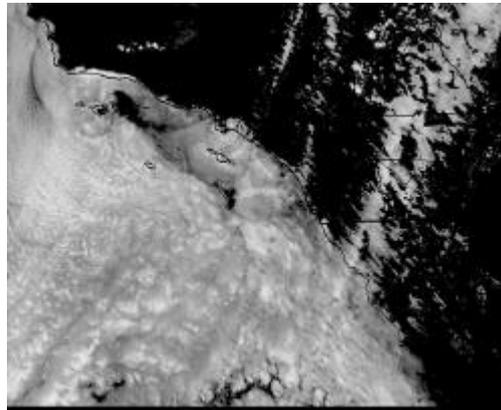
realização das previsões (PEDRO e COIMBRA, 2015), (PEDRO *et al.*, 2018). Há também trabalhos que buscam estimar a transmitância e a velocidade de nuvens por meio de imagens do céu para avaliar o impacto que estas grandezas exercem sobre a acurácia de previsões DNI (LI *et al.*, 2016). As imagens utilizadas são em sua maioria obtidas por câmeras de segurança apontadas para o céu ou imagens de satélites como indicado pelas Figuras 5 e 6.

Figura 5 – Imagem do céu obtida por uma câmera de segurança



Fonte: Li *et al.* (2016)

Figura 6 – Imagem obtida por satélite



Fonte: Nonnemacher e Coimbra (2014)

Este trabalho se enquadra na categoria que utiliza processamento de imagens para a extração de atributos e posterior adição dos mesmos em modelos de aprendizagem de máquina. A Figura 7 ilustra um exemplo de imagem que foi utilizada neste trabalho.

Figura 7 – Exemplo de imagem obtida neste trabalho



Fonte: Autor

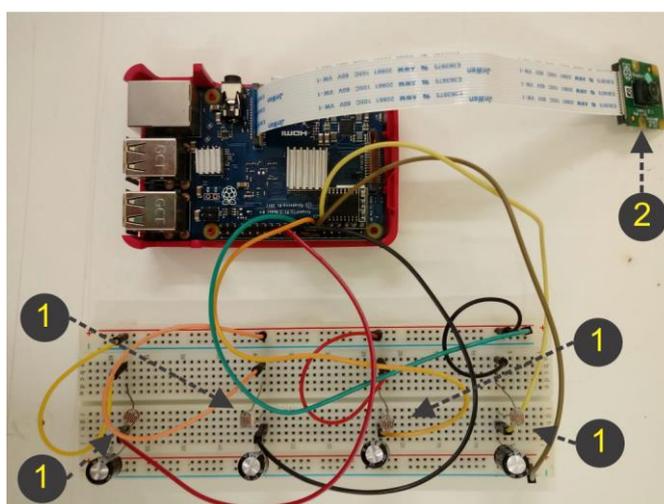
4 METODOLOGIA

Neste tópico é abordada a coleta dos dados, bem como explica-se o procedimento de treinamento dos algoritmos. Também será dado um melhor detalhamento sobre os modelos de aprendizagem de máquina utilizados, além de se ter uma pequena discussão sobre filtragens de imagens digitais e a apresentação das métricas de erro aplicadas para a avaliação da acurácia das previsões.

4.1 Dados

Os dados foram coletados no Laboratório de Energia Solar e Gás Natural (LESGN) do Centro de Tecnologia da Universidade Federal do Ceará, na cidade de Fortaleza-CE, Brasil $3^{\circ}43'6''S$ e $38^{\circ}32'36''O$, entre os dias de 03/06/2019 à 07/06/2019 e na faixa das 8:00 às 17:30. Os sensores LDR's e a câmera foram integrados por meio do uso de um *Raspberry Pi 3*, um computador que tem o tamanho de um cartão usado principalmente em projetos de programação, robótica e em iniciativas em geral com software e hardware livres. O sistema de aquisição já montado é representado na Figura 8, onde o número 1 indica os LDR's e o 2 a câmera.

Figura 8 – Sistema de aquisição montado



Fonte: Autor

Neste estudo, foram realizadas previsões de irradiância solar global para os horizontes de 10, 20 e 30 minutos a posteriori por meio da aplicação dos modelos de aprendizagem de máquina em quatro conjuntos de dados:

- Conjunto I: Um formado por atributos que correspondem a sinais obtidos por sensores de resistência dependente da luz (LDR's), onde foram utilizados quatro LDR's. Além disso, considerou-se a irradiância do instante atual como um preditor. Desta forma, o conjunto de dados ficou com 5 atributos e 1466 observações;
- Conjunto II: Formado pelos atributos do conjunto anterior, adicionado de descritores estatísticos (média, desvio padrão e entropia de Shannon) para os canais Vermelho, Verde e Azul (RGB) de imagens do céu, totalizando 14 atributos e 1466 observações;
- Conjunto III: Formado pelos atributos do primeiro conjunto de dados adicionado de descritores estatísticos (média, desvio padrão, entropia de Shannon) extraídos de imagens do céu em tons de cinza filtradas pela aplicação do filtro de suavização da mediana, assim totalizando 8 atributos e 1466 observações;
- Conjunto IV: Formado pelos atributos do primeiro conjunto de dados adicionado de descritores estatísticos (média, desvio padrão, entropia de Shannon) extraídos de imagens do céu em tons de cinza filtradas pela aplicação do filtro de aguçamento do laplaciano da gaussiana, assim totalizando 8 atributos e 1466 observações.

Posteriormente, os algoritmos foram aplicados nos mesmos conjuntos de dados diferindo apenas pela substituição dos atributos correspondentes aos sinais obtidos pelos sensores de luminosidade por funções de recorrência sobre os valores de irradiância de instantes anteriores.

As funções de recorrência operam sobre os valores de índice de clareza, K_t . Assim, tendo a série temporal de irradiância solar, determinou-se a série de K_t 's correspondente. Em seguida, calculou-se por meio das funções os valores dos novos preditores: $\mathbf{B}(t)$ (média retrógrada), $\mathbf{L}(t)$ (média atrasada), $\mathbf{V}(t)$ (variabilidade). Estes preditores são vetores e são dados pelas Equações (13), (14), (15).

$$B_i(t) = \frac{\sum_{t \in [t-i\delta, t]} K_t(t)}{N} \quad (13)$$

$$L_i(t) = \frac{\sum_{t \in]t-i\delta, t-(i-1)\delta]} K_t(t)}{N} \quad (14)$$

$$V_i = \sqrt{\frac{\sum_{t \in]t-i\delta, t]}{\Delta K_t^2(t)}}{N}} \quad (15)$$

Onde $\Delta K_t(t) = K_t(t) - K_t(t - \Delta t)$ e Δt é o horizonte de previsão considerado.

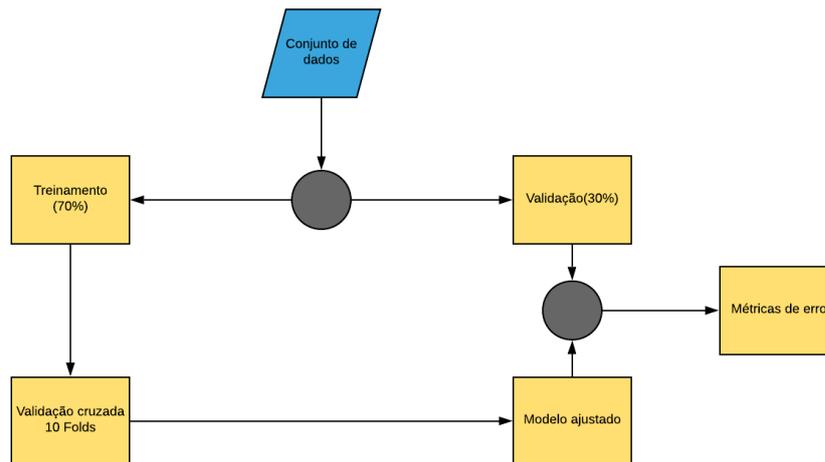
Seguindo a metodologia dos trabalhos de Pedro e Coimbra (2015), utilizou-se um $\delta = 5$ minutos, já o índice i varia de 1 a 3, indicando que as médias sobre os valores de K_t 's são determinadas, para a menor janela de tempo, tomando os valores de K_t para os cinco minutos anteriores ao instante em questão. Para a maior janela de tempo, as médias são realizadas sobre os valores de K_t de quinze minutos anteriores ao instante em questão. Por meio destes preditores foi adicionado informações sobre os valores de irradiância solar de instantes anteriores.

Assim os algoritmos foram aplicados em quatro novos conjuntos de dados preservando apenas a informação relativa às imagens do céu:

- Conjunto A: Um formado por atributos que correspondem aos vetores $\mathbf{L}(t)$, $\mathbf{B}(t)$ e $\mathbf{V}(t)$. Além disso, considerou-se a irradiância do instante atual como um preditor. Desta forma, o conjunto de dados ficou com 10 atributos e 1466 observações;
- Conjunto B: Formado pelos atributos do conjunto anterior, adicionado de descritores estatísticos (média, desvio padrão e entropia de Shannon) para os canais Vermelho, Verde e Azul (RGB) de imagens do céu, totalizando 19 atributos e 1466 observações;
- Conjunto C: Formado pelos atributos do primeiro conjunto de dados adicionado de descritores estatísticos (média, desvio padrão, entropia de Shannon) extraídos de imagens do céu em tons de cinza filtradas pela aplicação do filtro de suavização da mediana, assim totalizando 13 atributos e 1466 observações;
- Conjunto D: Formado pelos atributos do primeiro conjunto de dados adicionado de descritores estatísticos (média, desvio padrão, entropia de Shannon) extraídos de imagens do céu em tons de cinza filtradas pela aplicação do filtro de aguçamento do laplaciano da gaussiana, assim totalizando 13 atributos e 1466 observações.

Para o treinamento dos modelos, os conjuntos de dados foram divididos aleatoriamente de tal forma que 70% ficou para o treinamento dos algoritmos e o restante constituiu o conjunto de validação. O ajuste dos métodos foi feito por meio da aplicação de validação cruzada 10-Fold (JAMES *et al.*, 2013). Esta etapa de treinamento é representada pela Figura 9.

Figura 9 – Representação da etapa de treinamento



Fonte: Autor

4.2 Modelos utilizados para realizar a previsão

4.2.1 Modelo de baseline

O modelo de baseline escolhido atribui para uma nova observação a irradiância média do conjunto de treinamento (HYNDMAN e ATHANASOPOULOS, 2018). Seja $T = \{G_1, G_2, \dots, G_n\}$ os valores de irradiância do conjunto de treinamento. O valor atribuído para cada nova observação é dado pela Equação (16).

$$G = \frac{\sum_{i=1}^n G_i}{n} \quad (16)$$

Funciona como um modelo que fornece um desempenho mínimo, atuando como uma cota inferior para a acurácia dos demais modelos utilizados. Espera-se que o mesmo forneça os maiores valores para as métricas de erro RMSE e MAE.

4.2.2 *K-Vizinhos mais Próximos (KNN)*

Este conhecido método é geralmente aplicado em problemas de classificação, porém o mesmo admite uma versão para problemas de regressão, onde o algoritmo identifica os K vizinhos, que estão no conjunto de treinamento, mais próximos de uma nova observação por meio do uso de uma métrica de dissimilaridade. Na presente dissertação, utilizou-se a distância euclidiana, e atribui a média aritmética das irradiâncias de tais vizinhos. A irradiância estimada para a nova observação será como na Equação (17).

$$G = \frac{\sum_{i=1}^K G_i}{K} \quad (17)$$

Onde, G_i , com $i = \{1, 2, \dots, K\}$ são as irradiâncias dos K vizinhos mais próximos. O número de vizinhos K é, portanto, o único hiperparâmetro deste modelo.

4.2.3 *Regressão por Vetor Suporte (SVR)*

Este modelo por mais que seja uma versão de uma Máquina de Vetor Suporte (SVM), apresenta a diferença de ser utilizado em problemas de regressão, enquanto o SVM é um método comumente aplicado no contexto de classificação. Uma das grandes vantagens em se trabalhar com o SVR se deve à possibilidade de construção de funções não lineares complexas, além de se ter a garantia de otimalidade global via a solução de um programa quadrático convexo (QP), treinamento efetivo e, geralmente, apresentando um desempenho preditivo positivo com um reduzido risco de *overfitting* graças ao uso do princípio de minimização do risco estrutural. Neste estudo, o algoritmo SVR aplicado foi do tipo ε -SVR, onde ocorre a construção de um ε -tubo em que os erros são ignorados no seu interior, mas penalizados fora do mesmo (CARRASCO et al, 2019; DRUCKER *et al.*, 1997).

Dado $x_i \in \mathbb{R}^n$ e as saídas correspondentes $y_i \in \mathbb{R}$, para $i = \{1, 2, \dots, m\}$ o ε -SVR linear busca obter um hiperplano $f(x) = w^T x + b$ pela construção de um ε -tubo como explicado anteriormente; para tanto, faz-se necessária a solução do QP.

