

UNIVERSIDADE FEDERAL DE ALFENAS

IGOR CAETANO SILVA

**ANÁLISE E APERFEIÇOAMENTO DE MODELOS INTELIGENTES PARA
DETECCÃO DE LÂMPADAS DE ILUMINAÇÃO PÚBLICA**

Alfenas/MG

2022

IGOR CAETANO SILVA

**ANÁLISE E APERFEIÇOAMENTO DE MODELOS INTELIGENTES PARA
DETECCÃO DE LÂMPADAS DE ILUMINAÇÃO PÚBLICA**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas, MG, como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria. **Linha de Pesquisa:** Matemática Aplicada e Modelagem Matemática.

Orientador: Prof. Dr. Ricardo Menezes Salgado.

Alfenas/MG

2022

Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central

Silva, Igor Caetano.

Análise e aperfeiçoamento de modelos inteligentes para detecção de lâmpadas de iluminação pública / Igor Caetano Silva. - Alfenas, MG, 2022.
120 f. : il. -

Orientador(a): Ricardo Menezes Salgado.

Dissertação (Mestrado em Estatística Aplicada e Biometria) -
Universidade Federal de Alfenas, Alfenas, MG, 2022.
Bibliografia.

1. Aprendizagem de Máquina. 2. Iluminação Pública. 3. Detecção de Padrões. 4. Algoritmos de Classificação. 5. Redução de Custos. I. Salgado, Ricardo Menezes, orient. II. Título.

IGOR CAETANO SILVA

ANÁLISE E APERFEIÇOAMENTO DE MODELOS INTELIGENTES PARA DETECÇÃO DE LÂMPADAS DE ILUMINAÇÃO PÚBLICA

A Banca examinadora abaixo-assinada aprova a Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria.

Aprovada em: 29 de julho de 2022.

Prof. Dr. Ricardo Menezes Salgado
Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Prof. Dr. Takaaki Ohishi
Instituição: Universidade Estadual de Campinas - Unicamp

Prof. Dr. Flávio Miguel Varejão
Instituição: Universidade Federal do Espírito Santo - UFES



Documento assinado eletronicamente por **Ricardo Menezes Salgado, Professor do Magistério Superior**, em 03/08/2022, às 15:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Takaaki Ohishi, Usuário Externo**, em 03/08/2022, às 15:42, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Flávio Miguel Varejão, Usuário Externo**, em 03/08/2022, às 16:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0788150** e o código CRC **1A1954FD**.

AGRADECIMENTOS

Agradeço primeiramente à Universidade Federal de Alfenas e o Laboratório de Inteligência Computacional (LInC) pelos recursos disponibilizados sem os quais não seria possível realizar os experimentos necessários para elaboração desta dissertação.

A Ludmilla, Julia, Paula e, em especial, Maria por serem pilares tão importantes de apoio, incentivo e alegria.

Ao meu orientador Dr. Ricardo Menezes Salgado pelos ensinamentos, auxílio e paciência durante todo o desenvolvimento deste trabalho.

Ao Dr. Luiz Alberto Beijo pelos ensinamentos e compartilhamento de ideias na disciplina de Estágio Docência que contribuíram imensamente para meu crescimento intelectual e pessoal.

A CAPES e FAPEMIG pelo apoio financeiro. E por fim, a todos aqueles que de alguma forma contribuíram para essa realização.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

Há alguns anos, uma mudança na legislação passou a responsabilidade de gerenciamento e manutenção da rede de iluminação pública das companhias elétricas (agora responsáveis apenas pelo faturamento da energia utilizada) para as autoridades municipais. Essa mudança acarretou em diversos problemas de desinformação, nos quais muitas vezes a companhia elétrica não é notificada sobre mudanças na rede de iluminação pública. Para evitar perdas comerciais, as companhias passaram a enviar times de conferência manual, processo caro, moroso e pouco confiável. Neste sentido, este trabalho tem como objetivo aprimorar o estudo de detecção inteligente de lâmpadas de iluminação pública, através da otimização dos modelos propostos por Soares *et al.* (2015), capazes de classificar o tipo e potência de lâmpadas de pontos de iluminação pública de forma eficiente como uma solução alternativa para este problema. A proposta é utilizar algoritmos de diferentes níveis de complexidade (tanto de abordagem tradicional quanto aprendizado profundo), em conjunto com técnicas mais complexas de validação, seleção de características, transformação de dados e de otimização de hiperparâmetros. Os resultados mostram que os modelos com algoritmos mais complexos (máquina de vetores de suporte, XGBoost, floresta aleatória e *perceptron* multicamadas) conseguem atingir uma acurácia média final de 80-86% de acerto que ao serem comparados por testes *t de Student* não apresentaram evidências de diferença significativa a nível de 5%.

Palavras-chave: Aprendizagem de Máquina; Iluminação Pública; Detecção de Padrões;
Algoritmos de Classificação; Redução de Custos.

ABSTRACT

A few years ago, a change in legislation transferred the responsibility for managing and maintaining the public lighting network from electric companies (now responsible only for billing the energy used) to municipal authorities. This change led to several misinformation problems, in which the electric company is often not notified about changes in the public lighting network. To avoid commercial losses, companies started to send manual conference teams, an expensive, time-consuming and unreliable process. In this sense, this work aims to improve the study of intelligent detection of street lighting lamps, through the optimization of the models proposed by Soares *et al.* (2015), able to efficiently classify the type and power of street lighting point lamps as an alternative solution to this problem. The proposal is to use algorithms of different levels of complexity (from both traditional approach and deep learning), along with more complex techniques of validation, feature selection, data transformation and hyperparameter optimization. The results show that models with more complex algorithms (support vector machine, XGBoost, random forest and multilayer perceptron) manage to reach a final average accuracy of 80-86%, which when compared using Student's t tests did not show evidence of significant difference at the 5% level.

Keywords: Machine Learning; Street lighting; Pattern Recognition; Classification Algorithms; Cost Reduction.

LISTA DE FIGURAS

Figura 1 - Ambiente de laboratório semi-controlado usado para criação da base de dados.....	20
Figura 2 - Diagrama representativo da estrutura de classificação hierárquica.....	21
Figura 3 - Representação do processo de aquisição de dados, incluindo veículo de transporte, dispositivo para coleta e equipamento de iluminação pública.....	33
Figura 4 - Arquitetura do equipamento de coleta de dados.....	34
Figura 5 - Ângulos utilizados para o cálculo de matrizes de co-ocorrência.....	41
Figura 6 - Exemplo de um modelo de k-vizinhos mais próximos para classificação binária.....	45
Figura 7 - Exemplo de Árvore de decisão, demonstrando as etapas do processo de classificação de um modelo de árvore de decisão.....	46
Figura 8 - Visualização do processo de decisão do modelo Floresta Aleatória.....	48
Figura 9 - Representação do processo de iteração dos algoritmos de <i>gradient boosting</i>	49
Figura 10 - Representação de um algoritmo de máquina de vetores de suporte.....	50
Figura 11 - Exemplo de um modelo de <i>perceptron</i> multicamadas para classificação binária.....	51
Figura 12 - Representação de uma rede neural convolucional.....	52
Figura 13 - Exemplo de processo de validação cruzada.....	57
Figura 14 - Exemplo de processo de validação cruzada aninhada.....	57
Figura 15 - Foto de uma lâmpada de um poste de iluminação pública representada em cores (esquerda), escala cinza (centro) e escala binária (direita).....	62
Figura 16 - Valores de importância das variáveis para o <i>naive classifier</i> de Floresta Aleatória.....	67
Figura 17 – Representação estrutural geral da solução proposta.....	71
Figura 18 - Diagrama representativo da estrutura de classificação hierárquica.....	72
Figura 19 - Diagrama representativo da estrutura de classificação única.....	72
Figura 20 - Diagrama representativo da rotina de treinamento e validação dos experimentos.....	74
Figura 21 - Representação do funcionamento de uma rede neural convolucional.....	79
Figura 22 - Fotos de três tipos distintos de lâmpadas de pontos de iluminação pública: MV (esquerda), MH (centro) e HPS (direita).....	83
Figura 23 - Acurácia e perda do processo de treinamento e validação de uma das iterações	

da rede neural classificadora de tipo.	83
Figura 24 - Fotos de duas lâmpadas de 150W: MH (esquerda) e HPS (direita).	86
Figura 25 - Acurácia e perda do processo de treinamento e validação de uma das iterações da rede neural classificadora de potência.	87
Figura 26 - Visualização dos conjuntos de dados representando os três diferentes tipos de lâmpadas.	88
Figura 27 - Visualização dos conjuntos de dados representando as seis diferentes potências de lâmpada.	89
Figura 28 - Acurácia e perda do processo de treinamento e validação de uma das iterações da rede neural classificadora de única.	91
Figura 29 - Visualização dos conjuntos de dados representando os nove diferentes modelos de lâmpada.	92
Figura 30 - Histogramas das variáveis provenientes dos sensores.	115
Figura 31 - Histogramas das variáveis provenientes de imagens.	116

LISTA DE QUADROS

Quadro 1 - Importância média das variáveis para classificação de tipos.	109
Quadro 2 - Importância média das variáveis para classificação de potência.	110
Quadro 3 – Importância média das variáveis para classificação única.	111
Quadro 4 – <i>Ranking</i> de variáveis para classificação de tipos.....	112
Quadro 5 – <i>Ranking</i> de variáveis para classificação de potência.	113
Quadro 6 – <i>Ranking</i> de variáveis para classificação única.....	114
Quadro 7 - Resultados do teste t com classificadores de tipos de lâmpada.....	117
Quadro 8 - Resultados do teste t com classificadores de potência da lâmpada.....	118
Quadro 9 - Resultados do teste t com classificadores únicos.	119
Quadro 10 - Resultados do teste t com os classificadores de melhor performance.....	120

LISTA DE TABELAS

Tabela 1 - Acurácia média dos melhores classificadores dos testes em laboratório de Soares <i>et al.</i> (2015) sem uso da informação de altura do poste.....	21
Tabela 2 - Acurácia média dos melhores classificadores dos testes em laboratório de Soares <i>et al.</i> (2015) com uso da informação de altura do poste.	22
Tabela 3 - Acurácia média dos melhores classificadores dos testes em campo de Soares <i>et al.</i> (2015).....	23
Tabela 4 - Acurácia média dos melhores classificadores dos testes com a base de dados da câmera Canon de Broetto <i>et al.</i> (2013).....	24
Tabela 5 - Acurácia média dos melhores classificadores dos testes com a base de dados da câmera Manta de Broetto <i>et al.</i> (2013).	24
Tabela 6 - Resultados dos experimentos de Broetto e Varejão (2016).	25
Tabela 7 - Características fotométricas dos sensores radiométricos.	35
Tabela 8 - Resumo da distribuição de tipos de lâmpadas do banco de dados.	60
Tabela 9 - Resumo da distribuição de potências de lâmpadas do banco de dados.....	60
Tabela 10 - Resumo da distribuição de modelos únicos de lâmpada do banco de dados.	60
Tabela 11 - Estatísticas descritivas (Tendência Central e dispersão) das variáveis providas pelos sensores, em nm.	61
Tabela 12 - Estatísticas descritivas (Tendência Central e dispersão) das variáveis obtidas pela DFT.....	63
Tabela 13 - Estatísticas descritivas (Tendência Central e dispersão) das variáveis obtidas através dos momentos invariantes de Hu.	63
Tabela 14 - Estatísticas descritivas (Tendência Central e dispersão) das variáveis obtidas através dos descritores de Haralick.	64
Tabela 15 - Variáveis do conjunto D para classificação de tipos, potência e única de lâmpadas.	68
Tabela 16 - Intervalo de valores de hiperparâmetros para o algoritmo KNN.	75
Tabela 17 - Intervalo de valores de hiperparâmetros para o algoritmo DT.....	76
Tabela 18 - Intervalo de valores de hiperparâmetros para o algoritmo RF.	76
Tabela 19 - Intervalo de valores de hiperparâmetros para o algoritmo XGB.....	76
Tabela 20 - Intervalo de valores de hiperparâmetros para o algoritmo SVM.	76
Tabela 21 - Intervalo de valores de hiperparâmetros para o algoritmo MLP.....	76
Tabela 22 - Arquitetura da rede neural convolucional utilizada.	78

Tabela 23 - Acurácia média e Intervalo de Confiança (95%) dos classificadores de tipo de lâmpada nos conjuntos de dados A, B, C e D com uso de normalização.	81
Tabela 24 - Acurácia média e Intervalo de Confiança (95%) dos classificadores de potência lâmpada nos conjuntos de dados A, B, C e D com uso de normalização.	84
Tabela 25 - Acurácia média e Intervalo de Confiança (95%) dos classificadores únicos nos conjuntos de dados A, B, C e D com uso de normalização.	90
Tabela 26 - Acurácia média e Intervalo de Confiança (95%) dos classificadores de Tipo nos conjuntos de dados A, B, C e D com uso de padronização.	93
Tabela 27 - Acurácia média e Intervalo de Confiança (95%) dos classificadores de potência nos conjuntos de dados A, B, C e D com uso de padronização.	93
Tabela 28 - Acurácia média e Intervalo de Confiança (95%) dos classificadores únicos nos conjuntos de dados A, B, C e D com uso de padronização.	93
Tabela 29 - Acurácia média e Intervalo de Confiança (95%) dos melhores modelos.	96

LISTA DE SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNN	<i>Convolutional Neural Network</i>
DFT	<i>Discrete Fourier Transform</i>
DOE	<i>United States of America Department of Energy</i>
DSP	<i>Digital Signal Processor</i>
DT	<i>Decision Tree</i>
EDP	Energias de Portugal
EXIF	<i>Exchangeable Image File Format</i>
FAPEMIG	Fundação de Amparo à Pesquisa do Estado de Minas Gerais
FESF	<i>Features Extraction and Selection Framework</i>
GLCM	<i>Grey Level Co-occurrence Matrix</i>
GPS	<i>Global Positioning System</i>
GRASP	<i>Greedy Randomized Adaptative Search Procedures</i>
HPS	<i>High Pressure Sodium</i>
HUE	<i>hue, saturation and value</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IoT	<i>Internet of Things</i>
KNN	<i>K-nearest neighbors</i>
LED	<i>Light-Emitting Diode</i>
LInC	Laboratório de Inteligência Computacional
LTC	<i>Lamp Type Classifier</i>
LPC	<i>Lamp Power Classifier</i>
MLP	<i>Multilayer perceptron</i>
MV	<i>Mercury Vapor</i>
MH	<i>Metal Halide</i>
PCA	<i>Principal Component Analysis</i>
RGB	<i>Red, Green and Blue</i>
RF	<i>Random Forest</i>
SEDU	Secretaria de Estado do Desenvolvimento Urbano
SFFS	<i>Sequential Floating Forward Selection</i>
SFS	<i>Sequential Forward Selection</i>

SVM	<i>Support Vector Machine</i>
UC	<i>Unique Classifier</i>
UFES	Universidade Federal do Espírito Santo
UNICAMP	Universidade Estadual de Campinas
UNIFAL	Universidade Federal de Alfenas
XGB	XGBoost

SUMÁRIO

1	INTRODUÇÃO	17
1.1	PRINCIPAIS CONTRIBUIÇÕES DOS TRABALHOS QUE EMBASARAM ESSA PESQUISA	19
1.1.1	Experimentos realizados em laboratório	20
1.1.2	Experimentos realizados em campo	22
1.1.3	Processamento de imagens	23
1.1.4	Seleção de características	24
1.2	OBJETIVOS	25
1.2.1	Objetivo Geral	26
1.2.2	Objetivos Específicos	26
1.3	CONTEXTUALIZAÇÃO	26
1.4	TRABALHOS RELACIONADOS	28
1.5	JUSTIFICATIVA	30
1.6	ESTRUTURA DA DISSERTAÇÃO	31
2	REFERENCIAL TEÓRICO	32
2.1	ARQUITETURA DO SISTEMA	32
2.1.1	Visão geral do <i>hardware</i>	33
2.1.2	Deteção da assinatura de espectro	35
2.2	PROCESSAMENTO DE IMAGENS	37
2.2.1	Descritores de Fourier	38
2.2.2	Descritores de Hu	39
2.2.3	Descritores de Haralick	40
2.3	TÉCNICAS DE RECONHECIMENTO DE PADRÕES	43
2.3.1	K-Vizinhos mais próximos	44
2.3.2	Árvore de decisão	45
2.3.3	Floresta Aleatória	47
2.3.4	XGBoost	48
2.3.5	Máquina de vetores de suporte	50
2.3.6	Perceptron multicamadas	51
2.3.7	Redes neurais convolucionais	52
2.4	TÉCNICAS DE PROCESSAMENTO DE DADOS	54
2.4.1	Atributos artificiais	54

2.4.2	Seleção de características	55
2.4.3	Métodos de treinamento e validação	56
3	METODOLOGIA	59
3.1	BASE DE DADOS	59
3.1.1	Conjuntos de dados	65
3.1.2	Seleção de características	66
3.1.3	Transformação dos dados	69
3.2	CONSTRUÇÃO DE TREINAMENTO DOS MODELOS	70
3.3	TREINAMENTO DOS MODELOS	73
3.3.1	Loop interno	75
3.3.2	Experimentos	79
4	RESULTADOS E DISCUSSÃO	81
4.1	COMPARAÇÃO DE MODELOS COM ESTRUTURA DE CLASSIFICAÇÃO HIERÁRQUICA	81
4.1.1	Desempenho dos classificadores de tipo	81
4.1.2	Desempenho dos classificadores de potência	84
4.2	COMPARAÇÃO DE MODELOS COM ESTRUTURA DE CLASSIFICAÇÃO ÚNICA	90
4.3	RESULTADOS COM USO DE PADRONIZAÇÃO	92
4.4	COMPARAÇÃO ENTRE OS MODELOS	94
4.5	COMPARAÇÃO COM RESULTADOS DE SOARES <i>ET AL.</i> (2013, 2014, 2015), BROETTO <i>ET AL.</i> (2013) E BROETTO E VAREJÃO (2016)	96
5	CONCLUSÃO	98
	REFERÊNCIAS	100
	APÊNDICE A – IMPORTÂNCIA MÉDIA DAS VARIÁVEIS PARA CLASSIFICAÇÃO DE TIPOS	109
	APÊNDICE B - IMPORTÂNCIA MÉDIA DAS VARIÁVEIS PARA CLASSIFICAÇÃO DE POTÊNCIA	110
	APÊNDICE C - IMPORTÂNCIA MÉDIA DAS VARIÁVEIS PARA CLASSIFICAÇÃO ÚNICA	111
	APÊNDICE D – RANKING DE VARIÁVEIS PARA CLASSIFICAÇÃO DE TIPOS	112
	APÊNDICE E - RANKING DE VARIÁVEIS PARA CLASSIFICAÇÃO DE POTÊNCIA	113

APÊNDICE F - RANKING DE VARIÁVEIS PARA CLASSIFICAÇÃO ÚNICA.....	114
APÊNDICE G – HISTOGRAMAS DAS VARIÁVEIS PROVENIENTES DOS SENSORES	115
APÊNDICE H – HISTOGRAMAS DAS VARIÁVEIS PROVENIENTES DE IMAGENS	116
APÊNDICE I - RESULTADOS DOS TESTES T COM CLASSIFICADORES DE TIPOS DE LÂMPADA.....	117
APÊNDICE J - RESULTADOS DOS TESTES T COM CLASSIFICADORES DE POTÊNCIA DA LÂMPADA	118
APÊNDICE K - RESULTADOS DOS TESTES T COM CLASSIFICADORES DE CLASSIFICAÇÃO ÚNICA	118
APÊNDICE L - RESULTADOS DOS TESTES T COM CLASSIFICADORES DE MELHOR PERFORMANCE.....	119

1 INTRODUÇÃO

Iluminação pública é definida, segundo a Agência Nacional de Energia Elétrica (ANEEL, 2000), como serviço público que tem como propósito exclusivo prover luz (de forma contínua, periódica ou eventual), ou claridade artificial, à logradouros públicos no período noturno ou em escurecimentos diurnos ocasionais, incluso casos que iluminação permanente no período diurno seja necessária.

Este serviço abrange o fornecimento de energia para os mais diversos locais públicos, como: ruas, avenidas, estradas, praças, parques, pontos de transporte coletivo, monumentos históricos ou culturais, entre outros. Assim, este é essencial para a vida moderna (especialmente em centros urbanos) e está diretamente ligado a segurança pública, pois permite que os cidadãos aproveitem espaços públicos no período noturno, garante a segurança no tráfego, além de ajudar a melhorar a imagem do município, favorecer turismo e comércio local, entre diversos outros benefícios.

Conseqüentemente, segundo Silva (2006), a iluminação pública pode refletir as desigualdades socioespaciais verificadas no país, como: número de pontos de luz por habitante mais expressivos em alguns locais, ambientes sem iluminação, luminárias mais modernas sendo utilizadas apenas em grandes centros urbanos, cidades menores com tecnologias obsoletas ou em condições que não atingem os níveis mínimos de qualidade para conferir a segurança desejada, entre outros.

Dessa forma, a prestação deste serviço deve ser provida de forma a potencializar estes benefícios, reduzindo os custos bem como os impactos ambientais e à sociedade. A tarefa de gerenciamento, instalação e manutenção de pontos de iluminação públicas era, até 2010, das companhias de distribuição de energia. Entretanto, devido a mudanças na regulamentação governamental, suportadas pelo inciso V do artigo 30 da Constituição Federal (BRASIL, 1988), estas atividades passaram de forma integral a ficar a cargo do governo Municipal, ou a quem ele delegar a operação e monitoração do acervo. Assim, fica a cargo das companhias elétricas apenas o faturamento da energia consumida.

Segundo a Secretária de Estado do Desenvolvimento Urbano (SEDU, 2018) do Estado do Paraná, as responsabilidades sobre a iluminação pública, compartilhadas entre a distribuidora e o município, devem estar muito bem definidas, de modo que o ponto final da competência da companhia elétrica deve ser o ponto inicial de responsabilidade da administração municipal. O limiar de obrigação entre as duas partes é denominado ***Ponto de Entrega*** ou ***Ponto de Conexão***.

Porém, por se tratar de um serviço público que não é conduzido e organizado de forma centralizada, e sim uma responsabilidade individual de cada um dos, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE, 2017), 5.570 municípios do território brasileiro, há uma grande quantidade de métodos, técnicas e tecnologias que são utilizadas, além dos vários níveis de suporte e qualidade dos serviços fornecidos.

Neste cenário, não é raro ocorrer de municípios realizarem a instalação de novos circuitos de iluminação ou fazer alterações nas instalações antigas sem reportar devidamente as companhias elétricas. Como a cobrança pelo uso da energia ainda é realizado pelas distribuidoras, a omissão de informações sobre mudanças na rede por parte do município pode resultar em prejuízo, principalmente comercial, mas também técnico e de eficiência. Logo, o controle dos dispositivos de iluminação pública se tornou uma grande preocupação para os órgãos reguladores e empresas de distribuição de energia elétrica.

Para minimizar este problema, as companhias assumem a função de inspeção enviando técnicos para verificar se a rede de iluminação pública está sincronizada com a base de dados concessionária ou se é necessário realizar a atualização de alguns registros. Assim, a equipe de inspeção visita ruas e áreas públicas para identificar a existência de novos pontos de luz, bem como as principais características de pontos de iluminação pública, como: posicionamento, a tecnologia e potência da lâmpada. Neste processo, muitas vezes é necessário que o operador suba no poste (com auxílio de escadas apropriadas ou por meio de caminhões de lança articulada) e coletar as informações almejadas visualmente.

Salienta-se como fator agravante, a coexistência de inúmeros padrões em redes de iluminação pública, combinando lâmpadas de designs, tipos e potências diferentes, além dos diversos tipos de estradas públicas, o que torna o trabalho de identificação dos detalhes de uma instalação ainda mais complexo.

Desse modo, dependendo das qualificações do time de inspeção e se são funcionários da própria companhia ou terceirizados, o serviço pode não ser tão confiável, visto que, não há formas padronizadas de conferir a verificação feita. Em alguns casos, pode ser observado que os técnicos evitam as inspeções *in loco* (por ser um procedimento minucioso e demorado), preferindo informar uma condição aproximada da rede que resulta em dados incorretos e/ou omissão de novos pontos de iluminação. Em outras palavras, o processo de conferência é moroso, tem uma abordagem cara e apresenta efetividade duvidosa.

Por conseguinte, as condições descritas motivaram o desenvolvimento de um sistema, que incorpora elementos de *hardware* (sensores de luz, uma câmera digital e circuitos periféricos) e *software* (metodologia computacional), com alto grau de automatização,

planejado para determinar o tipo e voltagem das lâmpadas instaladas nos postes utilizados para iluminação pública. Soares *et al.* (2013, 2014, 2015) descrevem os principais elementos que constituem este aparelho, assim como, seu funcionamento em detalhes e resultados de testes em laboratório e campo.

Logo, este dispositivo eletrônico é capaz de adquirir informação luminosa do ambiente, esta que é processada por um algoritmo inteligente que a informa a localização, modelo e voltagem da lâmpada em análise. Na melhor das hipóteses, este poderia realizar uma atualização rápida dos registros das companhias elétricas, permitindo uma estimativa de perda causada por falta de informações corretas sobre consumo.

Nesse sentido, Soares *et al.* (2013, 2014, 2015) propôs modelos inteligentes para classificação de tipos e potência de lâmpadas utilizando informações provenientes de sensores e Broetto *et al.* (2013) propôs um modelo para classificação de tipos de lâmpadas com o uso exclusivo de informações providas de imagens. Entretanto, apesar de resultados pertinentes, apenas alguns algoritmos e métodos mais simples de validação foram utilizados nos experimentos.

Logo, isto motivou o presente trabalho a propor uma versão aprimorada do modelo proposto e complementar as ideias apresentadas anteriormente, através da combinação das descobertas dos trabalhos anteriores, uso de uma gama maior de algoritmos inteligentes (de diferentes abordagens e níveis de complexidades), processo de otimização de hiperparâmetros, um método de validação mais complexo, múltiplas estruturas de classificação e o uso conjunto de informações provenientes dos sensores e imagens para criar modelos com maior precisão e desempenho. A seguir, são descritas em detalhes as contribuições feitas pelos trabalhos anteriores que motivaram esta pesquisa.

1.1 PRINCIPAIS CONTRIBUIÇÕES DOS TRABALHOS QUE EMBASARAM ESSA PESQUISA

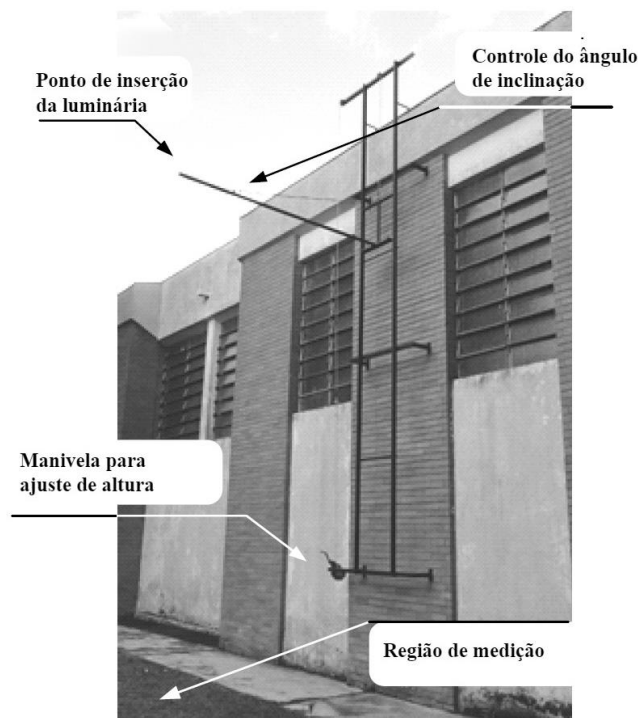
Parte da pesquisa mostrada neste trabalho foi proposta anteriormente em Soares *et al.* (2013), Soares *et al.* (2014) e Soares *et al.* (2015), além de um estudo detalhado em processamento de imagens de lâmpadas de iluminação pública apresentado por Broetto *et al.* (2013) e outro abordando processos de seleção de características em Broetto e Varejão (2016). Estes estudos foram realizados com o apoio da empresa de eletricidade EDP (Energias de Portugal) Brasil.

Os trabalhos de, Soares *et al.* (2013), Soares *et al.* (2014) e Soares *et al.* (2015), descrevem os elementos que constituem o sistema de coleta de dados formado por *hardware* e *software*, responsável por processar de forma *online* e *offline* as informações coletadas e prover a localização, tipo e potência das lâmpadas observadas, assim como, resultados de uma avaliação de performance proveniente de testes em laboratório e campo.

1.1.1 Experimentos realizados em laboratório

Para a coleta de dados dos experimentos realizados em laboratório, foi construído um ambiente semi-controlado, ilustrado na Figura 1, que permitisse ajuste de altura do local de coleta e ângulo de inclinação da haste que suporta a luminária observada. Neste espaço, foram conduzidos e registrados 324 experimentos sob condições controladas com diferentes luminárias (de diferentes tipos, potências e tecnologias), altura e ângulos de inclinação. Além disso, a metodologia de coleta foi executada em diferentes temperaturas e condições climáticas, permitindo que as experimentações funcionassem como uma simulação simplificada das condições reais, mesmo que a altura real de alguns postes de iluminação pública não pudesse ser reproduzida com a estrutura criada.

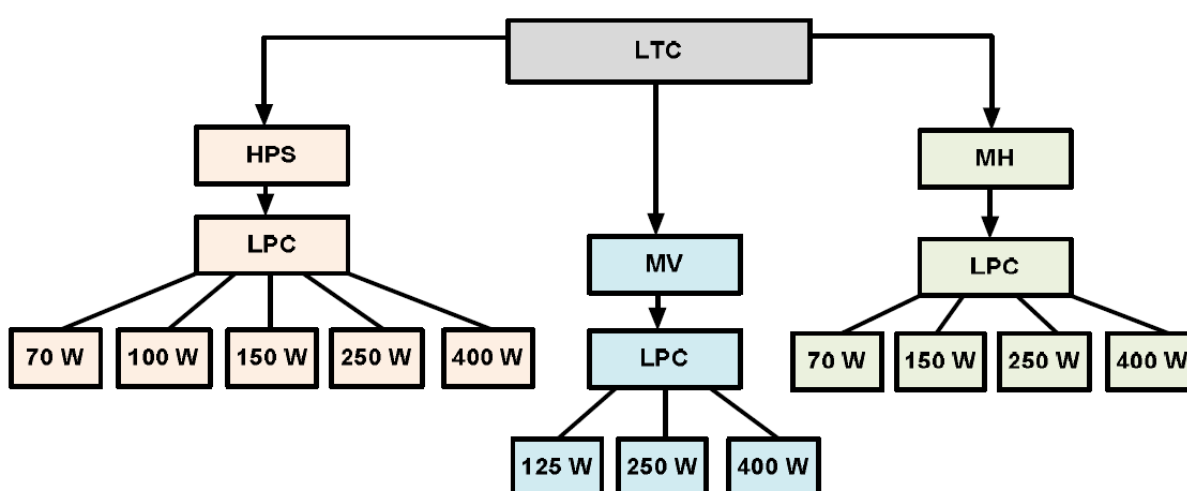
Figura 1 - Ambiente de laboratório semi-controlado usado para criação da base de dados.



Fonte: Adaptado de Soares *et al.* (2015).

A estrutura de classificação dos modelos propostos para classificação (Figura 2) das instâncias medidas em laboratório é dividida em duas partes: a classificação de lâmpadas e potência. Assim, os dados são processados primeiramente pelo classificador de tipo (*Lamp Type Classifier - LTC*), que os divide em 3 conjuntos: Lâmpada de vapor de mercúrio (*Mercury Vapor - MV*), lâmpada de vapor de sódio de alta pressão (*High Pressure Sodium - HPS*) e lâmpada de vapor metálico (*Metal Halide - MH*). Posteriormente, conforme a saída da primeira etapa, a instância é submetida a um classificador de potência (*Lamp Power Classifier - LPC*).

Figura 2 - Diagrama representativo da estrutura de classificação hierárquica.



Fonte: Soares *et al.* (2015).

Desta forma, utilizando esta metodologia, quatro classificadores são necessários e cada um é avaliado utilizando 3 diferentes algoritmos (J48, IBK e *Multilayer Perceptron*). Para observar a performance de cada classificador, dois tipos de experimentos foram realizados, um utilizando apenas dados provenientes dos sensores óticos e outro utilizando informações adicionais a respeito da altura do poste. Os melhores resultados, obtidos através das 324 instâncias (de diferentes tipos de lâmpadas, alturas, angulação e luminárias), para cada classificador em cada metodologia podem ser observados respectivamente nas Tabelas 1 e 2.

Tabela 1 - Acurácia média dos melhores classificadores dos testes em laboratório de Soares *et al.* (2015) sem uso da informação de altura do poste.

Classificador	Tipo de Lâmpada	Algoritmo	Acurácia Média
LTC	MV, MH e HPS	<i>Multilayer Perceptron</i>	0.975
LPC	MV	<i>Multilayer Perceptron</i>	0.988
LPC	MH	<i>Multilayer Perceptron</i>	0.862
LPC	HPS	<i>Multilayer Perceptron</i>	0.822

Fonte: Elaborada pelo autor (2022).

Tabela 2 - Acurácia média dos melhores classificadores dos testes em laboratório de Soares *et al.* (2015) com uso da informação de altura do poste.

Classificador	Tipo de Lâmpada	Algoritmo	Acurácia Média
LTC	MV, MH e HPS	<i>Multilayer Perceptron</i>	0.975
LPC	MV	<i>Multilayer Perceptron</i>	0.988
LPC	MH	<i>Multilayer Perceptron</i>	0.890
LPC	HPS	<i>Multilayer Perceptron</i>	0.896

Fonte: Elaborada pelo autor (2022).

Observa-se que a performance dos modelos LTC não tiveram sua média alterada de forma significativa pela adição da informação de altura do poste, com ambos atingindo a marca de 97.5% de assertividade. Contudo, o desempenho dos LPC para lâmpadas MV, MH e HPS tiveram um acréscimo significativo em suas médias com a adição da variável de altura do poste. Tal comportamento é justificado pelo fato da intensidade da radiação que é mensurada pelos sensores ser dependente da altura do poste em que se encontra a lâmpada observada. Logo, espera-se que melhores resultados sejam atingidos quando se conhece esta informação.

Outro ponto a ser notado nos resultados alcançados é a prevalência do algoritmo *Multilayer Perceptron*, o mais complexo entre os escolhidos, em todos os tipos de classificação.

1.1.2 Experimentos realizados em campo

Na realização dos experimentos em campo, foram coletadas informações de 276 pontos de iluminação distintos de diferentes tipos de lâmpadas e potências.

Os classificadores empregados foram os mesmos utilizados para os experimentos em laboratórios. Entretanto, ao invés de separar as instâncias em grupos de tecnologias similares por meio de um LTC, uma estratégia diferente de classificação foi utilizada, na qual um único classificador é responsável pela classificação de ambas as variáveis alvo (tipo e potência).

Outra ferramenta também utilizada para otimizar a assertividade dos modelos foi a aplicação de um método de seleção de características. Assim, cada classificador tem seu próprio subconjunto de características escolhidas por um *software* para maximizar seus resultados.

De forma análoga aos testes em ambiente controlado, dois tipos de experimentos foram realizados, um utilizando o método de seleção de características e outro utilizando todas as informações coletadas pelos sensores, o melhor resultado obtido em cada tipo pode ser observado na Tabela 3.

Tabela 3 - Acurácia média dos melhores classificadores dos testes em campo de Soares *et al.* (2015).

Seleção de características	Algoritmo	Acurácia Média
Não	<i>Multilayer Perceptron</i>	0.823
Sim	<i>Multilayer Perceptron</i>	0.830

Fonte: Elaborada pelo autor (2022).

O desempenho médio dos classificadores que usufruíram do processo de seleção de características é, em geral, superior aos que utilizaram a totalidade de dados coletados pelos sensores, devido a eliminação de informação redundante. E, assim como os melhores resultados nos testes em laboratório, o *Multilayer Perceptron* foi soberano entre os 3 algoritmos selecionados, atingindo as maiores taxas de acerto.

1.1.3 Processamento de imagens

Enquanto os trabalhos de Soares *et al.* (2013, 2014 e 2015) se concentram no uso das informações provenientes dos sensores, Broetto *et al.* (2013) se propõe a demonstrar como técnicas de mineração de dados combinadas com diferentes tipos de descritores de imagens (descritores de Fourier, descritores de Haralick e descritores de Hu) podem auxiliar na tarefa de classificar os tipos de lâmpadas de forma automatizada.

Para coleta de dados foram utilizadas duas câmeras, uma câmera profissional com alta taxa de zoom (Canon EOS Rebel T2i) e outra mais simples utilizada para aplicações visuais em computadores (Manta G-125B/C). Assim sendo, para realização dos experimentos foram criadas duas bases de dados, uma para cada câmera, com respectivamente 249 e 100 instâncias de duas classes distintas de luminárias.

Assim como os experimentos de Soares *et al.* (2013, 2014, 2015) foram utilizados três algoritmos do software Weka (J48, IB1 e *Multilayer Perceptron*). Entretanto, os experimentos foram divididos por base de dados e sete diferentes combinações de um, dois ou três dos grupos de descritores escolhidos. Os resultados dos melhores modelos em cada base de dados podem ser observados nas tabelas 4 e 5.

Tabela 4 - Acurácia média dos melhores classificadores dos testes com a base de dados da câmera Canon de Broetto *et al.* (2013).

Fourier	Haralick	Hu	Algoritmo	Acurácia Média
X			IB1	0.9919
	X		J48	0.9437
		X	J48	0.9116
X	X		IB1	0.9799
X		X	IB1	0.9839
	X	X	J48	0.9437
X	X	X	IB1	0.9759

Fonte: Elaborada pelo autor (2022).

Tabela 5 - Acurácia média dos melhores classificadores dos testes com a base de dados da câmera Manta de Broetto *et al.* (2013).

Fourier	Haralick	Hu	Algoritmo	Acurácia Média
X			IB1	1.0000
	X		<i>Multilayer Perceptron</i>	1.0000
		X	J48	0.9700
X	X		IB1	1.0000
X		X	IB1	1.0000
	X	X	<i>Multilayer Perceptron</i>	1.0000
X	X	X	IB1	1.0000

Fonte: Elaborada pelo autor (2022).

Os resultados obtidos foram descritos como satisfatórios, além de mostrar pouca alteração na mudança de câmera ou mesmo de algoritmo de classificação.

1.1.4 Seleção de características

Em Broetto e Varejão (2016), são comparados três diferentes algoritmos de seleção de características, utilizando dados coletados em campo por Soares *et al.* (2014, 2015). Neste, é descrito a aplicação de um *framework* de inteligência computacional que combina extração e seleção de características (*Features Extraction and Selection Framework* - FESF), no qual diferentes modelos de extração de variáveis podem representar informações diferentes e complementares sobre o problema, e algoritmos de seleção de variáveis eliminam informações extraídas que possam ser redundantes, inúteis ou ruidosas.

Ou seja, diferentes modelos de processamento de imagens foram aplicados nos dados visuais coletados em campo por Soares *et al.* (2014), descritores RGB e *Exchangeable Image File Format* (EXIF), juntamente com os dados fornecidos pelos sensores e descritores extraídos das imagens por Broetto *et al.* (2013), definiu-se um vetor de recursos que descrevem as

instâncias do banco de dados. Em seguida, o algoritmo de seleção de recursos reduz o tamanho deste vetor escolhendo um subconjunto que melhore a classificação do modelo.

Os resultados dos experimentos realizados mostraram que a aplicação deste *framework* melhorou a precisão da classificação do sistema. Foram utilizados três diferentes algoritmos de busca para seleção de características: *Sequential Forward Selection* (SFS), *Sequential Floating Forward Selection* (SFFS) e o *Greedy Randomized Adaptive Search Procedures* (GRASP). Dessa forma, gerou-se três instanciações de *framework* diferentes, as quais os resultados podem ser observados na Tabela 6.

Tabela 6 - Resultados dos experimentos de Broetto e Varejão (2016).

Fonte	Quantidade de variáveis	Informação	Acurácia
Hu	7	Textura	0.3636
Fourier	10	Forma	0.4209
Haralick	6	Textura	0.5354
EXIF	7	Técnica	0.5152
RGB	3	Cor	0.6296
Sensores	13	Sinal	0.7710
Todas	46	Diversas	0.8182
GRASP (Todas)	11	Diversas	0.8835
SFS (Todas)	15	Diversas	0.8989
SFFS (Todas)	9	Diversas	0.8956

Fonte: Adaptado de Broetto e Varejão (2016).

A partir destes resultados, conclui-se que o método SFS teve a melhor acurácia, ainda que SFFS tenha atingido um desempenho próximo com um custo computacional e conjunto de variáveis menor.

1.2 OBJETIVOS

Conforme descrito anteriormente, os trabalhos de Soares *et al.* (2013, 2014, 2015) utilizaram os dados provenientes dos sensores e Broetto *et al.* (2013) exclusivamente recursos extraídos de através de processamento de imagens. Dessa forma, acredita-se que há espaço para integrar as variáveis de ambas as fontes em novos modelos para otimizar as estruturas de classificação propostas, fazer uma exploração mais acentuada dos dados coletados em campo visando expandir, integrar e otimizar o trabalho iniciado por Soares *et al.* (2013) e Broetto *et al.* (2013).

1.2.1 Objetivo Geral

E, essa dissertação teve como objetivo geral aprimorar o estudo de detecção inteligente de lâmpadas de iluminação pública tornando a classificação mais precisa para minimizar o erro de classificação do tipo e potência. Os modelos foram construídos utilizando o banco de dados coletado em campo pelo trabalho de Soares *et al.* (2013, 2014, 2015).

1.2.2 Objetivos Específicos

O trabalho tem como objetivos específicos:

- (a) Propor novos modelos inteligentes que utilizam algoritmos de aprendizagem de máquina de abordagem tradicional de diversos níveis de complexidade;
- (b) Comparar a performance dos modelos usando uma metodologia experimental mais adequada;
- (c) Avaliar o desempenho de um modelo de aprendizado profundo.

1.3 CONTEXTUALIZAÇÃO

Com o crescimento contínuo da urbanização e da população que habitam os centros urbanos, a sociedade se estrutura de forma cada vez mais complexa, aumentando sua quantidade e formas de consumo, necessidades e dependências de serviços logísticos. Estas atividades são responsáveis pelo tratamento e abastecimento de água, comida e energia, sistema de transporte, tratamento de esgoto e outros processos que não são apenas essenciais ao mundo moderno como também estão inerentes ao funcionamento dele.

Dentre estes sistemas, tem-se a iluminação pública, responsável por eliminar a limitação do exercício de atividades em períodos noturnos por ausência de luz solar, promover comércio e turismo, fomentar sentimento de cidadania, além de estar diretamente ligada a áreas como segurança pública.

A iluminação já é utilizada pelo ser humano há milênios. Segundo Lopes (2002), para realizar atividades no período noturno, como exploração de cavernas e construção de abrigos, o homem já produzia e manejava o fogo para tal. Processo este que continuou a ser utilizado por mais alguns milênios de forma cada vez mais aprimorada, em conjunto com os avanços tecnológicos e utilização de combustíveis fósseis.

No século XIX, segundo Hémary (1993), foram criadas as redes de distribuição de gás, para a iluminação das cidades mais modernas de países industrializados. Nesse meio tempo, cidades do meio rural e países não industrializados ainda utilizavam o petróleo para tal.

Ainda no século XIX, através de estudos e experimentos com descargas elétricas, foi descoberto a possibilidade de utilizar eletricidade para produção de luz que conseqüentemente levou a criação da primeira lâmpada elétrica. Aos poucos, junto aos avanços tecnológicos nesta área, a iluminação elétrica passou a substituir a iluminação a combustível a nível comercial.

Com isso, permitiu-se que a população dos grandes centros urbanos pudesse realizar atividades durante o período noturno, algo antes inimaginável, agora se tornava uma demanda a ser assumida pelas autoridades governamentais.

Desde então, o sistema de iluminação pública passa por constantes mudanças e melhorias, com o advento das lâmpadas elétricas, estas em evolução até os dias atuais. Primeiramente, as lâmpadas incandescentes, as quais produzem luz esquentando o filamento de Tungstênio (DOE, 2012), seguidas por lâmpadas fluorescentes de baixa e alta pressão. Nestas, segundo Vito (2007), a luz é aumentada por uma descarga elétrica contendo um gás ou vapor interno, uma pequena quantidade de mercúrio é introduzida no tubo e um composto é usado para converter a luz ultravioleta em luz visível. As lâmpadas de vapor de sódio e vapor metálico funcionam de forma parecida, alterando apenas o elemento utilizado no equipamento.

Dado o constante avanço tecnológico, diversos modelos de lâmpadas foram criados, variando *design*, utilidade, potência e materiais utilizados. Assim, junto ao surgimento de outras tecnologias em paralelo, como a Internet das Coisas (*Internet of Things – IoT*), que permite a conexão de objetos à *internet*, diversas possibilidades para personalização, configuração e soluções para problemas de iluminação pública podem ser exploradas por múltiplas perspectivas.

Entretanto, a aparente infinidade de oportunidades vem com limitações, o alto custo de investimento necessário para aplicabilidade destas soluções em algo escala é um limitador constante ao considerar suas aplicações. Além disso, esta restrição só é agravada ao se considerar a natureza do investimento, que exige um grande esforço (mudança de infraestrutura, gerenciamento, processos, treinamento de funcionários, entre outros) e tem retorno financeiro apenas a longo prazo.

Outro fator bastante influente em estudos de caso de Iluminação Pública é a diversidade na metodologia utilizada pelas prefeituras e companhias. Ou seja, como cada instituição tem seu próprio modelo de gerenciamento, restrições e individualidades, é preciso levar estes em

consideração. Na próxima seção, estão descritos de forma breve alguns estudos desta área de pesquisa que refletem este raciocínio e contribuíram para o desenvolvimento desta pesquisa.

1.4 TRABALHOS RELACIONADOS

Além dos trabalhos de Soares *et al.* (2013, 2014 e 2015), Broetto *et al.* (2013) e Broetto e Varejão (2016), que contribuíram diretamente com os experimentos realizados neste trabalho, descritos em detalhes na Seção 1.1, outros trabalhos também influenciaram na abordagem deste estudo.

Dado a extensão do tema Iluminação Pública e sua amplitude de suas implicações, existem diversas abordagens dentro do meio científico. Boa parte tem como objetivo analisar o impacto no consumo de energia, custos de implementação e novas tecnologias. Uma das abordagens de melhoria frequentemente proposta é a otimização do gerenciamento da rede de iluminação pública.

Silva (2006), descreveu o sistema de iluminação pública no Brasil através de análises sobre seu alcance social, sua relação com outros índices de desenvolvimento local, potencial de melhoria na utilização de energia pela óptica de diferentes agentes do setor. Na qual concluiu uma forte relação entre indicadores como atendimento de serviço de iluminação pública e índices de atendimento de desenvolvimento local.

De forma similar, Santana (2010) avaliou o modelo de gerenciamento da companhia CITELUZ, que presta diversos serviços relacionados ao funcionamento da iluminação pública, e sua evolução sob aspectos legais e sociais após a reestruturação do Setor Elétrico. Assim, após a avaliação de diversos aspectos do método de gestão, concluiu-se que através de seu caráter otimizador de custos e controle nos setores administrativos, de engenharia, suprimentos, telemarketing, manutenção e construção de obras, este apresenta potencial para redução de custos e melhoria na qualidade dos processos e serviços de Iluminação Pública prestados pela companhia.

Radulovic, Skok e Kirincic (2011), analisaram o gerenciamento da iluminação pública em Rijeka, na Croácia, para determinar uma conexão entre a liberalização do mercado de energia e o desenvolvimento sustentável em áreas urbanas. Por fim, os resultados desta pesquisa indicaram uma forte conexão entre investimentos em gerenciamento energético de iluminação pública e sua influência positiva em indicadores como a emissão de dióxido de carbono.

Em Aguera (2015), foi apresentado um panorama claro e sucinto sobre o sistema de iluminação pública brasileiro, descrevendo diversas facetas do serviço no país, desde sua

história até seu funcionamento e custos. Este também aborda as vantagens e dificuldades que haviam de ser enfrentadas pelas prefeituras ao assumirem controle total sobre o gerenciamento da rede de iluminação pública após a mudança na legislação. Dentre estas dificuldades estão falta de conhecimento, técnico, dados, de infra estrutura, além dos diversos contratos do município com empresas terceirizadas que acarretam em maiores custos para gerenciar o setor, que conseqüentemente serão transferidos para a população.

Muitos autores também propõem novos métodos para otimização do gerenciamento e manutenção da rede de iluminação pública. Parmar (2019), propôs analisar como engenheiros e *designers* exploram a documentação e regulamentação de iluminação pública em Londres, e sua relação com as conseqüências criadas pela luz artificial da iluminação pública. O autor, através de questionários e entrevistas, comprovou não apenas a correlação entre os dois, mas também propôs uma metodologia de planejamento que não apenas busca otimizar a cooperação entre a equipe multidisciplinar necessária para desenvolver uma rede de iluminação pública e que facilitaria a detecção e mitigação de possíveis problemas de instalação em estágios iniciais de desenvolvimento.

Fardin, Marques e Có (2021), de forma análoga, desenvolveram um indicador de desempenho para avaliação de sistemas de iluminação pública (IDSIP) que leva em consideração não apenas o gasto de energia, mas também a qualidade de iluminação empregada em vias públicas. Após desenvolvimento e testes em Vitória, no estado do Espírito Santo, concluiu-se que uma melhor distribuição de pontos de iluminação junto a uma distribuição fotométrica adequada destes são mais eficientes no consumo de energia e nos benefícios providos pela iluminação pública no espaço urbano.

Oliveira *et al.* (2018), elaborou um modelo de gestão de manutenção de luminárias de iluminação pública para a cidade de Natal, no Rio Grande do Norte, que percorre desde o inventário do patrimônio público até o assessoramento especializado. Este método, apesar do investimento inicial elevado, seria capaz de gerar uma economia nos gastos com energia elétrica e aumento na vida útil das lâmpadas.

Outros estudos propõem mudanças ainda mais profundas, por meio da implementação de novas tecnologias. Maggi (2013), assim como diversas outras pesquisas, apresentou um estudo para projeto para substituição de luminárias tradicionais a base de vapor de sódio de alta pressão por luminárias de iluminação pública a base de LED (*Light-Emitting Diode*) mantendo os aspectos fotométricos exigidos pela legislação. Neste estudo, foi verificado que a substituição proposta pelo projeto pode garantir um nível de iluminação similar às lâmpadas tradicionais utilizando aproximadamente metade da potência elétrica.

Marino, Leccese e Pizzuti (2017), propuseram a implementação de um sistema inteligente de monitoramento preditivo e controle adaptativo de iluminação pública. Assim, coletaram informações sobre o tráfego de veículos por meio de uma câmera inteligente em uma rodovia, e utilizando métodos preditivos foi possível implementar uma estratégia de controle com base nas previsões de tráfego obtidas. A estratégia proposta mostrou um alto potencial de economia de energia sem afetar a segurança da população.

Nesse pensamento, Reinker (2019) propôs também uma conversão das luminárias de iluminação pública para lâmpadas a base de LED em *Rhode Island* nos Estados Unidos, porém de forma integrada a um sistema de transporte inteligente que permitiria que as luzes fossem ligadas apenas quando necessário, como por exemplo, tráfego de carros e pessoas durante a noite.

Por fim, Muse (2019), sumariza de forma clara as oportunidades criadas por novas tecnologias aplicáveis à Iluminação Pública, como LED e a IoT considerando aspectos sociais, econômicos e ambientais no cenário brasileiro com o objetivo de criar uma Matriz Multicritério para o auxílio em projetos que envolvam iluminação pública, de forma que estes possam atender as necessidades e interesses das populações.

Nota-se que apesar de todos os trabalhos visam a otimização de recursos, redução de custos, aumento da vida útil das luminárias e automatização de processos de manutenção, há uma diversidade muito grande de abordagens e contextos nos quais são elaborados. O que ressalta não apenas a amplitude do tema estudado e possibilidades de pesquisa, mas também sua complexidade.

1.5 JUSTIFICATIVA

Um denominador comum em praticamente todos os estudos observados é o alto investimento necessário para aplicar as soluções propostas. Recursos estes que não são exclusivamente referentes aos equipamentos, mas instalação destes, treinamento, mudança no gerenciamento da rede, entre outros. Assim, grande parte dos estudos são em pequena escala, mesmo em casos com soluções com elevado grau de comprovação como é o caso da economia proporcionada pelas lâmpadas à base de LED.

Portanto, através da aplicação de modelos com algoritmos de aprendizagem de máquina tradicionais, métodos de aprendizado profundo no problema proposto, juntamente com os ganhos informacionais e comerciais que estes podem trazer para o setor, espera-se que, assim

como os trabalhos dos autores citados anteriormente, esta pesquisa evidencie os ganhos que aplicações de novas tecnologias e métodos podem trazer.

Sob o ponto de vista comercial, uma das principais motivações da pesquisa, caso bem sucedido, o estudo pode resultar em uma solução mais barata, prática e eficiente que o método atual. Através da melhoria dos modelos propostos por Soares *et al.* (2015) será possível potencializar, com mais confiabilidade, os benefícios da integração de modelos de aprendizagem de máquina no processo de conferência, como: redução no desperdício de recursos, tempo, esforço, energia, gastos monetários e prejuízos, além de um aumento na acurácia das informações que compõe o banco de dados interno da companhia elétrica.

1.6 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está estruturada da seguinte forma: o primeiro capítulo é a introdução ao tema, também estão contidos os objetivos e a justificativa; o segundo capítulo descreve a revisão teórica para construção dos métodos, ferramentas e conceitos utilizados durante a pesquisa; no terceiro capítulo são abordados os materiais e métodos utilizados; no quarto capítulo são descritos e discutidos os resultados dos experimentos descritos na seção anterior e, por fim, no quinto capítulo são sumarizadas as conclusões do estudo.

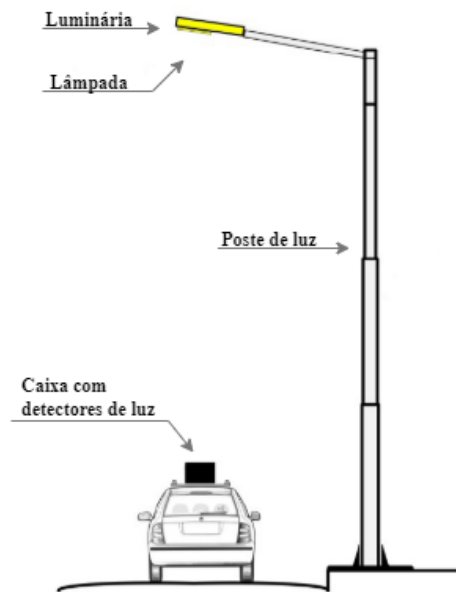
2 REFERENCIAL TEÓRICO

Esta seção tem por finalidade expor e analisar a bibliografia que fundamentam os conceitos e formulações utilizadas para a criação dos modelos de predição de tipo e potência de lâmpadas de iluminação públicas descritos nesta dissertação. Na Subseção 2.1 é descrito o sistema desenvolvido por Soares *et al.* (2013, 2014, 2015), o processo de coleta e armazenamento do banco de dados utilizado. Em seguida, na Subseção 2.2 são caracterizadas as técnicas de processamento utilizadas, e nas subseções 2.3 e 2.4 apresenta-se, respectivamente, uma noção de técnicas de reconhecimento de padrões, os diferentes algoritmos de aprendizagem de máquina e as técnicas de processamento de dados empregados na construção dos modelos propostos.

2.1 ARQUITETURA DO SISTEMA

O *hardware* que compõe o sistema proposto por Soares *et al.* (2013, 2014, 2015), pode ser descrito como um equipamento eletrônico que pode ser instalado em um veículo é capaz de coletar informação luminosa do ambiente (Figura 3). Os módulos eletrônicos que compõem o dispositivo são embutidos em um receptáculo projetado para acomodar corretamente todos os dispositivos eletrônicos. Este receptáculo oferece resistência à impactos, proteção contra umidade, poeira e especialmente a influência de outras fontes de luz que não estão sob avaliação no momento (como por exemplo, prevenir o reflexo indesejado de alguma radiação luminosa sobre os sensores).

Figura 3 - Representação do processo de aquisição de dados, incluindo veículo de transporte, dispositivo para coleta e equipamento de iluminação pública.



Fonte: Adaptado de Soares *et al.* (2014).

Durante a coleta, o veículo deve ser posicionado logo abaixo da luminária sob observação. Para isso, o motorista é auxiliado por uma série de sensores de luz dispostos de forma a fornecer as informações sobre o melhor posicionamento do equipamento.

A localização da lâmpada é dada imediatamente através de um dispositivo de GPS (*Global Positioning System*) que integra o sistema. Dada a natureza dos procedimentos de aquisição e processamento de dados, o sistema não informa imediatamente o tipo e potência da lâmpada observada. No entanto, os dados coletados são armazenados para serem processados posteriormente.

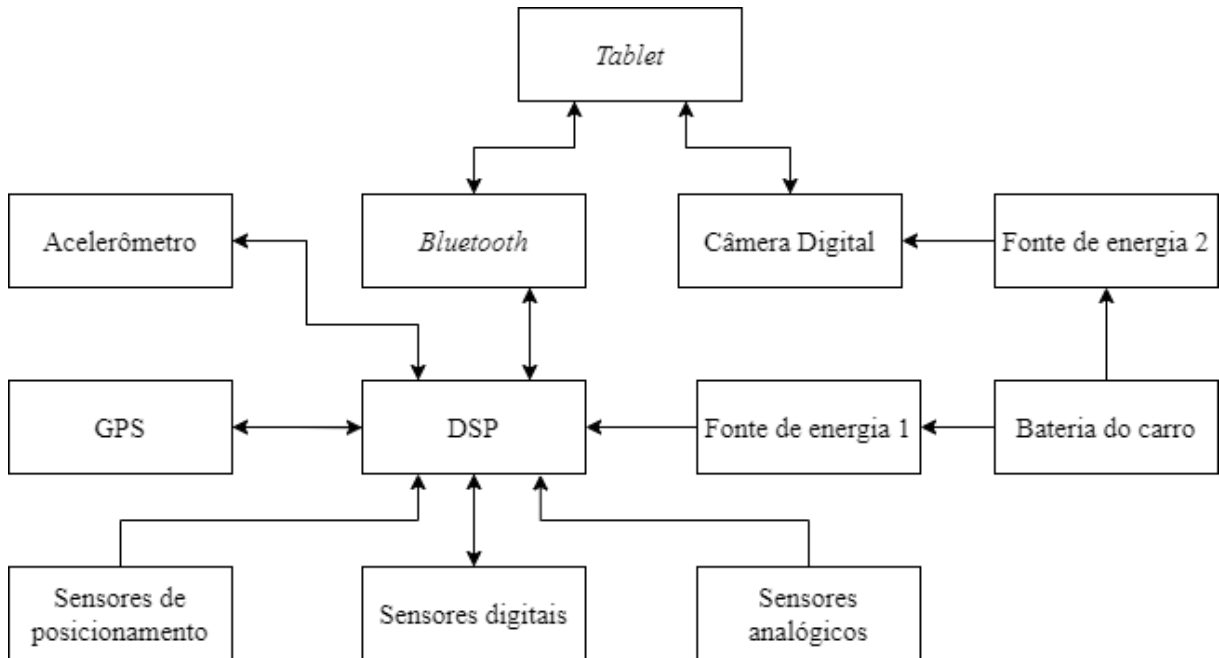
A rotina de pós-processamento é capaz de deduzir, por fim, o tipo e potência da lâmpada. Associado com os dados do GPS e uma foto da luminária (tirada por uma câmera digital que também compõe o equipamento), as informações sobre os postes de iluminação pública da área observada são dadas pelo sistema com boa confiabilidade.

2.1.1 Visão geral do *hardware*

O *hardware* utilizado para coleta dos dados consiste basicamente em sensores eletrônicos capazes de coletar diferentes informações sobre o equipamento de iluminação em

avaliação. A Figura 4 representa uma visão geral da arquitetura do equipamento, pensada para constituir os componentes necessários para alimentar o sistema computacional proposto.

Figura 4 - Arquitetura do equipamento de coleta de dados.



Fonte: Adaptado de Soares *et al.* (2014).

Os principais componentes de coleta de dados são um conjunto de sensores de luz analógicos e digitais (de baixo custo e pronto para uso), com diferentes respostas espectrais. A Tabela 7 resume as principais características fotométricas dos sensores escolhidos.

Além destes sensores principais, outros transdutores foram utilizados, como sensores de posicionamento (sensores de iluminação que informam o motorista do carro se o carro está na posição adequada abaixo da luminária), um acelerômetro (que calcula a inclinação do solo) e um dispositivo de GPS (responsável por informar a posição geográfica da luminária).

Tabela 7 - Características fotométricas dos sensores radiométricos.

Sensor	Comprimento de onda de sensibilidade máxima	Alcance espectral de sensibilidade
1	560 nm	460 nm ... 660 nm
2	810 nm	460 nm ... 1100 nm
3	940 nm	800 nm ... 1200 nm
4	570 nm	350 nm ... 1100 nm
5	570 nm	400 nm ... 900 nm
6	880 nm	730 nm ... 1100 nm
7	640 nm	300 nm ... 1150 nm
8	650 nm	300 nm ... 1100 nm
9	800 nm	500 nm ... 1100 nm
10	480 nm	380 nm ... 570 nm
11	540 nm	420 nm ... 630 nm
12	620 nm	580 nm ... 680 nm
13	670 nm	380 nm ... 680 nm

Fonte: Elaborada pelo autor (2021).

Todos os sensores são gerenciados por um processador de sinal digital (*Digital Signal Processor - DSP*), que coleta suas informações e as envia via *Bluetooth* para um sistema de armazenamento de dados externo alocado dentro do veículo, como um computador *tablet*. Uma câmera digital é colocada dentro do receptáculo do sistema, assim como da placa principal. Dessa forma, o equipamento consegue fornecer tanto imagens dos pontos de iluminação observados como dados quantitativos a respeito da luz que emitem.

2.1.2 Detecção da assinatura de espectro

Conforme comentado por Keefe (1980), a radiação emitida por um tipo de lâmpada específica é composta por diferentes linhas de espectro, ou seja, a composição química de cada lâmpada produz uma assinatura de espectro. Portanto, desde que as lâmpadas sob análise são quimicamente diferentes umas das outras, é possível determinar o tipo de lâmpada através do estudo e análise de seu espectro.

Para que esse processo de classificação pudesse ser feito, um conjunto de sensores radiométricos foram utilizados. A escolha de cada sensor foi baseada nas características de emissão dos tipos de lâmpada observados por Soares *et al.* (2013, 2014, 2015): MV, HPS e MH.

Basicamente, os sensores ópticos são transdutores que convertem a radiação sensibilizante (ponderada por sua curva de sensibilidade) em corrente elétrica (também chamada de fotocorrente). Assim, a saída do sensor óptico pode ser definida pela Equação 1.

$$M = K \int_0^{\infty} J(\lambda) S(\lambda) d\lambda \quad (1)$$

Onde K é a constante da proporcionalidade, $J(\lambda)$ é a distribuição espectral da lâmpada e $S(\lambda)$ é a sensibilidade do sensor. Logo, contanto que cada lâmpada tenha sua assinatura espectral e os sensores escolhidos tenham diferentes características de sensibilidade diferentes, os tipos de lâmpada podem ser deduzidos através do estudo das saídas dos sensores.

Para auxiliar no processo de escolha dos sensores, Soares *et al.* (2013, 2014, 2015) também realizaram medições radiométricas em uma esfera integradora para analisar os espectros de radiação emitidos pelos tipos de lâmpadas. Nestes experimentos, foram utilizados dois sensores: o sensor A que se mostrou mais sensível às radiações azul e verde e o sensor B, mais sensível à radiação vermelha e infravermelha. Dado os resultados particulares de cada sensor, notou-se que A é mais sensível à MV e MH, e B é mais sensível para a radiação de HPS. Portanto, são sensores que carregam informações distintas que podem ser utilizadas para identificar o tipo de lâmpada.

Conforme mencionado anteriormente, os sensores A e B são parte do grupo de dispositivos utilizados na coleta de informações sobre as lâmpadas. A escolha dos demais sensores foi feita de forma que a sensibilidade máxima de cada dispositivo esteja em um comprimento de onda diferente, estratégia que aumenta a capacidade de detecção do sistema proposto.

Outro fator que pode influenciar na resposta dos sensores é a magnitude das linhas de espectro, estas que são responsáveis pela informação da potência das lâmpadas. De qualquer forma, a magnitude é dependente também de outras variáveis como a altura do poste, ângulo de registro das informações, temperatura ambiente, vida útil da lâmpada e variações ocorridas na tensão de alimentação.

Dados os diversos fatores que podem afetar a coleta desta informação, a tarefa de reconhecer a potência das lâmpadas requer estratégias mais complexas. Além disso, a previsão teórica da influência de cada parâmetro supracitado na magnitude das linhas espectrais é uma

tarefa difícil, uma vez que todas essas variáveis são correlacionadas. Felizmente, técnicas de reconhecimento de padrões podem ser utilizadas para lidar com este perfil de problema.

2.2 PROCESSAMENTO DE IMAGENS

As tecnologias de visão computacional têm como principal objetivo reproduzir a eficácia da visão humana em tarefas como detecção e classificação de objetos e ambientes, organização de informações, entre outras. Entretanto, esta é uma tarefa complexa, visto que o mundo tem três dimensões e a grande maioria das ferramentas de imagem fornecem apenas representações em duas dimensões (SONKA; HLAVAC; BOYLE, 2014).

O processamento digital de imagens pode ser definido, segundo Pedrini (2021), como um conjunto de técnicas para capturar, representar e transformar imagens com o auxílio de um computador. Assim, a forma como uma máquina compreende uma figura pode ser descrita, conforme Silva (2017), como a relação entre uma imagem de entrada, e o modelo previamente estabelecido na observação do mundo real.

A passagem da imagem para o modelo reduz a quantidade de informação para fabricar informações pertinentes. Este procedimento de processamento da imagem é dividido em diversas etapas e vários níveis de representação (iniciando-se na imagem original e finalizando na imagem interpretada. Portanto, conforme Sonka, Hlavac e Boyle (2014), a visão computacional através da criação de algoritmos define a relação entre os níveis da imagem.