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \quad & \frac{1}{2} \|w\|^2 + Ce^T (\xi + \xi^*) \\ \text{s. t.} \quad & y - (Aw + be) \leq \varepsilon e + \xi \\ & (Aw + be) - y \leq \varepsilon e + \xi^* \end{aligned}$$

Onde, $A = [x_1 \ x_2 \ \dots \ x_m]^T \in \mathbb{R}^{m \times n}$, $y = (y_1, y_2, \dots, y_m) \in \mathbb{R}^m$, $e \in \mathbb{R}^m$ é um vetor de uns, $\xi, \xi^* \in \mathbb{R}^m$ são variáveis de folga que penalizam as observações que estão fora do ε -tubo e $C > 0$ é um fator que regulariza o *trade-off* de maior ou menor flexibilidade do modelo, como pode-se notar tanto a função objetivo como as funções restrições de desigualdade são convexas, portanto o QP é convexo e pode ser resolvido pela aplicação de algum método eficiente para solução de programas convexas, como o método dos pontos interiores (MEHROTRA, 1992).

O ε -SVR pode ser estendido para uma configuração não linear pelo uso de *kernels*. Isto pode ser feito pela aplicação de um artifício no problema dual correspondente (DRUCKER *et al.*, 1997). A versão de *kernels* para o ε -SVR é dada por

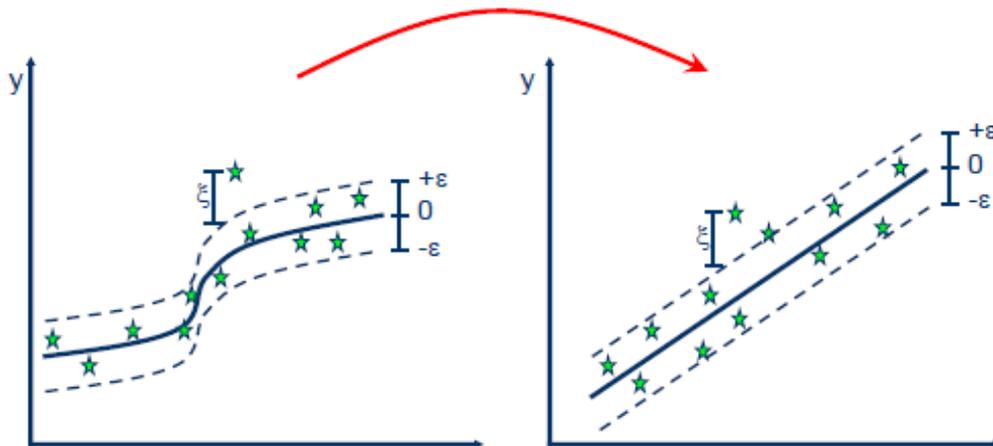
$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & \frac{1}{2} (\alpha - \alpha^*)^T K(A, A^T) (\alpha - \alpha^*) - y^T (\alpha - \alpha^*) + \varepsilon e^T (\alpha + \alpha^*) \\ \text{s. t.} \quad & e^T (\alpha - \alpha^*) = 0 \\ & 0 \leq \alpha, \alpha^* \leq C e \end{aligned}$$

Onde $K(A, A^T) \in \mathbb{R}^{m \times m}$ representa a matriz de funções *kernel* da forma $k_{is} = \kappa(x_i, x_s)$, $\kappa: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ e α, α^* são vetores de variáveis duais (multiplicadores de Lagrange do problema primal). O *kernel* utilizado neste estudo foi o radial gaussiano, que apresenta as funções de *kernel* da forma dada pela Equação (18).

$$\kappa(x_i, x_s) = \exp\left(-\frac{\|x_i - x_s\|^2}{2\sigma^2}\right) \quad (18)$$

Neste trabalho, tomou-se $\varepsilon = 0,1$ e os parâmetros σ e C são obtidos por validação cruzada 10-Fold no conjunto de treinamento. Uma representação esquemática do uso de *kernels* para se trabalhar com funções não lineares é ilustrado na Figura 10.

Figura 10 – Representação esquemática do uso de kernels em SVR



Fonte: Disponível em <https://www.saedsayad.com/support_vector_machine_reg.htm> Acesso em Junho 2019

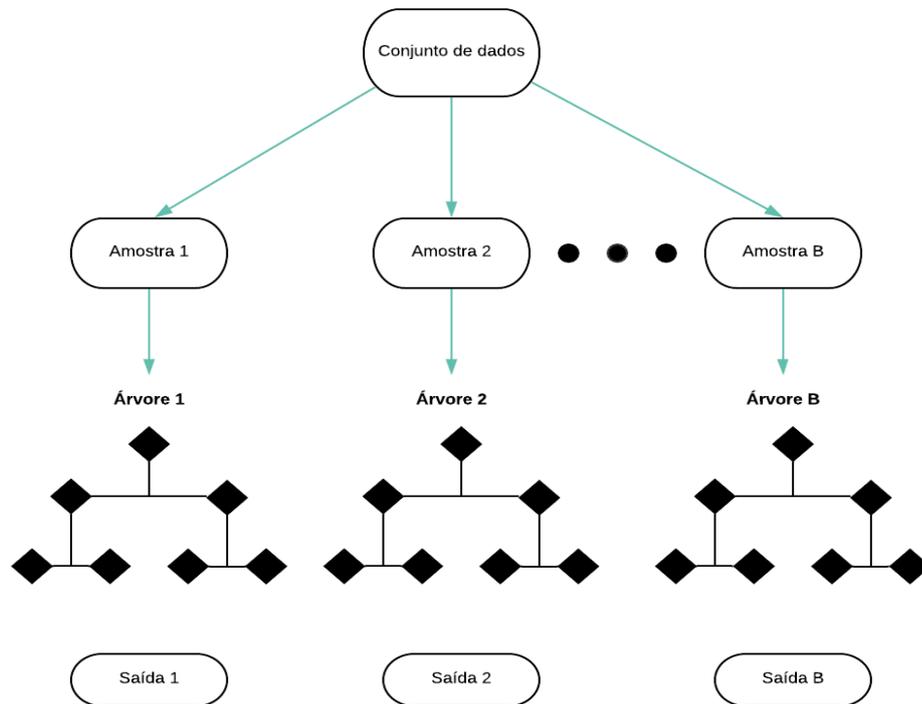
4.2.4 *Árvore de decisão com a aplicação de um procedimento de Bagging*

As árvores de decisão envolvem a segmentação do espaço de atributos por meio da aplicação de regras, as quais podem ser resumidas em um formato de árvore, proporcionando o desenvolvimento de métodos simples e de fácil interpretação. Entretanto, estas tipicamente não são competitivas quando se compara a sua acurácia de previsão com as de outros modelos clássicos de aprendizagem supervisionada. Desta forma, existem algumas abordagens que envolvem a formação de múltiplas árvores que são combinadas para conduzir a uma única previsão, algo similar ao que ocorre nas máquinas de comitê (DIETTERICH, 2000).

Tal combinação pode resultar em melhorias significativas na acurácia de previsão, mas com uma perda na capacidade de interpretação (JAMES *et al.*, 2013). Uma destas metodologias de combinação de árvores é o *Bootstrap Aggregation (Bagging)*, que é um método *Ensemble* que combina as saídas de múltiplos algoritmos de máquina em um único valor. Esta metodologia é geralmente utilizada com o intuito de redução da variância de modelos de alta variância, como os das árvores de decisão (JAMES *et al.*, 2013; KUHN e JOHNSON, 2013).

O primeiro passo neste procedimento é a realização de um *Bootstrap* no conjunto de dados considerado, que consiste na formação de várias amostras através da realização de múltiplas amostragens com reposição no conjunto de dados original. Posteriormente, para cada amostra desenvolve-se uma árvore de decisão e, por fim, as saídas resultantes de cada árvore serão combinadas por uma média aritmética para a obtenção do resultado final. Todo este procedimento está representado na Figura 11.

Figura 11 – Representação esquemática de um *Bagging* aplicado em árvores de decisão



Fonte: Autor

Onde tem-se o vetor

$$O = \{\text{Saída 1}, \text{Saída 2}, \dots, \text{Saída B}\} = \{f_1^*(x), f_2^*(x), \dots, f_B^*(x)\}.$$

Assim, a saída resultante do *Bagging* será dada por pela Equação (19).

$$f^*(x) = \frac{\sum_{i=1}^B f_i^*(x)}{B} \quad (19)$$

Neste caso, percebe-se que o único hiperparâmetro presente neste método é o número de árvores considerada, onde pode-se assumir um valor arbitrariamente grande para garantir bons resultados, mas tendo sempre cuidado para que o tempo computacional não se torne elevado. Para o presente estudo, considerou-se 500 árvores, que foi o valor que proporcionou um melhor balanceamento entre desempenho e tempo computacional.

4.2.5 Florestas Aleatórias

O algoritmo de Florestas Aleatórias é análogo ao *Bagging*, mas com o uso de uma pequena modificação visando descorrelacionar as árvores (JAMES *et al.*, 2013). Uma vez que, no *Bagging* para cada amostra ocorre o desenvolvimento de uma árvore por meio da

segmentação do mesmo espaço de preditores, este procedimento tende a ocasionar a obtenção de respostas que apresentam um elevado grau de correlação.

No geral, combinações de previsões de vários modelos fornecem melhores resultados quando as saídas dos sub-modelos são descorrelacionadas ou na melhor das hipóteses apresentam um nível fraco de correlação. Basicamente, a proposta do algoritmo é a de que para a formação de cada árvore seja feita a divisão do espaço de m atributos selecionados aleatoriamente de um total de p preditores, onde, tipicamente, seleciona-se $m = \sqrt{p}$. Entretanto, por meio do processo de aprendizagem é possível obter o valor de m ótimo.

4.2.6 Máquina de Aprendizagem Mínima (MLM)

O algoritmo MLM foi originalmente apresentado em trabalhos recentes (DE SOUZA JÚNIOR *et al.*, 2015). A metodologia consiste na construção de um mapeamento linear (regressão linear múltipla) entre as matrizes distâncias de entrada e saída. Posteriormente, o mapeamento linear ajustado (ótimo) é utilizado para estimar as distâncias de K pontos de referência no espaço das saídas em relação à resposta que deseja-se determinar. Por fim, a saída é estimada via a solução de um problema de multilateração. A formulação geral do MLM é perfeitamente capaz de trabalhar com problemas de regressão não linear, bem como no contexto em que o espaço das saídas é multidimensional (DE SOUZA JÚNIOR *et al.*, 2015).

Seja um conjunto com N observações com atributos $X = \{x_i\}$, $i = \{1, 2, \dots, N\}$, $x_i \in \mathbb{R}^D$ e o conjunto com as saídas correspondentes $Y = \{y_i\}$, $i = \{1, 2, \dots, N\}$, $y_i \in \mathbb{R}^S$. Assume-se a existência de um mapeamento contínuo entre o espaço de entrada e o de saída $f: X \rightarrow \Omega$ dado pela Equação (20).

$$y = f(x) + \varepsilon \quad (20)$$

Com y e x sendo vetores colunas e $\varepsilon \in \mathbb{R}^{N \times S}$ a matriz de resíduos. O MLM opera em duas etapas.

1. Ajuste de uma regressão linear múltipla para obter matrizes distância de saída a partir de matrizes distância de entrada;
2. Estima a resposta desejada por meio da solução de um problema de multilateração.