Dessa forma, para que a imagem se torne apropriada e propícia para uma prática específica, é imprescindível que ela passe por métodos de aprimoramento de imagem. Este processo permite que a imagem esteja mais adequada às subseqüentes técnicas de processamento, como por exemplo melhor visualização de características ou restaurar imagens deterioradas (CHAIRA, 2015).

Uma vez que a imagem original foi processada e está pronta para ser computada de forma eficaz, é iniciado o processo de descrição da imagem. Esse método, também conhecido como “extração de características”, tem como objetivo extrair dados que irão resultar em informações relevantes e de qualidade que irão ser importantes para a finalidade estabelecida (como por exemplo, a discriminação entre classes e objetos).

Uma das técnicas tipicamente utilizadas neste processo são os descritores, que utilizam métodos matemáticos para representar a forma de um objeto ou região através de um vetor de características. Nas subseções seguintes são apresentados os três descritores de imagens

utilizados em Broetto *et al.* (2013), Broetto e Varejão (2016) e também nos experimentos realizados nesta dissertação (descritores de Fourier, Hu e Haralick).

2.2.1 Descritores de Fourier

Amplamente conhecida e utilizada para processamento de sinais (PAPOULIS, 1962), a Transformada de Fourier é uma operação matemática que muda os sinais do domínio do tempo (função temporal) para o da frequência. Porém, a aplicação desta técnica não é limitada a processamento de sinais digitais, e pode ser utilizada em aplicações como análise, filtragem, reconstrução e compressão de imagens.

De acordo com Fisher *et al.* (2004), é uma importante ferramenta de processamento de imagens e pode ser utilizada para decompor a imagem em seus componentes de seno e cosseno. Através desta transformação a imagem é remodelada do domínio espacial para o da frequência, no qual cada frequência é a representação particular de um ponto no domínio espacial.

Com isso, os pixels do contorno de um objeto podem ser representados como um sinal. Embora o número de coeficientes gerados pela transformação seja grande, um subconjunto muito menor destes coeficientes é suficiente para capturar seus contornos únicos.

Na Transformada Discreta de Fourier (*Discrete Fourier Transform - DFT*), tipicamente utilizada para análise de espectro de sinais, dadas N amostras de um sinal no domínio do tempo, denotadas por $f(k)$, $k = 0, 1, 2, \dots, N-1$; a DFT é dada por um conjunto de N amostras de sinais no domínio da frequência, descritas por $f(n)$, $n = 0, 1, 2, \dots, N-1$, e definidas pela Equação 2.

$$F(n) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) e^{-j2\pi nk/N} \quad (2)$$

Assim, cada $f(k)$ é um número complexo que codifica a amplitude e a fase de um componente senoidal complexo ($e^{-j2\pi nk/N}$) da função $F(n)$. Os coeficientes que resultam desta metodologia formam os descritores do contorno do objeto em seu domínio de frequência. Além disso, enquanto os descritores de baixa frequência contêm informações sobre as características gerais do contorno, os com alta frequência contêm informações sobre detalhes da forma do objeto.

Embora o número de coeficientes gerados de transformação seja geralmente grande, um subconjunto muito menor desses coeficientes é suficiente para capturar os contornos únicos da imagem.

2.2.2 Descritores de Hu

Outro método usado para extração de características de uma imagem são os momentos da imagem, ou momentos estatísticos. Esses momentos e suas funções podem ser caracterizados como valores numéricos calculados a partir de uma imagem pré-segmentada para fornecer breves descrições da distribuição espacial dos pontos contidos em uma imagem em duas dimensões. A fórmula geral é dada pela Equação 3:

$$M_{ij} = \sum_x \sum_y x^i y^j I(x, y) \quad (3)$$

Por outro lado, a definição geral do momento da imagem digital possui poucas informações importantes sobre um conjunto de imagens, uma vez que é sensível à escala, rotação e translação. No entanto, é possível realizar algumas transformações matemáticas nesta fórmula para garantir a invariância nesses aspectos.

A primeira transformação assegura invariância sobre a translação e gera o chamado momento central da imagem, definido pela Equação 4.

$$\mu_{ij} = \sum_x \sum_y (x - x^i)^i (y - y^j)^j I(x, y) \quad (4)$$

A definição feita anteriormente requer computação do centroide do objeto. A partir da definição do momento central da imagem pode-se obter a definição de momentos invariantes em escala, mostrado na Equação 5.

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00} \left(1 + \frac{i+j}{2}\right)} \quad (5)$$

Por fim, um conjunto de sete momentos invariantes são utilizados para descrever a imagem. Este conjunto de momentos invariantes, mostrados nas equações 6, 7, 8, 9, 10, 11 e 12, são conhecidos como o conjunto de momentos invariantes de Hu (HU, 1962).

$$I_1 = \eta_{20} + \eta_{02} \quad (6)$$

$$I_2 = (\eta_{20} - \eta_{02}) + (2\eta_{11}) \quad (7)$$

$$I_3 = (\eta_{30} - 3\eta_{12}) + (3\eta_{21} - \eta_{03}) \quad (8)$$

$$I_4 = (\eta_{30} - \eta_{12}) + (\eta_{21} + \eta_{03}) \quad (9)$$

$$I_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (10)$$

$$I_6 = (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (11)$$

$$I_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (12)$$

A maior desvantagem deste modelo é que não há garantias de que os sete momentos formem um conjunto pleno de descritores para os padrões de entrada (KULKARNI; 1994). Contudo, de acordo com Huang e Leng (2010), no caso de imagens com pouco ruído, os momentos invariantes não se desorganizam em presença de variações de translação, rotação e escala.

2.2.3 Descritores de Haralick

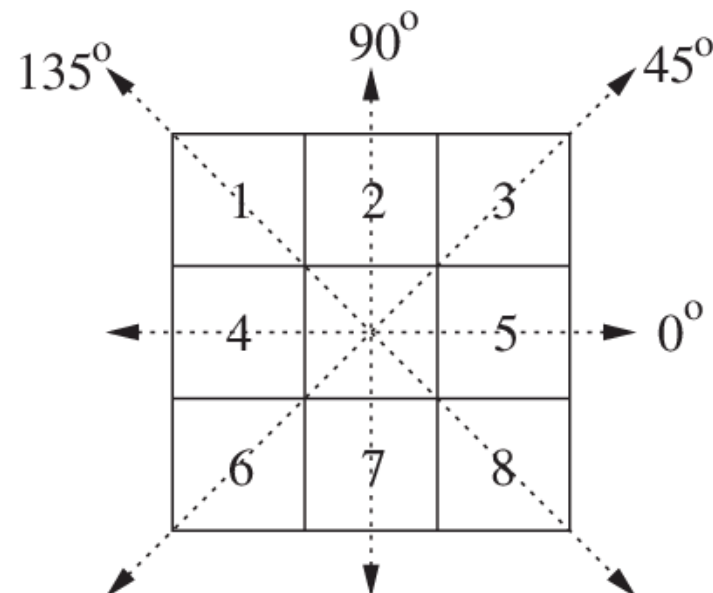
Segundo Haralick, Shanmugam e Dinstein (1973), os descritores de texturas de Haralick usam uma abordagem estatística para descrever as texturas da imagem com base na distribuição e relacionamento da escala de cinza da imagem. Esta abordagem extrai informações úteis

(uniformidade, densidade, aspereza, regularidade, intensidade, entre outras) através da chamada Matriz de Co-ocorrência de Nível de Cinza (*Grey Level Co-occurrence Matrix – GLCM*).

A GLCM é uma tabulação do número de diferentes combinações de valores de intensidade ocorrendo em uma imagem. Toda informação é representada na forma de uma matriz de frequências relativas $P(i, j, d, D)$, em que dois componentes de textura adjuntos (pixels), separados por uma distância d na direção D , com níveis de cinza i e j .

Assim, cada termo a_{ij} é representado pela matriz o número de vezes em que ocorreu alguma transição no nível de cinza entre os pixels i e j considerando a distância d e a direção D . A sequência de transição dos ângulos é feita obrigatoriamente em relação ao componente central, conforme mostra a Figura 5.

Figura 5 - Ângulos utilizados para o cálculo de matrizes de co-ocorrência.



Fonte: SCHWARTZ (2005).

Estas matrizes, sempre simétricas, são normalizadas através da divisão de cada um de seus elementos pelo número total de pixels vizinhos utilizados em sua computação.

Conforme descrito em trabalhos como Nascimento (2003), Weszka, Dyers e Rosenfeld (1976), Connors e Harlow (1980), após a elaboração da GLCM, utilizando-se de uma abordagem estatística pode-se definir a relação e distribuição da escala de cinza da imagem para descrever suas texturas e extrair características proveitosas.

Dessa maneira, o modelo original proposto por Haralick, Shanmugam e Dinstein (1973) é composto por um conjunto de 14 descritores (Segundo Momento Angular, Contraste, Correlação, Soma dos Quadrados (Variância), Momento de Diferença Inverso, Média da Soma;

Variância da Soma, Entropia da Soma, Entropia, Variância da Diferença, Entropia da Diferença, Medidas de Informação da Correlação (2) e Coeficiente de Correlação Máximo).

Nesta dissertação, assim como Chikamai *et al.* (2013), foram utilizados um conjunto de seis descritores de Haralick, mostrados nas equações 13, 14, 15, 16, 17 e 18. Cinco destes descritores (Contraste, Correlação, Energia, Entropia e Homogeneidade) são um subconjunto do modelo original proposto por Haralick, Shanmugam e Dinstein (1973). O sexto descritor é a Máxima Probabilidade, também proposto por Haralick (1979) em um trabalho posterior.

$$\text{Contraste} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i-j)^2 P(i, j, d, D) \quad (13)$$

$$\text{Correlação} = \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} i j P(i, j, d, D) - \mu_x \mu_y}{\delta_x \delta_y}, \text{ sendo:} \quad (14)$$

$$\mu_x = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} i P(i, j, d, D); \quad \mu_y = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} j P(i, j, d, D);$$

$$\delta_x = \sqrt{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} i^2 P(i, j, d, D) - \mu_x^2}; \quad \delta_y = \sqrt{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} j^2 P(i, j, d, D) - \mu_y^2}$$

$$\text{Energia} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} [P(i, j, d, D)]^2 \quad (15)$$

$$\text{Entropia} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} P(i, j, d, D) \log_2 [P(i, j, d, D)] \quad (16)$$

$$\text{Homogeneidade} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \frac{1}{1 + (i-j)^2} P(i, j, d, D) \quad (17)$$

$$\text{Máxima probabilidade} = \text{MAX}[P(i, j, d, D)] \quad (18)$$

Cada um dos descritores escolhidos representam uma informação específica sobre a textura da imagem, conforme descritos abaixo:

- a) Contraste: Variação dos níveis de cinza na imagem;
- b) Correlação: Grau de correspondência entre os pixels e seus vizinhos;

- c) Energia: Uniformidade da textura;
- d) Entropia: Desordem contida na textura;
- e) Homogeneidade: Distribuição dos pixels, em específico, a proximidade da distribuição dos elementos em relação à diagonal da GLCM;
- f) Máxima Probabilidade: Direção mais importante da textura a ser examinada.

Tal qual modelo de descritores de Hu, os descritores de Haralick extraem informações de textura de um objeto. Estes descritores em particular fornecem uma informação intuitiva, que pode ser muito útil quando análises são realizadas a fim de obter características marcantes de imagens de postes com luminárias públicas.

2.3 TÉCNICAS DE RECONHECIMENTO DE PADRÕES

Devido à dificuldade de inferir teoricamente as classes de lâmpadas, com base nos dados das informações radiométricas e de imagem, o processo de classificar o tipo de lâmpada e potência, a melhor alternativa para este processo é abordá-lo como um problema de reconhecimento de padrões.

Reconhecimento de padrões é a área de pesquisa que tem por objetivo a classificação de objetos (padrões) em um número de classes (THEODORIDIS, 1999). Segundo Huspi (2019), em ciência da computação e aprendizagem de máquina, termo também pode ser aplicado a uma tecnologia capaz de combinar de forma sistemática informações armazenadas em um banco de dados com informações de *input* e categorizá-las em classes. Além do mais, problemas de categorização compõe grande parte dos problemas de aprendizagem de máquina.

O aprendizado de máquina tem como seu principal foco duas questões correlatas: Como um sistema computacional pode ser aprimorado automaticamente através de experiências e quais são as leis estatísticas e computacionais que governam os conhecimentos de aprendizado, incluindo computadores, humanos e organizações (JORDAN; MITCHELL, 2015). E no cerne destes modelos computacionais estão os algoritmos, estes que irão definir as regras, como o modelo irá processar os dados para extrair respostas automáticas para novos pedidos.

Dada a amplitude das duas questões, assim como a vastidão de problemas que podem ser solucionados por técnicas de aprendizagem de máquina, existem diversas maneiras e métodos, algoritmos, que podem ser utilizados para cada tipo de problema e suas individualidades. Este trabalho em particular, é focado em um problema de classificação e, consequentemente, em algoritmos de classificação.

Um classificador pode ser compreendido como função matemática que mapeia os dados de entrada para uma determinada categoria. Algoritmos de aprendizagem de máquina implementam classificadores analisando um conjunto de instâncias de entrada de dados cuja associação de categoria é conhecida (KOTSIANTIS, 2007).

Dentre os diversos algoritmos de classificação existentes, foram selecionados para os experimentos descritos nesta dissertação seis algoritmos de abordagem tradicional (K-Vizinhos mais próximos, Árvore de Decisão, Floresta Aleatória, XGBoost, Máquina de Vetores de Suporte e *Perceptron* Multicamadas) e um algoritmo de aprendizado profundo (Rede Neural Convolutiva).

Os critérios para seleção de algoritmos foram a diversidade de abordagem e complexidade. Assim, tem-se uma seleção de algoritmos bem estabelecidos com diferentes atuações para performance de classificação: K-Vizinhos mais próximos utiliza distância entre instâncias, Árvore de Decisão, Floresta Aleatória e XGBoost utilizam árvores, Máquina de Vetores de Suporte utiliza hiperplanos e *Perceptron* Multicamadas utiliza redes neurais.

Logo, será possível construir desde modelos simples (também conhecidos como “*lazy learners*”) usando o K-Vizinhos mais próximos até modelos mais complexos como uma rede neural convolutiva. Dessa forma, espera-se que esta gama de variações ofereça *insights* sobre quais abordagens têm melhor desempenho no problema estudado. Nas subseções seguintes os algoritmos selecionados são apresentados em detalhes.

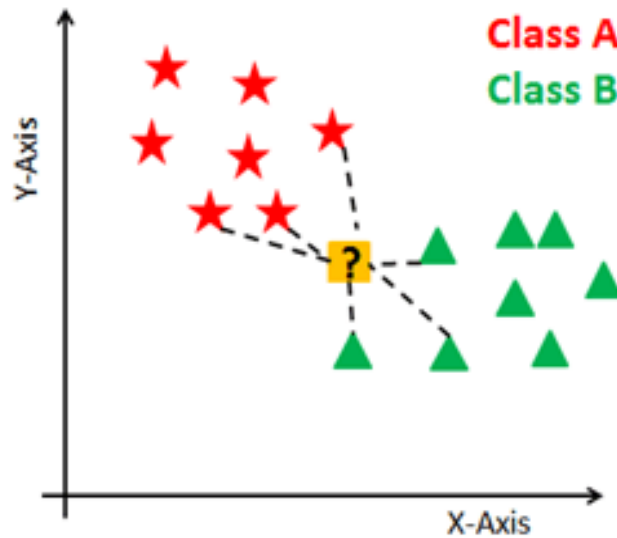
2.3.1 K-Vizinhos mais próximos

Lazy learners, ou algoritmos de aprendizado tardio, podem ser definidos como métodos de aprendizagem que apenas realizam o processamento dos dados de treinamento quando há uma necessidade de consulta a ser respondida (WITTEN; FRANK; HALL, 2011). Geralmente, os dados são armazenados na memória para encontrar informações que são relevantes na base para responder uma demanda particular. Um dos *lazy learners* mais utilizados é o K-vizinhos mais próximos (*K-nearest neighbors* - KNN).

O algoritmo KNN é um método não paramétrico de classificação de objetos, criado por Evelyn e Hodges (1951), que se baseia nos exemplos de treinamentos mais próximos no espaço de variáveis. Nesse método, uma instância é classificada pela maioria de “votos” de seus vizinhos, ou seja, o objeto é classificado de acordo com a classe mais comum nos k vizinhos mais próximos (k é um número positivo, geralmente pequeno). A forma mais comum de identificação dos vizinhos próximos é utilizando a distância euclidiana entre os objetos a serem

classificados e seus vizinhos. A Figura 6 mostra um exemplo simples de algoritmo de classificação k-vizinhos mais próximos.

Figura 6 - Exemplo de um modelo de k-vizinhos mais próximos para classificação binária.



Fonte: NAVIANI (2018).

A performance deste algoritmo está diretamente relacionada com o número k de vizinhos determinado para a execução do mesmo, caso um número inadequado seja selecionado para k pode acarretar na ocorrência de *overfitting/underfitting*. Segundo Panesar (2019), *overfitting* (sobreajuste) ocorre quando o modelo tem bons resultados nos dados de treinamento, entretanto é menos preciso com padrões desconhecidos, já o *underfitting* (subajuste) acontece quando o modelo não é capaz de capturar os sinais dos dados e obtém uma baixa performance em ambos os conjuntos de treino e validação. Entretanto, uma das desvantagens deste algoritmo, é o alto custo computacional para processamento de dados com alta dimensionalidade.

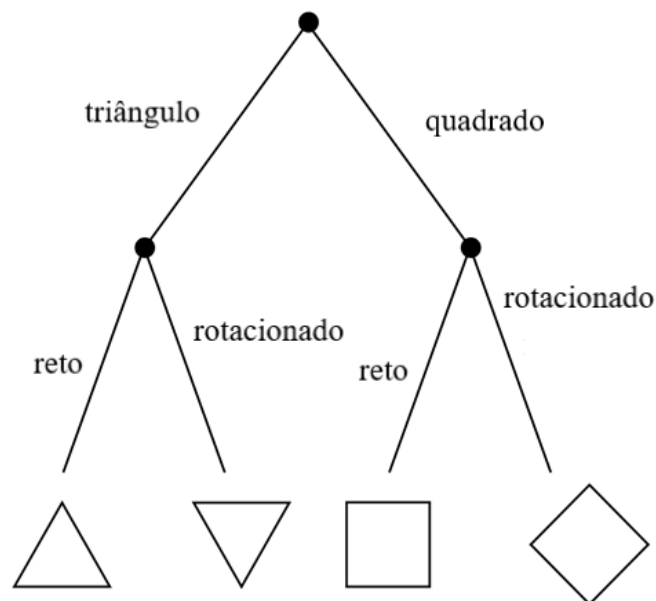
2.3.2 Árvore de decisão

Árvore de decisão (*Decision Tree - DT*) é um algoritmo de aprendizado de máquina supervisionado amplamente utilizado, devido sua estrutura de simples compreensão e bom desempenho, além de servir como base para algoritmos mais complexos como, por exemplo, o Floresta Aleatória.

Uma árvore de decisão é construída através de uma abordagem algorítmica que identifica formas de dividir o conjunto de dados de forma sistemática. Ou seja, todo conjunto de instâncias observadas é dividido em dois ou mais subconjuntos. E as divisões subsequentes são feitas por nós de decisão, que irão dividir sub-nós em novos sub-nós. E por fim, o nó terminal, o último nó da árvore, contém o valor previsto de classificação (AYYADEVARA, 2018).

Na Figura 7 é possível observar o fluxo do processo de decisão do algoritmo em um problema de classificação supervisionado de figuras geométricas. Nela, temos a raiz (também conhecida como nó inicial) que divide os dados entre figuras em formato triangular e quadrangular. Em seguida, tem-se um sub-nó (também chamado de nó de decisão) que faz a distinção de figuras retas e rotacionadas. Por fim, tem-se as folhas que tem nós chegando até elas, mas não partindo delas, o que indica o fim da árvore e que se atingiu a classificação final. Ou seja, no exemplo discutido as folhas indicam a classificação da instância em uma das quatro possíveis classes (quadrado reto, quadrado rotacionado, triângulo reto e triângulo rotacionado) conforme as divisões realizadas pelos nós anteriores.

Figura 7 - Exemplo de Árvore de decisão, demonstrando as etapas do processo de classificação de um modelo de árvore de decisão.



Fonte: Adaptado de Plenio e Vitelli (2001).

Dessa maneira, nota-se que a estrutura do algoritmo é bastante intuitiva e similar à de um fluxograma, na qual o fluxo se inicia na raiz e se ramifica por um ou mais níveis até que mais nenhuma repartição possa ser estabelecida. Portanto, o aprendizado e performance de

modelos de DT estão diretamente ligados à sua profundidade, entretanto, segundo Rokach e Maimon (2015), árvores muito profundas podem memorizar todos os dados de treinamento gerando o problema de *overfitting/underfitting*.

2.3.3 Floresta Aleatória

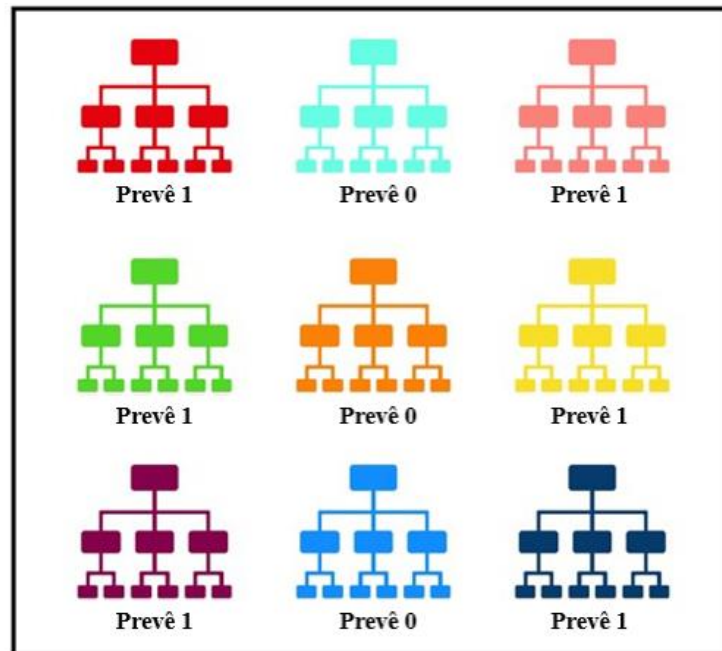
A grande maioria dos problemas relacionados a Aprendizado de Máquina podem ser divididos em dois grupos, o de classificação (no qual se prediz a que classe certa observação pertence) e regressão (no qual se prediz o valor de certa instância), e um dos modelos utilizados em ambos os casos é o Floresta Aleatória, em diversas aplicações.

O Floresta Aleatória (*Random Forest* - RF) foi criado por Leo Breiman (2001) e é definido por ele como um algoritmo, passível de ser utilizado tanto para regressão quanto classificação, que usa o método ensemble, adjeção de aspectos de modelos construídos com o objetivo de criar um de melhor desempenho, em suas camadas finais em árvore.

Basicamente, o algoritmo consiste na criação de árvores de decisão individuais, no qual cada uma é cultivada de acordo com um parâmetro aleatório (BHARATHIDASON; VENKATAESWARAN, 2014).

Neste modelo, um subconjunto de recursos preditivos é considerado durante cada divisão selecionada aleatoriamente, assim, a decisão das árvores leva a uma previsão conjunta, calculando a média de todas as previsões individuais (BREIMAN, 2001). Logo, o resultado mais frequente se torna a previsão do modelo, conforme demonstrado na Figura 8.

Figura 8 - Visualização do processo de decisão do modelo Floresta Aleatória.



CONTAGEM: Seis 1's e 3 0's.
PREDIÇÃO: 1.

Fonte: Adaptado de Yiu (2019).

Legenda: No grupo de 9 árvores, 6 previram o resultado 1 e 3 previram o resultado 0. Com isso, a predição final do modelo é 1.

A maior vantagem do algoritmo floresta aleatória é o fato das árvores protegerem-se de seus erros individuais. Por cada árvore ser um modelo individual, algumas podem estar certas e outras incorretas em suas previsões, porém como um modelo ensemble e prevendo como um grupo, é possível obter um rendimento melhor que com suas previsões individuais.

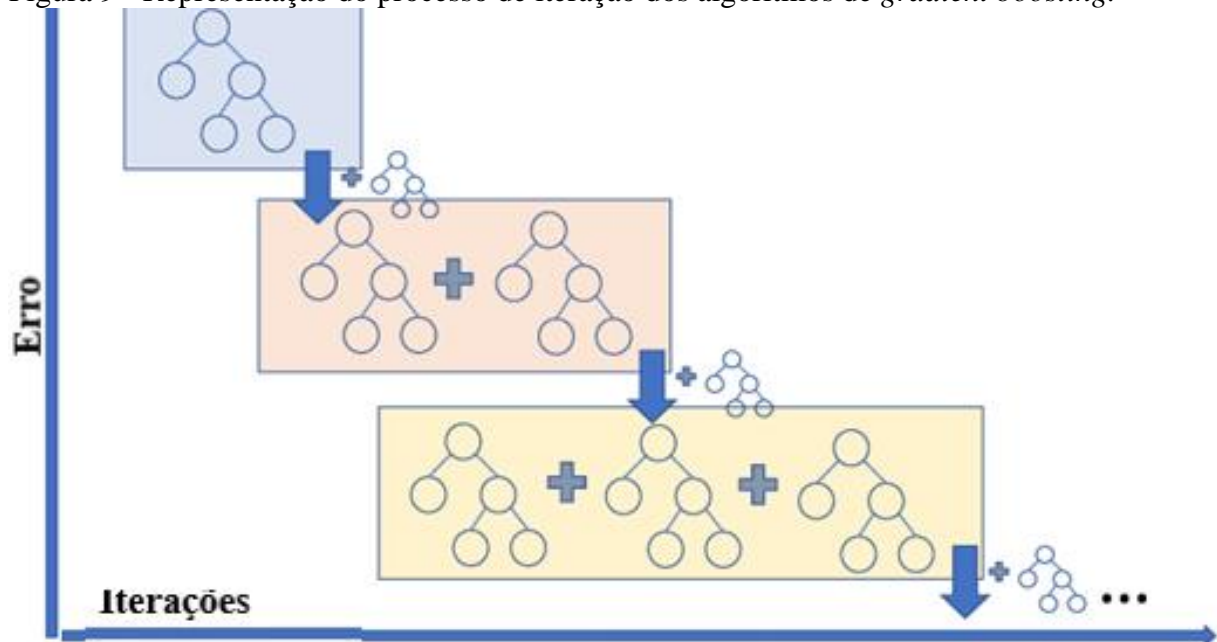
Dentre alguns dos diversos meios de aplicação deste método de aprendizagem de máquina, tem-se: astronomia (GAO *et al.*, 2009), ecologia (CUTLER *et al.*, 2007), agricultura (LIMA *et al.*; 2022) e biotecnologia (BOULESTREIX *et al.*, 2012). Fazendo deste modelo um excelente ponto de partida para o problema proposto.

2.3.4 XGBoost

Métodos *ensemble* muitas vezes são usados para reduzir viés (erro decorrente da diferença entre as previsões médias e os valores reais a serem preditos) nos erros de predição do modelo, porém há também métodos *ensemble* como *boosting*, que criam membros *ensemble* sequencialmente ao invés de paralelamente, como realizado no modelo Floresta Aleatória. Dessa forma, os modelos mais novos com os erros de previsão dos elementos anteriores.

Já o *gradient boosting* é uma abordagem na qual novos modelos são treinados para prever os resíduos dos modelos anteriores. Mais especificamente o algoritmo de *boosting* executa iterações n vezes para aprender a função que faz previsões minimizando a função de perda, a cada iteração é adicionado um novo estimador para tentar corrigir a predição de cada instância (MITCHELL; FRANK, 2017), conforme pode ser visualizado na Figura 9.

Figura 9 - Representação do processo de iteração dos algoritmos de *gradient boosting*.



Fonte: Adaptado de Baturynska e Martinsen (2021).

O XGBoost (*Extreme Gradient Boost - XGB*), introduzido por Friedman (2001) também dá nome a uma biblioteca de código aberto que permite combinar o *gradient boosting* e o Floresta Aleatória, sendo assim, mais complexo e superior a diversos modelos de classificação. O modelo foi utilizado em diversas soluções vencedoras em portais de competições de Ciência de Dados, como Kaggle (17 de 29 soluções vencedoras em 2015) e KDDCup, as 10 melhores soluções em 2015 usaram XGBoost (BEKKERMAN, 2015).

Apesar das similaridades com o método Floresta Aleatória, suas diferenças distinguem muito o processo de classificação principalmente em como as árvores são construídas e como os resultados são combinados.

Conforme Glen (2019), o fato de o XGBoost construir as árvores de forma sequenciada e não paralela como o Floresta Aleatória permite que a adição de novas árvores melhore as deficiências presentes nas já existentes. Nesta lógica, os resultados são combinados durante o processo de criação das árvores e não no final do processo (por “regra da maioria”). Portanto,

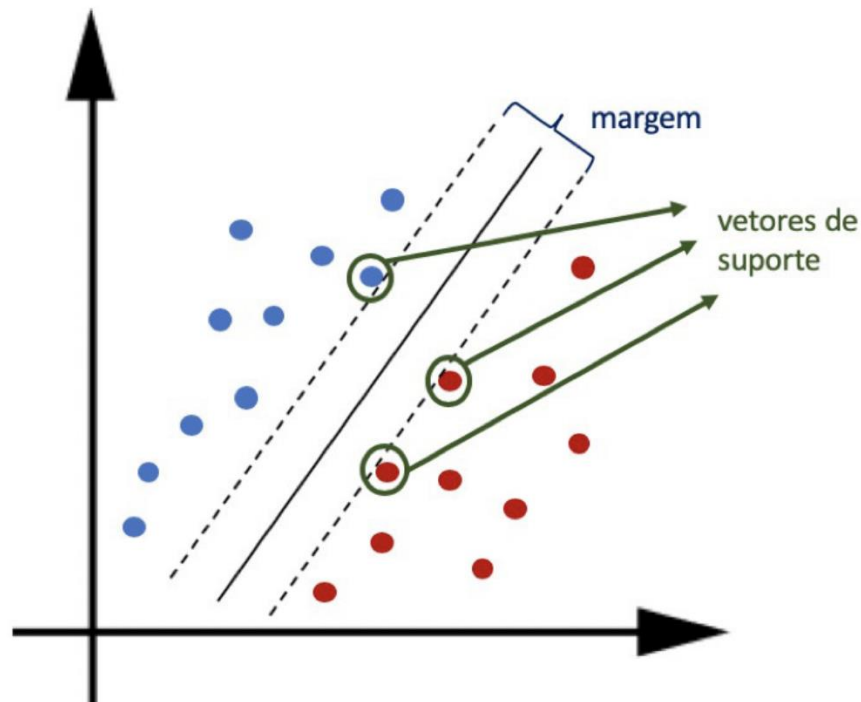
apesar de todas as vantagens, o método é mais complexo e pode causar *overfitting* em bancos de dados com muitos “ruídos”.

2.3.5 Máquina de vetores de suporte

Máquina de vetores de suporte (*Support vector machine – SVM*) é um algoritmo de aprendizado supervisionado proposto por Cortes e Vapnik (1995), que pode ser utilizado para regressão e classificação, e tem como objetivo encontrar um hiperplano que melhor separe as variáveis em diferentes domínios. De acordo com Yadav (2018), o hiperplano é uma função utilizada para diferenciar características.

De forma similar ao algoritmo KNN, este algoritmo se baseia nas métricas das instâncias do banco de dados para realizar previsões. Basicamente, conforme Meyer (2017), o algoritmo define um hiperplano entre classes maximizando a margem de distâncias entre os pontos mais próximos das mesmas. Os pontos que se encontram no limiar dessa divisão são chamados de vetores de suporte, e o meio dessa margem é o hiperplano ótimo que será usado para classificação.

Figura 10 - Representação de um algoritmo de máquina de vetores de suporte.



Fonte: Escovedo e Koshiyama (2020).

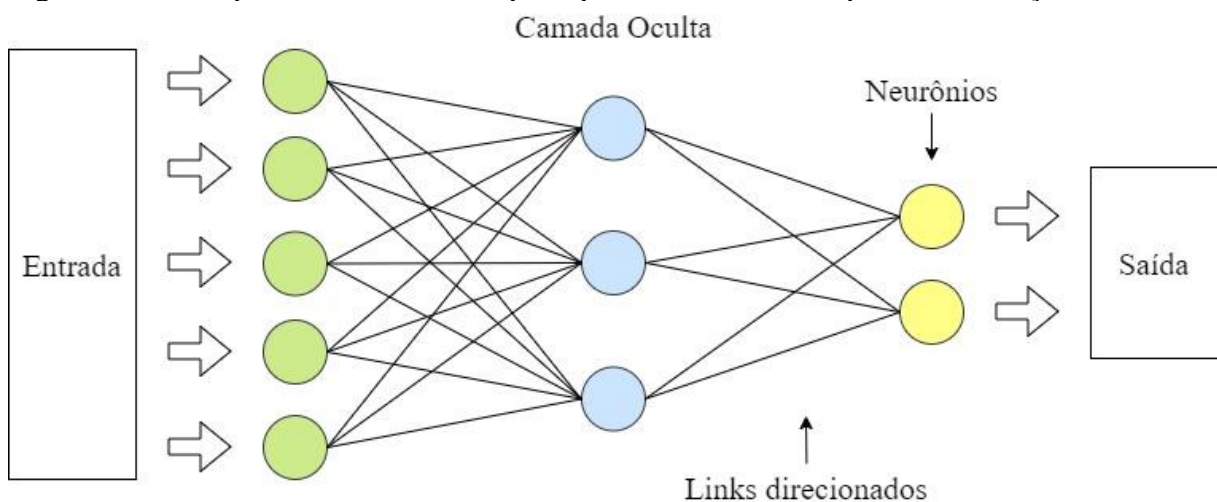
A máquina de vetores de suporte emprega a ideia de *kernel* para obter o melhor hiperplano. Segundo Blyth e Robertson (2005), *kernel* é um espaço nulo entre dois espaços

vetoriais. Já o gênero de transformação que será usado para obtenção do *kernel* depende da natureza das instâncias de treinamento e pode ser de diversos tipos: linear, radial, polinomial, entre outros.

2.3.6 *Perceptron* multicamadas

O *perceptron* multicamadas (*Multilayer perceptron* - MLP), conforme Bishop (1995), é uma rede neural em camadas na qual as informações fluem de forma unidirecional da camada de entrada para a camada de saída, passando pelas camadas ocultas uma a uma. Cada conexão entre os neurônios tem seu próprio peso atribuído, como pode ser observado na Figura 11. No entanto, os pesos atribuídos às conexões podem ser corrigidos através da propagação dos erros de camada a camada, iniciando na camada de saída até a camada de entrada.

Figura 11 - Exemplo de um modelo de *perceptron* multicamadas para classificação binária.



Fonte: Elaborada pelo autor (2021).

Perceptrons para a mesma camada têm a mesma função de ativação, geralmente uma sigmoide para as camadas ocultas. Dependendo de sua aplicação, a camada de saída pode também ser uma função sigmoide ou linear.

O desempenho do modelo de MLP está diretamente relacionado não só às variáveis, mas ao número de camadas ocultas, qualidade dos dados de treinamento e hiperparâmetros como taxa de aprendizados, responsável por controlar a mudança de peso e número de iterações a serem feitas.

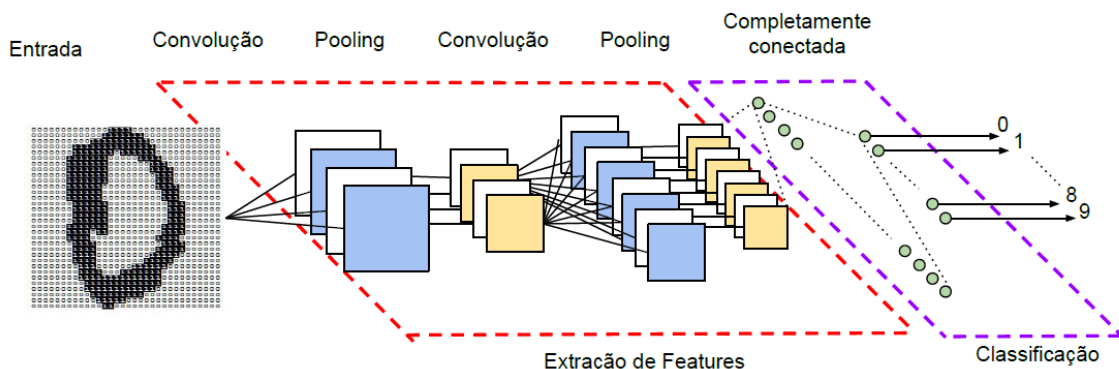
É possível obter resultados satisfatórios com apenas um neurônio, porém é possível combiná-los nessa estrutura de camadas, cada uma com um número específico de neurônios,

formando uma rede neural denominada *perceptron* multicamadas, tornando-a capaz de aprender relações cada vez mais complexas (LEITE, 2018). Porém, ao se usar mais camadas, o modelo pode tender a sobreajustar os dados de treinamento.

2.3.7 Redes neurais convolucionais

Uma variação do algoritmo MLP, baseada no processo biológico de processamento de imagens, é a rede neural convolucional (*Convolutional neural network* - CNN). Segundo Parkhi *et al.* (2015), de forma análoga à algoritmos tradicionais para processamento de dados visuais, a CNN tem a capacidade de aplicar filtros em imagens, preservando a associação de vizinhança entre seus pixels ao longo do processamento de toda sua rede. Assim, este tipo de rede é usado em diversos problemas de reconhecimento e detecção de padrões em imagens e vídeos. Na Figura 12 é possível observar a representação de funcionamento de uma CNN.

Figura 12 - Representação de uma rede neural convolucional.



Fonte: Adaptado de Vargas *et al.* (2016).

Este tipo de rede é composto por diversas partes com funções distintas. Geralmente, os dados de entrada passam por camadas de convolução, formadas por diversos neurônios, nas quais cada neurônio aplica um filtro em uma parte distinta da imagem, ou seja, cada neurônio é ligado a um grupo de *pixels* da camada anterior e para cada ligação é designado um peso. Estes pesos podem ser compreendidos como uma matriz que representa o filtro que será aplicado na camada de convolução, e em conjunto com as entradas de um neurônio produzem uma saída que será a entrada na camada seguinte.

Diferente do algoritmo MLP, no qual um neurônio é plenamente conectado e tem conexão com todos os neurônios da camada predecessora, na CNN apenas parte das entradas é

ligada a cada neurônio. Através desta alteração de arquitetura, a CNN faz análise de campos receptivos locais. Logo, os neurônios que compartilham a mesma camada e que juntos cobrem uma parcela da imagem são agrupados de forma que possibilite a aplicação de um mesmo filtro em diferentes posições da imagem na qual é feito o compartilhamento dos pesos durante o treinamento da rede. Ao compartilhar os pesos é possível reduzir a quantidade de parâmetros a serem assimilados e o tempo gasto para treinamento.

Outras variáveis importantes para uma camada de uma CNN são: a magnitude do filtro a ser aplicado está diretamente ligado ao tamanho da vizinhança que cada neurônio da camada irá processar e o *stride* (passo), que determina qual a distância entre cada janela de pixels e o tamanho da próxima camada na mesma unidade.

Conforme Zeiler e Fergus (2014), por se tratar de um processo automatizado, diversos estudos foram realizados para esclarecer o processo de aprendizado destas redes. Pois, diferente da abordagem convencional de visão computacional, em uma CNN não são definidos os filtros ou variáveis a serem usadas no processo de treinamento, apenas a arquitetura dos filtros. Assim, durante o treinamento os pesos são alterados até que se encontre os melhores valores dos filtros para a base de dados usada. Dessa forma, uma das maneiras de observar o desempenho dos processos de aprendizado é verificar as ativações dos filtros no final do treinamento (ZEILER & FERGUS, 2014).

Muitas vezes em redes com o objetivo de detectar objetos, as camadas iniciais tem como foco a análise do contorno utilizando filtros de arestas e bordas. Posteriormente, em análises mais acentuadas, identificam-se segmentos de interesse e particularidades mais complexas. Com isso, tem-se dentro da rede uma hierarquia de características. Além disso, as CNN têm a capacidade de criar filtros nD , uma vez que, podem misturar diversos mapas de variáveis simultaneamente e, conseqüentemente, extrair variáveis cada vez mais complexas, diferente de redes neurais tradicionais. Dessa maneira, tem-se filtros complexos que contam com o máximo de informações originadas a partir do conjunto de treino.

Após a convolução, é comum a aplicação de uma função de ativação, inserida em cada um dos neurônios e tem como objetivo transformar os dados recebidos. Geralmente, estas funções são não-lineares para que as aplicações subsequentes dos dados transformados tornem as categorias de saídas linearmente separáveis. Outra camada bastante utilizada é a camada de *pooling* (agrupamento), que tem como objetivo diminuir a dimensão dos dados dentro da rede, para que haja maior agilidade no treinamento e criar invariância espacial.

Atualmente existem diversas arquiteturas de CNN que combinam as funções de transformação de sinal já citadas e diversas outras, como por exemplo a LeNet-5. A LeNet-5

possui duas camadas de convolução, *pooling* e mais uma camada de convolução (LECUN *et al.*, 1998). Já a arquitetura GoogLeNet tem cinco camadas de convolução e uma camada de *pooling* (SZEGEDY *et al.*, 2015). Como resultado, tem-se um grupo de variáveis específico para o problema no qual a rede foi treinada. Desse modo, as camadas seguintes fazem o trabalho de classificação ou regressão dos dados. Esta separação de papel das camadas entre extração de variáveis e classificação/regressão é mostrado na Figura 12.

Portanto, segundo Vargas *et al.* (2016), se o propósito da rede é realizar uma classificação, acrescenta-se após o grupo de camadas de convolução e *pooling*, pelo menos uma camada totalmente conectada para designar o caminho de decisão com base nas saídas das camadas anteriores. E por fim, é aplicado a função de classificação, diretamente ligada ao resultado da rede. Uma das funções mais utilizadas para este propósito é a função SoftMax, por ser bastante simples e geralmente ter bons resultados.

2.4 TÉCNICAS DE PROCESSAMENTO DE DADOS

Junto ao imprescindível uso dos algoritmos de aprendizagem de máquina, os modelos utilizados neste estudo também fazem uso de técnicas de processamento de dados para otimizar seu aprendizado. Nas subseções a seguir, são apresentados os métodos utilizados.

2.4.1 Atributos artificiais

Atributos artificiais são atributos criados a partir das instâncias do banco de dados, muitas vezes através de medidas descritivas (média, mediana, moda, desvio padrão, entre outras). Segundo Patel (2021), estes atributos podem ser utilizados pelos algoritmos para melhora de performance e obtenção de melhores resultados.

A criação de variáveis no campo de Aprendizagem de Máquina é considerada mais uma arte do que uma ciência, pois implica na intervenção direta do indivíduo misturando criativamente atributos já existentes (Mueller, Massaron; 2016). Essa tarefa muitas vezes é executada através de operações aritméticas, divisões condicionais por categorias, ou mesmo utilizando transformações e funções mais complexas para gerar novas características com maior poder preditivo que as originais.

Dessa forma, quanto mais conhecimento se têm sobre o banco de dados estudado e sobre o comportamento de suas instâncias e particularidades, maior será a capacidade do pesquisador (ou analista) criar atributos que evidenciam certas características e comportamentos dos dados

que conseqüentemente possam afetar positivamente no processo de aprendizagem e desempenho do modelo.

O processo de construção de atributos, segundo Smith e Bull (2003), tem como principal objetivo descobrir relações implícitas através da inserção de novos recursos. Em contraponto, métodos de seleção de características tem como meta refinar as variáveis utilizadas para que potenciais ruídos e ambigüidades sejam reduzidas ou removidas.

2.4.2 Seleção de características

Seleção de características é um processo de escolha de um subconjunto de atributos relevantes para o uso da construção de classificadores (KUDO; SKLANSKY, 2000). A suposição central é que os dados geralmente contêm recursos redundantes, irrelevantes e ruidosos. Recursos redundantes são aqueles que não fornecem mais informações que os recursos selecionados atualmente. Por outro lado, recursos irrelevantes fornecem informações inúteis em qualquer contexto, enquanto recursos ruidosos geralmente diminuem o desempenho dos classificadores. Portanto, além de permitir tempos de aprendizagem mais curtos, a seleção de recursos também melhora o desempenho de classificação e generalização, reduzindo o *overfitting*.

Em resumo, um algoritmo de seleção de características é basicamente composto por um critério de seleção e uma estratégia de busca. O critério de seleção pontua os diferentes subconjuntos de características e estratégia de busca propõe novos subconjuntos de variáveis. Existem muitas estratégias de pesquisa diferentes para propor subconjuntos de características. O algoritmo mais simples é testar cada subconjunto possível encontrando aquele que maximiza a precisão. Uma vez que essa técnica é uma espécie de busca exaustiva e, às vezes, é computacionalmente inviável, geralmente são escolhidas estratégias de busca subótimas.

Para esta pesquisa, devido o número considerável de experimentos e modelos utilizados, métodos que exigem uma quantidade alta de iterações e variações de subconjuntos de características foram descartados devido ao longo tempo que seria necessário para repetir (e o alto poder computacional necessário para executar tal rotina repetidas vezes), foram utilizados métodos de filtro.

Conforme Gupta (2020), métodos de filtro captam as propriedades intrínsecas de recursos medidos por meio de estatísticas univariadas ao invés do desempenho de um processo de validação cruzada como as técnicas tradicionais. Assim, tem-se métodos mais rápidos, menos caros computacionalmente e úteis ao se trabalhar com dados de alta dimensão.

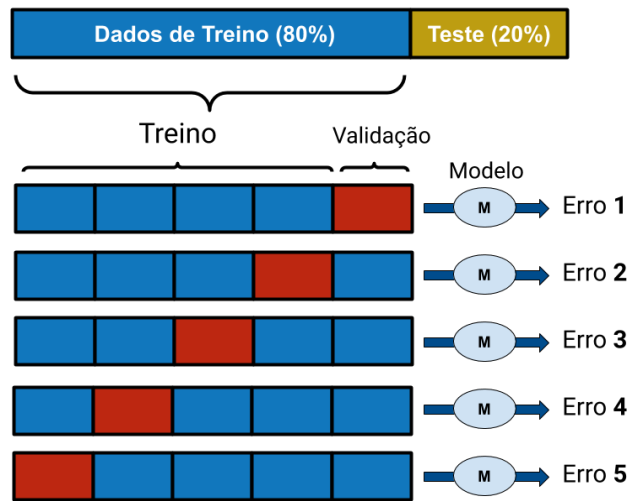
Dessa forma, a seleção de características foi feita através de um método de filtro, análise do ganho informacional de cada variável no contexto da variável alvo, informação obtida através do cálculo da redução na entropia da transformação de um conjunto de dados.

2.4.3 Métodos de treinamento e validação

É comum medir o desempenho de um classificador em termos de precisão. Como o classificador prevê a classe de cada instância, a Acurácia é a proporção de instâncias classificadas corretamente sobre todo o conjunto de instâncias, muitas vezes representada em forma de porcentagem (%). Em geral, a construção de classificadores e a avaliação de sua Acurácia podem ser divididas em duas etapas: treinamento, validação e teste. A etapa de treinamento é usada para definir os parâmetros do algoritmo do classificador, enquanto a etapa de teste é usada para avaliar o desempenho do classificador. Ambas as etapas usam um conjunto de dados rotulados, que possui as entradas (no caso deste trabalho, dados de sensores e descritores de imagem) e as saídas (neste estudo em específico, a potência e tipo de lâmpadas) conhecidas *a priori*.

Frequentemente a forma de dividir os dados para avaliar o desempenho leva a resultados demasiadamente otimistas (JAPKOWICZ; SHAH, 2011). O conjunto de dados de treinamento, validação e de teste devem ser diferentes um do outro para que a quantificação da capacidade de generalização do classificador seja possível. Um bom meio-termo entre a significância estatística e a complexidade computacional é usar a validação cruzada *k-fold* estratificada, na qual o conjunto de dados é dividido em *k* subconjuntos, cada subconjunto é usado uma vez para teste e *k-1* vezes para treinamento (MASON *et al.*, 2014).

Figura 13 - Exemplo de processo de validação cruzada.

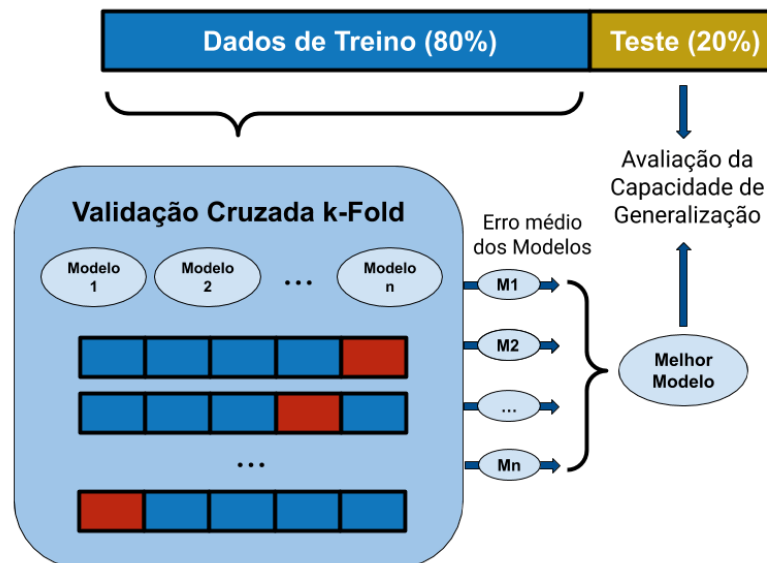


Fonte: SCACCIA (2020).

Este procedimento é particularmente útil quando é necessário comparar o desempenho de vários classificadores em um cenário com um conjunto de dados de treinamento, validação e teste. Uma prática comum e bastante eficiente, é unir o processo de validação cruzada com o de otimização de hiper parâmetros repetindo-o n vezes.

Chamado de validação cruzada aninhada, este processo consiste em escolher uma pasta k de teste a cada execução formando um *loop* externo e executar a validação cruzada k -fold com o restante das pastas do conjunto de dados, formando um loop interno. Assim, os processos de validações cruzadas são executados de forma aninhada.

Figura 14 - Exemplo de processo de validação cruzada aninhada.



Fonte: SCACCIA (2020).

A otimização de parâmetros é feita no loop interno, utilizando pesquisa aleatória, bayesiana ou em grade. Estas funções de procura retornam um conjunto de hiper parâmetros de bom desempenho que podemos utilizar para configurar nosso modelo.

Segundo Browniee (2020) a otimização de parâmetros pode causar *overfitting* e nos prover uma avaliação otimista de um modelo que não deveria ser utilizado, porém combiná-la com a validação cruzada aninhada é uma forma de reduzir as bias do processo.

3 METODOLOGIA

Nesta seção serão descritos os métodos e materiais utilizados na construção dos modelos propostos nesta dissertação. Na Subseção 3.1 é descrito o banco de dados usado, e nas subseções 3.2 e 3.3 é retratado, respectivamente, as estruturas de classificação utilizadas e a metodologia de treinamento empregada.

3.1 BASE DE DADOS

Conforme descrito por Soares *et al.* (2013, 2014, 2015), seus experimentos foram realizados usando dados coletados em laboratório e campo. Reforça-se também que o estudo teve como foco apenas três tipos de tecnologia de lâmpadas: MV, MH e HPS.

Contudo, o presente estudo apenas descreve resultados obtidos através apenas do banco de dados coletado em campo, pois dessa forma, permite-se a atuação dos modelos no processo de classificação do tipo e potência de lâmpadas em exemplos mais complexos que ilustram melhor a realidade das ruas e avenidas, cenário do problema proposto.

As medições de campo foram realizadas no estado do Espírito Santo, no período de agosto de 2013 a setembro de 2013 exclusivamente no período da noite (entre 18:00 e 22:00 horas), em cidades nas quais a rede de iluminação pública é fornecida pela distribuidora local de energia. Foram coletadas 297 ocorrências de 3 diferentes tipos de lâmpada e 6 potências, totalizando 9 em categorias de lâmpadas distintas observadas.

Salienta-se que Soares *et al.* (2013, 2014, 2015) utilizou em seus experimentos com dados coletados em campo apenas 276 instâncias, enquanto neste trabalho, assim como Broetto e Varejão (2016), foi utilizado o conjunto completo formado por 297 instâncias.

Na Tabela 8, é possível observar a distribuição das instâncias em relação aos tipos de lâmpadas (HPS, MH ou MV). Primeiramente, nota-se que apesar dos dados estarem desbalanceados em relação aos modelos, a representação de cada lâmpada no banco de dados está diretamente ligada à variação de potências de cada tipo (conforme Tabela 10).

Tabela 8 - Resumo da distribuição de tipos de lâmpadas do banco de dados.

Classe	Tipo	Potência(s)	Quantidade	%
1	MV	125 W	21	7,07
2	MH	150, 250 e 400 W	109	36,70
3	HPS	70, 100, 150, 250 e 400 W	167	56,23
Total			297	100,0

Fonte: Elaborada pelo autor (2021).

Da mesma forma, na Tabela 9, observa-se uma distribuição similar em relação à distribuição de potências (70, 100, 125, 150, 250 e 400 Watts), fator diretamente ligado ao consumo de energia da lâmpada. Na tabela é possível verificar que as potências com maior representatividade são as potências que são comuns para mais de um tipo de lâmpada (150, 250 e 400 Watts) .

Tabela 9 - Resumo da distribuição de potências de lâmpadas do banco de dados.

Classe	Potência	Tipo(s)	Quantidade	%
1	70 W	HPS	30	10,10
2	100 W	HPS	32	10,77
3	125 W	MV	21	7,07
4	150 W	MH, HPS	58	19,53
5	250 W	MH, HPS	82	27,61
6	400 W	MH, HPS	74	24,92
Total			297	100,0

Fonte: Elaborada pelo autor (2021).

Na Tabela 10, é possível notar um balanceamento percentual razoável (aproximadamente 10%) entre as configurações únicas de lâmpada. Logo, é possível constatar que não há a prevalência de uma classe apenas em relação a distribuição de modelos de lâmpadas.

Tabela 10 - Resumo da distribuição de modelos únicos de lâmpada do banco de dados.

Classe	Descrição	Potência	Quantidade	%
1	MV	125 W	21	7,1
2	MH	150 W	23	7,7
3	MH	250 W	49	16,5
4	MH	400 W	37	12,5
5	HPS	70 W	30	10,1
6	HPS	100 W	32	10,8
7	HPS	150 W	35	11,8
8	HPS	250 W	33	11,1
9	HPS	400 W	37	12,5
Total			297	100,0

Fonte: Elaborada pelo autor (2021).

Naturalmente, através das informações coletadas pelo sistema sobre estas instâncias, foram construídos dois conjuntos de variáveis, um com base nos dados coletados pelos sensores, outro a partir das imagens registradas pela câmera.

As diferentes respostas espectrais obtidas pelo conjunto de sensores analógicos e digitais do sistema, coletadas conforme descrito no Capítulo 2.1, foram agrupadas em uma coleção de treze variáveis, uma para cada sensor conforme Tabela 7, descritas na Tabela 11.

Tabela 11 - Estatísticas descritivas (Tendência Central e dispersão) das variáveis providas pelos sensores, em nm.

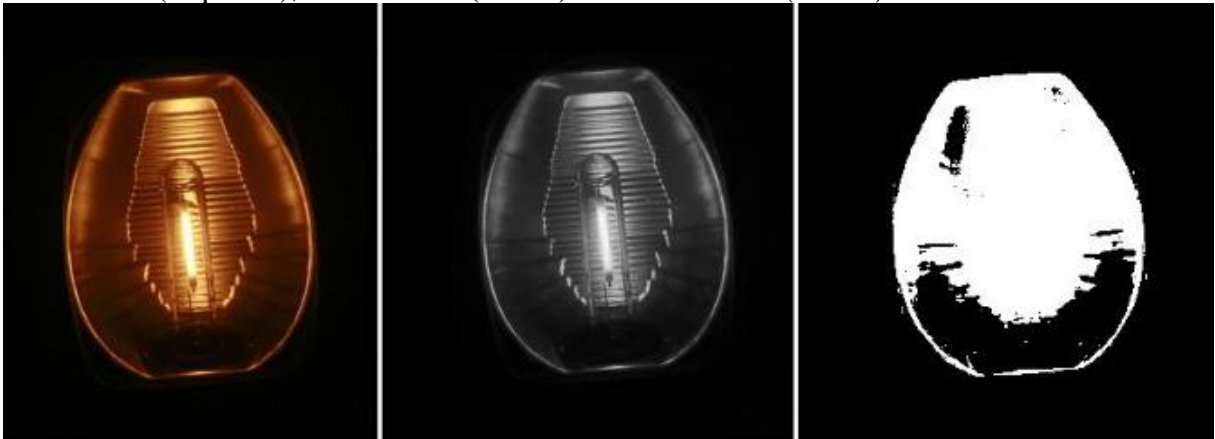
Variável	Mínimo	Mediana	Máximo	Média	Desvio Padrão
sfh5711	1.054,00	2.427,00	3.236,00	2.272,43	450,12
sfh3400	129,00	789,00	3.276,00	833,39	550,13
tsl261r	338,00	1.411,00	4.095,00	1.875,61	1.263,33
sfh3410	86,00	686,00	2.668,00	689,12	440,98
sfh2430	29,00	664,00	2.863,00	655,23	464,40
bp104fas	124,00	1.647,00	4.095,00	1.749,88	1.101,68
tsl13s	270,00	1.689,00	3.370,00	1.610,99	840,82
tsl2569t_c	350,00	5.401,00	31.722,00	6.171,05	4.956,53
tsl2569t_ir	172,00	5.300,00	20.274,00	5.011,53	3.381,70
tcs3414_r	37,00	1.422,00	9.060,00	2.031,05	1.743,20
tcs3414_g	541,00	12.348,00	51.748,00	12.260,03	8.528,30
tcs3414_b	116,00	10.670,00	46.249,00	10.789,68	7.687,85
tcs3414_c	46,00	2.724,00	13.650,00	3.025,51	2.260,16

Fonte: Elaborada pelo autor (2022).

Para construir o segundo conjunto de variáveis através das imagens digitais obtidas antes do processo de extração de características, técnicas de pré-processamento foram utilizadas para transformar fotografias coloridas em fotos preto e branco.

Primeiramente, as imagens (coloridas) foram convertidas para escala cinza. A representação RGB (vermelho, verde e azul) foi transformada em uma representação em HSV (matiz, saturação e valor). O valor, que representa o brilho da imagem, foi utilizado como a intensidade do cinza na imagem em escala cinza. Este método encontra um valor limiar apropriado para dividir a imagem entre dois conjuntos de *pixels*, o primeiro sendo compostos pelos *pixels* com intensidade de cinza menor ou igual que o valor limiar escolhido, e o segundo que consiste nos valores com intensidade de cinza maior que o valor do limiar. Uma vez que estes conjuntos são conhecidos, tem-se uma imagem binária. A Figura 15 mostra o exemplo de uma imagem de lâmpada de um poste público com suas representações em cores, escala cinza e preto e branco.

Figura 15 - Foto de uma lâmpada de um poste de iluminação pública representada em cores (esquerda), escala cinza (centro) e escala binária (direita).



Fonte: Adaptado de Soares *et al.* (2014).

A partir das imagens binárias obtidas, foram calculados os descritores de Fourier, já os descritores de Haralick e os momentos invariantes de Hu foram determinados a partir das imagens em escala cinza. Os primeiros descritores elaborados foram os de Fourier que, uma vez que, fornecem informações quantitativas sobre o contorno do objeto. Logo, podem dar uma descrição representativa das luminárias instaladas em postes de iluminação. Como cada tipo de luminária inclui um conjunto finito de modelos e tecnologias de lâmpadas predeterminadas, compreender o modelo de luminária implicitamente fornece informações sobre quais tipos de lâmpadas podem estar instaladas em determinado poste de iluminação.

Neste problema em específico, a abordagem é particularmente interessante, conforme representamos os *pixels* de contorno de um objeto como um sinal. Logo, é possível definir a Equação 19 que descreve os pontos de contorno do objeto segmentado, na qual $t = 1, 2, \dots, T$ e T é o número de pontos compondo o contorno.

$$P(t) = (x(t), y(t)). \quad (19)$$

Foram utilizados 512 pontos para o contorno do objeto, e uma vez aplicado a DFT, foram considerados os primeiros dez coeficientes de Fourier normalizados, conforme Granlund (1972) e Crimmins (1982), para a criação de variáveis (df01, df02, df03, df04, df05, df06, df07, df08, df09 e df10). A distribuição dos descritores de Fourier é descrita na Tabela 12.

Tabela 12 - Estatísticas descritivas (Tendência Central e dispersão) das variáveis obtidas pela DFT.

Variável	Mínimo	Mediana	Máximo	Média	Desvio Padrão
df01	0,00081	0,03110	0,28060	0,04021	0,03792
df02	0,00176	0,04415	0,16938	0,04986	0,03034
df03	0,00041	0,02342	0,12957	0,02570	0,01685
df04	0,00141	0,01362	0,10554	0,01690	0,01329
df05	0,00036	0,01049	0,09875	0,01326	0,01090
df06	0,00089	0,00821	0,07145	0,01056	0,01014
df07	0,00026	0,00712	0,06231	0,00984	0,00885
df08	0,00049	0,00617	0,06139	0,00846	0,00823
df09	0,00013	0,00537	0,05649	0,00805	0,00765
df10	0,00066	0,00423	0,05425	0,00635	0,00696

Fonte: Elaborada pelo autor (2022).

Posteriormente, como descritos nos subitens 2.2.2 e 2.2.3, foram calculados, também conforme Broetto *et al.* (2013), os descritores de Hu (i1, i2, i3, i4, i5, i6 e i7) e Haralick (Contraste, Correlação, Energia, Entropia, Homogeneidade e Probabilidade Máxima). A distribuição das variáveis obtidas pelos métodos de Hu e Haralick são descritas nas tabelas 13 e 14 respectivamente.

Tabela 13 - Estatísticas descritivas (Tendência Central e dispersão) das variáveis obtidas através dos momentos invariantes de Hu.

Variável	Mínimo	Mediana	Máximo	Média	Desvio Padrão
i1	0,16264	0,17728	0,33395	0,18154	0,01868
i2	0,00000	0,00086	0,00918	0,00128	0,00130
i3	0,00000	0,00009	0,00814	0,00033	0,00093
i4	0,00000	0,00000	0,00607	0,00015	0,00060
i5	0,00000	0,00000	0,00004	0,00000	0,00000
i6	-0,00003	0,00000	0,00024	0,00000	0,00002
i7	-0,00001	0,00000	0,00000	0,00000	0,00000

Fonte: Elaborada pelo autor (2022).

Tabela 14 - Estatísticas descritivas (Tendência Central e dispersão) das variáveis obtidas através dos descritores de Haralick.

Variável	Mínimo	Mediana	Máximo	Média	Desvio Padrão
Contraste	-1047,4100	-19,54000	-3,50000	-28,55000	62,63000
Correlação	0,00008	0,00016	0,00036	0,00017	0,00004
Energia	0,01000	1,31000	3,40000	1,37000	1,04000
Entropia	4,68000	19,82000	36,89000	20,67000	8,96000
Homogeneidade	2,22000	3,40000	3,84000	3,36000	0,27000
Probabilidade Máxima	0,02000	0,57000	0,92000	0,53000	0,26000

Fonte: Elaborada pelo autor (2022).

Dessa forma, tem-se um total de 36 variáveis à disposição dos experimentos, sendo: 13 provenientes dos sensores e 23 provenientes das imagens (10 descritores de Fourier, 7 descritores de Hu e 6 descritores de Haralick).

Por conseguinte, visando expandir a quantidade e qualidade da informação disponível para o aprendizado dos modelos de classificação, além das 36 variáveis obtidas, foram criados sete atributos artificiais (média, mediana, desvio padrão, variância, coeficiente de variação, curtose e assimetria) em todos os conjuntos. Neste grupo, cada nova característica é construída através da aplicação de uma estatística descritiva às instâncias do banco de dados.

A escolha de utilizar estatísticas descritivas se dá justamente pela sua função de auxiliar na descrição e compreensão de um conjunto de dados e sua distribuição. Foram escolhidas duas medidas de tendência central, média (definida pela Equação 20) e mediana (valor de centro dos valores observados ordenados), que nos dá informações a respeito do “centro” da distribuição dos valores das instâncias.

Dentre os cinco atributos restantes, três são parâmetros de dispersão absoluta que dimensionam o afastamento médio dos valores de um conjunto de seu valor médio, sendo eles: desvio padrão, variância e coeficiente de variação (definidos respectivamente pelas equações 21, 22 e 23).

Os últimos dois são medidas de dispersão relativa, curtose (calculada através da Equação 24) e assimetria (definida pela Equação 25), que descrevem o grau de achatamento e desvio da simetria de uma distribuição nesta ordem.

$$\mu = \frac{\sum x_i}{N} \quad (20)$$

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \quad (21)$$

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N} \quad (22)$$

$$CV = 100 \frac{\sigma}{\mu} \quad (23)$$

$$Curtose = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - \mu}{\sigma} \right]^4 \quad (24)$$

$$Assimetria = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - \mu}{\sigma} \right]^4 \quad (25)$$

Nas quais N é o número de observações e $i = 1, 2, \dots, N$.

3.1.1 Conjuntos de dados

A partir da totalidade de 43 variáveis obtidas (36 provenientes da coleta de informações pelo *hardware* e 7 de atributos artificiais), foram criados quatro conjuntos de dados para a realização dos experimentos com o intuito de observar quais informações são relevantes para as classificações almejadas e qual a melhor combinação provém o melhor resultado.

O primeiro conjunto, que será chamado de **conjunto A**, é formado apenas por informações dos sensores. O segundo, **denominado B**, é composto apenas pelos descritores de imagens (Fourier, Hu e Haralick, descritos anteriormente).

O terceiro, **conjunto C**, é um conjunto completo, gerado por todas as informações coletadas pelo *hardware* (imagens e sensores). Por fim, o **agrupamento D** consiste em um subconjunto de C, no qual apenas as 25 variáveis mais relevantes para o treinamento do modelo foram mantidas. O processo de seleção destas características é descrito na próxima seção.