- Regressão Linear Múltipla entre matrizes distância

Esta etapa inicia-se com a escolha aleatória de algumas observações do conjunto de dados onde as mesmas serão denominadas pontos de referência. Vale ressaltar que considera-se para cada ponto de referência os seu atributos e suas saídas correspondentes de forma separada, formando assim os conjuntos $\mu = \{m_k\}$, $k = \{1, 2, \dots, K\}$, $\mu \subset X$ que são constituídos pelos atributos dos pontos de referência e o conjunto $T = \{t_k\}$, $k = \{1, 2, \dots, K\}$, $T \subset Y$ das saídas correspondentes. Define-se a matriz $D_x \in \mathbb{R}^{N \times K}$, tal que sua k -ésima coluna representa as distâncias $d(x_i, m_k)$, $i = \{1, 2, \dots, N\}$ entre os atributos x_i e os preditores do k -ésimo ponto de referência. De forma análoga, tem-se a matriz $\Delta_y \in \mathbb{R}^{N \times K}$, tal que sua k -ésima coluna representa as distâncias $\delta(y_i, t_k)$, $i = \{1, 2, \dots, N\}$ entre a saída y_i e a resposta do k -ésimo ponto de referência. Basicamente, assume-se que há uma relação linear entre as matrizes distância da entrada e matrizes distância de saída, obtendo assim um problema de regressão linear múltipla, como representado pela Equação (21).

$$\Delta_y = D_x B + E \quad (21)$$

$B \in \mathbb{R}^{K \times K}$ sendo a matriz dos coeficientes. O ajuste deste modelo pode ser feito por meio da solução de um problema de mínimos quadrados, ou seja, pela minimização da função custo, dada pela Equação (22)

$$RSS(B) = tr \left((\Delta_y - D_x B)^T (\Delta_y - D_x B) \right) \quad (22)$$

Para o trabalho em questão, o número de observações é tal que $N > K$. Desta forma, a solução por mínimos quadrados fornece a Equação (23).

$$B^* = (D_x^T D_x)^{-1} D_x^T \Delta_y \quad (23)$$

Para o cálculo da inversa da Equação (23), recomenda-se determinar a pseudo-inversa de Moore-Penrose (BEN-ISRAEL e GREVILLE, 2003; PENROSE, 1955). Com isto, para uma observação de teste com atributos $x \in \mathbb{R}^D$, calcula-se as distâncias em relação aos vetores $\{m_k\}$, $k = \{1, 2, \dots, K\}$ obtendo assim o vetor $d(x, \mu) = [d(x, m_1), d(x, m_2), \dots, d(x, m_K)]$, onde por meio do modelo de regressão linear múltipla ajustado, pode-se estimar as distâncias entre a

resposta desejada y e as saídas conhecidas $\{t_k\}$, $k = \{1, 2, \dots, K\}$ dos pontos de referência, como ilustrado na Equação (24).

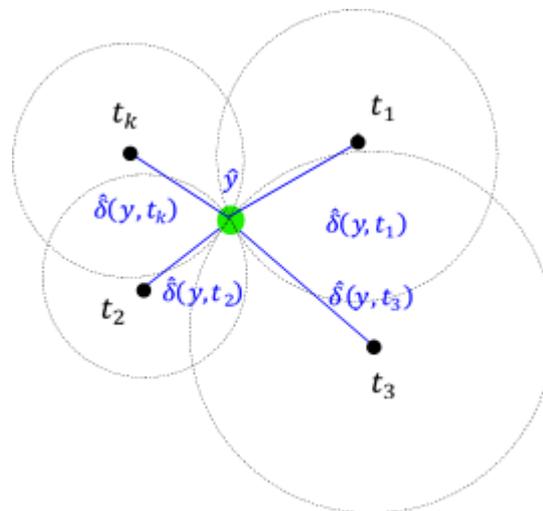
$$\delta^*(y, T) = d(x, \mu)B^* \quad (24)$$

Obtendo assim o vetor $\delta^*(y, T) = [\delta^*(y, t_1), \delta^*(y, t_2), \dots, \delta^*(y, t_K)]$

- Estimativa das Respostas

Estimar as respostas conhecendo as saídas $\{t_k\}$, $k = \{1, 2, \dots, K\}$ dos pontos de referência e as estimativas das distâncias $\delta^*(y, T) = [\delta^*(y, t_1), \delta^*(y, t_2), \dots, \delta^*(y, t_K)]$ pode ser visto como um problema de multilateração (NIEWIADOMSKA-SZYNKIEWICZ e MARKS, 2009). Existem muitas estratégias que podem ser adotadas para a solução de tal problema, e a mais utilizada busca determinar \hat{y} por meio de uma interpretação geométrica, que gera K equações não lineares relacionadas à consideração de se ter K hiperesferas S -dimensional centradas nos t_k e passando por \hat{y} com raios iguais a $\delta^*(y, t_k)$, que para o caso bidimensional é como ilustrado na Figura 12.

Figura 12 – Interpretação geométrica para a solução do problema de multilateração



Fonte: Autor

Desta forma, tem-se

$$(\hat{y} - t_k)^T (\hat{y} - t_k) = (\delta^*(y, t_k))^2 \quad (25)$$

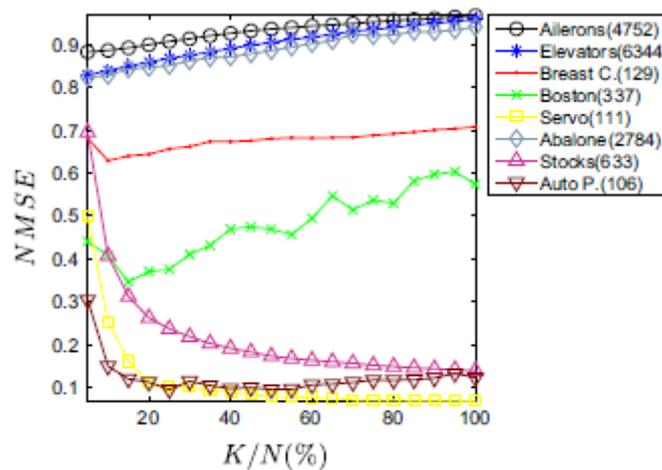
Assim, a resposta é estimada pela minimização da função objetivo dada pela Equação (26)

$$J(y^*) = \sum_{k=1}^K ((\hat{y} - t_k)^T (\hat{y} - t_k) - (\delta^*(y, t_k)))^2 \quad (26)$$

Por ser um problema de mínimos quadrados não linear, recomenda-se o uso do algoritmo de Levenberg-Marquardt (MARQUARDT, 1963; MORÉ, 1978) para a solução de tal problema de otimização. Assim, percebe-se que o único hiperparâmetro presente no modelo é o número de pontos de referência K .

No trabalho de apresentação do método (DE SOUZA JÚNIOR *et al.*, 2015), os autores aplicaram o MLM em doze conjuntos de dados considerados de *benchmarking* disponibilizados no repositório da Universidade da Califórnia (www.ics.uci.edu/~mllearn/), tanto para propósitos de regressão como de classificação. No contexto de regressão, havia 8 conjuntos de dados, onde para todos a seleção do número ótimo de pontos de referência se deu por validação cruzada 10-*Fold* no conjunto de treinamento. Os resultados são reportados na Figura 13, onde o *NMSE* representa a métrica *Normalized Mean Squared Error* que é similar ao *RMSE* e o número de pontos de referência é normalizado pelo número de observações no conjunto de treinamento.

Figura 13 – Busca por hiperparâmetro no MLM para outros conjuntos de dados



Fonte: De Souza Junior *et al.* (2015)

Os números entre parênteses na legenda da Figura 13 representam o número de observações no conjunto de treinamento em questão. Assim, evidencia-se que de fato o comportamento na seleção do número de pontos de referência é distinto de acordo com o conjunto de dados em que o MLM é aplicado.

4.3 Filtragem espacial de imagens digitais

Nesta seção, apresenta-se uma pequena fundamentação com respeito à filtragem espacial de imagens digitais, tratando um pouco sobre os conceitos de convolução e correlação espacial, bem como tem-se um detalhamento sobre os filtros (também chamados de máscaras, *kernels*, *templates* e janelas) de suavização da mediana e o filtro de aguçamento do laplaciano da gaussiana.

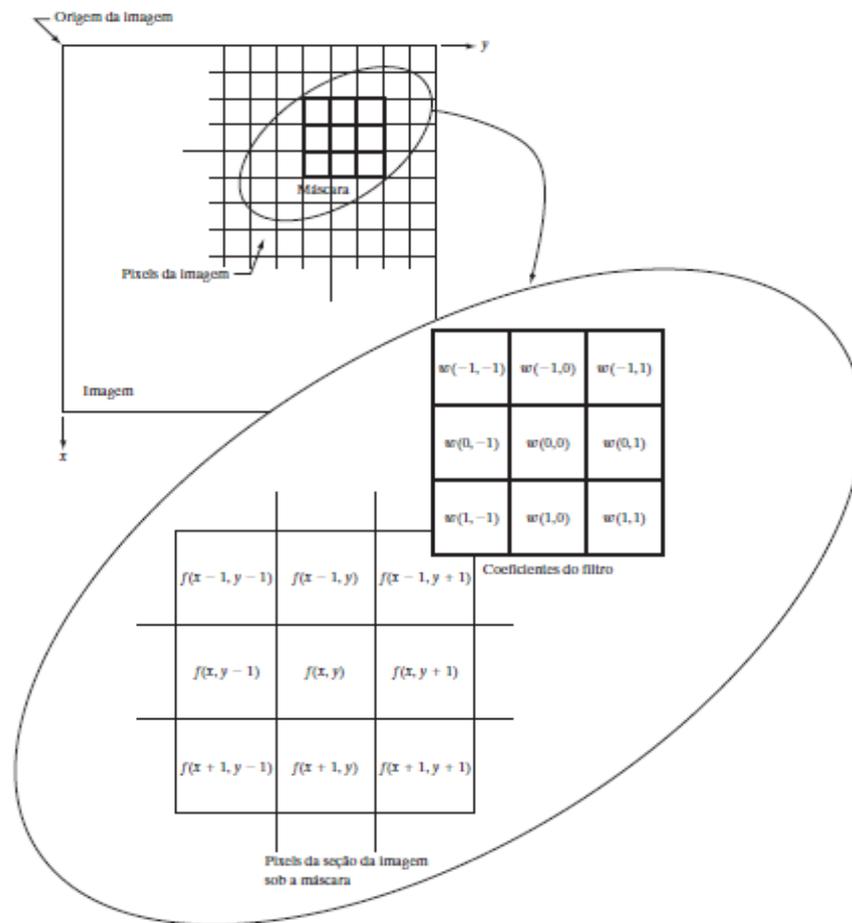
4.3.1 Convolução e Correlação espacial

Um filtro espacial consiste em: (1) uma vizinhança (normalmente um pequeno retângulo), (2) uma operação predefinida realizada sobre os pixels da imagem incluídos na vizinhança. A filtragem cria um novo pixel com coordenadas iguais às coordenadas do centro da vizinhança e cujo valor é o resultado da operação de filtragem.

Caso a operação realizada sobre os pixels da imagem seja linear, o filtro é chamado de filtro espacial linear (GONZALEZ, 2009). Vale destacar que o pixel filtrado normalmente é atribuído a uma posição correspondente em uma nova imagem criada para receber os resultados da filtragem.

Raramente, ocorre de os pixels filtrados substituírem os valores da posição correspondente na imagem original, já que isso alteraria o conteúdo da imagem, enquanto a filtragem ainda continua sendo realizada. A imagem filtrada é gerada à medida que o centro do filtro percorre cada pixel da imagem de entrada (GONZALEZ, 2009). Este procedimento é representado para um filtro 3×3 na Figura 14.

Figura 14 – Representação do processo de filtragem espacial



Fonte: GONZALEZ (2009)

Em qualquer ponto (x, y) da imagem, a resposta, $g(x, y)$, do filtro, é a soma dos produtos dos coeficientes do filtro com os pixels da imagem englobados pelo mesmo:

$$g(x, y) = w(-1, -1)f(x - 1, y - 1) + w(-1, 0)f(x - 1, y) + \dots + w(0, 0)f(x, y) + \dots + w(1, 1)f(x + 1, y + 1)$$

Em geral, a filtragem espacial linear de uma imagem com dimensões $M \times N$ com um filtro de dimensões $m \times n$ é dada pela Equação (27)

$$g(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t)f(x + s, y + t) \quad (27)$$

Onde x e y variam de forma que cada pixel em w percorre todos os pixels em f e $a = \frac{m-1}{2}$ e $b = \frac{n-1}{2}$. Desta forma, o procedimento de mover uma máscara pela imagem e calcular a

soma dos produtos em cada posição, exatamente como explicado anteriormente caracteriza a operação de correlação. No caso da convolução tem-se um procedimento análogo, exceto ao fato de o filtro ser rotacionado de 180° graus antes de se realizar a operação de filtragem, assim a correlação e a convolução de um filtro $w(x, y)$ de tamanho $m \times n$ com uma imagem $f(x, y)$ de tamanho $M \times N$ são dados pelas Equações (28) e (29).

$$w(x, y) \odot f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x + s, y + t) \quad (28)$$

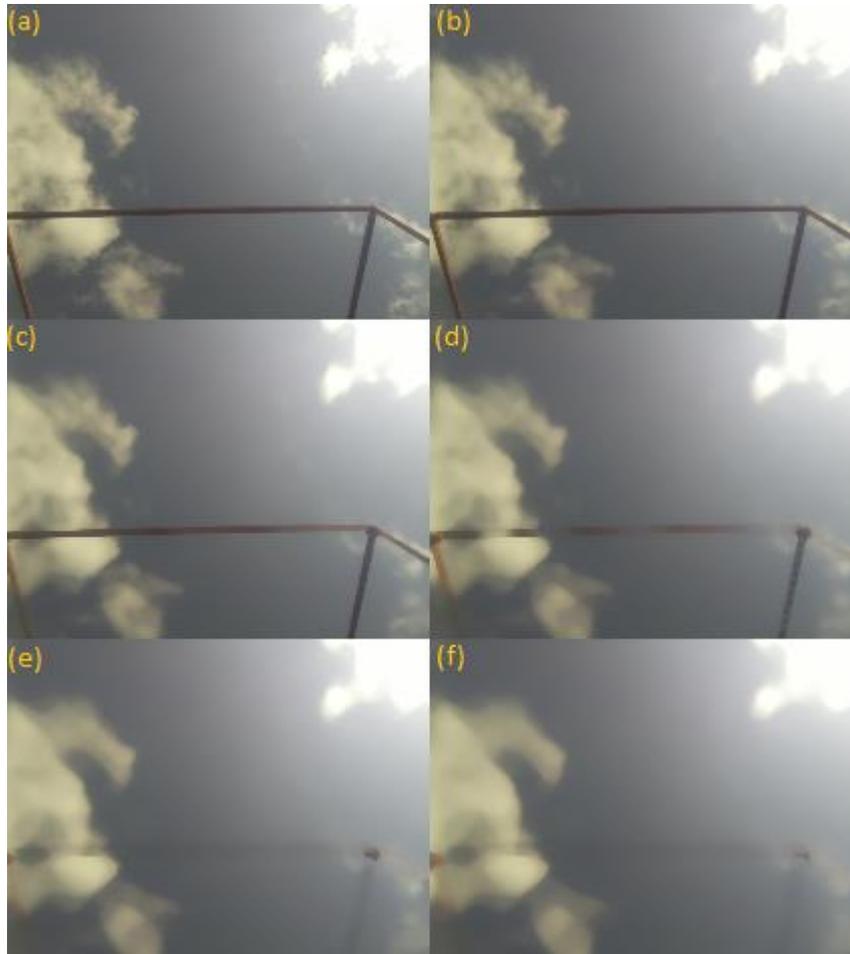
$$w(x, y) \star f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x - s, y - t) \quad (29)$$

Para a realização das filtragens espaciais de imagens neste trabalho foi utilizada a operação de convolução, e a dimensão das máscaras aplicadas foi de 5×5 , onde este tamanho foi escolhido por ser aquele que proporcionou um melhor impacto sobre as imagens sem que houvesse grande perda de informação relativa à imagem original.

4.3.2 Filtro de suavização da mediana

Filtros de suavização ou filtros passa-baixa caracterizam-se por atenuarem regiões de alta frequência na imagem, regiões de alta variabilidade nas intensidades de cinza, tais como, ruídos e contornos. O aspecto geral proporcionado pela aplicação destas máscaras nas imagens é o borramento das mesmas. O filtro de suavização da mediana atribui para os pixels da imagem filtrada a mediana dos pixels englobados pela máscara e, de uma forma geral, aumentando o tamanho do *kernel* o nível de borramento na imagem filtrada se torna mais acentuado, como ilustra a Figura 15 para a aplicação do filtro da mediana com *kernels* quadrados de tamanho 3, 5, 7, 9 e 11 em uma imagem utilizada neste trabalho.

Figura 15 – (a) Imagem original, (b) Imagem filtrada com mediana de tamanho 3, (c) Imagem filtrada com mediana de tamanho 5, (d) Imagem filtrada com mediana de tamanho 7, (e) Imagem filtrada com mediana de tamanho 9, (f) Imagem filtrada com mediana de tamanho 11



Fonte: Autor

Como se pode perceber pela Figura 15, de fato um filtro de suavização atenua regiões de alta frequência, como se nota pelo artefato metálico que basicamente desaparece quando se intensifica o efeito do filtro da mediana pelo aumento da dimensão de sua máscara.

4.3.3 Filtro de aguçamento do laplaciano da gaussiana (LoG)

Filtros de aguçamento ou filtros passa-alta apresentam uma funcionalidade contrária aos filtros de suavização. Neste caso, tem-se uma atenuação nas regiões de baixa frequência, regiões de pequena variabilidade na intensidades de cinza dos pixels, e um realce nas regiões de alta frequência. Estes filtros são muito utilizados para propósitos de detecção de bordas. Neste estudo considerou-se o filtro de aguçamento do laplaciano da gaussiana cuja máscara é obtida pela aplicação do operador laplaciano, como dado pela Equação (30).

$$\nabla^2 f(x, y) = \frac{\partial^2}{\partial x^2} f(x, y) + \frac{\partial^2}{\partial y^2} f(x, y) \quad (30)$$

Este operador deve ser aplicado na função gaussiana bidimensional que pela omissão do fator de escala pode ser representada pela Equação (31).

$$g(x, y) = \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (31)$$

Daí resulta na expressão dada pela Equação (32).

$$\nabla^2 g(x, y) = \frac{x^2+y^2-2\sigma^2}{\sigma^4} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (32)$$

Pela substituição das coordenadas correspondentes na Equação (32), é possível obter a máscara do laplaciano da gaussiana, e pela convolução desta com a imagem obter a imagem filtrada (MLSNA e RODRIGUEZ, 2009). A principal diferença deste filtro para os demais filtros de aguçamento é que o mesmo apresenta uma etapa de suavização anterior ao realce da alta frequência, evitando uma intensificação de ruídos que possam estar presentes na imagem original. Para ilustrar o efeito da aplicação deste filtro como um método de detecção de bordas, a Figura 16 mostra uma imagem utilizada neste trabalho processada com o filtro do laplaciano da gaussiana utilizando um *kernel* quadrado com dimensões 3, 5, 7, 9 e 11.

Figura 16 - (a) Imagem original, (b) Imagem filtrada com LoG de tamanho 3, (c) Imagem filtrada com LoG de tamanho 5, (d) Imagem filtrada com LoG de tamanho 7, (e) Imagem filtrada com LoG de tamanho 9, (f) Imagem filtrada com LoG de tamanho 11



Fonte: Autor

Neste caso, percebe-se que não há grandes alterações nas imagens filtradas quando se altera a dimensão do kernel.

4.3.4 Descritores estatísticos

Os descritores estatísticos utilizados foram a média aritmética, desvio padrão e entropia de Shannon, que foram extraídos das imagens coloridas para os canais RGB (vermelho, verde e azul). Estes são definidos pelas Equações (33), (34) e (35), respectivamente.

$$\mu = \frac{\sum_{i=1}^N v_i}{N} \quad (33)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (v_i - \mu)^2}{N}} \quad (34)$$

$$e = - \sum_{\substack{i=1 \\ p_i \neq 0}}^{N_B} p_i \log_2 p_i \quad (35)$$

Onde v_i representa as intensidades de cinza dos pixels em um dado canal, N representa o número total de pixels em um dado canal, p_i é a frequência relativa de um i -ésimo nível de cinza de um dado canal e $N_B = 256$, uma vez que as imagens utilizadas neste trabalho são de 8-bits.

4.4 Métricas de erro

As métricas de erro utilizadas para realizar a avaliação do desempenho dos modelos de aprendizagem de máquina aplicados são apresentadas nesta seção.

4.4.1 Erro Médio Absoluto (MAE)

O MAE calcula a média das diferenças absolutas entre o valor previsto, \hat{y}_i , e o valor real, y_i , isso é, não leva em consideração se o erro é para mais ou para menos e para as diferenças absolutas não são atribuídos pesos, como indica a Equação (36).

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (36)$$

4.4.2 Erro Médio por Viés (MBE)

O MBE se assemelha ao cálculo do MAE, mas se diferencia deste por considerar o sinal do erro, isto é, não calcula o valor absoluto das diferenças. Desse modo, requer prudência em sua análise, uma vez que permite a compensação de erros (erros com sinais distintos). A mesma é dada pela Equação (37).

$$MBE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (37)$$

4.4.3 Raiz do Erro Quadrático Médio (RMSE)

O RMSE calcula a magnitude média do erro pela raiz quadrada da média dos quadrados dos erros. Desse modo, atribui um peso maior aos erros de maior magnitude, e peso menor aos erros de menor magnitude. É obtido na mesma unidade da variável em análise, sendo definida pela Equação (38).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (38)$$

4.4.4 Raiz do Erro Quadrático Médio Relativo (rRMSE)

O rRMSE é a razão entre o RMSE e a média dos valores da variável em análise, como dado pela Equação (39).

$$rRMSE = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}}{\frac{\sum_{i=1}^n y_i}{n}} \quad (39)$$

Tal métrica foi utilizada porque fornece faixas de classificação do desempenho das previsões: $rRMSE < 10\%$ excelente, $10\% < rRMSE < 20\%$ bom, $20\% < rRMSE < 30\%$ razoável, $rRMSE > 30\%$ ruim (LI *et al.*, 2013).

4.4.5 Habilidade de Previsão (s)

Esta métrica avalia o desempenho do modelo aplicado quando comparado com a acurácia do modelo de baseline, sendo definido pela Equação (40).

$$s = 1 - \frac{RMSE}{RMSE_B} \quad (40)$$

Onde, o $RMSE$ representa a métrica de erro para o modelo considerado e o $RMSE_B$ representa a métrica de erro para o modelo de baseline.

5 RESULTADOS E DISCUSSÃO

Nesta seção, são apresentados inicialmente os resultados relacionados à etapa de treinamento dos modelos. Vale ressaltar que para cada conjunto de dados e cada horizonte de previsão foram obtidos ajustes ótimos para cada modelo por meio da etapa de treinamento, ou seja, houve a obtenção de um grande volume de resultados relacionados à determinação de hiperparâmetros ótimos para os métodos. Assim, nesta seção está detalhada a busca pelos parâmetros livres de cada modelo apenas para as previsões de 30 minutos a posteriori para o conjunto de dados formado pelos sinais dos LDR's, em conjunto com os descritores estatísticos extraídos das imagens do céu sem a realização de filtragem. Os demais resultados relacionados à etapa de treinamento para os demais conjuntos de dados e horizontes de previsão são obtidos de forma análoga. Posteriormente, apresentam-se os valores das métricas de erro obtidas pela aplicação dos algoritmos ajustados no conjunto de validação.

5.1 Busca por hiperparâmetros

5.1.1 SVR

Como já havia sido discutido anteriormente, tomou-se $\varepsilon = 0,1$ e o parâmetro σ foi determinado inicialmente resultando em $\sigma = 0,1573075$. Em seguida, o parâmetro C foi determinado pela aplicação de validação cruzada 10-Fold no conjunto de treinamento. Os resultados são reportados na Tabela 3.

Tabela 3 – Busca por hiperparâmetro SVR

| C | $RMSE$ (W/m ²) | MAE (W/m ²) |
|----------|----------------------------|---------------------------|
| 0,25 | 163,72 | 110,19 |
| 0,50 | 159,53 | 105,35 |
| 1 | 158,05 | 102,62 |
| 2 | 158,31 | 102,25 |
| 4 | 159,42 | 102,68 |
| 8 | 161,22 | 103,45 |
| 16 | 162,75 | 104,07 |
| 32 | 167,61 | 107,16 |
| 64 | 172,42 | 110,24 |
| 128 | 180,58 | 115,77 |

Fonte: Autor

Desta forma, percebe-se que o parâmetro C que forneceu o menor $RMSE$ na etapa de treinamento foi $C_{ótimo} = 1$ sendo este o utilizado quando o modelo foi aplicado no conjunto de validação. Nota-se também que dentro da faixa de valores de C utilizados, o valor de 1 é um dos menores, além disso, a partir deste valor o aumento em C conduz a um aumento no $RMSE$, ou seja, um aumento no nível de flexibilidade do modelo ocasiona uma piora no desempenho de previsão. Este fato pode estar relacionado a um maior risco de *overfitting*, quando se aumenta a flexibilidade do modelo para o conjunto de dados considerado. Esta elevação proporciona menores bias, porém aumenta a variância do método.

5.1.2 Florestas Aleatórias

Neste caso, o hiperparâmetro a ser considerado corresponde à dimensão do espaço de preditores que será segmentado. Novamente, a métrica escolhida para a seleção do parâmetro livre ótimo foi o $RMSE$, e os resultados são reportados na Tabela 4. Assim, pelo treinamento obteve-se $m_{ótimo} = 6$.

Tabela 4 – Busca por hiperparâmetros no algoritmo RF

| m | $RMSE$ (W/m ²) | MAE (W/m ²) |
|-----|----------------------------|---------------------------|
| 2 | 149,27 | 101,67 |
| 3 | 148,95 | 100,74 |
| 4 | 148,72 | 100,12 |
| 6 | 148,46 | 99,32 |
| 7 | 148,71 | 99,18 |
| 8 | 148,95 | 99,19 |
| 10 | 149,25 | 99,07 |
| 11 | 149,53 | 99,07 |
| 12 | 149,55 | 99,06 |
| 14 | 150,26 | 99,13 |

Fonte: Autor

Neste caso, nota-se uma menor sensibilidade do modelo à variação da dimensão do espaço de preditores, como se percebe pela pequena variação dos valores das métricas de erro com a alteração do parâmetro m .