Em contrapartida, os experimentos realizados com algoritmo de aprendizado profundo utilizarão exclusivamente as imagens originais registradas por Soares *et al.* (2014, 2015), das

quais o próprio algoritmo irá extrair suas variáveis explicativas, via processos internos da técnica.

3.1.2 Seleção de características

Existem diversos métodos de se selecionar um subconjunto de variáveis, a complexidade destes, do processo de seleção em si e da escolha de qual é a melhor alternativa para dado problema é influenciada por vários fatores número de variáveis, poder computacional disponível, complexidade do modelo e problema, entre outros.

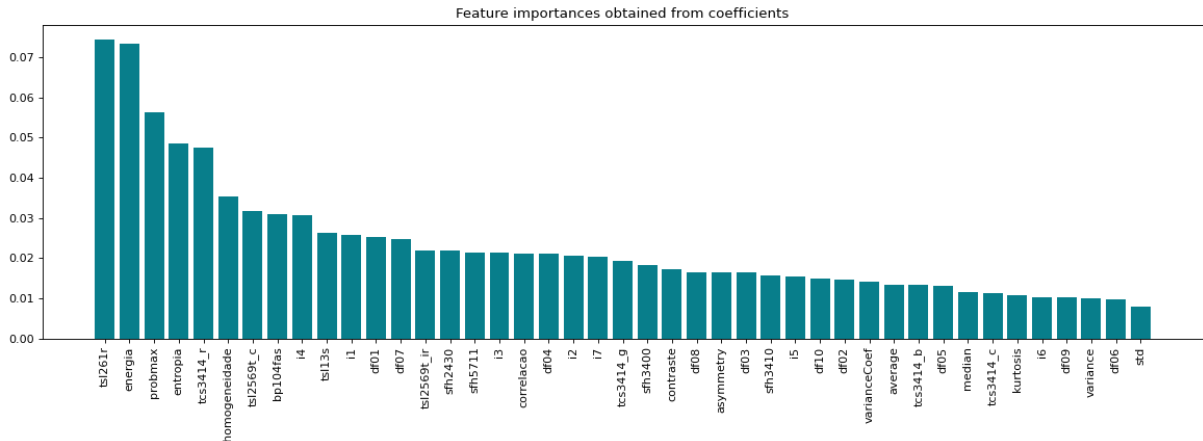
No problema em que este estudo se propõe a estudar, dado o número considerável de variáveis (43), a utilização de diversos modelos (6) e abordagens de classificação (hierárquica e única, que serão abordadas em detalhes na Subseção 3.2), métodos complexos e que exigem um número alto de interações e conseqüentemente um poder computacional proporcional às condições descritas foram descartadas.

Logo, optou-se por um método mais simples e menos moroso, a seleção de características foi feita através treinamento de *naive classifier* e análise de quais características eram mais relevantes para estes. Um *naive classifier*, segundo Brownlee (2019), é um modelo de classificação simples que assume nada ou quase nada sobre o problema e sua performance serve como uma base de comparação para outros modelos validados em uma base dados que permita comparação entre eles. Nesta dissertação foram treinados *naive classifiers*, modelos com os hiperparâmetros em seus valores *default* (predefinidos), com quatro dos seis algoritmos escolhidos (Árvore de Decisão, Floresta de Aleatória, XGBoost e Máquina de Vetores de Suporte) em cada uma das abordagens de classificação (tipo, potência e única). Apenas quatro foram utilizados pois as bibliotecas destes disponibilizam a informação desejada através dos atributos *.feature_importances_* (para os algoritmos DT, RF e XGB) e *.coef_* (para o algoritmo SVM).

No atributo *.feature_importances_*, a importância do recurso é calculada como a diminuição na impureza do nó de uma árvore ponderada pela probabilidade de atingir esse nó (calculada pelo número de amostras que chegam ao nó, dividido pelo número total de amostras). Assim, quanto maior o valor, maior a importância da variável. Já os “pesos” obtidos por *.coef_*, representam as coordenadas vetoriais que são ortogonais ao hiperplano, logo, conforme sua direção indica a classe prevista, o tamanho absoluto dos coeficientes em relação um ao outro pode ser utilizado para determinar a importância de cada característica na tarefa de classificação proposta.

Desta forma, utilizando como entrada informações providas por ambos sensores e imagens (totalizando 36 variáveis) foram criados cinco *naive classifier* (diferenciados pelo valor da semente aleatória utilizada) para cada um dos quatro algoritmos. Estes modelos foram treinados para classificação de tipos de lâmpadas e, posteriormente, foi extraída a classificação de importância das variáveis, ilustrada pela Figura 16, para um dos modelos elaborados.

Figura 16 - Valores de importância das variáveis para o *naive classifier* de Floresta Aleatória.



Fonte: Elaborada pelo autor (2022).

Após coletados os valores de importância das variáveis para os cinco classificadores de cada algoritmo e, em seguida, calculou-se uma média para os valores de importância de cada variável conforme apêndice A. Com base nestes resultados, as características foram ranqueadas da com maior média de importância para a com menor média de importância, como mostra o apêndice D. Por fim, foram somadas as posições no *ranking* de cada uma das variáveis para cada um dos algoritmos (apêndice D), e as 25 características com os menores valores foram selecionadas (ou seja, as que estiveram com maior frequência nas primeiras posições), formando o conjunto de D para classificação de tipos de lâmpadas.

Dessa forma, ao somar as importâncias num primeiro momento é possível ter uma média de importância de uma variável para o treinamento com cinco diferentes divisões do banco de dados para um mesmo algoritmo. E, em seguida, utilizando a posição do *ranking* para seleção é possível contornar as diferentes medidas de importância utilizadas, visto que *.feature_importances_* e *.coef_* medem a importância de variáveis de forma distinta, o número de 25 variáveis foi escolhido empiricamente, visando eliminar variáveis desnecessárias, mas preservando o máximo de características que possam prover alguma informação para o treinamento dos modelos.

O processo descrito foi repetido para a classificação de potência e única (resultados representados nos apêndices B, C, E e F). Assim, formou-se três grupos de 25 variáveis cada, como mostra a Tabela 15.

Tabela 15 - Variáveis do conjunto D para classificação de tipos, potência e única de lâmpadas.

Tipo	Potência	Única
tsl261r	tsl261r	tsl261r
bp104fas	bp104fas	bp104fas
Coeficiente de variação	Contraste	Contraste
Contraste	Correlação	Correlação
Correlação	df02	Desvio Padrão
df01	df03	df02
df03	df04	df03
df04	df06	df04
df07	df08	df06
Energia	df09	df09
Entropia	df10	df10
Homogeneidade	Energia	Energia
i1	Entropia	Entropia
i2	Homogeneidade	Homogeneidade
i3	i1	i1
i4	i4	i4
i7	Mediana	i7
Mediana	Probabilidade Máxima	Média
Probabilidade Máxima	sfh2430	Probabilidade Máxima
sfh5711	sfh3400	sfh3400
tcs3414_c	tcs3414_c	tcs3414_c
tcs3414_r	tcs3414_r	tcs3414_r
tsl13s	tsl13s	tsl13s
tsl2569t_c	tsl2569t_c	tsl2569t_c
tsl2569t_ir	Variância	tsl2569t_ir

Fonte: Elaborada pelo autor (2022).

Apesar de se tratar de três classificações distintas, com diferentes números de classes e até mesmo padrões diferentes a serem detectados, os conjuntos de variáveis obtidos são similares. Dentre os três conjuntos, todos compartilham 17 características em comum (tsl261r, bp104fas, Contraste, Correlação, df03, df04, Energia, Entropia, Homogeneidade, i1, i4, Média, Probabilidade Máxima, tcs3414_c, tcs3414_r, tsl13s e tsl2569t_c), além de compartilhar até sete variáveis com um dos dois demais grupos.

A partir desta observação, e do fato que as variáveis únicas de cada grupo estavam em grande parte a poucas posições de distância da 25ª colocação, inferiu-se um comportamento similar ao representado na Figura 16.

Ou seja, uma porção de variáveis contém informações que agregam de forma significativa para o aprendizado dos modelos em todas as abordagens (composta principalmente por características providas de sensores), uma segunda parcela que provê elucidação de forma mais branda e com pouca diferença de importância entre si para os algoritmos, e a terceira parte formada por características que adicionam pouco a ponto de serem irrelevantes.

Portanto, o agrupamento D será formado por três conjuntos distintos, apesar de similares, formados por 25 variáveis e que serão utilizados conforme o objetivo da classificação. Portanto, posteriormente ao se referir que o conjunto D foi utilizado para um experimento de classificação de tipos de lâmpadas, por exemplo, fica implícito que se utilizou o conjunto de variáveis do conjunto D selecionado para classificação de tipos de lâmpadas.

3.1.3 Transformação dos dados

Após a definição dos conjuntos de dados, antes que fossem utilizados como entradas nos modelos elaborados para experimentação, estes foram submetidos ao processo de normalização. Dado a alta amplitude de algumas variáveis e a discrepância entre os valores de cada uma das características, viu-se necessário transformar os dados para evitar que experimentos resultassem em modelos viesados.

Especialmente em algoritmos são combinados de forma não-linear, como nas Redes Neurais MLP, a normalização dos dados é estritamente necessária para padronização das variáveis de entrada. Caso contrário, no caso do MLP, as funções de ativação podem operar em intervalos de saturação que acarretam na paralisia do aprendizado do modelo. Portanto, foram utilizados dois tipos de transformação: normalização e padronização.

A normalização tem como objetivo transformar todas as características em uma mesma ordem de grandeza, garantindo que os dados sejam semelhantes em todos os registros, mantendo sua integridade.

Em vista disso, os dados de todos os conjuntos foram normalizados entre os valores 0 e 1, através da Equação 26.

$$Z = \frac{[X - MIN(X)]}{[MAX(X) - MIN(X)]} \quad (26)$$

Na qual, Z é o valor normalizado, x é o valor a ser normalizado, $\text{MIN}(X)$ e $\text{MAX}(X)$ são respectivamente o valor mínimo e máximo dos dados da variável a qual o valor a ser normalizado pertence.

Nos experimentos que utilizam o algoritmo de classificação SVM, a normalização foi feita de forma diferente. Devido a uma demora exagerada na convergência do modelo durante os experimentos iniciais, foi necessária uma transformação adicional nos dados para acelerar este processo. A transformação logarítmica, definida pela Equação 27, tem como objetivo reduzir a importância relativa de altos valores, facilitando a generalização do modelo.

$$y = \ln(x) \quad (27)$$

Após a adição da transformação logarítmica, os modelos com o algoritmo de Máquina de Vetores de Suporte passaram a ser executados em tempo hábil. Experimentos com outros algoritmos foram feitos utilizando a transformação logarítmica, mas não houve nenhuma melhora significativa na acurácia que justificasse a repetição de todos os experimentos.

Já a padronização tem como função transformar os dados de forma que tenham distribuição normal padrão, ou seja, tenham uma média de 0 e um desvio padrão de 1 (variância unitária) através da Equação 28.

$$Z = \frac{x - \mu}{\sigma} \quad (28)$$

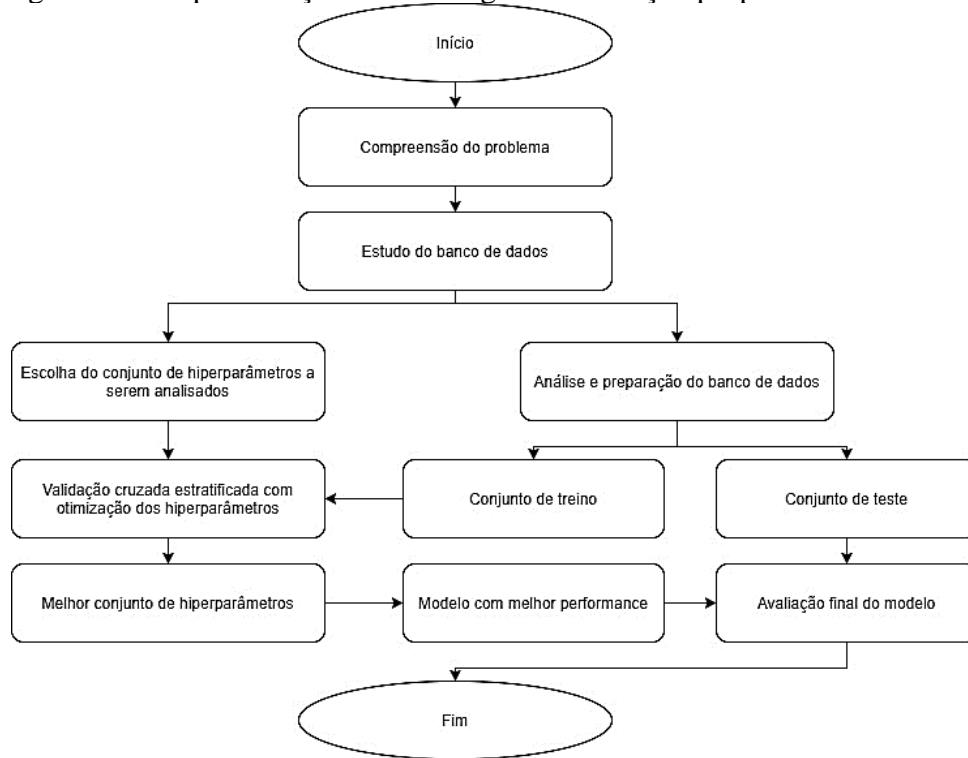
Entretanto, assim como a normalização, os modelos de SVM apresentaram problemas para convergir durante o processo de treinamento. Logo, também foi necessária uma transformação logarítmica também para estes experimentos. Porém, como padronização não permite a especificação da amplitude dos valores transformados tem-se a ocorrência de valores negativos. Com isso, para subverter este problema na transformação logarítmica, utilizou-se a Equação 29, proposta por Busse e Hefeker (2007).

$$y = \ln(x + \sqrt{x^2 + 1}) \quad (29)$$

3.2 CONSTRUÇÃO DE TREINAMENTO DOS MODELOS

Por se tratar de um estudo de um problema de reconhecimento de padrões e envolver modelos de aprendizagem de máquina, a pesquisa irá se desenvolver de forma bastante similar aos ciclos de vida de projetos de Ciência de Dados e Aprendizagem de Máquina, conforme mostra a Figura 17.

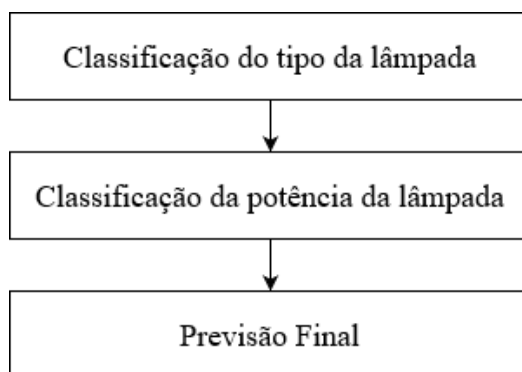
Figura 17 – Representação estrutural geral da solução proposta.



Fonte: Elaborada pelo autor (2021).

De forma similar a Soares *et al.* (2015), os experimentos foram realizados em duas abordagens de classificação distintas. Na primeira, chamada de classificação hierárquica, o processo é dividido em duas camadas (conforme Figura 18), no qual os dados de entrada são inicialmente submetidos a um primeiro classificador, que é exclusivamente treinado e responsável por definir o tipo de lâmpada (mercúrio, metálica ou sódio). Esta predição é adicionada aos dados de entrada do segundo classificador, este incumbido e treinado para definir a potência da lâmpada (70, 100, 125, 150, 250 ou 400). Por fim, as saídas destes classificadores são combinadas em uma predição final da classe da lâmpada.

Figura 18 - Diagrama representativo da estrutura de classificação hierárquica.

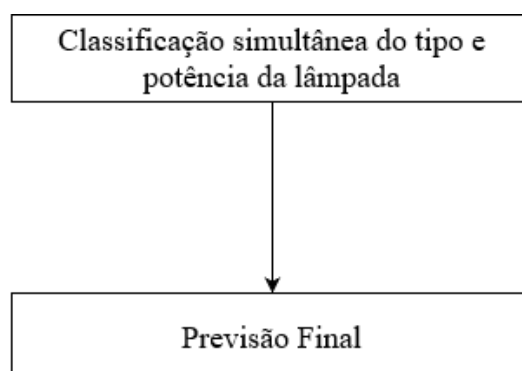


Fonte: Elaborada pelo autor (2022).

Optou-se por uma hierarquia na qual o tipo é classificado primeiro, e posteriormente a potência, visto que a classificação de tipo é mais simples (padrão mais fácil de ser detectado, composto por apenas três classes, entre outras) e sua saída pode auxiliar no aprendizado do classificador de potência. Outro fator que incentiva esta sequência hierárquica é que como nem toda potência está presente em todo tipo de lâmpada, erros como classificar a potência de 150W a uma lâmpada de mercúrio, passa a ser evitado com mais facilidade.

A segunda abordagem, chamada de classificação única, é composta apenas por um classificador, este responsável pela classificação do tipo e potência da lâmpada simultaneamente, conforme ilustrado pela Figura 19.

Figura 19 - Diagrama representativo da estrutura de classificação única.



Fonte: Elaborada pelo autor (2022).

A estrutura proposta por Soares *et al.* (2013, 2014, 2015) também foi analisada, ilustrada na Figura 2, na qual o tipo de lâmpada é classificado na primeira camada, e posteriormente, três

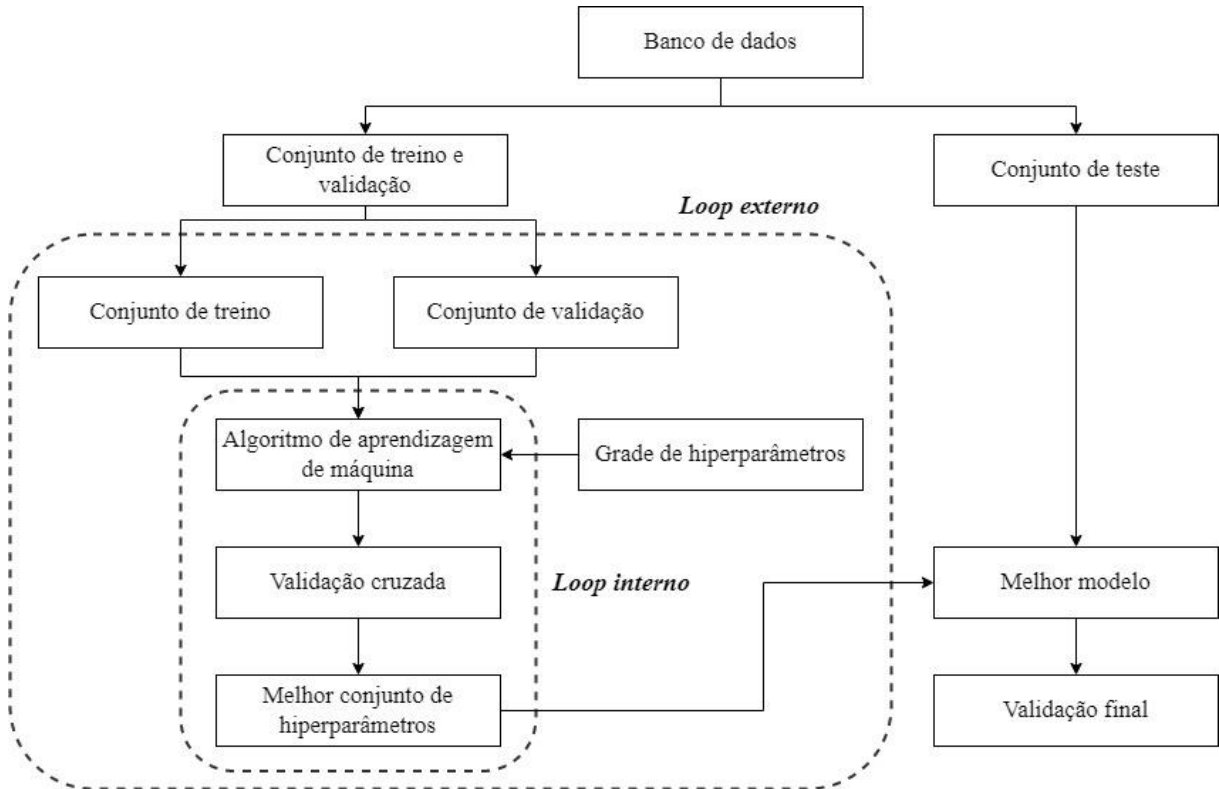
classificadores (um para cada tipo de lâmpada) são treinados exclusivamente para prever a potência de cada tipo de lâmpada. Entretanto, nos testes iniciais a mesma apresentou resultados inferiores ou iguais à abordagem hierárquica escolhida. E, assim, mesmo podendo ser utilizada para meios comparativos, dado o uso computacional dos experimentos e tempo necessário para execução das rotinas que serão descritas nas próximas seções, esta dissertação optou por se ater às duas abordagens que demonstraram maior potencial de eficácia no problema proposto.

Dessa forma, como ambas as estruturas estabelecidas serão testadas com os seis algoritmos de aprendizado supervisionado de abordagem tradicional previamente citados (KNN, DT, RF, SVM, XGB e MLP) e uma rede neural convolucional, a grande diferença entre elas é que na hierárquica, tem-se um modelo específico para classificação das lâmpadas em tipos (3 classes) e, em seguida, um segundo modelo para definir a potência da lâmpada (6 classes), e na única, um modelo apenas é responsável para classificação da lâmpada por seu tipo e potência simultaneamente (9 classes). Assim, mediante o exposto, as diferentes abordagens podem dar *insights* sobre quais padrões e métodos são mais relevantes para cada algoritmo.

3.3 TREINAMENTO DOS MODELOS

A rotina de treinamento dos modelos, ilustrada anteriormente na Figura 17, foi elaborada com base no método de validação cruzada aninhada para otimizar a comparação estatística dos modelos, composta por dois *loops*, um externo e um interno. O processo é detalhado na Figura 20.

Figura 20 - Diagrama representativo da rotina de treinamento e validação dos experimentos.



Fonte: Elaborada pelo autor (2022).

No *loop* externo, através do método *k-fold*, o banco de dados é dividido em 10 pastas, das quais nove compõem o conjunto de treino e validação e uma partição constitui o conjunto de teste. Em seguida, o *loop* interno é executado, iterando todas as possíveis combinações de parâmetros da grade dada, e seleciona o arranjo de valores de hiperparâmetros do modelo com melhor desempenho. Por fim, um modelo com a melhor configuração encontrada é treinado e validado com a pasta de teste.

A métrica de erro utilizada nos processos de validação (no conjunto de validação e conjunto de teste) foi a acurácia que contabiliza o número total de predições corretas no total de instâncias avaliadas. Apesar de uma métrica simples, por se tratar de um banco de dados balanceado e de um problema que não necessita de penalidades diferentes para acertos e erros de classes específicas, a métrica é suficiente para se avaliar de forma geral o desempenho dos modelos na tarefa de classificação. A fórmula para cálculo da acurácia pode ser observada na Equação 30.

$$Acurácia = \frac{\text{Número de predições corretas}}{\text{Número total de predições feitas}} \quad (30)$$

Posteriormente, os melhores parâmetros e a acurácia no conjunto teste são salvos, e o *loop* externo se repete até que cada uma das 10 pastas tenha sido conjunto teste neste processo.

3.3.1 *Loop* interno

O *loop* interno funciona de forma similar ao *loop* externo, de forma que, os dados que o *loop* externo separa como conjunto de treino e validação são divididos em 10 pastas estratificadas, nas quais 9 serão utilizadas para treino e uma para validação. E, analogamente, esse laço se repete até que cada uma das partes tenha sido o conjunto de validação.

Todavia, a grande diferença entre eles está em que o *loop* externo tem como objetivo treinar e testar o melhor modelo em diferentes divisões do banco de dados, e o laço interno tem como finalidade encontrar a melhor combinação de hiperparâmetros para o modelo através da validação cruzada.

Para encontrar a melhor configuração de modelo foi utilizada a abordagem mais direta para otimização de hiperparâmetros, o *Grid Search* (PEDREGOSA, 2011). Basicamente, uma lista de valores candidatos para cada hiperparâmetros é definida e avaliada. O nome “*grid*” vem do fato que os possíveis candidatos criados por meio de variações dos hiperparâmetros são combinados, e a àquele que apresentar a melhor performance no conjunto de validação é selecionado.

Todavia, quanto maior o número de candidatos e, conseqüentemente, o espaço de busca, mais moroso será este processo. Assim, para cada um dos seis algoritmos utilizados, foram definidos de 3 a 8 hiperparâmetros e seus respectivos candidatos, conforme mostram as Tabelas 16, 17, 18, 19, 20 e 21.

Tabela 16 - Intervalo de valores de hiperparâmetros para o algoritmo KNN.

Hiperparâmetro	Intervalo de valores
<i>n_neighbors</i>	1, 3, 5, 7, 9
<i>weights</i>	<i>uniform, distance</i>
<i>algorithm</i>	<i>auto, ball_tree, kd_tree, brute</i>

Fonte: Elaborada pelo autor (2021).

Tabela 17 - Intervalo de valores de hiperparâmetros para o algoritmo DT.

Hiperparâmetro	Intervalo de valores
<i>max_depth</i>	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 15, 20, 25, 30
<i>max_features</i>	10, 11, 13, 15, 17, 19, 20, 21, 23, 25
<i>min_samples_split</i>	32, 64, 128, 256
<i>min_samples_leaf</i>	32, 64, 128, 256

Fonte: Elaborada pelo autor (2021).

Tabela 18 - Intervalo de valores de hiperparâmetros para o algoritmo RF.

Hiperparâmetro	Intervalo de valores
<i>max_depth</i>	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 15, 20, 25, 30
<i>max_features</i>	10, 15, 20, 25
<i>n_estimators</i>	50, 100, 300, 500, 1000, 2000, 3000

Fonte: Elaborada pelo autor (2021).

Tabela 19 - Intervalo de valores de hiperparâmetros para o algoritmo XGB.

Hiperparâmetro	Intervalo de valores
<i>max_depth</i>	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 15, 20, 25, 30
<i>learning_rate</i>	0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95
<i>n_estimators</i>	5000

Fonte: Elaborada pelo autor (2021).

Tabela 20 - Intervalo de valores de hiperparâmetros para o algoritmo SVM.

Hiperparâmetro	Intervalo de valores
<i>C</i>	0.1, 1, 10, 30, 50, 70, 90, 100, 300, 500, 700, 1000, 2000
<i>gamma</i>	1, .1, .01, .001
<i>kernel</i>	<i>rbf, poly, sigmoid</i>

Fonte: Elaborada pelo autor (2021).

Tabela 21 - Intervalo de valores de hiperparâmetros para o algoritmo MLP.

Hiperparâmetro	Intervalo de valores
<i>activation</i>	<i>identity, logistic, tanh, relu</i>
<i>solver</i>	<i>sgd, adam</i>
<i>learning_rate</i>	<i>constant, invscaling, adaptive</i>
<i>max_iter</i>	5000
<i>tol</i>	0.001
<i>learning_rate_init</i>	0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95
<i>momentum</i>	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
<i>hidden_layer_sizes</i>	(9,), (14,), (15,), (16,), (17,), (18,), (19,), (20,), (21,), (24,), (25,), (27,), (28,), (30,)

Fonte: Elaborada pelo autor (2021).

No algoritmo k-vizinhos mais próximos foram analisados três hiperparâmetros: *n_neighbors* (número de vizinhos mais próximos a serem considerados), *weights* (função de peso a ser utilizada nas predições) e *algorithm* (algoritmo utilizado para calcular quais são os vizinhos mais próximos).

Nas árvores de decisão, quatro hiperparâmetros foram objetos de busca: *max_depth* (a profundidade máxima da árvore), *max_features* (número máximo de variáveis considerado para procurar as melhores divisões), *min_samples_split* (número mínimo de amostras para divisão de um nó interno) e *min_samples_leaf* (o número mínimo de amostras para compor um nó folha).

O algoritmo de floresta aleatória foi composto por três hiperparâmetros: *max_depth*, *n_estimators* (número de árvores que compõem a floresta) e *max_features*. De forma similar, o XGBoost é configurado por: *max_depth*, *n_estimators* e *learning_rate* (taxa/velocidade em que o modelo aprende a cada iteração).

Para experimentos com máquina de vetores de suporte, três hiperparâmetros foram avaliados: *C* (parâmetro de regularização), *kernel* (especifica o tipo de *kernel* que será utilizado no algoritmo) e *gamma* (coeficiente *kernel*).

A grade do algoritmo de redes neurais *perceptron* multicamadas foi composto por seis hiperparâmetros: *activation* (função de ativação para a camada oculta), *solver* (especifica o algoritmo para otimização de peso entre os nós), *learning_rate*, *learning_rate_init* (a taxa de aprendizado inicial utilizada), *momentum* (momentum para atualização do gradiente descendente), *hidden_layer_sizes* (número de neurônios na camada oculta), *max_iter* (número máximo de iterações para convergência) e *tol* (tolerância para otimização).

Dessa maneira, após o *Grid Search* testar todas as combinações de hiperparâmetros, a combinação com melhor performance de acurácia no conjunto de validação é selecionada para treinar um modelo no *loop* externo.

Já os experimentos com algoritmos aprendizado profundo não farão uso de técnicas de otimização de hiperparâmetros por limitação de poder computacional. Estes serão realizados por uma única CNN, com 3 camadas de convolução, cada uma seguida de uma camada de *pooling* (a arquitetura da rede é representada na Tabela 22). Por se tratar de algoritmos mais complexos, é possível que os modelos criados a partir destes consigam detectar novos padrões e identificar mais facilmente as diferentes classes de classificação única e de potência de lâmpadas. Outra vantagem que estes modelos podem oferecer é o processamento de imagens, que permitirá o uso das fotos coloridas para extração de variáveis que podem fornecer novas informações sobre as luminárias.

Tabela 22 - Arquitetura da rede neural convolucional utilizada.

Camada	Tipo	Mapa de recursos	kernel
0	<i>Input</i>	576 x 576 x 3	-
1	Convolução ReLu	-	3
2	Max Pool	288 x 288 x 6	-
3	Convolução ReLu	-	3
4	Max Pool	144 x 144 x 16	-
5	Convolução ReLu	-	3
6	Max Pool	72 x 72 x 16	-
7	Camada totalmente conectada ReLu	1024	-
8	Camada totalmente conectada ReLu	512	-
9	<i>Output</i>	3, 6 ou 9 (tipo, potência ou única)	-

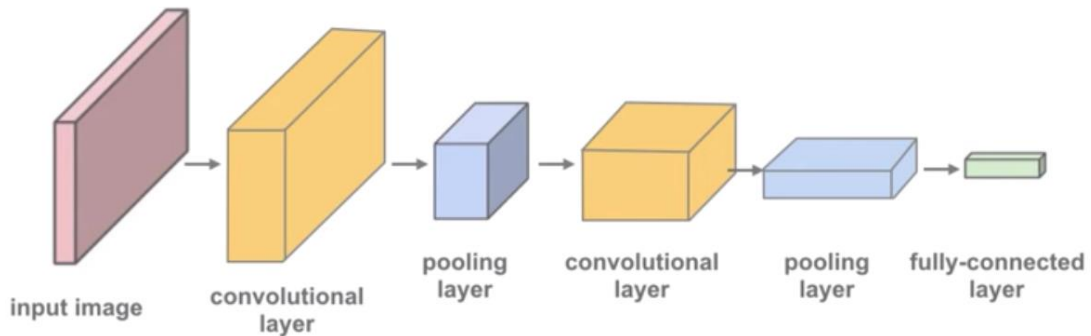
Fonte: Elaborada pelo autor (2021).

A rede usa a função de convolução Conv2d da biblioteca Torch da linguagem de programação Python. A convolução 2D, conforme Shafkat (2018), utiliza uma operação relativamente simples, iniciando um *kernel* (uma matriz de pesos) que desliza sobre a imagem de duas dimensões utilizada como entrada, realizando multiplicações elemento a elemento com a janela de *pixels* que está sendo analisada e, posteriormente, somando os resultados em um único *pixel* de saída. Também serão utilizadas camadas de *pooling* para diminuir progressivamente a altura e largura dos tensores de saída de cada camada convolucional.

Logo, a camada de convolução irá transformar a imagem de 3 canais em um mapa de variáveis de 16 canais, e a camada de *pooling* (que utiliza a função MaxPool2d, também da biblioteca *Torch*) divide sua altura e largura pela metade. Para encadear as camadas e funções de ativação (ReLu) em uma única arquitetura foi utilizada a função nn.Sequential.

O mapa de características irá então ficar menor à medida que mais camadas forem adicionadas, até que finalmente restará um pequeno mapa de variáveis, que será transformado em um vetor. A Figura 21 ilustra a disposição de camadas descrita.

Figura 21 - Representação do funcionamento de uma rede neural convolucional.



Fonte: Adaptado de Camacho (2018).

Algumas camadas totalmente conectadas foram adicionadas no final, de forma a obter um vetor de tamanho igual ao número de classes (3 para classificação de tipos, 6 para classificação de potências e 9 para classificação de modelos únicos) de cada imagem. Neste vetor, cada elemento será interpretado como a probabilidade de a imagem pertencer a cada uma das possíveis classes e, conseqüentemente, a classe referente ao maior valor de probabilidade será escolhida como predição do modelo para a imagem avaliada.