5.1.3 Árvore de regressão com a aplicação de um procedimento de Bagging

Como já havia sido abordado, o método que utiliza *Bagging* apresenta como único hiperparâmetro o número de árvores a ser considerado, tal parâmetro não é crítico uma vez que um valor elevado de árvores não irá conduzir a problemas relacionados à *overfitting* (JAMES *et al.*, 2013). Entretanto, um número exagerado pode proporcionar um alto tempo de processamento. Neste trabalho, para todos os conjuntos de dados e horizontes de previsão foi utilizadas 500 árvores, que foi o valor com o melhor balanceamento entre desempenho e tempo computacional.

5.1.4 KNN

Para este método, o único hiperparâmetro a ser considerado é o número de vizinhos K , os resultados para a busca do melhor ajuste são ilustrados na Tabela 5.

Tabela 5 – Busca por hiperparâmetros no KNN

| K | $RMSE$ | MAE |
|-----|--------|--------|
| 5 | 162,12 | 106,75 |
| 7 | 161,40 | 108,04 |
| 9 | 160,87 | 108,45 |
| 11 | 160,43 | 109,48 |
| 13 | 160,52 | 110,65 |
| 15 | 161,11 | 112,08 |
| 17 | 161,44 | 113,36 |
| 19 | 161,71 | 114,30 |
| 21 | 161,99 | 115,17 |
| 23 | 162,48 | 115,97 |

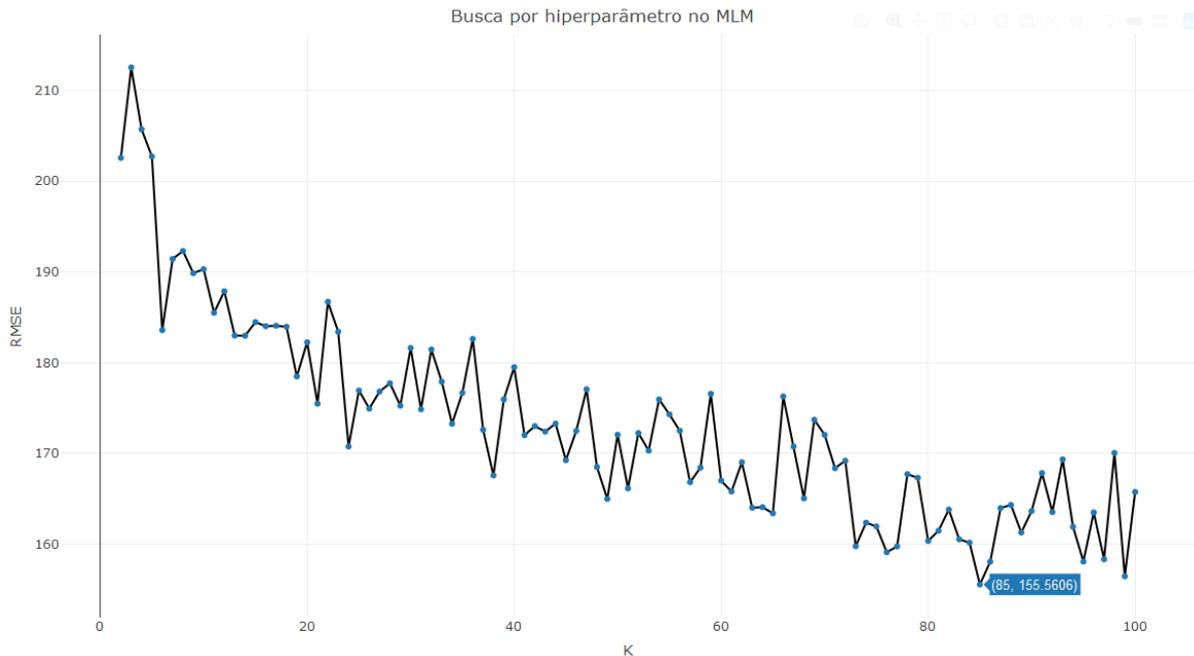
Fonte: Autor

Percebe-se que até o valor de $K_{ótimo} = 11$ há uma queda no $RMSE$. Em seguida, após este valor ocorre um aumento gradual na métrica de erro, indicando que uma redução na flexibilidade do modelo, ou seja, uma redução na variância com uma consequente ampliação do bias proporciona uma piora no desempenho de previsão deste método para o conjunto de dados considerado.

5.1.5 MLM

Similar ao método KNN, o único hiperparâmetro para este modelo é o número de pontos de referência K , o qual obteve-se o valor de $K_{\text{ótimo}} = 85$. Os resultados obtidos por validação cruzada 10-Fold no conjunto de treinamento são observados na Figura 17.

Figura 17 – Busca por hiperparâmetro no MLM



Fonte: Autor

Desta forma, percebe-se que no geral o aumento no valor de K proporcionou uma redução no valor do $RMSE$. Este comportamento não é geral e dependendo do conjunto de dados em que o MLM é aplicado as tendências podem ser alteradas.

5.2 Resultados obtidos pela aplicação dos modelos ajustados no conjunto de validação

Nesta seção, são apresentados os valores das métricas de erro obtidos pela aplicação dos algoritmos treinados nos conjuntos de validação. Primeiramente, são apresentados os resultados relacionados ao uso dos algoritmos no conjunto com atributos formados apenas por sinais dos LDR's. Os mesmos são reportados na Tabela 6 para um horizonte de previsão de 10 minutos a posteriori, na Tabela 7 para 20 minutos a posteriori e na Tabela 8 para 30 minutos a posteriori.

Tabela 6 – Resultados para o conjunto I

| Conjunto I para 10 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 4,97 | 221,57 | 261,57 | 42,82 | 0 |
| SVR | -17,52 | 89,11 | 149,13 | 24,60 | 43,01 | -22,82 | 104,63 | 167,23 | 27,37 | 36,09 |
| RF | 1,11 | 46,87 | 72,22 | 11,91 | 72,40 | 3,55 | 103 | 159,43 | 26,09 | 39,08 |
| Bagging | 0,19 | 107,03 | 151,58 | 25,01 | 42,08 | 1,36 | 116,55 | 165,14 | 27,02 | 36,89 |
| KNN | 5,09 | 98,36 | 149,34 | 24,64 | 42,93 | 3,74 | 102,41 | 158,05 | 25,86 | 39,60 |
| MLM | -0,78 | 103,94 | 151,28 | 24,96 | 42,19 | -1,57 | 106,26 | 163,03 | 26,68 | 37,70 |

Fonte: Autor

Tabela 7 – Resultados para o conjunto I

| Conjunto I para 20 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 16,47 | 222,28 | 262,11 | 42,84 | 0 |
| SVR | -6,31 | 99,59 | 156,69 | 26,32 | 40,22 | -2,44 | 100,94 | 157,79 | 25,79 | 39,80 |
| RF | 0,71 | 48,09 | 72,05 | 12,10 | 72,51 | 13,27 | 97,46 | 150,18 | 24,55 | 42,71 |
| Bagging | -0,21 | 115,07 | 158,92 | 26,70 | 39,37 | 12,10 | 114,67 | 161,02 | 26,32 | 38,57 |
| KNN | 6,90 | 96,54 | 146,33 | 24,58 | 44,17 | 13,87 | 102,59 | 157,59 | 25,76 | 39,88 |
| MLM | 8,68 | 111,65 | 157,89 | 25,81 | 39,76 | 1,42 | 112,13 | 161,36 | 27,10 | 38,44 |

Fonte: Autor

Tabela 8 – Resultados para o conjunto I

| Conjunto I para 30 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 20,44 | 233,49 | 274,15 | 45,33 | 0 |
| SVR | -8,59 | 91,61 | 148,19 | 25,36 | 45,94 | 1,18 | 108,18 | 165,72 | 27,40 | 39,55 |
| RF | 1,20 | 46,33 | 70,37 | 12,04 | 74,33 | 19,18 | 103,72 | 156,35 | 25,85 | 42,97 |
| Bagging | -0,21 | 115,07 | 158,92 | 26,70 | 39,37 | 14,15 | 121,57 | 170,75 | 28,23 | 37,72 |
| KNN | 5,97 | 93,22 | 142,34 | 24,36 | 48,08 | 19,51 | 111,71 | 169,36 | 28,01 | 38,22 |
| MLM | 1,02 | 108,03 | 152,71 | 26,13 | 44,30 | 13,99 | 120,01 | 166,96 | 27,61 | 39,10 |

Fonte: Autor

Pela avaliação dos resultados e comparando com aqueles encontrados nos trabalhos de Pedro e Coimbra (2015) e Pedro *et al.* (2018), percebe-se que atributos relacionados a sinais de LDR's não conseguem capturar de forma adequada o comportamento da irradiância solar. Isto se deve a uma falta de sensibilidade dos sensores LDR's. Em casos de início e fim de dia os valores de irradiância solar caem significativamente, enquanto os valores reportados pelos sensores de luminosidade ainda estão elevados. Possíveis soluções para este problema seriam realizar alterações no circuito elétrico, como a adição de potenciômetros e/ou adicionar filtros ópticos na tentativa de amenizar o nível de luminosidade e proporcionar uma melhora na sensibilidade dos sensores.

Outro fator relevante que teve forte influência para a obtenção destes e dos demais resultados foi que considerou-se apenas o valor da irradiância solar do instante atual como preditor e, como já foi discutido anteriormente, para que se tenha um bom desempenho na previsão de irradiância solar é necessário que se adicione a informação sobre a série histórica de irradiância solar, uma vez que se sabe que as condições climáticas locais favorecem o desenvolvimento de séries de irradiância que são fortemente autocorrelacionadas. Assim, uma boa estratégia para incrementar a acurácia das previsões é a adição de preditores relacionados a valores de irradiância passada, como considerar os valores de irradiância dos 30 minutos anteriores.

Os resultados relacionados ao uso dos algoritmos no conjunto com atributos formados apenas por sinais dos LDR's e os descritores estatísticos das imagens sem a realização da filtragem, conjunto II, são reportados na Tabela 9 para um horizonte de previsão de 10 minutos

a posteriori, na Tabela 10 para 20 minutos a posteriori e na Tabela 11 para 30 minutos a posteriori.

Tabela 9 – Resultados para o conjunto II

| Conjunto II para 10 minutos a posteriori | | | | | | | | | | |
|--|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 4,97 | 221,57 | 261,68 | 42,82 | 0 |
| SVR | -22,19 | 84,08 | 140,58 | 23,19 | 46,28 | -26,78 | 103,50 | 165,23 | 27,04 | 36,86 |
| RF | 0,73 | 41,43 | 64,24 | 10,60 | 75,45 | -2,37 | 97,65 | 154,70 | 25,32 | 40,88 |
| Bagging | 0,86 | 100,75 | 144,22 | 23,79 | 44,89 | -2,17 | 106,51 | 156,38 | 25,59 | 40,24 |
| KNN | -4,03 | 99,47 | 148,08 | 24,43 | 43,41 | -6,20 | 105,80 | 161,90 | 26,49 | 38,13 |
| MLM | -0,42 | 96,33 | 142,38 | 23,49 | 45,59 | -4,31 | 108,08 | 162,61 | 26,61 | 37,86 |

Fonte: Autor

Tabela 10 – Resultados para o conjunto II

| Conjunto II para 20 minutos a posteriori | | | | | | | | | | |
|--|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 16,47 | 222,28 | 262,11 | 42,84 | 0 |
| SVR | -18,53 | 87,85 | 128,74 | 23,98 | 45,54 | -11,63 | 95,72 | 149,35 | 24,41 | 43,02 |
| RF | 0,82 | 44,76 | 68,04 | 11,43 | 74,04 | 8,67 | 97,83 | 144,23 | 23,58 | 44,97 |
| Bagging | -1,45 | 110,18 | 150,35 | 25,26 | 42,64 | 5,47 | 108,82 | 152,37 | 24,91 | 41,87 |
| KNN | -3,55 | 100,34 | 147,72 | 24,81 | 43,64 | 1,56 | 102,85 | 151,77 | 24,81 | 42,10 |
| MLM | -0,61 | 97,14 | 144,31 | 24,24 | 44,94 | 5,47 | 104,98 | 152,23 | 24,88 | 41,92 |

Fonte: Autor

Tabela 11 – Resultados para o conjunto II

| Conjunto II para 30 minutos a posteriori | | | | | | | | | | |
|--|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 20,44 | 233,49 | 274,15 | 45,33 | 0 |
| SVR | -16,23 | 87,37 | 141,20 | 24,17 | 48,49 | -3,93 | 102,19 | 152,28 | 25,18 | 44,45 |
| RF | 1,27 | 41,07 | 62,30 | 10,66 | 77,27 | 15,15 | 98,26 | 147,82 | 24,44 | 46,08 |
| Bagging | -0,55 | 109,78 | 150,69 | 25,79 | 45,04 | 8,47 | 115,39 | 159,99 | 26,46 | 41,64 |
| KNN | -2,10 | 98,41 | 146,19 | 25,02 | 46,68 | 8,70 | 104,93 | 154,23 | 25,50 | 43,74 |
| MLM | -0,38 | 101,27 | 145,35 | 24,87 | 45,98 | 14,69 | 110,38 | 159,49 | 26,37 | 41,82 |

Fonte: Autor

Observando os resultados percebe-se que a introdução dos descritores estatísticos extraídos das imagens como atributos nos algoritmos, proporcionou melhorias nos resultados para todos os métodos e horizontes de previsão considerados. Estes resultados concordam com aqueles encontrados em trabalhos recentes (PEDRO e COIMBRA, 2015) e (PEDRO *et al.*, 2018). De fato, a introdução de informações relacionadas à imagens do céu contribui para uma melhor generalização dos algoritmos e acredita-se que a introdução de novos atributos extraídos destas imagens irá contribuir para uma melhora ainda maior no desempenho dos modelos.