3.3.2 Experimentos

A rotina de treinamento descrita será executada para toda combinação entre abordagem de classificação (hierárquica e única), algoritmo de aprendizado supervisionado (KNN, DT, RF, XGB, SVM e MLP), transformação dos dados (normalização e padronização) e conjunto de dados (A, B, C e D). Com isso, como cada rotina retornará 10 conjuntos de hiperparâmetros, um para cada pasta de teste do *loop* externo, o processo será repetido três vezes com sementes aleatórias diferentes para obtenção de uma amostra razoável para comparação e escolha da melhor solução.

De forma análoga, os experimentos com algoritmos de aprendizado profundo serão executados para todas ambas estruturas de classificação utilizando apenas as imagens originais coletadas em campo como entrada, e será repetido com 3 sementes aleatórias, resultando também em 10 resultados com pastas de teste no *loop* externo para cada semente. Entretanto, não será feito nenhum processo de hiperparâmetros, variação no conjunto de dados.

Por fim, para comparação de resultados entre os modelos, para garantir que há de fato evidências de superioridade ou não de um sobre o outro, será utilizado o teste de t-pareado de Student para amostras dependentes (visto que serão utilizadas as mesmas sementes para todos

os experimentos para que os dados sejam os mesmos para todos os algoritmos). Logo, será considerado um nível de significância de 95% e terá como hipótese nula: a média de acurácia do algoritmo A é menor ou igual à média de acurácia do algoritmo B.

4 RESULTADOS E DISCUSSÃO

Esta seção relata os resultados obtidos pelos modelos compostos por algoritmos de abordagem tradicional e aprendizado profundo propostos nesta dissertação. A apuração dos modelos que utilizaram os dados normalizados é realizada individualmente para as estruturas hierárquica e única, respectivamente, nas subseções 4.1 e 4.2. Na Subseção 4.3, são descritas as apurações dos modelos que utilizaram dados padronizados. E por fim, nas subseções 4.4 e 4.5 são descritas a comparação entre modelos e com os resultados obtidos pelos trabalhos anteriores nesta ordem.

4.1 COMPARAÇÃO DE MODELOS COM ESTRUTURA DE CLASSIFICAÇÃO HIERÁRQUICA

Conforme descrito anteriormente, os modelos com classificação hierárquica são divididos em duas camadas, uma responsável por prever o tipo da lâmpada analisada e a segunda camada, encarregada de prever a potência da mesma. Assim, como cada camada é composta por um algoritmo de classificação, estes serão avaliados inicialmente de forma individual e posteriormente de em conjunto.

4.1.1 Desempenho dos classificadores de tipo

Através dos resultados de acurácia obtidos dos modelos de predição de tipo de lâmpada, foi calculada a acurácia média e seu intervalo de confiança (95%) a fim de compreender e comparar o desempenho de cada algoritmo nos diferentes conjuntos de dados. A apuração pode ser observada na Tabela 23.

Tabela 23 - Acurácia média e Intervalo de Confiança (95%) dos classificadores de tipo de lâmpada nos conjuntos de dados A, B, C e D com uso de normalização.

Modelo	A	B	C	D
KNN	0,81 (0,78 - 0,84)	0,57 (0,54 - 0,61)	0,74 (0,72 - 0,77)	0,99 (0,99 - 0,99)
DT	0,79 (0,76 - 0,81)	0,67 (0,64 - 0,70)	0,82 (0,79 - 0,85)	0,82 (0,81 - 0,83)
SVM	1,00 (0,99 - 1,00)	0,85 (0,82 - 0,87)	0,99 (0,99 - 1,00)	0,99 (0,99 - 1,00)
RF	0,98 (0,97 - 0,99)	0,84 (0,82 - 0,87)	0,98 (0,97 - 0,99)	0,98 (0,97 - 0,99)
XGB	0,97 (0,96 - 0,98)	0,86 (0,84 - 0,88)	0,98 (0,97 - 0,99)	0,98 (0,97 - 0,99)
MLP	0,93 (0,91 - 0,96)	0,70 (0,67 - 0,74)	0,97 (0,95 - 0,99)	0,99 (0,98 - 1,00)

Fonte: Elaborada pelo autor (2021).

Nota-se que 12 dos 24 modelos tiveram acurácia média maior que 95% (com desempenho formatado em negrito na tabela), dos quais 9 utilizaram o banco de dados C e D como entrada (ambos compostos por informações provenientes de ambos sensores e imagens).

Além disso, ao comparar os conjuntos A e B, que utilizam exclusiva e respectivamente dados dos sensores e imagens, foi possível perceber uma discrepância na importância de informação para aprendizado dos modelos. Foi possível observar que os algoritmos mais complexos conseguem bons resultados (acurácia média acima de 90%) através do conjunto A, mas não utilizando o conjunto B.

Porém, apesar dos descritores de imagens não serem suficientes para que os algoritmos atinjam excelência, ao observar os resultados com os conjuntos C e D percebe-se que informações dos sensores e imagens podem funcionar de forma complementar. Os algoritmos KNN, DT e MLP tiveram um aumento considerável na média ao utilizar grupos que utilizam variáveis provenientes de ambas as fontes.

Não somente isso, mas ao utilizar o conjunto D, um subconjunto do grupo C, em todos os algoritmos houve um aumento ou a mesma média de acurácia. Pode-se concluir que apesar dos ganhos ao utilizar diferentes tipos de informações, algumas variáveis têm um valor agregado maior para o treinamento dos modelos para classificação do tipo de lâmpada.

Em específico no algoritmo KNN, é possível observar que há presença de ruído nos dados do conjunto C, visto que ao utilizar o grupo D como entrada há um aumento de 25% na acurácia média. Ou seja, algumas das variáveis não contribuem ou prejudicam o aprendizado de certos algoritmos.

Entretanto, outro fator a ser mencionado é que todos os algoritmos, com exceção de DT, conseguiram atingir uma média igual ou acima de 97% em pelo menos um banco de dados. Ou seja, os dados colhidos a partir de sensores escolhidos especificamente para emissão radioativa de cada tipo de lâmpada observado, assim como, a distinta diferença visual entre os tipos de lâmpadas (Figura 22), de fato refletem a distinção entre as três classes.

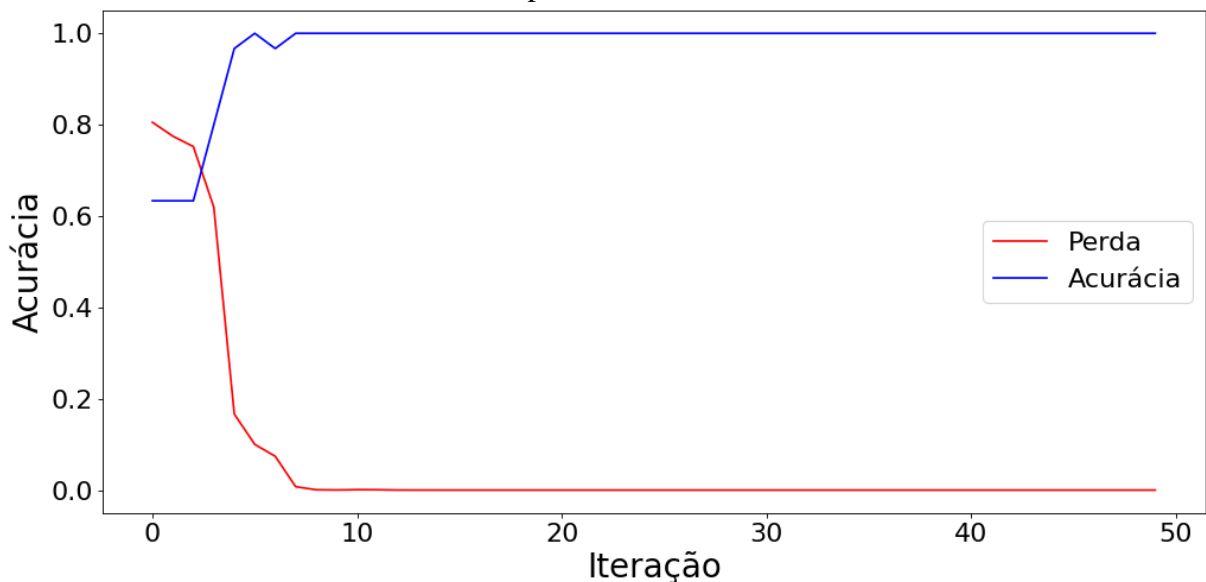
Figura 22 - Fotos de três tipos distintos de lâmpadas de pontos de iluminação pública: MV (esquerda), MH (centro) e HPS (direita).



Fonte: Adaptado de Soares et al. (2014).

Outro fator que reforça a vantagem que a distinção visual entre os três tipos de lâmpada tem neste tipo de classificação são os resultados dos experimentos com aprendizado profundo. No qual a CNN proposta funcionou de forma excepcional, conforme comportamento de uma repetição do modelo (representativa de todas as iterações realizadas) no qual, ilustrado na Figura 23, a cada *epoch* durante o treinamento a acurácia aumenta e a perda diminui no processo de validação.

Figura 23 - Acurácia e perda do processo de treinamento e validação de uma das iterações da rede neural classificadora de tipo.



Fonte: Elaborada pelo autor (2022).

Os resultados obtidos pela CNN com os conjuntos de teste foram: acurácia média de 0,99 com Intervalo de Confiança (95%) de 0,99 – 1,00 e perda média de 0,03 com Intervalo de Confiança (95%) de 0,00 – 0,05. A rede em todas as iterações conseguiu chegar aos melhores

desempenhos obtidos pelos algoritmos de abordagem tradicional e superou todos os resultados obtidos com uso exclusivo de descritores das imagens.

4.1.2 Desempenho dos classificadores de potência

Posteriormente, os resultados dos experimentos com os modelos da camada de previsão de potência também foram apurados, descritos na Tabela 24. Entretanto, o desempenho dos algoritmos foi inferior, como esperado, às métricas atingidas na primeira camada.

Tabela 24 - Acurácia média e Intervalo de Confiança (95%) dos classificadores de potência de lâmpada nos conjuntos de dados A, B, C e D com uso de normalização.

Modelo	A	B	C	D
KNN	0.55 (0.51 - 0.58)	0.33 (0.31 - 0.34)	0.47 (0.44 - 0.50)	0.79 (0.79 - 0.80)
DT	0.59 (0.55 - 0.62)	0.48 (0.46 - 0.50)	0.56 (0.53 - 0.59)	0.58 (0.58 - 0.59)
SVM	0.86 (0.84 - 0.88)	0.59 (0.57 - 0.61)	0.81 (0.79 - 0.84)	0.85 (0.82 - 0.87)
RF	0.77 (0.74 - 0.80)	0.64 (0.60 - 0.68)	0.80 (0.78 - 0.83)	0.82 (0.79 - 0.84)
XGB	0.76 (0.73 - 0.79)	0.62 (0.58 - 0.66)	0.80 (0.78 - 0.83)	0.81 (0.79 - 0.84)
MLP	0.66 (0.62 - 0.69)	0.47 (0.43 - 0.51)	0.75 (0.72 - 0.78)	0.84 (0.82 - 0.87)

Fonte: Elaborada pelo autor (2021).

Nota-se que apenas 8 dos 24 modelos tiveram a acurácia média igual ou superior a 80% (com desempenho formatado em negrito na tabela), e destes, somente 2 chegaram à marca de 85%. As melhores métricas foram atingidas pelo algoritmo SVM nos conjuntos A e D, seguido pelo MLP com 84% também no conjunto D.

Modelos com estruturas mais simples, como KNN e DT, que já apresentaram os resultados mais limitados na classificação do tipo, aqui tiveram ainda mais dificuldades para detectar padrões, principalmente nos conjuntos de dados menores e com menos fontes de informações.

Por conseguinte, assim como na primeira camada, percebe-se que apesar de, em geral, os algoritmos de abordagem tradicional performarem melhor nos conjuntos com informações de múltiplas fontes, dados providos de sensores demonstram também ter um valor para o treinamento dos modelos na classificação de potência.

O conjunto B, assim como na primeira camada, foi o grupo de variáveis que proporcionou o pior resultado para todos os algoritmos. Entretanto, continua provendo informações complementares importantes, ao passo que na maior parte dos casos, os resultados obtidos pelos modelos no conjunto A são otimizados ao utilizar o conjunto C ou D.

É válido mencionar que, nos experimentos realizados com os algoritmos de classificação de potência, considerou-se que a primeira camada acertou todas as classificações de tipos de lâmpadas, ou seja, foram utilizados os valores reais de tipo no conjunto de entrada na segunda camada.

Esta escolha se justifica primeiramente devido ao grande número de combinações possíveis para criação de modelos de estrutura hierárquica (considerando todas as variações de conjuntos de dados e algoritmos totalizam-se 576 modelos). Em segundo, à alta acurácia atingida por muitos dos modelos de classificação de tipos (10 atingindo o intervalo de 98-100% de acurácia média), e os tipos de erros observados.

Nos modelos de alto desempenho na classificação de tipos, apenas uma ou duas instâncias são classificadas de forma incorreta no conjunto de testes. Foi observado que estas instâncias eram preditas de forma errônea pelos algoritmos da segunda camada, mesmo utilizando a informação de tipo correta. Em outras palavras, embora a predição da potência se beneficie da informação do tipo no processo de classificação da potência, o erro de classificação nunca está diretamente ligado a esta.

Para verificar essa lógica, foram realizados testes com algoritmos de ambas as camadas funcionando de forma conjunta, conforme descrito na Figura 18. Assim, combinou-se os algoritmos com maiores acurácias em ambos os tipos de classificação (como por exemplo, um modelo utilizando o algoritmo KNN utilizando o conjunto D e SVM utilizando o conjunto A na primeira e segunda camada respectivamente). Em conformidade com a lógica acima, todos os resultados apresentaram média de acurácia igual à do algoritmo de classificação de potência.

Em outras palavras, dada ao alto desempenho dos algoritmos na primeira camada, o desempenho da estrutura hierárquica está diretamente ligado à capacidade do algoritmo da segunda camada em classificar as potências.

A alta complexidade na detecção de potências já era esperada devido ao maior número de classes (seis diferentes potências) e as dificuldades em se coletar dados de forma acurada acerca da assinatura de espectro das lâmpadas, como descrito anteriormente. De forma geral, algumas das causas do aumento da complexidade podem ser vistas ao se comparar a correlação teórica entre as variáveis e as classes a serem previstas.

Diferente da classificação de tipos, na qual a emissão de radioatividade medida pelos sensores está diretamente ligada à composição química da lâmpada e em aparência todas as classes eram distintas entre si, na predição de potências, além das influências de fatores externos na coleta dos dados relacionados ao espectro das lâmpadas, e não há uma distinção clara visualmente clara entre as classes para se extrair das imagens (Figura 24). Assim, de fato, os

dados dos sensores e imagens não apresentam comportamento tão distinto em relação às diferentes classes de potências como nos diferentes tipos de lâmpadas.

Figura 24 - Fotos de duas lâmpadas de 150W: MH (esquerda) e HPS (direita).

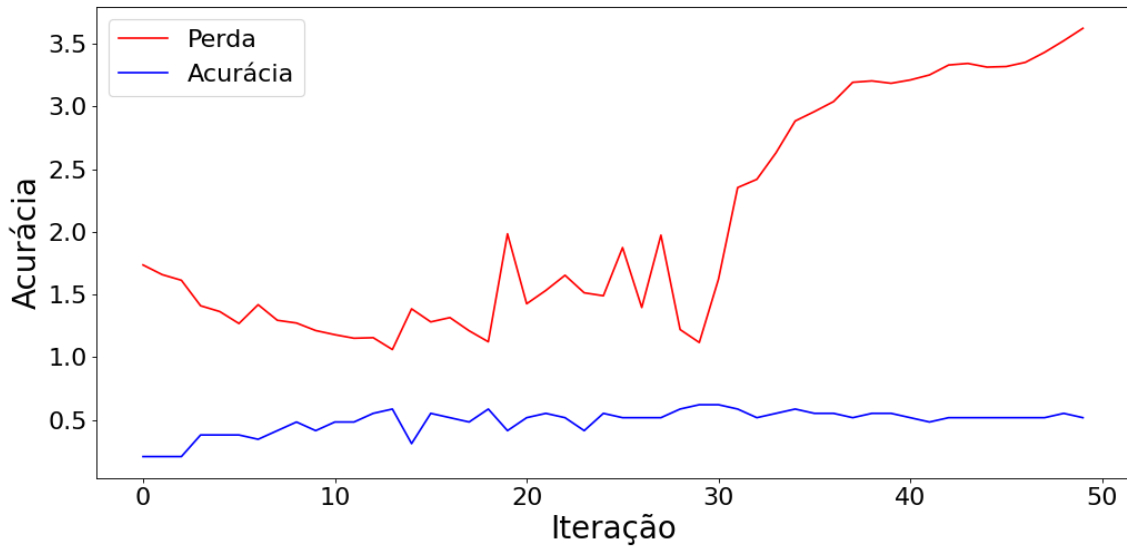


Fonte: Adaptado de Soares et al. (2014).

No entanto, as estratégias utilizadas, como a criação do grupo D e o posicionamento desta camada subsequente à classificação de tipos para que esta primeira predição pudesse auxiliar na predição de potência, de fato otimizaram a acurácia dos modelos comparado a testes empíricos realizados inicialmente.

De forma similar, os experimentos com aprendizado profundo também tiveram um desempenho inferior, conforme ilustra a Figura 25, que por utilizar apenas as imagens e as classes não terem a distinção visual às classes de tipos dispõe de bem menos informações para distinguir uma classe da outra. A CNN atingiu uma acurácia média de 0,50 com Intervalo de Confiança (95%) de 0,46 – 0,54 e perda média de 4,27 com Intervalo de Confiança (95%) de 3,59 – 4,96.

Figura 25 - Acurácia e perda do processo de treinamento e validação de uma das iterações da rede neural classificadora de potência.

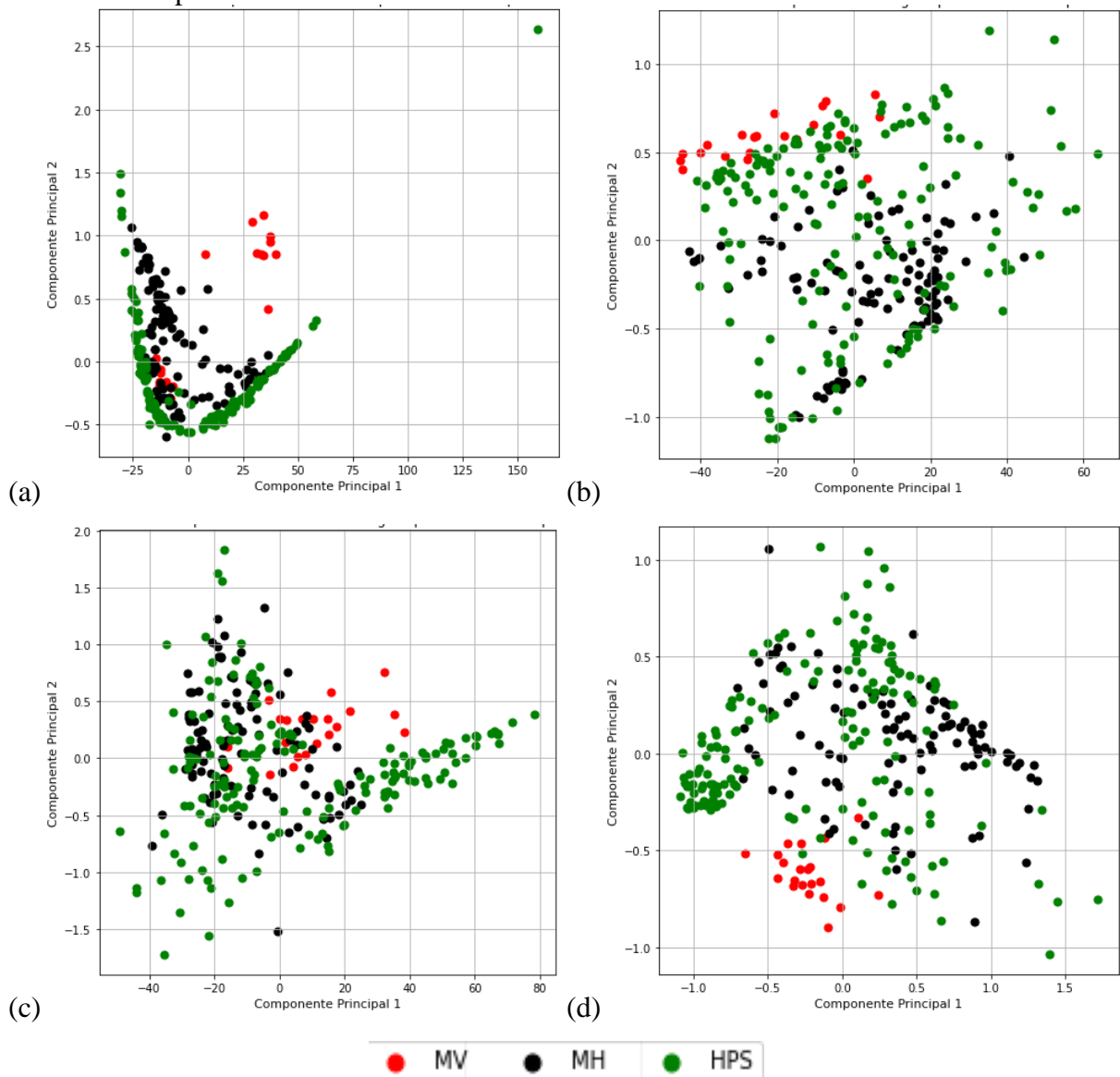


Fonte: Elaborada pelo autor (2021).

Outro fator interessante na performance da CNN é a inconstância de acerto por classe a cada iteração. Em grande parte das repetições realizadas, é possível observar que para certa divisão de dados, o algoritmo consegue acertar com excelência todas as instâncias de uma classe, apenas para em outra divisão errar a predição de todas as instâncias desta mesma classe. Este fenômeno ocorre com todas as categorias de potências e acredita-se que a causa seja a similaridade entre as classes que só agravado pelo uso exclusivo de informações visuais.

Para visualizar o comportamento dos dados em relação às classes a serem previstas, foi utilizado o PCA (*Principal Component Analysis*) para reduzir a dimensão dos conjuntos de dados para 2 para elaboração de gráficos. Na Figura 26, pode-se observar como os dados estão dispostos em relação ao tipo de lâmpada. Apesar de não haver uma distinção ideal entre todas as classes, é possível observar que há uma separação notável entre as classes, especialmente nos conjuntos com variáveis providas pelos sensores e com variáveis selecionadas.

Figura 26 - Visualização dos conjuntos de dados representando os três diferentes tipos de lâmpadas.



Fonte: Elaborada pelo autor (2021).

Legenda: a) PCA do conjunto de dados A em relação ao tipo de lâmpada.

b) PCA do conjunto de dados B em relação ao tipo de lâmpada.

c) PCA do conjunto de dados C em relação ao tipo de lâmpada.

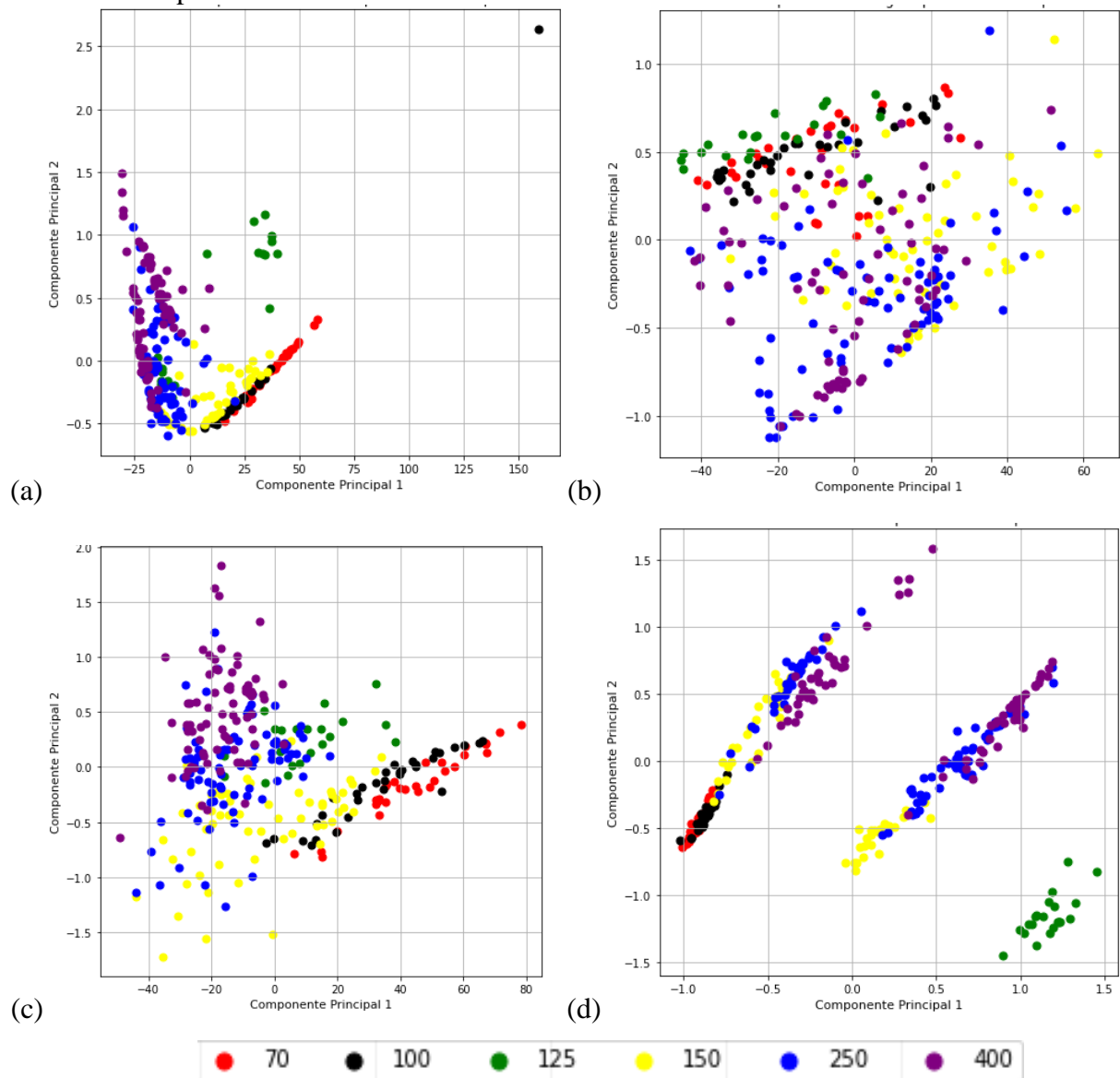
d) PCA do conjunto de dados D em relação ao tipo de lâmpada.

Na Figura 27, tem-se representação dos dados em relação às classes de potência. É possível observar que a disposição dos dados em relação a potência é mais caótica, sem divisão clara dos conjuntos e sobreposição de classes.

As consequências desta desordem são vistas nos erros da classificação de tipos na qual praticamente a totalidade de erros estão entre as classes mais similares (MH e HPS). Já no processo de predição de potências, isso se torna ainda mais grave, devido ao número maior de classes e dois tipos de lâmpadas dividirem as classes de 150 e 250W. Neste caso, a exceção é a

classe de 125W, observar itens a e d da Figura 28, que é exclusivamente do tipo MV. Logo, ao utilizar o tipo de lâmpada na segunda camada de classificação, tem-se uma facilidade maior para distinguir esta classe.

Figura 27 - Visualização dos conjuntos de dados representando as seis diferentes potências de lâmpada.



Fonte: Elaborada pelo autor (2021).

Legenda: a) PCA do conjunto de dados A em relação à potência da lâmpada.

b) PCA do conjunto de dados B em relação à potência da lâmpada.

c) PCA do conjunto de dados C em relação à potência da lâmpada.

d) PCA do conjunto de dados D em relação à potência da lâmpada.

Em suma, os algoritmos têm melhor desempenho e versatilidade na classificação de tipos, atingindo um bom desempenho em diferentes conjuntos de dados e complexidade de

métodos, já na classificação de potência é notável um salto de dificuldade no processo de detecção de padrões que limita o desempenho do modelo de estrutura hierárquica em geral.

4.2 COMPARAÇÃO DE MODELOS COM ESTRUTURA DE CLASSIFICAÇÃO ÚNICA

Nos modelos com estrutura de classificação única, um único algoritmo fica responsável pela classificação conjunto de tipo e potência, identificando o arquétipo de cada lâmpada. Os resultados dos experimentos com esta estrutura de modelo estão descritos na Tabela 25, de maneira similar à seção anterior.

Tabela 25 - Acurácia média e Intervalo de Confiança (95%) dos classificadores únicos nos conjuntos de dados A, B, C e D com uso de normalização.

Modelo	A	B	C	D
KNN	0.51 (0.48 - 0.55)	0.27 (0.24 - 0.30)	0.42 (0.39 - 0.46)	0.78 (0.77 - 0.78)
DT	0.50 (0.47 - 0.53)	0.39 (0.36 - 0.42)	0.52 (0.48 - 0.55)	0.49 (0.48 - 0.50)
SVM	0.86 (0.83 - 0.88)	0.61 (0.59 - 0.63)	0.83 (0.81 - 0.85)	0.86 (0.84 - 0.88)
RF	0.79 (0.76 - 0.82)	0.65 (0.61 - 0.69)	0.83 (0.80 - 0.86)	0.84 (0.80 - 0.85)
XGB	0.76 (0.73 - 0.78)	0.62 (0.59 - 0.64)	0.82 (0.79 - 0.85)	0.83 (0.80 - 0.86)
MLP	0.67 (0.63 - 0.70)	0.51 (0.47 - 0.54)	0.79 (0.76 - 0.82)	0.84 (0.82 - 0.87)

Fonte: Elaborada pelo autor (2021).

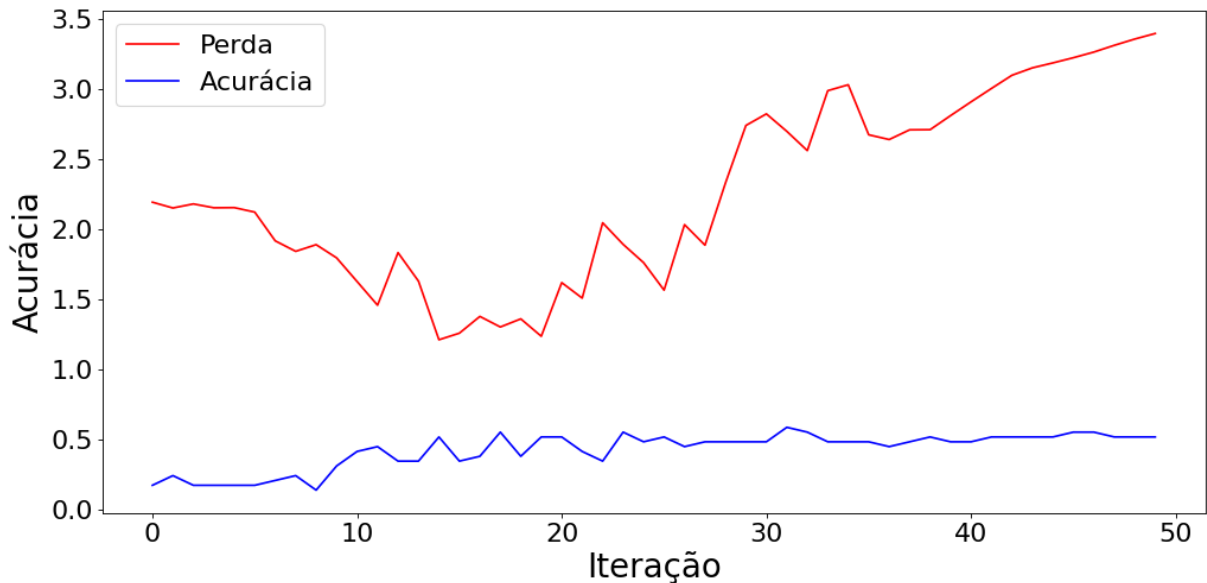
Inopinadamente, observa-se em todos os modelos de classificação única, um resultado similar aos algoritmos com mesma configuração na segunda camada da estrutura hierárquica. De maneira idêntica, 8 dos 24 modelos atingiram uma média de acurácia igual ou superior à 80% (com desempenho formatado em negrito na tabela) e os três melhores rendimentos foram obtidos pelas mesmas configurações de modelos (respectivamente SVM com o conjunto A, SVM e MLP utilizando o conjunto D).

Esta semelhança pode ser explicada pela complexidade de classificação da potência explicada anteriormente. Como grande parte dos algoritmos de classificação de tipos de lâmpadas funcionam com praticamente 100% de precisão, a performance do modelo de classificação hierárquica é limitada pela acurácia da classificação de potência.

Assim, essa similaridade entre os resultados reforça a limitação causada pelos padrões, ou ausência destes, para classificação de potência. Pois, mesmo com um aumento no número de classes, de seis para nove, que reflete em um número maior de padrões a ser reconhecido pelo modelo, com a similaridade de classes, cria-se a necessidade de uma sensibilidade maior do algoritmo para detectar sutilezas nos dados, os resultados foram aproximadamente os mesmos.