Os resultados dos trabalhos de Alonso-Montesinos *et al.* (2015) foram um pouco piores do que os encontrados em Pedro e Coimbra (2015) e Pedro *et al.* (2018) de fato a previsão fornecida por Alonso-Montesinos *et al.* (2015) se baseia na detecção de vetores de movimento de nuvens (CMV), que em alguns casos utilizam o método da correlação cruzada máxima para realizar essa determinação. Não há a utilização de informações relacionadas aos valores passados de irradiância solar, mostrando novamente a importância de se adicionar tal conhecimento para os modelos por meios da consideração de novos preditores.

Assim acredita-se que melhorias significativas nas acurácias de previsão podem ocorrer desde que se considerem os valores de irradiância solar de instantes passados como atributos. Os resultados relacionados ao uso dos algoritmos no conjunto III são reportados na Tabela 12 para um horizonte de previsão de 10 minutos a posteriori, na Tabela 13 para 20 minutos a posteriori e na Tabela 14 para 30 minutos a posteriori e os resultados para aplicação dos métodos no conjunto IV são ilustrados na Tabela 15 para um horizonte de previsão de 10 minutos a posteriori, na Tabela 16 para 20 minutos a posteriori e na Tabela 17 para 30 minutos a posteriori.

Tabela 12 – Resultados para o conjunto III

| Conjunto III para 10 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 4,97 | 221,57 | 261,68 | 42,82 | 0 |
| SVR | -21,52 | 86,40 | 143,89 | 23,74 | 45,01 | -24,45 | 104,54 | 166,86 | 27,31 | 36,23 |
| RF | 0,85 | 43,32 | 67,68 | 11,17 | 74,14 | -1,47 | 96,75 | 153,40 | 25,10 | 41,38 |
| Bagging | 1,51 | 103,26 | 146,81 | 24,22 | 43,90 | -1,15 | 108,81 | 158,63 | 25,96 | 39,38 |
| KNN | -3,48 | 103,68 | 152,08 | 25,09 | 41,89 | -5,47 | 107,33 | 161,91 | 26,49 | 38,13 |
| MLM | -1,59 | 97,57 | 144,78 | 23,89 | 44,67 | 1,72 | 107,19 | 162,75 | 26,63 | 37,81 |

Fonte: Autor

Tabela 13 – Resultados para o conjunto III

| Conjunto III para 20 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 16,47 | 222,28 | 262,11 | 42,84 | 0 |
| SVR | -18,27 | 88,53 | 143,43 | 24,09 | 45,28 | -11,86 | 96,07 | 151,62 | 24,78 | 42,15 |
| RF | 1,09 | 44,63 | 68,15 | 11,45 | 74,00 | 10,75 | 93,66 | 141,10 | 23,06 | 46,17 |
| Bagging | -0,69 | 113,26 | 151,79 | 25,50 | 42,09 | 5,37 | 111,33 | 152,26 | 24,89 | 41,91 |
| KNN | -0,11 | 98,53 | 146,57 | 24,62 | 44,08 | 6,38 | 101,71 | 151,51 | 24,77 | 42,20 |
| MLM | -0,72 | 98,20 | 144,44 | 24,26 | 44,89 | 6,30 | 104,89 | 150,61 | 24,62 | 42,54 |

Fonte: Autor

Tabela 14 – Resultados para o conjunto III

| Conjunto III para 30 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 20,44 | 233,49 | 274,15 | 45,33 | 0 |
| SVR | -15,58 | 87,76 | 140,99 | 24,13 | 48,57 | -2,13 | 103,77 | 155,04 | 25,64 | 43,45 |
| RF | 1,00 | 43,0 | 65,14 | 11,15 | 76,23 | 15,23 | 97,24 | 145,17 | 24,00 | 47,05 |
| Bagging | -0,54 | 109,27 | 150,79 | 25,81 | 45,00 | 9,34 | 114,24 | 158,55 | 26,22 | 42,17 |
| KNN | 2,94 | 89,96 | 136,60 | 23,38 | 50,17 | 12,71 | 103,95 | 157,36 | 26,02 | 42,60 |
| MLM | -0,14 | 100,23 | 146,02 | 25,00 | 46,74 | 13,14 | 105,90 | 154,26 | 25,51 | 43,73 |

Fonte: Autor

Tabela 15 – Resultados para o conjunto IV

| Conjunto IV para 10 minutos a posteriori | | | | | | | | | | |
|--|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 4,97 | 221,57 | 261,68 | 42,82 | 0 |
| SVR | -17,07 | 84,30 | 142,37 | 23,49 | 45,59 | -18,77 | 103,67 | 164,79 | 26,97 | 37,03 |
| RF | 1,57 | 49,82 | 74,57 | 12,30 | 71,51 | 2,52 | 104,27 | 157,68 | 25,80 | 39,74 |
| Bagging | -1,15 | 106,08 | 151,83 | 25,05 | 41,98 | -0,36 | 114,86 | 164,19 | 26,87 | 37,25 |
| KNN | 5,68 | 98,47 | 150,31 | 24,80 | 42,56 | 4,70 | 106,69 | 165,53 | 27,09 | 36,74 |
| MLM | -0,90 | 102,23 | 149,50 | 24,67 | 42,87 | 3,32 | 111,32 | 166,64 | 27,27 | 36,32 |

Fonte: Autor

Tabela 16 – Resultados para o conjunto IV

| Conjunto IV para 20 minutos a posteriori | | | | | | | | | | |
|--|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 16,47 | 222,28 | 262,11 | 42,84 | 0 |
| SVR | -7,16 | 99,36 | 157,84 | 26,51 | 39,78 | 0,73 | 103,96 | 160,77 | 26,28 | 38,67 |
| RF | 1,73 | 50,61 | 75,92 | 12,75 | 71,04 | 14,92 | 98,57 | 148,62 | 24,29 | 43,30 |
| Bagging | -0,11 | 111,33 | 156,72 | 26,33 | 40,21 | 11,93 | 115,14 | 162,79 | 26,61 | 37,89 |
| KNN | 4,79 | 107,88 | 161,03 | 27,05 | 38,56 | 15,30 | 109,81 | 164,05 | 26,82 | 37,41 |
| MLM | 4,01 | 118,44 | 166,48 | 27,97 | 36,49 | 18,15 | 120,98 | 168,77 | 27,59 | 35,61 |

Fonte: Autor

Tabela 17 – Resultados para o conjunto IV

| Conjunto IV para 30 minutos a posteriori | | | | | | | | | | |
|--|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 20,44 | 233,49 | 274,15 | 45,33 | 0 |
| SVR | -10,58 | 91,89 | 145,36 | 24,88 | 46,98 | 8,64 | 110,77 | 164,87 | 27,26 | 39,86 |
| RF | 0,73 | 49,65 | 72,33 | 12,38 | 73,62 | 22,85 | 105,00 | 155,44 | 25,70 | 43,30 |
| Bagging | -0,53 | 111,80 | 154,10 | 26,37 | 43,79 | 13,25 | 118,38 | 168,16 | 27,80 | 38,66 |
| KNN | 4,23 | 94,64 | 143,49 | 24,56 | 47,66 | 21,17 | 108,28 | 163,11 | 26,97 | 40,50 |
| MLM | 0,63 | 109,13 | 153,00 | 26,18 | 44,19 | 18,35 | 118,70 | 164,70 | 27,23 | 39,92 |

Fonte: Autor

Os resultados para a aplicação dos filtros indicam que a filtragem pelo *kernel* da mediana não altera significativamente a acurácia dos métodos quando se compara com o caso em que não há filtragem. Esta mudança pode ser positiva ou negativa, mas no geral não é abrupta. Um comportamento geral percebido foi a melhora no desempenho do algoritmo Florestas Aleatórias para todos os horizontes de previsão relativos aos resultados sem aplicação de filtro.

Desta forma, percebe-se que tal modelo é mais sensível aos atributos obtidos das imagens e que a atenuação de ruídos nas imagens, que representam pontos ou regiões de alta frequência na mesma, contribui positivamente para a acurácia deste método. O intuito de se utilizar um filtro de suavização se deve ao fato de que as imagens foram capturadas em campo aberto, o que aumenta as chances de se ter ruídos nas mesmas.

Os resultados relacionados ao conjunto IV, no geral, ilustram uma piora no desempenho dos modelos quando comparados com aqueles encontrados para as imagens não filtradas. A razão para o uso do filtro LoG foi a de que o mesmo funciona como detector de bordas, assim haveria a possibilidade de identificar a ocorrência de nuvens, que é uma das principais razões para o caráter intermitente da irradiância solar. Acreditava-se que esta informação seria útil para a aprendizagem dos modelos, o que ainda não se provou verdade.

Os resultados obtidos pela aplicação dos algoritmos de aprendizagem nos conjuntos A, B, C e D são reportados nas Tabelas 18, 19 e 20 para o conjunto A. Nas Tabelas 21, 22 e 23 para o conjunto B. Nas Tabelas 24, 25 e 26 para o conjunto C e, por fim, nas Tabelas 27, 28 e 29 para o conjunto D.

Tabela 18 – Resultados para o conjunto A

| Conjunto A para 10 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | -6,48 | 219,83 | 258,50 | 42,68 | 0 |
| SVR | -13,32 | 54,28 | 101,94 | 16,65 | 60,57 | -6,66 | 82,95 | 139,95 | 23,10 | 45,86 |
| RF | 0,33 | 44,50 | 72,64 | 11,87 | 71,90 | 10,31 | 79,61 | 123,84 | 20,44 | 52,09 |
| Bagging | -1,07 | 84,91 | 126,72 | 20,70 | 50,98 | 6,03 | 96,12 | 141,09 | 23,29 | 45,42 |
| KNN | -1,43 | 68,12 | 110,30 | 18,02 | 57,33 | 7,35 | 81,02 | 129,57 | 21,39 | 49,88 |
| MLM | -7,75 | 92,36 | 141,64 | 23,14 | 45,21 | 1,41 | 94,95 | 147,92 | 24,42 | 42,78 |

Fonte: Autor

Tabela 19 – Resultados para o conjunto A

| Conjunto A para 20 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | -0,53 | 226,98 | 267,59 | 44,47 | 0 |
| SVR | -13,53 | 61,66 | 108,34 | 17,99 | 59,52 | -10,38 | 83,61 | 137,74 | 22,89 | 48,53 |
| RF | -0,01 | 43,34 | 69,41 | 11,52 | 74,06 | 8,05 | 81,42 | 124,64 | 20,71 | 53,42 |
| Bagging | -0,32 | 89,62 | 134,04 | 22,26 | 49,91 | 8,46 | 91,17 | 137,75 | 22,89 | 48,52 |
| KNN | -0,83 | 77,69 | 121,40 | 20,16 | 54,63 | 8,42 | 88,72 | 139,66 | 23,21 | 47,81 |
| MLM | -2,39 | 85,67 | 133,08 | 22,09 | 50,27 | 8,82 | 93,39 | 143,24 | 23,80 | 46,47 |

Fonte: Autor

Tabela 20 – Resultados para o conjunto A

| Conjunto A para 30 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|--------------|----------------------------|----------------------------|-----------------------------|--------------|--------------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | -2,32 | 238,12 | 278,63 | 46,92 | 0 |
| SVR | -10,51 | 72,95 | 121,26 | 20,50 | 56,48 | -4,08 | 85,22 | 135,99 | 22,90 | 51,19 |
| RF | -0,40 | 37,65 | 63,25 | 10,69 | 77,30 | 8,27 | 73,91 | 120,70 | 20,32 | 56,68 |
| Bagging | -0,14 | 93,08 | 134,30 | 22,70 | 51,80 | 10,64 | 97,29 | 143,93 | 24,24 | 48,34 |
| KNN | -3,05 | 67,69 | 110,48 | 18,68 | 60,35 | 5,44 | 86,79 | 136,01 | 22,90 | 51,18 |
| MLM | 0,05 | 85,33 | 130,43 | 22,09 | 53,19 | 7,61 | 98,31 | 145,53 | 24,50 | 47,77 |