Esta suposição é reforçada também pelo desempenho dos experimentos de classificação única realizados com a CNN similares à classificação de potência, ilustrado na Figura 28. A rede atingiu acurácia média de 0,62 com Intervalo de Confiança (95%) de 0,43 – 0,80 e perda média de 3,83 com Intervalo de Confiança (95%) de 3,13 – 4,52.

Figura 28 - Acurácia e perda do processo de treinamento e validação de uma das iterações da rede neural classificadora de única.

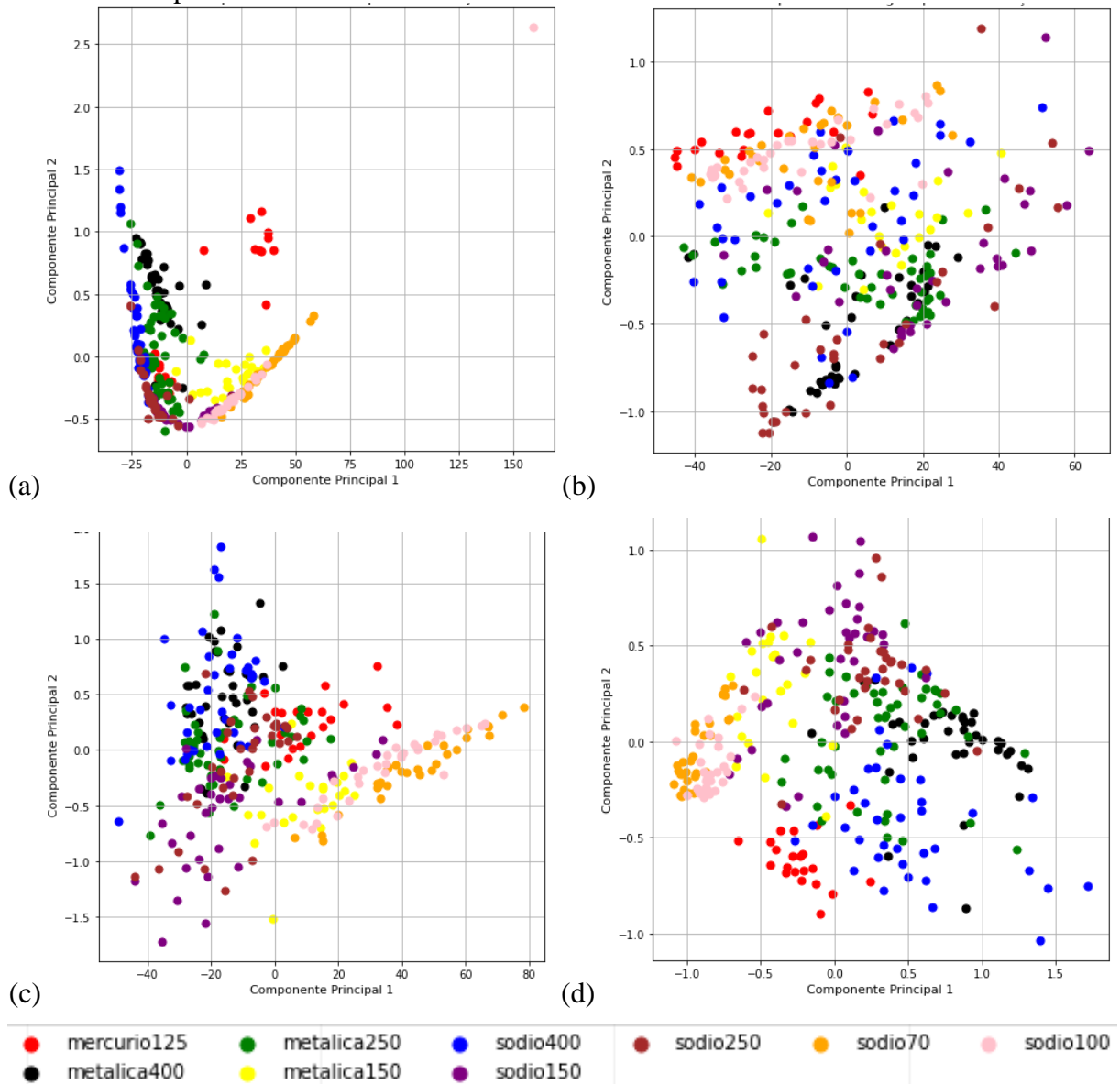


Fonte: Elaborada pelo autor (2022).

Apesar do aumento na acurácia média em comparação com os resultados da CNN na classificação de potência e dos modelos com algoritmos tradicionais que utilizam exclusivamente descritores de imagem, ainda assim é um desempenho inferior aos melhores resultados dos modelos que utilizam com outros grupos de dados, agravado pela inconsistência no aprendizado de classificação de classes com diferentes repartições da base de dados.

Ou seja, conclui-se que acontece a mesma limitação enfrentada pela segunda camada da estrutura hierárquica. Conforme é possível observar na Figura 29, os dados em relação às classes únicas têm uma relação similar dos dados em relação às classes de potência. Com exceção da classe da lâmpada de mercúrio de 125W, como é possível observar mais claramente nos gráficos referentes aos conjuntos A e D, as demais classes não tem uma distinção muito clara.

Figura 29 - Visualização dos conjuntos de dados representando os nove diferentes modelos de lâmpada.



Fonte: Elaborada pelo autor (2021).

Legenda: a) PCA do conjunto de dados com *features* de sensores em relação às classes de classificação única.

b) PCA do conjunto de dados com *features* de imagens em relação às classes de classificação única.

c) PCA do conjunto de dados completo em relação às classes de classificação única.

d) PCA do conjunto de *features* selecionadas em relação às classes de classificação única.

Portanto, apesar dos empecilhos, a estrutura de classificação única se mostra uma alternativa tão eficiente quanto a, previamente proposta, estrutura de classificação hierárquica, ambas com acurácias aproximadamente idênticas.

4.3 RESULTADOS COM USO DE PADRONIZAÇÃO

Em contrapartida, os modelos que utilizaram a padronização como método de transformação dos dados tiveram um desempenho inferior, descritos nas tabelas 26, 27 e 28 em comparação aos resultados obtidos com uso da normalização apresentados nas seções anteriores.

Tabela 26 - Acurácia média e Intervalo de Confiança (95%) dos classificadores de tipo nos conjuntos de dados A, B, C e D com uso de padronização.

Modelo	A	B	C	D
KNN	0,81 (0,78 – 0,84)	0,56 (0,54 – 0,59)	0,55 (0,52 – 0,58)	0,83 (0,80 – 0,85)
DT	0,79 (0,76 – 0,81)	0,67 (0,64 – 0,70)	0,70 (0,67 – 0,74)	0,74 (0,71 – 0,77)
SVM	1,00 (0,99 - 1,00)	0,81 (0,78 – 0,84)	0,82 (0,79 – 0,82)	0,84 (0,82 – 0,87)
RF	0,98 (0,97 – 0,99)	0,77 (0,74 – 0,79)	0,86 (0,83 – 0,88)	0,86 (0,84 – 0,89)
XGB	0,78 (0,75 – 0,80)	0,75 (0,73 – 0,78)	0,87 (0,85 – 0,89)	0,86 (0,84 – 0,88)
MLP	0,93 (0,91 – 0,96)	0,67 (0,65 – 0,70)	0,76 (0,73 – 0,78)	0,85 (0,82 – 0,88)

Fonte: Elaborada pelo autor (2022).

Tabela 27 - Acurácia média e Intervalo de Confiança (95%) dos classificadores de potência nos conjuntos de dados A, B, C e D com uso de padronização.

Modelo	A	B	C	D
KNN	0,34 (0,32 – 0,37)	0,24 (0,21 – 0,27)	0,27 (0,24 – 0,30)	0,79 (0,79 – 0,80)
DT	0,36 (0,34 – 0,39)	0,35 (0,32 – 0,37)	0,37 (0,33 – 0,40)	0,39 (0,37 – 0,41)
SVM	0,69 (0,65 – 0,72)	0,45 (0,43 – 0,48)	0,55 (0,52 – 0,58)	0,53 (0,50 – 0,57)
RF	0,42 (0,38 – 0,45)	0,45 (0,43 – 0,48)	0,50 (0,47 – 0,53)	0,54 (0,51 – 0,58)
XGB	0,43 (0,39 – 0,47)	0,43 (0,40 – 0,47)	0,50 (0,47 – 0,53)	0,54 (0,51 – 0,57)
MLP	0,40 (0,36 – 0,43)	0,34 (0,31 – 0,37)	0,41 (0,37 – 0,44)	0,55 (0,52 – 0,58)

Fonte: Elaborada pelo autor (2022).

Tabela 28 - Acurácia média e Intervalo de Confiança (95%) dos classificadores únicos nos conjuntos de dados A, B, C e D com uso de padronização.

Modelo	A	B	C	D
KNN	0,51 (0,48 – 0,55)	0,17 (0,15 – 0,20)	0,19 (0,16 – 0,22)	0,42 (0,39 – 0,45)
DT	0,50 (0,47 – 0,53)	0,30 (0,27 – 0,33)	0,33 (0,29 – 0,36)	0,33 (0,30 – 0,37)
SVM	0,86 (0,83 - 0,88)	0,44 (0,41 – 0,47)	0,43 (0,40 – 0,46)	0,50 (0,47 – 0,53)
RF	0,79 (0,76 – 0,82)	0,39 (0,37 – 0,41)	0,49 (0,46 – 0,52)	0,46 (0,43 – 0,49)
XGB	0,76 (0,76 – 0,78)	0,39 (0,36 – 0,42)	0,47 (0,44 – 0,50)	0,49 (0,46 – 0,52)
MLP	0,67 (0,63 – 0,70)	0,23 (0,20 – 0,26)	0,33 (0,29 – 0,36)	0,47 (0,44 – 0,50)

Fonte: Elaborada pelo autor (2022).

Nesse sentido, nota-se que apenas 3 modelos se igualam aos resultados obtidos com normalização, dois classificadores de tipo e um de classificação única. O restante dos modelos todos apresentaram uma queda brusca na performance de predição correta de classes.

Uma possível razão para isto é a diversidade de distribuições das variáveis, pois de acordo com Chadha (2021), a padronização dos dados é preferível e mais eficaz para variáveis

com curva de sino, ou seja, com distribuição normal. E, conforme os apêndices G e H, é possível observar que nenhuma das variáveis apresenta a distribuição desejada para a transformação.

Portanto, devido ao baixo desempenho destes modelos, os resultados obtidos com padronização não serão considerados durante a comparação de modelos na próxima seção.

4.4 COMPARAÇÃO ENTRE OS MODELOS

Tipicamente a avaliação e comparação de modelos são feitas através do resultado obtido pelo defronte das estatísticas descritivas das performances em métodos de validação cruzada. Entretanto, para garantir a escolha do melhor modelo, de fato, e evitar escolher um modelo que por chances estatísticas teve melhor desempenho, se faz necessário o uso de métodos estatísticos.

Neste caso, foi utilizado teste t de Student pareado e unilateral de amostras dependentes para comparação das médias de acurácia e verificar se há ou não evidências para rejeitar ou não a hipótese de um modelo ser superior a outro à nível de 5% de significância.

Para facilitar a leitura, o nome dos modelos nos quadros foi abreviado de forma que a primeira letra representa o conjunto de dados que o modelo utiliza (s para conjunto de dados A, i para conjunto de variáveis B, c para o banco de dados C e f para o conjunto D), a segunda representa a classificação que o modelo faz (t para classificação do tipo, p para classificação da potência e u para classificação única), e por fim, separado por “_” a sigla do algoritmo utilizado.

No Apêndice I, estão descritos os resultados dos testes t comparando todos os modelos de classificação do tipo de lâmpada. Neste, mostra-se que de fato os melhores modelos são os que utilizaram o conjunto de dados A (SVM), completo (SVM, RF, XGB e MLP) e especialmente o grupo de características D (todos, com exceção do DT).

Entretanto, apesar destes terem uma certa diferença de acurácia entre si (de 1-2%) como pode ser visto na Tabela 23, pelos testes estatísticos performados, não há evidência de que haja diferença significativa entre as médias de acurácia em nível de 5% de significância.

Já no Apêndice J, estão detalhados os resultados dos testes t para os modelos de classificação de potência. Este que mostra que os modelos que utilizaram o conjunto D que tiveram os melhores resultados (SVM, RF, XGB e MLP) e o modelo SVM com os dados de A que atingiu a maior média de acurácia (86%), de fato são superiores aos demais modelos, porém não há evidências de diferenças significativas na média de acurácia destes a nível de 5% de significância.

Portanto, a escolha dos melhores modelos para esta estrutura hierárquica fica sujeito aos critérios de aplicação. Para a primeira etapa de classificação, tem-se 11 modelos com acurácia acima de 95% para os quais não há evidência de diferença em suas médias de performance, logo, a escolha de qual modelo será mais adequado pode ser baseada em outros critérios, tais como: custo computacional, complexidade, quantidade de variáveis de entrada, entre outros. Na segunda etapa, para classificação de potência, também se tem 5 modelos que não apresentam diferença significativa entre suas médias de acurácia a nível de 5%.

A escolha mais óbvia seria utilizar os modelos `st_svm` e o modelo `sp_svm`, pois foram os modelos com maior média de acurácia em suas categorias, utilizam um número menor de variáveis, permite trabalhar com um algoritmo apenas, além de ter um custo computacional menor que parte dos algoritmos propostos. Entretanto, como não há diferença significativa na acurácia, pode-se escolher dentre os modelos com melhores resultados os que melhor se adaptarem aos critérios de uma possível aplicação real.

No Apêndice K, estão descritos os resultados dos testes para a estrutura classificação única. E, de forma similar às estatísticas descritivas, estes se assemelham bastante aos resultados dos classificadores de potência, sendo os modelos com performance superiores: `fu_svm`, `fu_rf`, `fu_xgb`, `fu_mlp` e `su_svm`. Da mesma forma, estes não apresentam diferença significativa entre suas médias de acurácia a nível de 5%, mesmo `su_svm` apresentando a média de acurácia superior de 86%. Assim, definir qual o melhor modelo para classificação única também não se restringe apenas a performance do modelo.

Por fim, foram feitos testes t para comparar os melhores modelos de cada estrutura entre si (`sp_svm`, `fp_svm`, `fp_rf`, `fp_xgb`, `fp_mlp`, `su_svm`, `cu_rf`, `cu_xgb`, `fu_svm`, `fu_rf`, `fu_xgb` e `fu_mlp`). No Apêndice L estão descritos os resultados de todos os testes realizados, nele é possível observar que apenas 3 modelos (`fp_rf`, `fp_xgb` e `cu_xgb`) apresentaram diferença inferior significativa em comparação aos modelos `sp_svm` e `fu_svm`. Nos demais modelos, não há evidências de diferenças significativas em suas médias de acurácia.

Assim, nota-se pela Tabela 29 que todos os modelos selecionados são formados por algoritmos mais complexos (SVM, RF, XGB e MLP), que dois terços dos modelos são de estrutura de classificação única, dois terços são modelos que utilizam o banco de dados D e 4 dos 9 modelos são formados pelo algoritmo SVM.

Resultados estes que reforçam as observações feitas anteriormente, como o efeito do uso de grupos de características que combinam variáveis originadas de ambos sensores e imagens, da eliminação de variáveis redundantes e ruídos nos dados, a necessidade de

algoritmos mais complexos para detecção das classes de potências e de classificação única, entre outras.

Tabela 29 - Acurácia média e Intervalo de Confiança (95%) dos melhores modelos.

Modelo	Acurácia média (Intervalo de Confiança)
sp_svm	0.86 (0.84 - 0.88)
fp_svm	0.85 (0.82 - 0.87)
fp_mlp	0.84 (0.82 - 0.87)
su_svm	0.86 (0.83 - 0.88)
cu_rf	0.83 (0.80 - 0.86)
fu_svm	0.86 (0.84 - 0.88)
fu_rf	0.84 (0.80 - 0.85)
fu_xgb	0.83 (0.80 - 0.86)
fu_mlp	0.84 (0.82 - 0.87)

Fonte: Elaborada pelo autor (2021).

Dessa forma, assim como as comparações entre modelos anteriores, a definição de qual modelo é melhor irá depender das condições de aplicação prática. Dito isso, a opção mais segura talvez seja utilizar modelos com algoritmo SVM, visto a versatilidade no problema proposto, apresentando resultados satisfatórios em diferentes tipos de classificação com diversos conjuntos de dados.

Pode-se argumentar também uma certa vantagem dos modelos de classificação única por depender apenas da variação e incerteza de um único modelo de classificação. Entretanto, como não há uma métrica de comparação eliminatória para todos os cenários, não é possível afirmar a superioridade de um modelo sobre o outro.

4.5 COMPARAÇÃO COM RESULTADOS DE SOARES *ET AL.* (2013, 2014, 2015), BROETTO *ET AL.* (2013) E BROETTO E VAREJÃO (2016)

Apesar dos trabalhos anteriores se tratarem de conjuntos diferentes de instâncias, estrutura de classificação diferente e algumas informações desconhecidas sobre a rotina de treinamento dos modelos (divisão dos dados, tamanho dos conjuntos de treinamento, validação e teste, entre outros), há similaridades suficientes entre as metodologias utilizadas para viabilizar uma comparação direta de performance.

Ao comparar acurácia média com Soares *et al.* (2014, 2015), mesmo com um processo de validação mais rigoroso, observa-se que a CNN e 10 dos classificadores de tipo de abordagem tradicional, de diferentes níveis de complexidade) ficaram acima do melhor

percentual obtido anteriormente (0,97) por Soares *et al* (2014, 2015). Já na classificação única, em conjuntos de dados sem seleção de características, apenas o modelo com o algoritmo SVM se sobressaiu aos 0,823 obtidos anteriormente, e com seleção de características, dois modelos superaram a média de 0,830.

Por conseguinte, ao comparar os modelos obtidos com Broetto *et al.* (2013), mesmo com um número maior de classes, foi possível atingir resultados similares utilizando apenas informações providas de imagens com a CNN proposta.

Outro ponto a ser apontado é a diferença entre as estruturas hierárquicas, pois utilizando apenas um modelo para classificação de potência (e não 3, um para cada classe de tipo) foi possível atingir resultados equivalentes aos apresentados pelos estudos predecessores.

Já Broetto e Varejão (2016) conseguiram resultados superiores chegando a 89,56% com modelos que utilizaram algoritmos de seleção de características. Entretanto, assim como Soares *et al.* (2014, 2015), utilizou-se apenas validação cruzada como método de validação.

Entretanto, os resultados obtidos reafirmam muitas das conclusões feitas previamente, principalmente por Soares *et al.* (2014, 2015). Mesmo com um método de validação superior e resultados equivalentes ou acima, os modelos tiveram os mesmos problemas para detectar padrões das classes de potências e na classificação única. Mas, ao mesmo tempo, nota-se com os resultados dos experimentos em laboratório realizados por Soares *et al.* (2013, 2014, 2015) que com informações que estejam diretamente correlacionadas com a potência, como a altura do poste, tiveram uma acurácia melhor, pode-se aumentar o desempenho dos classificadores de potências. Assim como Broetto *et al.* (2013) apontou, os descritores de imagem também foram relevantes e funcionam como informação auxiliar para o treinamento dos modelos de classificação.

E, por fim, Broetto *et al.* (2016), apesar de não haver comprovação que a diferença entre as acurácias é significativa, indica que o uso de técnicas mais complexas e robustas de seleção de características podem auxiliar no desempenho do modelo.

5 CONCLUSÃO

Este estudo descreve a metodologia de aprimoramento do modelo inteligente de identificação de luminárias de iluminação pública proposto por Soares *et al.* (2014, 2015), que tem como objetivo reduzir perdas comerciais substanciais por parte da companhia de distribuição de energia causadas pela quantidade substancial de informações que são passadas incorretamente ou não são informadas por parte das prefeituras.

Assim, por se tratar de um problema de detecção de padrões, a primeira abordagem consistiu em testar a performance de seis algoritmos tradicionais de aprendizagem de máquina (KNN, DT, RF, SVM, XGB e MLP) com os dados de luminárias de iluminação pública coletadas em campo.

Foram realizados experimentos com diferentes conjuntos de dados, utilizando exclusivamente dados obtidos pelos sensores, apenas dados de imagens e utilizando dados dos sensores e das imagens. Um quarto conjunto de dados também foi utilizado, este obtido através de seleção de características nos dados de sensores e imagens com o objetivo de eliminar dados irrelevantes, redundantes e ruídos. Além disso, foram performados experimentos com aprendizado profundo utilizando uma CNN que usa como entrada apenas imagens.

Após a realização de todos os experimentos, e comparação entre os mesmos utilizando teste t pareado unilateral de Student para duas amostras dependentes, foram obtidos 9 modelos que a nível de 5% de significância não apresentam diferença em suas acurácias médias (83-86%). Nestes experimentos se destaca o algoritmo SVM com sua flexibilidade de aplicação e resultados notáveis nos 3 tipos de classificação, nos 4 conjuntos de dados e em ambas as arquiteturas testadas.

Os resultados dos experimentos que usaram apenas dados dos sensores foram consideravelmente superiores aos que utilizaram apenas dados das imagens. Isto indica que, nesse caso, os dados dos sensores têm mais informações que facilitam o discernimento entre as classes. Algo que é agravado pela limitação de tamanho do banco de dados (apenas 297 instâncias), tamanho inferior ao indicado para modelos mais complexos como CNN e dificulta, ou mesmo impossibilita, o modelo de detectar certas nuances e peculiaridades que distingue as classes a serem preditas.

Entretanto, os resultados de experimentos que utilizam tanto dados dos sensores e das imagens foram melhores que os de modelos que utilizam informações de uma fonte exclusiva, o que indica a existência de informações complementares nos dados dos sensores e das imagens. Pode ser observado também que a seleção de características, de forma geral, melhorou a

performance dos modelos, além de ser o conjunto de dados de 6 dos 9 melhores modelos obtidos.

Além disso, apesar da variedade de complexidade dos modelos que conseguiram bons resultados com a classificação de tipos, na classificação de potência encontrou-se uma dificuldade sistemática na detecção de padrões que também é mencionado nos trabalhos publicados anteriormente. Para contornar este problema, e possíveis temas para trabalhos futuros, pode-se buscar uma expansão do conjunto de dados através de coleta de dados de novas instâncias e também obtenção de dados de novas variáveis com informações relevantes para as classes de potências, como a altura do poste que foi utilizada nos estudos em laboratório de Soares *et al.* (2013, 2014, 2015), análise dos tipos de lâmpadas que se encontram em pontos de iluminação vizinhos (visto que lâmpadas de pontos próximos geralmente tem a mesma configuração), construção de modelos com algoritmos e estruturas mais complexos combinados com técnicas mais robustas de seleção de variáveis.

REFERÊNCIAS

- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (BRASIL). **Manual para elaboração do Programa Anual de Combate ao Desperdício de Energia Elétrica da ANEEL**. Brasília, DF: ANEEL, 1999. Disponível em: www.aneel.gov.br. Acesso em: 13 jun. 2021.
- AGUERA, Roger Silveira. **Cenário brasileiro da iluminação pública**. 2015. 68 f. Monografia (Graduação em Engenharia Elétrica) - Universidade de São Paulo, SP, 2015. Disponível em: <http://www.tcc.sc.usp.br/tce/disponiveis/18/180500/tce-19042016-113314/?&lang=br>. Acesso em: 13 jun. 2021.
- AMORIM, Leandro Gomes de Paiva. **Utilização de sistemas dedicados e protocolos de rede aplicados à eficiência energética da iluminação pública**. 2011. 109 f. Dissertação (Mestrado em Ciências) - Programa de Pós-graduação em engenharia Elétrica e Computação, Universidade Federal do Rio Grande do Norte, Natal, RN, 2011. Disponível em: https://repositorio.ufrn.br/jspui/bitstream/123456789/15374/1/Leonardo_GPA DISSERT.pdf. Acesso em: 13 jun. 2021.
- AYYADEVARA, V. K. **Pro machine learning algorithms: a hands-on approach to implementing algorithms in Python and R**. New York: Apress Media, 2018. Disponível em: https://www.researchgate.net/publication/326084761_Pro_Machine_Learning_Algorithms. Acesso em: 10 nov. 2021.
- BATURYNSKA, I.; MARTINSEN, K. Prediction of geometry deviations in additive manufactured parts: comparison of linear regression with machine learning algorithms. **Journal of Intelligent Manufacturing**, [s. l.], v. 32, n.1, p. 179-200, 2021. Disponível em: <https://link.springer.com/content/pdf/10.1007/s10845-020-01567-0.pdf>. Acesso em: 10 out. 2021.
- BEKKERMAN, R. **The present and the future of the kdd cup competition: an outsider's perspective**. LinkedIn, 25 ago. 2015. Disponível em: <https://www.linkedin.com/pulse/present-future-kdd-cup-competition-outsiders-ron-bekkerman/>. Acesso em: 10 out. 2021.
- BHARATHIDASON, S.; VENKATAESWARAN, J. Improving classification accuracy based on random forest model with uncorrelated high performing trees. **International Journal of Computer Applications**, [s. l.], v. 101, n. 13, p. 26-30, 2014. Disponível em: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.800.1781&rep=rep1&type=pdf>. Acesso em: 10 out. 2021.
- BISHOP, C. **Neural networks for pattern recognition**. Oxford: Oxford University Press, 1995.
- BLYTH, T. S.; ROBERTSON E. F. **Basic linear algebra**. 2. ed. London: Springer - Verlag, 2005.
- BOLLINGER, Joshua David. **Applications of solar energy to power stand-alone area and street lighting**. Dissertação (Mestrado em Engenharia Elétrica) - University of Missouri-Rolla, Rolla, MO, 2007. Disponível em:

https://scholarsmine.mst.edu/cgi/viewcontent.cgi?article=6505&context=masters_theses. Acesso em: 13 ago. 2021.

BOULESTEIX, A. L. *et al.* Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. **Data Mining and Knowledge Discovery**, [s. l.], v. 2, n. 6, 2012. Disponível em: <https://www.semanticscholar.org/paper/Overview-of-random-forest-methodology-and-practical-Boulesteix-Janitza/b779b41898931373f6e1f302b0a0d9a38313956a>. Acesso em: 13 ago. 2021.

BRASIL. **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal, 1988.

BREIMAN, L. Random forests. **Machine learning**, [s. l.], v. 45, n. 1, p. 5–32, 2001. Disponível em: <https://link.springer.com/article/10.1023/A:1010933404324>. Acesso em: 01 mar. 2022.

BROETTO, R. S. *et al.* **Automated image recognition of public lighting luminaries**. In: Electronics, Circuits, and Systems (ICECS), 2013 IEEE 20th International Conference on, p. 657–660, 2013.

BROETTO, R. S.; VAREJÃO, F. M. **Heterogeneous feature models and feature selection applied to detection of street lighting lamps types and wattages**. In: IECON 2016, p. 933-938, 2016.

BROWNLEE, J. **Development and evaluation of naïve classifier strategies using probability**. Machine Learning Mastery, Probability, set. 2019. Disponível em: <https://machinelearningmastery.com/how-to-develop-and-evaluate-naive-classifier-strategies-using-probability/>. Acesso em: 01 mar. 2022.

BROWNLEE, J. **Nested cross-validation for machine learning with python**. Python Machine Learning, 29 jul. 2020. Disponível em: <https://machinelearningmastery.com/nested-cross-validation-for-machine-learning-with-python/>. Acesso em: 13 ago. 2021.

BUSE, M.; HEFEKER, C. Political risk, institutions and foreign direct investment. **European Journal of Political Economy**, [s. l.], v. 23, n. 2, p. 397-415, 2007. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0176268006000267>. Acesso em: 13 ago. 2021.

CAMACHO, C. **Convolutional Neural Networks: machine and deep learning educator**. 2018. Disponível em: https://cezannec.github.io/Convolutional_Neural_Networks. Acesso em: 01 abr. 2022.

CHADHA, A. S. **Normalization and standardization mean: when to normalize data and when to standardize data?**. Towards Data Science, jul. 2021. Disponível em: <https://akalbir.medium.com/what-does-feature-scaling-mean-when-to-normalize-data-and-when-to-standardize-data-c3de654405ed>. Acesso em: 01 abr. 2022.

CHAIRA, T. **Medical image processing: advanced fuzzy set theoretic techniques**. CRC Press, 2015.

CHIKAMAI *et al.* Combining Feature Methods for Content-based Classification of Mammogram Images. **International Journal of Computers, Communications and Control (IJCCC)**, Jul. 2013

CONNORS, R.; HARLOW, C. A theoretical comparison of texture algorithms. **IEE Trans Pattern Anal Mach. Intell**, [s. l.], v. 2, n. 3, p. 204-222, 1980. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/21868894/>. Acesso em: 01 abr. 2022.

CORTES, C.; VAPNIK, V. Support-vector network. **Machine Learning**, [s. l.], v. 20, p. 1-25, 1995. Disponível em: <https://link.springer.com/article/10.1007/BF00994018>. Acesso em: 1 ago. 2021.

CRIMMINS, T. A complete set of Fourier descriptors for two-dimensional shapes. **IEE Trans. Syst.**, [s. l.], v. 12, n. 6, p.848-855, 1982. Disponível em: <https://ieeexplore.ieee.org/document/4308918>. Acesso em: 1 ago. 2021.

CUTLER, R *et al.* Random forests for classification in ecology. **Ecological Society of America**, [s. l.], v. 88, n. 11, p. 2783–2792, 2007. Disponível em: <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/07-0539.1>. Acesso em: 01 abr. 2022.

DAO, H. **Image Classification Using Convolutional Neural Networks**. Information Technology, Oulu, 2020. Disponível em: <https://www.theseus.fi/bitstream/handle/10024/276486/thesis%20-%20Hung%20Dao%20-%20final.pdf?sequence=2&isAllowed=y>. Acesso em: 05 jan. 2022.

Hung Dao Bachelor's Thesis Spring 2020 Information Technology Oulu University of Applied Sciences

DEVIJVER, P. A.; KITTLER, J. **Pattern Recognition: a statistical approach**. Prentice Hall, 1972.

DOE - UNITED STATES OF AMERICA DEPARTMENT OF ENERGY. Life-cycle assessment of energy and environmental impacts of LED lighting products. Solid-State Lighting Program: **U.S. Department of Energy**, 2012. Disponível em: http://apps1.eere.energy.gov/buildings/publications/pdfs/ssl/2012_led_lca-pt2.pdf . Acesso em: 1 ago. 2021.

ESCOVEDO, T.; KOSHIYAMA, A. S. **Introdução a Data Science: algoritmos de Machine Learning e métodos de análise**. São Paulo: Ed. Casa do Código, 2020.

EVELYN, F.; HODGES; JOSEPH, L. **Discriminatory analysis nonparametric discrimination: consistency proprieties**. USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

FARDIN, M. P.; MARQUES, J. C.; CÓ, M. A. IDSIP: Proposta de Indicador de Desempenho em Sistemas de Iluminação Pública. **Ifes Ciência**, [s. l.], v. 7, n.1, p. 1-12, 2021. Disponível em: <https://ojs.ifes.edu.br/index.php/ric/article/view/959>. Acesso em: 13 ago. 2021.

FISHER, R. *et al.* **Fourier Transform**. 2004. Disponível em: https://homepages.inf.ed.ac.uk/rbf/HIPR2/hipr_top.htm. Acesso em: 21 jan. 2022.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, [s. l.], v. 29, n. 5, p. 1189–1232, 2001. Disponível em: <https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boostingmachine/10.1214/aos/1013203451.full>. Acesso em: 13 ago. 2021.

GAO, D. *et al.* Random Forest algorithm for classification of multi-wavelength data. **Research in Astronomy and Astrophysics**, [s. l.], v. 9, n. 2, p. 220- 226, 2009. Disponível em: https://www.researchgate.net/publication/231050565_Random_forest_algorithm_for_classification_of_multiwavelength_data. Acesso em: 21 jan. 2022.

GLEN, S. **Decision tree vs random forest vs gradient boosting machines**. Towards Data Science, jul. 2019. Disponível em: <https://www.datasciencecentral.com/decision-tree-vs-random-forest-vs-boosted-trees-explained/>. Acesso em: 15 nov. 2021.

GRANLUND, G. H. Fourier preprocessing for hand printed character recognition. **IEE Transactions on Computers**, [s. l.], v. C-21, n. 2, p. 195-201, 1972. Disponível em: <https://ieeexplore.ieee.org/abstract/document/5008926>. Acesso em: 15 nov. 2021.

GUPTA, A. **Feature selection techniques in machine learning**. Analytics Vidhya, out. 2020.

HARALICK, R. M. Statistical and structural approaches to texture. **IEEE**, [s. l.], v. 67, n. 5, p. 453-801, 1979. Disponível em: <https://ieeexplore.ieee.org/document/1455597>. Acesso em: 15 nov. 2021.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. **IEEE Trans, Systems, Man, and Cybernetics**, [s. l.], v. 3, n. 6, p. 610-621, 1973. Disponível em: https://www.researchgate.net/publication/302341151_Textural_Features_for_Image_Classification. Acesso em: 2 out. 2021

HÉMERY, D. **Uma história da energia**. Brasília, DF: Editora da Universidade de Brasília, 1993.

HU, M. K. Pattern recognition by invariant moments. **IEEE Trans. Inf. Theory**, v. 8, n. 2, p. 179-187, 1962. Disponível em: <http://www.sci.utah.edu/~gerig/CS7960-S2010/handouts/Hu.pdf>. Acesso em: 2 out. 2021

HUANG, Z.; LENG, J. Analysis of Hu's moment invariants on image scaling and rotation. **International Conference on Computer Engineering and Technology (ICCET)**, 2010.