Fonte: Autor

Tabela 21 – Resultados para o conjunto B

| Conjunto B para 10 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|--------------|----------------------------|----------------------------|-----------------------------|--------------|--------------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | -6,48 | 219,83 | 258,50 | 42,68 | 0 |
| SVR | -13,76 | 58,36 | 106,92 | 17,47 | 58,64 | -5,53 | 87,70 | 144,86 | 23,92 | 43,96 |
| RF | 0,41 | 36,00 | 57,87 | 9,45 | 77,61 | 9,70 | 82,70 | 128,57 | 21,23 | 50,26 |
| Bagging | -1,51 | 83,55 | 125,89 | 20,56 | 51,30 | 6,18 | 93,87 | 138,93 | 22,94 | 46,26 |
| KNN | -0,19 | 68,99 | 113,91 | 18,61 | 55,93 | 1,69 | 84,29 | 135,80 | 22,42 | 47,47 |
| MLM | -2,22 | 87,51 | 136,62 | 22,32 | 47,15 | 3,55 | 98,20 | 147,48 | 24,35 | 42,95 |

Fonte: Autor

Tabela 22 – Resultados para o conjunto B

| Conjunto B para 20 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|--------------|----------------------------|----------------------------|-----------------------------|--------------|--------------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | -0,53 | 226,98 | 267,59 | 44,47 | 0 |
| SVR | -11,94 | 55,45 | 99,54 | 16,53 | 62,80 | -3,99 | 85,27 | 135,47 | 22,51 | 49,38 |
| RF | -0,33 | 39,47 | 61,47 | 10,21 | 77,03 | 8,89 | 82,64 | 125,51 | 20,86 | 53,10 |
| Bagging | -2,82 | 85,31 | 126,66 | 21,03 | 52,67 | 5,36 | 90,33 | 135,02 | 22,44 | 49,54 |
| KNN | -1,07 | 79,39 | 123,89 | 20,57 | 53,70 | 3,33 | 85,25 | 133,12 | 22,12 | 50,25 |
| MLM | -5,47 | 87,46 | 132,49 | 22,00 | 50,49 | 6,10 | 94,00 | 143,38 | 23,83 | 46,42 |

Fonte: Autor

Tabela 23 – Resultados para o conjunto B

| Conjunto B para 30 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|--------------|----------------------------|----------------------------|-----------------------------|--------------|--------------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 2,32 | 238,12 | 278,63 | 46,92 | 0 |
| SVR | -16,19 | 78,57 | 130,37 | 22,04 | 53,21 | -4,32 | 87,31 | 138,31 | 23,29 | 50,36 |
| RF | 0,18 | 36,20 | 57,69 | 9,75 | 79,29 | 9,96 | 79,84 | 127,73 | 21,51 | 54,16 |
| Bagging | -2,03 | 89,28 | 131,14 | 22,17 | 52,93 | 8,87 | 94,67 | 141,72 | 23,86 | 49,14 |
| KNN | -5,43 | 74,88 | 121,19 | 20,49 | 56,50 | 3,17 | 88,44 | 138,33 | 23,29 | 50,35 |
| MLM | -9,40 | 92,54 | 139,91 | 23,65 | 49,79 | 9,05 | 100,28 | 148,02 | 24,92 | 46,88 |

Fonte: Autor

Tabela 24 – Resultados para o conjunto C

| Conjunto C para 10 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|--------------|----------------------------|----------------------------|-----------------------------|--------------|--------------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | -6,48 | 219,83 | 258,50 | 42,68 | 0 |
| SVR | -13,31 | 52,54 | 98,04 | 16,01 | 62,07 | -10,28 | 86,59 | 144,97 | 23,93 | 43,92 |
| RF | 0,33 | 40,53 | 64,04 | 10,46 | 75,23 | 9,75 | 81,98 | 125,71 | 20,75 | 51,37 |
| Bagging | -0,96 | 83,71 | 125,91 | 20,57 | 51,29 | 4,87 | 94,78 | 140,35 | 23,17 | 45,70 |
| KNN | -0,33 | 68,80 | 112,99 | 18,46 | 56,29 | -1,50 | 82,99 | 130,60 | 21,56 | 49,48 |
| MLM | -6,55 | 92,76 | 143,28 | 23,40 | 44,57 | 5,74 | 95,56 | 144,83 | 23,91 | 43,97 |

Fonte: Autor

Tabela 25 – Resultados para o conjunto C

| Conjunto C para 20 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|--------------|----------------------------|----------------------------|-----------------------------|--------------|--------------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | -0,53 | 226,98 | 267,59 | 44,47 | 0 |
| SVR | -13,07 | 55,43 | 100,34 | 16,66 | 62,50 | -5,50 | 85,37 | 138,00 | 22,93 | 48,43 |
| RF | -0,86 | 39,59 | 62,07 | 10,31 | 76,81 | 8,45 | 80,71 | 122,90 | 20,42 | 54,07 |
| Bagging | -2,82 | 85,31 | 126,66 | 21,03 | 52,67 | 7,61 | 91,37 | 137,00 | 22,77 | 48,80 |
| KNN | -0,79 | 75,85 | 120,02 | 19,93 | 55,15 | 5,29 | 84,94 | 133,35 | 22,16 | 50,17 |
| MLM | -2,52 | 88,92 | 136,60 | 22,68 | 48,95 | 2,31 | 92,29 | 141,80 | 23,56 | 47,01 |

Fonte: Autor

Tabela 26 – Resultados para o conjunto C

| Conjunto C para 30 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|--------------|----------------------------|----------------------------|-----------------------------|--------------|--------------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 2,32 | 238,12 | 278,63 | 46,92 | 0 |
| SVR | -12,64 | 65,02 | 114,64 | 19,38 | 58,86 | -6,19 | 84,15 | 137,13 | 23,09 | 50,78 |
| RF | -0,64 | 39,56 | 63,16 | 10,68 | 77,33 | 8,25 | 77,53 | 123,19 | 20,74 | 55,79 |
| Bagging | -1,65 | 89,66 | 133,22 | 22,52 | 52,19 | 9,30 | 92,53 | 139,60 | 23,51 | 49,90 |
| KNN | -1,61 | 67,24 | 109,68 | 18,54 | 60,34 | 8,18 | 85,41 | 134,34 | 22,62 | 51,79 |
| MLM | -2,50 | 88,26 | 134,25 | 22,69 | 51,82 | 11,18 | 97,56 | 145,56 | 24,51 | 47,76 |

Fonte: Autor

Tabela 27 – Resultados para o conjunto D

| Conjunto D para 10 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | -6,48 | 219,83 | 258,50 | 42,68 | 0 |
| SVR | -19,17 | 69,34 | 123,24 | 20,13 | 52,34 | -14,65 | 86,90 | 143,06 | 23,62 | 44,66 |
| RF | 0,67 | 43,04 | 68,04 | 11,11 | 73,68 | 9,79 | 80,83 | 125,24 | 20,68 | 51,55 |
| Bagging | -0,16 | 85,47 | 128,37 | 20,97 | 50,34 | 6,91 | 96,03 | 140,74 | 23,23 | 45,56 |
| KNN | -0,15 | 77,20 | 123,02 | 20,01 | 52,41 | 4,33 | 86,14 | 135,75 | 22,41 | 47,48 |
| MLM | -7,13 | 89,95 | 139,91 | 22,85 | 45,88 | 4,65 | 96,31 | 145,44 | 24,01 | 43,74 |

Fonte: Autor

Tabela 28 – Resultados para o conjunto D

| Conjunto D para 20 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | -0,53 | 226,98 | 267,59 | 44,47 | 0 |
| SVR | -20,62 | 73,30 | 123,75 | 20,55 | 53,75 | -13,18 | 84,76 | 137,57 | 22,86 | 48,59 |
| RF | -0,10 | 40,04 | 63,28 | 10,51 | 76,35 | 10,00 | 80,78 | 124,47 | 20,68 | 53,49 |
| Bagging | -0,86 | 87,58 | 130,83 | 21,72 | 51,11 | 9,24 | 91,17 | 137,14 | 22,79 | 48,75 |
| KNN | 1,33 | 79,69 | 124,08 | 20,60 | 53,63 | 6,50 | 89,71 | 136,75 | 22,72 | 48,90 |
| MLM | -6,22 | 91,30 | 138,88 | 23,06 | 48,10 | 1,77 | 98,46 | 147,73 | 24,55 | 44,79 |

Fonte: Autor

Tabela 29 – Resultados para o conjunto D

| Conjunto D para 30 minutos a posteriori | | | | | | | | | | |
|---|----------------------------|----------------------------|-----------------------------|--------------|-------|----------------------------|----------------------------|-----------------------------|--------------|-------|
| Métodos | Conjunto de treinamento | | | | | Conjunto de validação | | | | |
| | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) | MBE (W/m ²) | MAE (W/m ²) | RMSE (W/m ²) | rRMSE (%) | s (%) |
| Baseline | - | - | - | - | - | 2,32 | 238,12 | 278,63 | 46,92 | 0 |
| SVR | -14,77 | 72,18 | 121,13 | 20,48 | 56,53 | -2,02 | 87,87 | 141,28 | 23,79 | 49,30 |
| RF | -0,21 | 36,35 | 59,53 | 10,06 | 78,63 | 8,70 | 75,74 | 123,61 | 20,81 | 55,64 |
| Bagging | 2,29 | 91,41 | 132,76 | 22,44 | 52,35 | 13,60 | 97,40 | 145,04 | 24,42 | 47,95 |
| KNN | -5,04 | 87,07 | 135,87 | 22,97 | 51,24 | 4,66 | 92,20 | 142,31 | 23,96 | 48,93 |
| MLM | -6,90 | 91,08 | 138,66 | 23,44 | 50,23 | 11,99 | 99,72 | 148,01 | 24,92 | 46,88 |

Fonte: Autor

Pela observação dos resultados, é possível notar que ao utilizar os preditores recursivos **B**, **L** e **V** o desempenho dos algoritmos foi superior ao caso em que se aplicou os mesmos nos conjuntos I, II, III e IV, que não utilizam informações relativas aos valores de irradiância solar de instantes anteriores. Indicando, que as condições climáticas locais favorecem ao desenvolvimento de séries de irradiância solar que são fortemente autocorrelacionadas e que para se obter uma boa acurácia previsão de irradiância é sempre importante adicionar informação passada da série temporal de irradiância solar como preditores para os modelos.

Também é possível notar que as acurácias dos modelos ficam menos sensíveis à presença de informações extraídas de imagens do céu. De fato, há uma queda no desempenho dos métodos, com exceção do modelo de *Bagging*, quando se adiciona descritores estatísticos extraídos das imagens como preditores. Isto pode ser justificado pelo fato de que os preditores recursivos, como os resultados indicam, são estatisticamente mais significativos do que os sinais obtidos pelos LDR's, esta maior relevância para explicar a variância da irradiância pode ser grande o suficiente para que informações extraídas de imagens do céu passem a funcionar como obstáculos no processo de aprendizagem dos métodos, enquanto para os sinais de LDR's que apresentam menor relevância estatística para a explicação da resposta, as informações obtidas das imagens auxiliam na aprendizagem.

Com relação a filtragem das imagens, percebe-se que no geral o pré-processamento das imagens trouxe benefícios para o desempenho dos modelos quando comparado com o caso de utilizar informações de imagens não filtradas. As mudanças mais perceptíveis ocorrem no algoritmo das Florestas Aleatórias que obteve melhores acurácias quando se utilizou o filtro da mediana, confirmando que este método apresenta maior sensibilidade à presença de ruídos nas imagens. O uso do filtro LoG, rara algumas exceções, não trouxe maiores benefícios do que o filtro da mediana, mas é interessante notar que não há uma piora no desempenho dos algoritmos quando se tem o seu uso, o que difere do caso em que se utiliza os conjuntos I, II, III e IV.

Para trabalhos futuros, uma boa estratégia seria realizar decomposições nas imagens por meio do uso da transformada *wavelet* Haar (STOLLNITZ *et al.*, 1996) e calcular para as sub-imagens de aproximação e de detalhes a energia dos coeficientes wavelets obtidos e utilizar as mesmas como atributos para os métodos.

Todos os códigos implementados e utilizados para a obtenção dos resultados encontram-se no repositório <https://github.com/Felipe-Marinho94/Previs-o-de-irradia-o-solar>

6 CONCLUSÃO

Na dissertação, realizou-se a previsão de irradiância solar global horizontal para os horizontes de 10, 20 e 30 minutos a posteriori pela aplicação de modelos de aprendizagem de máquina em conjuntos de dados constituídos por sinais coletados por sensores de resistência dependente da luz e descritores estatísticos (média, desvio padrão e entropia de Shannon) extraídos de imagens do céu capturadas por uma câmera, onde tais sensores foram integrados por meio do uso de um *Raspberry Pi 3*.

Os algoritmos de aprendizagem de máquina utilizados foram: Máquina de Aprendizagem Mínima (MLM), Regressão por Vetor Suporte (SVR), K-Vizinhos mais Próximos (KNN), Florestas Aleatórias (RF) e uma árvore de regressão com a aplicação de um procedimento de *Bagging*. Posteriormente, os algoritmos foram aplicados nos mesmos conjuntos de dados diferindo apenas pela substituição dos atributos correspondentes aos sinais obtidos pelos sensores de luminosidade por funções de recorrência sobre os valores de irradiância de instantes anteriores (Conjuntos A, B, C e D).

Os resultados indicam que o uso de sinais de sensores LDR's como os únicos preditores de um conjunto de dados não fornecem uma boa representação do efeito estocástico da irradiância solar, observou-se também que para os conjuntos de dados com sinais de LDR's, a adição de descritores estatísticos extraídos de imagens do céu, como preditores proporciona melhorias consideráveis nos desempenhos dos modelos utilizados. Este aumento na acurácia de previsão é observado tanto no caso de imagens não filtradas como no caso em que as mesmas são filtradas. O uso do filtro da mediana aumentou o desempenho do algoritmo de Florestas Aleatórias, mas isto não foi uma regra geral para os demais métodos.

Também verifica-se uma melhora significativa na acurácia de previsão quando se utiliza os conjuntos A, B, C e D, confirmando que as séries de irradiância são fortemente autorrelacionadas. Para esta caso, constatou-se que o uso de imagens do céu não contribui para uma melhor generalização dos algoritmos de aprendizagem.

REFERÊNCIAS

- ALONSO-MONTESINOS, J.; BATLLES, F. J.; PORTILLO, C. Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images. **Energy Conversion and Management**, v. 105, p. 1166-1177, 2015.
- ALSHARIF, M. H.; YOUNES, M. K.; KIM, J. Time series arima model for prediction of daily and monthly average global solar radiation: The case study of seoul, south korea. **Symmetry**, v. 11, n. 2, p. 240, 2019.
- ARYAPUTERA, A. W.; YANG, D.; WALSH, W. M. Day-ahead solar irradiance forecasting in a tropical environment. **Journal of Solar Energy Engineering**, v. 137, n. 5, p. 051009, 2015.
- BADESCU, V.; GUEYMARD, C.; CHEVAL, S.; OPREA, C.; MADLINA, B.; DUMITRESCU, A.; LACOBESCU, F.; MILOS, I.; RADA, C. Computing global and diffuse solar hourly irradiation on clear sky. Review and testing of 54 models. **Renewable and Sustainable Energy Reviews**, v. 16, n. 3, p. 1636-1656, 2012.
- BENALI, L.; NOTTON, G.; FOUILLOY, A.; VOYANT, C.; DIZENE, R. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. **Renewable energy**, v. 132, p. 871-884, 2019.
- BEN-ISRAEL, A.; GREVILLE, Thomas N. E. **Generalized inverses: theory and applications**. Springer Science & Business Media, 2003.
- BJERKNES, V. The problem of weather prediction, as seen from the standpoints of mechanics and physics. **Meteorologische Zeitschrift**, **21**, **1**, v. 7, 1904.
- CARRASCO, M.; LÓPEZ, J.; MALDONADO, S. Epsilon-nonparallel support vector regression. **Applied Intelligence**, p. 1-14.
- CHARNEY, J. G. On the scale of atmospheric motions. 1948.
- CHARNEY, J. G.; FJÖRTOFT, R.; NEUMANN, J. V. Numerical integration of the barotropic vorticity equation. **Tellus**, v. 2, n. 4, p. 237-254, 1950.
- COLAK, I.; YESILBUDAK, M.; GENÇ, N.; BAYINDIR, R. Multi-period prediction of solar radiation using ARMA and ARIMA models. In: **2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)**. IEEE, 2015. p. 1045-1049.
- DE SOUZA JÚNIOR, A. H.; CORONA, F.; BARRETO, G. A.; MICHE, Y.; LENDASSE, A. Minimal learning machine: a novel supervised distance-based approach for regression and classification. **Neurocomputing**, v. 164, p. 34-44, 2015.
- DIETTERICH, T. G. Ensemble methods in machine learning. In: **International workshop on multiple classifier systems**. Springer, Berlin, Heidelberg, 2000. p. 1-15.

- DONG, Z.; YANG, D.; REINDL, T.; WALSH, W. M. Short-term solar irradiance forecasting using exponential smoothing state space model. **Energy**, v. 55, p. 1104-1113, 2013.
- DRUCKER, H.; BURGESS, C. J. C.; KAUFMAN, L.; SMOLA, A.; VAPNIK, V. Support vector regression machines. In: **Advances in neural information processing systems**. 1997. p. 155-161.
- DUFFIE, J. A.; BECKMAN, W. A. **Solar engineering of thermal processes**. John Wiley & Sons, 2013.
- EHRENDORFER, M. **Spectral numerical weather prediction models**. SIAM, 2012.
- ELIASSEN, E.; MACHENHAUER, B.; RASMUSSEN, E. **On a numerical method for integration of the hydrodynamical equations with a spectral representation of the horizontal fields**. Kobenhavns Universitet, Institut for Teoretisk Meteorologi, 1970.
- GEORGE, B. **Time Series Analysis: Forecasting & Control**, 3/e. Pearson Education India, 1994.
- GONZALEZ, R. C.; WOODS, R. C. **Processamento digital de imagens**. Pearson Educación, 2009.
- HARVEY, A. C. **Forecasting, structural time series models and the Kalman filter**. Cambridge university press, 1990.
- HAYKIN, S. **Neural networks and learning machines/Simon Haykin**. New York: Prentice Hall, 2009.
- HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, v. 12, n. 1, p. 55-67, 1970.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. OTexts, 2018.
- INMAN, R. H.; PEDRO, H. T.C.; COIMBRA, C. F.M. Solar forecasting methods for renewable energy integration. **Progress in energy and combustion science**, v. 39, n. 6, p. 535-576, 2013.
- JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern recognition letters**, v. 31, n. 8, p. 651-666, 2010.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning**. New York: springer, 2013.
- JOHNSON, S. C. Hierarchical clustering schemes. **Psychometrika**, v. 32, n. 3, p. 241-254, 1967.
- KUHN, M.; JOHNSON, K. **Applied predictive modeling**. New York: Springer, 2013.

- LI, M.; TANG, X.; WU, W.; LIU, H. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. **Energy conversion and management**, v. 70, p. 139-148, 2013.
- LI, M.; CHU, Y.; PEDRO, H. T. C.; COIMBRA, C. F. M. Quantitative evaluation of the impact of cloud transmittance and cloud velocity on the accuracy of short-term DNI forecasts. **Renewable energy**, v. 86, p. 1362-1371, 2016.
- LINARES-RODRÍGUEZ, A.; RUIZ-ARIAS, J. A.; POZO-VÁZQUEZ, D.; TOVAR-PESCADOR, J. Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks. **Energy**, v. 36, n. 8, p. 5356-5365, 2011.
- MARQUEZ, R.; COIMBRA, C. F.M. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. **Solar Energy**, v. 85, n. 5, p. 746-756, 2011.
- MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters. **Journal of the society for Industrial and Applied Mathematics**, v. 11, n. 2, p. 431-441, 1963.
- MARSH, L. C.; CORMIER, D. R. **Spline regression models**. Sage, 2001.
- MEHROTRA, S. On the implementation of a primal-dual interior point method. **SIAM Journal on optimization**, v. 2, n. 4, p. 575-601, 1992.
- MEJIA, J. F.; GIORDANO, M.; WILCOX, E. Conditional summertime day-ahead solar irradiance forecast. **Solar Energy**, v. 163, p. 610-622, 2018.
- MLSNA, P. A.; RODRIGUEZ, J. J. Gradient and laplacian edge detection. In: **The Essential Guide to Image Processing**. Academic Press, 2009. p. 495-524.
- MORÉ, J. J. The Levenberg-Marquardt algorithm: implementation and theory. In: **Numerical analysis**. Springer, Berlin, Heidelberg, 1978. p. 105-116.
- MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. In: **Análise de séries temporais**. 2006.
- MYERS, R. H. **Classical and modern regression with applications**. Belmont, CA: Duxbury press, 1990.
- NONNENMACHER, L.; KAUR, A.; COIMBRA, C. F. M. Day-ahead resource forecasting for concentrated solar power integration. **Renewable energy**, v. 86, p. 866-876, 2016.
- NIEWIADOMSKA-SZYNKIEWICZ, E.; MARKS, M. Optimization schemes for wireless sensor network localization. **International Journal of Applied Mathematics and Computer Science**, v. 19, n. 2, p. 291-302, 2009.

- ORD, J. K.; KOEHLER, A. B.; SNYDER, R. D. Estimation and prediction for a class of dynamic nonlinear statistical models. **Journal of the American Statistical Association**, v. 92, n. 440, p. 1621-1629, 1997.
- ORSZAG, S. A. Transform method for the calculation of vector-coupled sums: Application to the spectral form of the vorticity equation. **Journal of the Atmospheric Sciences**, v. 27, n. 6, p. 890-895, 1970.
- PEDREGAL, D. J.; YOUNG, P. C. Statistical approaches to modelling and forecasting time series. **Companion to economic forecasting**, p. 69-104, 2002.
- PEDRO, H. T.C.; COIMBRA, C. F.M. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. **Renewable Energy**, v. 80, p. 770-782, 2015.
- PEDRO, H. T. C.; COIMBRA, C. F. M.; DAVID, D.; LAURET P. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. **Renewable Energy**, v. 123, p. 191-203, 2018.
- PEDRO, H. T. C.; INMAN, R. H.; COIMBRA, C. F. M. Mathematical methods for optimized solar forecasting. In: **Renewable Energy Forecasting**. Woodhead Publishing, 2017. p. 111-152.
- PENROSE, R. A generalized inverse for matrices. In: **Mathematical proceedings of the Cambridge philosophical society**. Cambridge University Press, 1955. p. 406-413.
- PHILLIPS, N. A. The general circulation of the atmosphere: A numerical experiment. **Quarterly Journal of the Royal Meteorological Society**, v. 82, n. 352, p. 123-164, 1956.
- PLATZMAN, G. W. The spectral form of the vorticity equation. **Journal of Meteorology**, v. 17, n. 6, p. 635-644, 1960.
- QING, X.; NIU, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. **Energy**, v. 148, p. 461-468, 2018.
- ROBERT, A.; YEE, T L.; RITCHIE, H. A semi-Lagrangian and semi-implicit numerical integration scheme for multilevel atmospheric models. **Monthly Weather Review**, v. 113, n. 3, p. 388-394, 1985.
- ROUSSOPOULOS, N.; KELLEY, S.; VINCENT, F. Nearest neighbor queries. In: **ACM sigmod record**. ACM, 1995. p. 71-79.
- SILBERMAN, I. Planetary waves in the atmosphere. **Journal of Meteorology**, v. 11, n. 1, p. 27-34, 1954.
- SMAGORINSKY, J. On the numerical integration of the primitive equations of motion for baroclinic flow in a closed region. **Monthly Weather Review**, v. 86, n. 12, p. 457-466, 1958.
- STOLLNITZ, E. J.; DEROSE, T. D.; SALESIN, D. H. **Wavelets for computer graphics: theory and applications**. Morgan Kaufmann, 1996.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 58, n. 1, p. 267-288, 1996.

TRAPERO, J. R.; KOURENTZES, N.; MARTIN, A. Short-term solar irradiation forecasting based on dynamic harmonic regression. **Energy**, v. 84, p. 289-295, 2015.

WALL JR, J. E.; WILLISKY, A. S.; SANDELL JR, N. R. On the fixed-interval smoothing problem. **Stochastics: An International Journal of Probability and Stochastic Processes**, v. 5, n. 1-2, p. 1-41, 1981.

WELCH, G.; BISHOP, G. An introduction to the Kalman filter. 1995.

WEST, M.; HARRISON, J. **Bayesian forecasting and dynamic models**. Springer Science & Business Media, 2006.

YAGLI, G. M.; YANG, D.; SRINIVASAN, D. Automatic hourly solar forecasting using machine learning models. **Renewable and Sustainable Energy Reviews**, v. 105, p. 487-498, 2019.

YOUNG, P. C.; PEDREGAL, D. J.; WLODEK. Dynamic harmonic regression. **Journal of forecasting**, v. 18, n. 6, p. 369-394, 1999.

ZHAO, X.; WEI, H.; WANG, H.; ZHU, T.; ZHANG, K. 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction. **Solar Energy**, v. 181, p. 510-518, 2019.