HUSPI. What is pattern recognition in machine learning?, **HUSPI**, nov. 2019. Disponível em: <https://huspi.com/blog-open/pattern-recognition-in-machine-learning/>. Acesso em: 12 dez. 2021

IBGE. **Panorama: População**. IBGE, 2017. Disponível em: <https://cidades.ibge.gov.br/brasil/panorama>. Acesso em: 2 out. 2021.

JAPKOWICZ, N.; SHAH, M. **Evaluating learning algorithms: a classification perspective**. Cambridge University Press, 2011.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: trends, perspectives, and prospects. **Science**, [s. l.], v. 349, n. 255, p. 255-260, 2015. Disponível em: <https://www.science.org/doi/10.1126/science.aaa8415>. Acesso em: 1 ago. 2021.

KEEFFE, W. M. Recent progress in metal halide discharge-lamp research. **IEE Proceedings A Physical Science: Measurement and Instrumentation, Management and Education - Reviews**, v. 127, v. 3, p. 181-189, 1980. Disponível em: <https://digital-library.theiet.org/content/journals/10.1049/ip-a-1.1980.0029>. Acesso em: 05 jun. 2022.

KINGSFORD, C.; SALZBERG, S. L. What are Decision Trees?. **Nature Biotechnology**, [s. l.], v. 9, n. 26, p. 1011-1013, 2008. Disponível em: <https://www.nature.com/articles/nbt0908-1011>. Acesso em: 1 ago. 2021.

KNIGHT, K; RICK, E. **Artificial intelligence**. McGraw-Hill, 1991.

KOTSIANTIS, S. B. Supervised machine learning: a review of classification techniques. **Informatica**, [s. l.], n. 31, p. 249-268, 2007. Disponível em: [https://datajobs.com/data-science-repo/Supervised-Learning-\[SB-Kotsiantis\].pdf](https://datajobs.com/data-science-repo/Supervised-Learning-[SB-Kotsiantis].pdf). Acesso em: 05 jun. 2022.

KUDO, M.; SKLANSKY, J. Comparison of algorithms that select features for pattern classifiers. **Pattern recognition**, [s. l.], v. 33, n. 1, p. 25-41, 2000. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0031320399000412>. Acesso em: 20 ago. 2021.

KULUKARNI, A. D. **Artificial neural networks for image understanding**. Nova Iorque: Van Nostrand Reinhold, 1994.

LECUN, Y. *et al.* Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, p. 2278–2324, 1998. Disponível em: http://vision.stanford.edu/cs598_spring07/papers/Lecun98.pdf. Acesso em: 20 ago. 2021.

LEITE, T. M. **Redes neurais, perceptron multicamadas e o algoritmo backpropagation**. Ensina, 10 mai. 2018. Disponível em: <https://medium.com/ensina-ai/redes-neurais-perceptron-multicamadas-e-o-algoritmo-backpropagation-eaf89778f5b8>. Acesso em: 20 ago. 2021.

LIMA, E. S. *et al.* Random Forest model to predict the height of eucalyptus. **Scientific Paper, Special Issue: Artificial Intelligence, Eng. agríc. (Online)**, n. 42, 2022. Disponível em: <https://doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210153/2022>. Acesso em: 05 jun. 2022.

LOPES, S. B.; MARTINS, J. M. V. **Eficiência energética em sistemas de iluminação pública**. 2002. Universidade de São Paulo, São Paulo, 2002.

LU, Y.; SONG, Y. Decision tree methods: applications for classification and prediction. **Shanghai Arch Psychiatry**, [s. l.], v. 27, n. 2, p. 130-135, 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/>. Acesso em: 25 jan. 2022.

MAGGI, Tiago. **Estudo e implementação de uma luminária de iluminação pública à base de LEDs**. 2013. 243 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Santa Maria, Rio Grande do Sul, 2013. Disponível em: <https://repositorio.ufsm.br/bitstream/handle/1/8519/MAGGI%2c%20TIAGO.pdf?sequence=1&isAllowed=y>. Acesso em 13 jun. 2021.

MARINO, F.; LECCESE, F.; PIZZUTI, S. Adaptive street lighting predictive control. **Energy Procedia**, [s. l.], v. 111, p.790-799, 2017. Disponível em: https://www.researchgate.net/publication/315684728_Adaptive_Street_Lighting_Predictive_Control. Acesso em: 20 ago. 2021.

MASON, A.; MUKHOPADHUAY, S. C.; JAYASUNDERA, K. Sensing technology: current status and future trends III. **Springer**, [s. l.], v. 11, 2014. Disponível em: <https://researchers.mq.edu.au/en/publications/sensing-technology-current-status-and-future-trends-iii>. Acesso em: 25 jan. 2022.

MEYER, D. **Support Vector Machines**. FH Technikum Wien, 1 fev. 2017. Disponível em: <https://mran.microsoft.com/snapshot/2018-06-22/web/packages/e1071/vignettes/svmdoc.pdf>. Acesso em: 12 set. 2021.

MITCHELL, R.; FRANK, E. Accelerating the XGBoost algorithm using GPU computing. **PeerJ Computer Science**, [s. l.], v. 3, 2017. Disponível em: <https://peerj.com/articles/cs-127/>. Acesso em: 20 set. 2021.

MUELLER, J. P.; MASSARON, L. **Machine learning: creating your own features in data**. Dummies, Machine Learning Articles, jun. 2016. Disponível em: <https://www.dummies.com/article/technology/information-technology/ai/machine-learning/machine-learning-creating-features-data-226788/#tab1>. Acesso em: 25 jan. 2022.

MUSE, L. P. **Iluminação Pública no contexto das cidades inteligentes**. Universidade Federal do Rio de Janeiro, Programa de Engenharia Urbana, Rio de Janeiro, 2019.

NAVIANI, A. **KNN classification using scikit-learn**. 2018. Disponível em: <https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>. Acesso em: 25 jun. 2021.

NASCIMENTO, João Paulo Ribeiro. **Análise e classificação de imagens baseadas em características de textura utilizando matrizes de co-ocorrência**. 2003. 116 f. Dissertação (Mestrado em Informática) - Universidade Federal do Paraná, Curitiba, 2003. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/25099/D%20-%20NASCIMENTO%2c%20JOAO%20PAULO%20RIBEIRO%20DO.pdf?sequence=1&isAllowed=y>. Acesso em: 3 fev. 2022.

OLIVEIRA, M. Z. C. O. *et al.* New approaches in street lighting maintenance: the case of Natal, Brazil. **IEEE**, [s. l.], 2018. Disponível em: <https://ieeexplore.ieee.org/document/8395838>. Acesso em: 30 jul. 2021.

OLIVEIRA, Jeymisson. **Redes neurais artificiais**. 2011. 156 f. Dissertação (Mestrado em Modelagem e Otimização) - Universidade Federal de Goiás- Regional Catalão, Catalão, Goiás,

2011. Disponível em: <http://www.dsc.ufcg.edu.br/~pet/jornal/setembro2011/materias/informatica.html>. Acesso em: 25 jun. 2021.

PANESAR, A. **Machine learning and ai for healthcare: big data for improved health outcomes**. Coventry: Apress, 2019.

PAPOULIS, A. **Fourier integral and its applications**. Nova Iorque, NY: McGraw-Hill, 1962.

PARKHI *et al.* **Deep face recognition**. Visual Geometry Group Department of Engineering Science University of Oxford, 2015.

PARMAR, Ajay. **Road lighting and future factors**. 2019. 183 f. Dissertação (Mestrado) - School of Science and Technology, Middlesex University London, 2019. Disponível em: <https://eprints.mdx.ac.uk/30868/1/AParmar%20thesis.pdf>. Acesso em: 13 jun. 2021.

PATEL, H. What is feature engineering: importance, tools and techniques for machine learning. **Towards Data Science**, 30 ago. 2021. Disponível em: <https://towardsdatascience.com/what-is-feature-engineering-importance-tools-and-techniques-for-machine-learning-2080b0269f10>. Acesso em: 21 set. 2021.

PEDREGOSA *et al.* Scikit-learn: machine learning in python. **Journal of Machine Learning Research**, [s. l.], v. 12, p. 2825-2830, 2011. Disponível em: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https://githubhel> p.com. Acesso em: 30 jul. 2021.

PEDRINI, H. Introdução ao processamento digital de imagem. **Instituto de Computação UNICAMP**. 2021. Disponível em: https://www.researchgate.net/publication/26976120_Segmentacao_de_imagens_baseada_em_dependencia_espacial_utilizando_campo_aleatorio_de_Markov_associado_com_caracteristicas_de_texturas. Acesso em: 3 fev. 2022

PLENIO, M. B.; VITELLI, V. The physics of forgetting: landauer's erasure principle and information theory. **Imperial College**, Reino Unido, mar. 2001. Disponível em: <https://arxiv.org/pdf/quant-ph/0103108.pdf>. Acesso em: 13 jun. 2021.

RABAZA, O. *et al.* A simple method for designing eficiente public lighting, based on new parameter relationships. **Expert Systems with Applications**, [s. l.], v. 40, n. 18, p. 7305-7315, 2013. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0957417413005162>. Acesso em: 1 ago. 2021.

RADULOVIC, D.; SKOK, S.; KIRINCIC, V. Energy efficiency public lighting management in the cities. **Science Direct**, abr. 2011. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0360544210005761>. Acesso em: 3 ago. 2021.

REINKER, J. Energy efficiency streetlight conversion: Integrating LEDs and Intelligent Transportation Systems. 2019. Dissertação (Mestrado) - University of Rhode Island, Rhode Island, 2019.

RIDLER, T. W.; CALVARD, S. **Picture thresholding using an iterative selection method.** IEEE Trans. Syst. Man Cybern. SMC-8, p. 630–632, 1978.

ROKACH, L.; MAIMON, O. **Data mining with decision trees: theory and applications.** 2. ed. New Jersey: World Scientific Publishing Co. Pte. Ltd., 2015.

SANTANA, Rosa Maria Bomfim. **Iluminação pública: uma abordagem gerencial.** 2010. 94 f. Dissertação (Mestrado em Regulação da Indústria de Energia) - Universidade Salvador, Salvador, 2010. Disponível em: https://hosting.iar.unicamp.br/lab/luz/ld/Arquitetural/Pesquisa/iluminacao_publica_uma_abordagem_gerencial.pdf. Acesso em: 3 ago. 2021.

SCACCIA, K. Validação Cruzada Aninhada com Scikit-learn. **Data ML**, 11 set. 2020. Disponível em: <https://dataml.com.br/validacao-cruzada-aninhada-com-scikit-learn/>. Acesso em: 15 set. 2021.

SCHWARTZ, William Robson. Segmentação de imagens baseada em dependência espacial utilizando campo aleatório de Markov associado com características de textura. 2005. 120 f. Dissertação (Mestrado) - Universidade Federal de Minas Gerais, 2005. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/1597/Disserta.pdf?sequence=1&isAllowed=y>. Acesso em: 01 abr. 2022.

SEDU. **Manual de Iluminação Pública.** Copel, 2018. Disponível em: www.copel.com. Acesso em: 20 set. 2021.

SHAFKAT, I. **Intuitively understanding convolutions for deep learning.** Towards Data Science. Disponível em: <https://towardsdatascience.com/intuitively-understanding-convolutions-for-deep-learning-1f6f42faee1>. Acesso em: 01 abr. 2022.

SILVA, Lourenço Lustosa Fróes. **Iluminação pública no Brasil: aspectos energéticos e institucionais.** 2006. 172 f. Dissertação (Mestrado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006. Disponível em: <http://antigo.ppe.ufrj.br/ppes/production/tesis/llfroes.pdf>. Acesso em: 25 jun. 2021.

SILVA, Thays Ferreira da. **Análise de descritores de imagem e classificadores aplicados para detecção de lesões em imagens de mamografia digital.** 2017. 114 f. Monografia (Bacharelado em Ciência da Computação) - Universidade Federal Rural de Pernambuco, Recife, 2017. Disponível em: <http://www.bcc.ufrpe.br/sites/ww3.bcc.ufrpe.br/files/Thays%20Silva.pdf>. Acesso em: 05 fev. 2022.

SMITH, M. G; BULL, L. **Feature construction and selection using genetic programming and a genetic algorithm.** In: RYAN, C. *et al.* Lecture Notes in Computer Science. Essex, UK: Springer, 2003. Disponível em: https://www.researchgate.net/publication/216301461_Feature_Construction_and_Selection_Using_Genetic_Programming_and_a_Genetic_Algorithm. Acesso em: 05 fev. 2022.

SOARES, G. M. *et al.* **Detection of street lighting bulbs information to minimize comercial losses.** In: Sensing Technology (ICST), Seventh International Conference on, p. 895–900, dez. 2013.

SOARES, G. M. *et al.* **Performance evaluation of a sensor-based system devised to minimize commercial losses in street lighting networks.** In: 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, p. 336–341, mai. 2014.

SOARES, G. M. *et al.* **On the use of light sensors and pattern recognition techniques for automated detection of street lighting lamps.** In: Sensing Technology: Current Status and Future Trends III, p. 67-92, 2015.

SONKA, M.; HLAVAC, V.; BOYLE, R. **Image processing, analysis, and machine vision.** Cengage Learning, 2014.

SZEGEDY, C. *et al.* **Going deeper with convolutions.** CVPR 2015, 2015.

THEODORIDIS, S. **Pattern Recognition.** Academic Press, 1999.

VARGAS, A. C. G. *et al.* **Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres.** In: Proceedings of the XXIX Conference on Graphics, Patterns and Images, p. 1–4, 2016.

VASCONCELOS, S. **Matrizes de co-ocorrência.** Instituto Federal Fluminense, Instituto de Computação, 2020.

VITO. Eco-design study Domestic Lighting. **Preparatory studies for eco-design requirements of EuPs**, vol. 19, 2007. Disponível em: https://www.applia-europe.eu/images/Library/Preparatory_study_for_external_power_supplies_01-2007_compressed.pdf. Acesso em: 13 set. 2021.

WESZKA, J.; DYER, C.; ROSEFIELD, A. A comparative study of texture measures for terrain classification. **IEE Trans. Syst., Man, Cybern.**, v. SMC-6, n.4, p. 269-285, abr. 1976.

WITTEN, I.H.; FRANK, E.; HALL, M.A. **Data mining: practical machine learning tools and techniques.** 3. ed. Burlington: Morgan Kaufmann Publishers, 2011.

YADAV, A. Support Vector Machines (SVM). **Towards Data Science**, 20 out. 2018. Disponível em: <https://towardsdatascience.com/support-vector-machines-svm-c9ef22815589>. Acesso em: 3 out. 2021.

YIU, T. **Understanding Random Forest.** Towards Data Science, jun., 2019. Disponível em: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>. Acesso em: 7 mar. 2021.

ZEILER, M. D.; FERGUS, R. **Visualizing and understanding convolutional networks.** Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, Proceedings, Part I. Springer International Publishing, 2014, p. 818–833, set. 2014.

**APÊNDICE A – IMPORTÂNCIA MÉDIA DAS VARIÁVEIS PARA
CLASSIFICAÇÃO DE TIPOS**

Quadro 1 - Importância média das variáveis para classificação de tipos.

Variável	DT	RF	XGB	SVM
Assimetria	0,00418	0,01361	0,00067	-0,02829
Média	0,00000	0,01365	0,01454	0,04309
bp104fas	0,02431	0,02667	0,03065	0,20226
Contraste	0,02185	0,02543	0,01596	-0,10235
Correlação	0,00466	0,01799	0,01668	0,21206
df01	0,00000	0,02165	0,01790	0,21791
df02	0,01826	0,01677	0,01483	-0,81653
df03	0,01827	0,01774	0,01596	0,05962
df04	0,02299	0,01813	0,01269	0,11247
df05	0,00157	0,01245	0,00711	-0,24258
df06	0,01001	0,01110	0,00862	0,27186
df07	0,03718	0,02203	0,01366	-0,37854
df08	0,00538	0,01488	0,01189	0,02784
df09	0,00300	0,01566	0,01996	-0,02351
df10	0,01466	0,01343	0,01214	0,00276
Energia	0,05088	0,06245	0,00723	0,01587
Entropia	0,16685	0,05766	0,07974	-0,01928
Homogeneidade	0,00433	0,03440	0,01882	0,07091
i1	0,00771	0,02363	0,01271	0,66361
i2	0,00618	0,01841	0,04302	-0,51434
i3	0,02718	0,02158	0,01502	0,03143
i4	0,01484	0,02621	0,00946	0,10661
i5	0,00461	0,01801	0,00621	-0,11505
i6	0,00497	0,01301	0,00752	0,08788
i7	0,04279	0,01943	0,01961	0,14317
Curtose	0,00903	0,01307	0,00855	0,08460
Mediana	0,00923	0,01398	0,02125	0,36134
Probabilidade Máxima	0,00157	0,05531	0,08234	0,00095
sfh2430	0,00000	0,02019	0,01432	-0,07097
sfh3400	0,00294	0,01801	0,02119	-0,01189
sfh3410	0,00152	0,01522	0,01642	0,02877
sfh5711	0,03658	0,01736	0,13456	0,02428
Desvio Padrão	0,01138	0,01146	0,01117	-0,20200
tcs3414_b	0,00000	0,01589	0,01327	0,06947
tcs3414_c	0,00237	0,01593	0,02575	0,05489
tcs3414_g	0,04169	0,01890	0,00000	-0,01780
tcs3414_r	0,01254	0,04453	0,00951	0,42291
tsl13s	0,00000	0,02221	0,02833	0,34874
tsl2569t_c	0,06879	0,03293	0,04814	-0,20549
tsl2569t_ir	0,01632	0,01923	0,04491	-0,09973
tsl261r	0,25224	0,08493	0,05851	0,89895
Variância	0,00433	0,00974	0,01229	-0,34577
Coeficiente de variação	0,01283	0,01514	0,01688	-0,00016

Fonte: Elaborada pelo autor (2022).

APÊNDICE B - IMPORTÂNCIA MÉDIA DAS VARIÁVEIS PARA CLASSIFICAÇÃO DE POTÊNCIA

Quadro 2 - Importância média das variáveis para classificação de potência.

Variável	DT	RF	XGB	SVM
Assimetria	0,01213	0,01804	0,00409	-0,19146
Média	0,01125	0,01713	0,02210	0,14865
bp104fas	0,04649	0,02853	0,02796	0,29771
Contraste	0,03196	0,02154	0,02115	-0,12627
Correlação	0,03926	0,02575	0,02132	-0,18891
df01	0,01051	0,01542	0,01362	0,14102
df02	0,01535	0,02366	0,02065	0,50256
df03	0,02011	0,02091	0,02370	-0,04844
df04	0,01195	0,02368	0,01677	0,70683
df05	0,02186	0,02058	0,01907	0,00072
df06	0,00728	0,02199	0,02447	0,32853
df07	0,04535	0,02546	0,01614	-0,31701
df08	0,01702	0,02116	0,02006	0,58600
df09	0,02778	0,02494	0,03494	0,19253
df10	0,03686	0,02233	0,01628	0,47340
Energia	0,04311	0,03578	0,02162	0,03243
Entropia	0,04994	0,03734	0,03869	-0,13916
Homogeneidade	0,04758	0,03364	0,05544	-0,10897
i1	0,03111	0,02845	0,02191	0,07153
i2	0,01927	0,02355	0,01894	-0,16130
i3	0,00602	0,01876	0,01134	0,16847
i4	0,03166	0,02259	0,02015	0,14854
i5	0,01490	0,01700	0,01786	0,00438
i6	0,02422	0,02080	0,01669	-0,16034
i7	0,02049	0,02043	0,01995	0,04074
Curtose	0,00800	0,01659	0,01705	0,00835
Mediana	0,01962	0,02263	0,01566	0,33448
Probabilidade Máxima	0,01836	0,03778	0,05256	0,17580
sfh2430	0,01031	0,02086	0,03714	0,03305
sfh3400	0,01380	0,02211	0,02641	0,51419
sfh3410	0,01447	0,01947	0,02516	0,07608
sfh5711	0,00518	0,01820	0,01573	-0,09388
Desvio Padrão	0,01315	0,01911	0,02643	0,04117
tcs3414_b	0,01130	0,01832	0,02032	0,35295
tcs3414_c	0,00587	0,02630	0,03009	0,59756
tcs3414_g	0,00305	0,01619	0,01243	0,08688
tcs3414_r	0,02906	0,03075	0,01356	-0,01732
tsl13s	0,00160	0,01964	0,02905	0,25026
tsl2569t_c	0,05284	0,02723	0,02764	0,33814
tsl2569t_ir	0,02318	0,01643	0,02292	-0,24911
tsl261r	0,08710	0,04321	0,04267	0,10267
Variância	0,02500	0,01738	0,02106	0,22110
Coefficiente de variação	0,01469	0,01837	0,01922	0,00009

Fonte: Elaborada pelo autor (2022).

APÊNDICE C - IMPORTÂNCIA MÉDIA DAS VARIÁVEIS PARA CLASSIFICAÇÃO ÚNICA

Quadro 3 – Importância média das variáveis para classificação única.

Variável	DT	RF	XGB	SVM
Assimetria	0,01965	0,01790	0,00944	-0,04106
Média	0,00327	0,01928	0,02819	0,02280
bp104fas	0,03078	0,03180	0,03495	0,14976
Contraste	0,02872	0,02264	0,02782	-0,01141
Correlação	0,02760	0,02473	0,01667	0,19563
df01	0,00249	0,01571	0,01179	0,26155
df02	0,03071	0,02165	0,02185	-0,68856
df03	0,02641	0,01894	0,01876	0,08714
df04	0,02368	0,01802	0,01139	0,14680
df05	0,00953	0,01735	0,00880	-0,24657
df06	0,01012	0,01909	0,01681	0,28162
df07	0,01977	0,02206	0,01563	-0,41157
df08	0,01286	0,01750	0,01345	0,03268
df09	0,03316	0,02032	0,02217	-0,11181
df10	0,02587	0,01761	0,01861	0,01181
Energia	0,04122	0,04636	0,05538	-0,04757
Entropia	0,02034	0,04485	0,05129	0,00311
Homogeneidade	0,03119	0,03232	0,03031	0,03948
i1	0,02749	0,02951	0,03284	0,57826
i2	0,01256	0,02194	0,02048	-0,43885
i3	0,02257	0,01650	0,01452	0,10244
i4	0,03632	0,02281	0,01728	0,09443
i5	0,00914	0,01734	0,01310	-0,11278
i6	0,02029	0,01783	0,01724	-0,10205
i7	0,02753	0,02202	0,01838	0,07225
Curtose	0,01539	0,01496	0,01260	0,06467
Mediana	0,01132	0,01776	0,01468	0,35168
Probabilidade Máxima	0,07792	0,04730	0,05531	-0,03081
sfh2430	0,01851	0,01906	0,01763	-0,10556
sfh3400	0,02517	0,02153	0,03021	-0,02313
sfh3410	0,00706	0,01950	0,00900	-0,00233
sfh5711	0,00742	0,01851	0,02630	-0,06343
Desvio Padrão	0,02849	0,01959	0,02693	-0,19976
tcs3414_b	0,00143	0,01766	0,02237	0,07408
tcs3414_c	0,00691	0,02507	0,01587	0,06294
tcs3414_g	0,00480	0,01413	0,00909	-0,06161
tcs3414_r	0,01837	0,03473	0,02718	0,44733
tsl13s	0,01069	0,01861	0,03259	0,19773
tsl2569t_c	0,03777	0,02954	0,03587	-0,27916
tsl2569t_ir	0,03826	0,01959	0,02452	-0,11596
tsl261r	0,10727	0,05031	0,05033	0,83508
Variância	0,01577	0,01872	0,02498	-0,00014
Coeficiente de variação	0,01419	0,01734	0,01743	-0,00014

Fonte: Elaborada pelo autor (2022).

APÊNDICE D – RANKING DE VARIÁVEIS PARA CLASSIFICAÇÃO DE TIPOS

Quadro 4 – *Ranking* de variáveis para classificação de tipos.

VARIÁVEIS	DT	RF	XGB	SVM	SOMA
tsl261r	1	1	4	1	7
bp104fas	10	8	8	9	35
Entropia	2	3	3	30	38
i7	5	17	14	10	46
tsl2569t_c	3	7	5	38	53
sfh5711	8	26	1	23	58
tcs3414_r	19	5	34	3	61
i1	24	11	28	2	65
i3	9	15	22	20	66
Homogeneidade	30	6	15	15	66
Probabilidade Máxima	36	4	2	26	68
Energia	4	2	39	24	69
tsl13s	43	12	9	5	69
Mediana	22	34	11	4	71
i4	16	9	35	12	72
df04	11	21	29	11	72
tsl2569t_ir	15	18	6	34	73
df03	13	25	21	17	76
Contraste	12	10	20	35	77
df01	40	14	16	7	77
Correlação	28	24	18	8	78
df07	7	13	26	41	87
tcs3414_c	35	28	10	18	91
i2	25	20	7	42	94
Coefficiente de variação	18	32	17	27	94
sfh3400	34	22	12	28	96
tcs3414_g	6	19	43	29	97
df06	21	42	36	6	105
df02	14	27	23	43	107
df09	33	30	13	31	107
sfh3410	38	31	19	21	109
df10	17	37	31	25	110
Curtose	23	38	37	14	112
df08	26	33	32	22	113
tcs3414_b	42	29	27	16	114
sfh2430	41	16	25	33	115
i6	27	39	38	13	117
Média	39	35	24	19	117
i5	29	23	41	36	129
Desvio Padrão	20	41	33	37	131
Assimetria	32	36	42	32	142
Variância	31	43	30	40	144
df05	37	40	40	39	156

Fonte: Elaborada pelo autor (2022).

APÊNDICE E - RANKING DE VARIÁVEIS PARA CLASSIFICAÇÃO DE POTÊNCIA

Quadro 5 – Ranking de variáveis para classificação de potência.

VARIÁVEIS	DT	RF	XGB	SVM	SOMA
tsl261r	1	1	3	20	25
tsl2569t_c	2	9	10	8	29
bp104fas	5	7	9	11	32
Probabilidade Máxima	23	2	2	15	42
Homogeneidade	4	5	1	35	45
df09	14	13	6	14	47
Entropia	3	3	4	37	47
Energia	7	4	19	27	57
tcs3414_c	40	10	7	2	59
i1	12	8	18	23	61
sfh3400	29	20	12	4	65
df02	25	15	23	5	68
df10	9	19	35	6	69
i4	11	18	25	18	72
df08	24	23	26	3	76
Correlação	8	11	20	40	79
df04	32	14	33	1	80
df06	38	21	14	10	83
Mediana	21	17	38	9	85
Variância	15	37	22	13	87
Contraste	10	22	21	36	89
tcs3414_r	13	6	40	32	91
tsl13s	43	29	8	12	92
sfh2430	36	25	5	26	92
df03	20	24	15	33	92
sfh3410	28	30	13	22	93
Desvio Padrão	30	31	11	24	96
df07	6	12	36	43	97
tcs3414_b	33	34	24	7	98
i7	19	28	27	25	99
df05	18	27	29	30	104
Média	34	38	17	17	106
i2	22	16	30	39	107
i6	16	26	34	38	114
tsl2569t_ir	17	41	16	42	116
Coefficiente de variação	27	33	28	31	119
i5	26	39	31	29	125
i3	39	32	42	16	129
df01	35	43	39	19	136
Curtose	37	40	32	28	137
tcs3414_g	42	42	41	21	146
sfh5711	41	35	37	34	147
Assimetria	31	36	43	41	151

Fonte: Elaborada pelo autor (2022).

APÊNDICE F - RANKING DE VARIÁVEIS PARA CLASSIFICAÇÃO ÚNICA

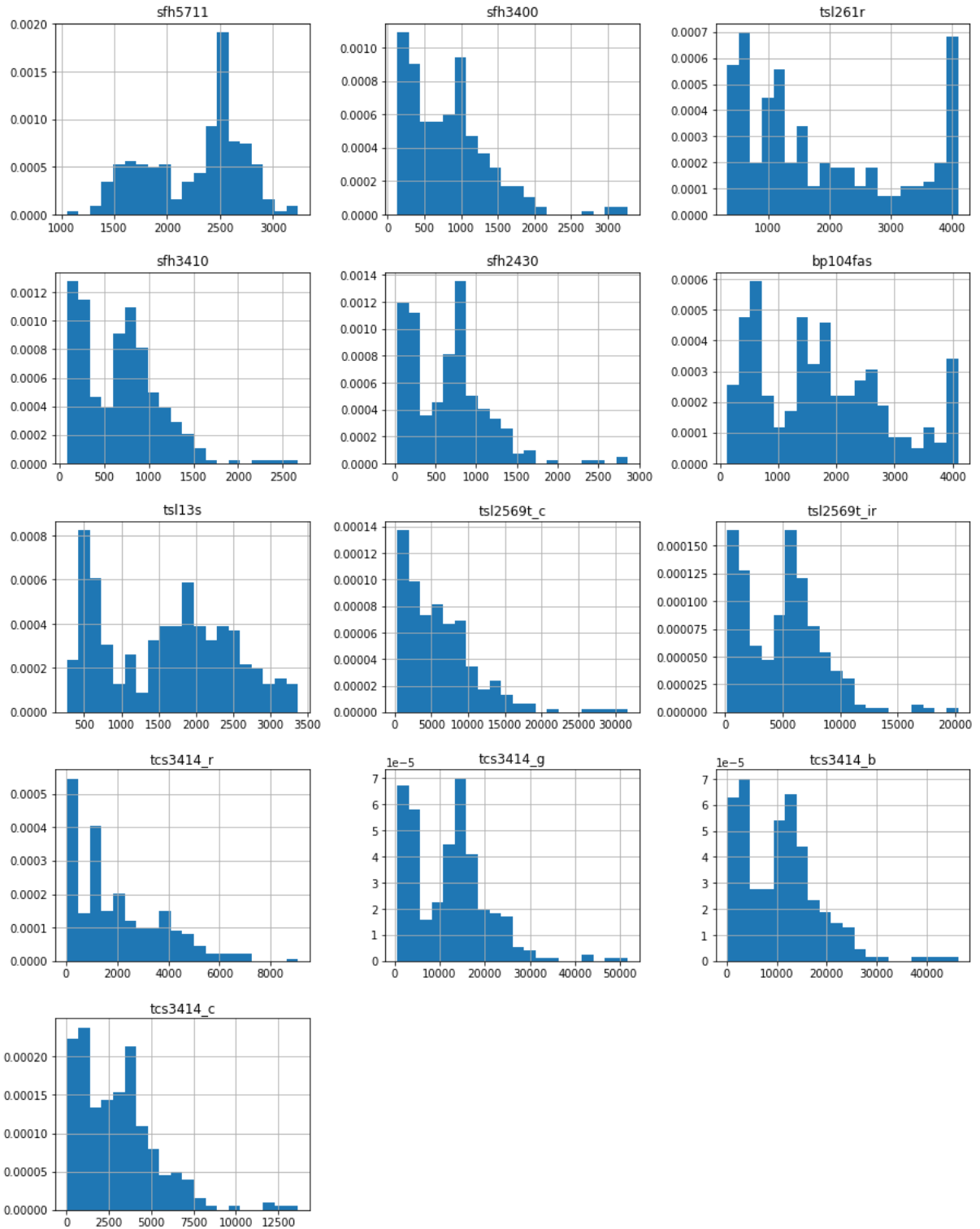
Quadro 6 – *Ranking* de variáveis para classificação única.

VARIÁVEIS	DT	RF	XGB	SVM	SOMA
tsl261r	1	1	4	1	7
bp104fas	9	7	6	9	31
Probabilidade Máxima	2	2	2	26	32
i1	15	9	7	2	33
Energia	3	3	1	28	35
Homogeneidade	8	6	9	18	41
tcs3414_r	26	5	13	3	47
Entropia	21	4	3	22	50
tsl2569t_c	5	8	5	38	56
i4	6	12	27	12	57
Contraste	11	13	12	24	60
Correlação	13	11	30	8	62
i7	14	15	24	15	68
sfh3400	18	18	10	25	71
tsl13s	33	28	8	7	76
tsl2569t_ir	4	20	17	35	76
df03	16	26	22	13	77
df09	7	19	19	33	78
Desvio Padrão	12	21	14	36	83
df02	10	17	20	42	89
df06	34	24	29	5	92
Média	41	23	11	20	95
df10	17	35	23	21	96
tcs3414_c	39	10	31	17	97
df04	19	30	39	10	98
Mediana	32	33	33	4	102
i3	20	40	34	11	105
sfh2430	25	25	25	32	107
tcs3414_b	43	34	18	14	109
Variância	27	27	16	39	109
df07	23	14	32	40	109
i2	31	16	21	41	109
sfh5711	37	29	15	30	111
i6	22	32	28	31	113
df08	30	36	35	19	120
Assimetria	24	31	40	27	122
Curtose	28	42	37	16	123
sfh3410	38	22	42	23	125
df01	42	41	38	6	127
Coefficiente de variação	29	39	26	43	137
i5	36	38	36	34	144
df05	35	37	43	37	152
tcs3414_g	40	43	41	29	153

Fonte: Elaborada pelo autor (2022).

APÊNDICE G – HISTOGRAMAS DAS VARIÁVEIS PROVENIENTES DOS SENSORES

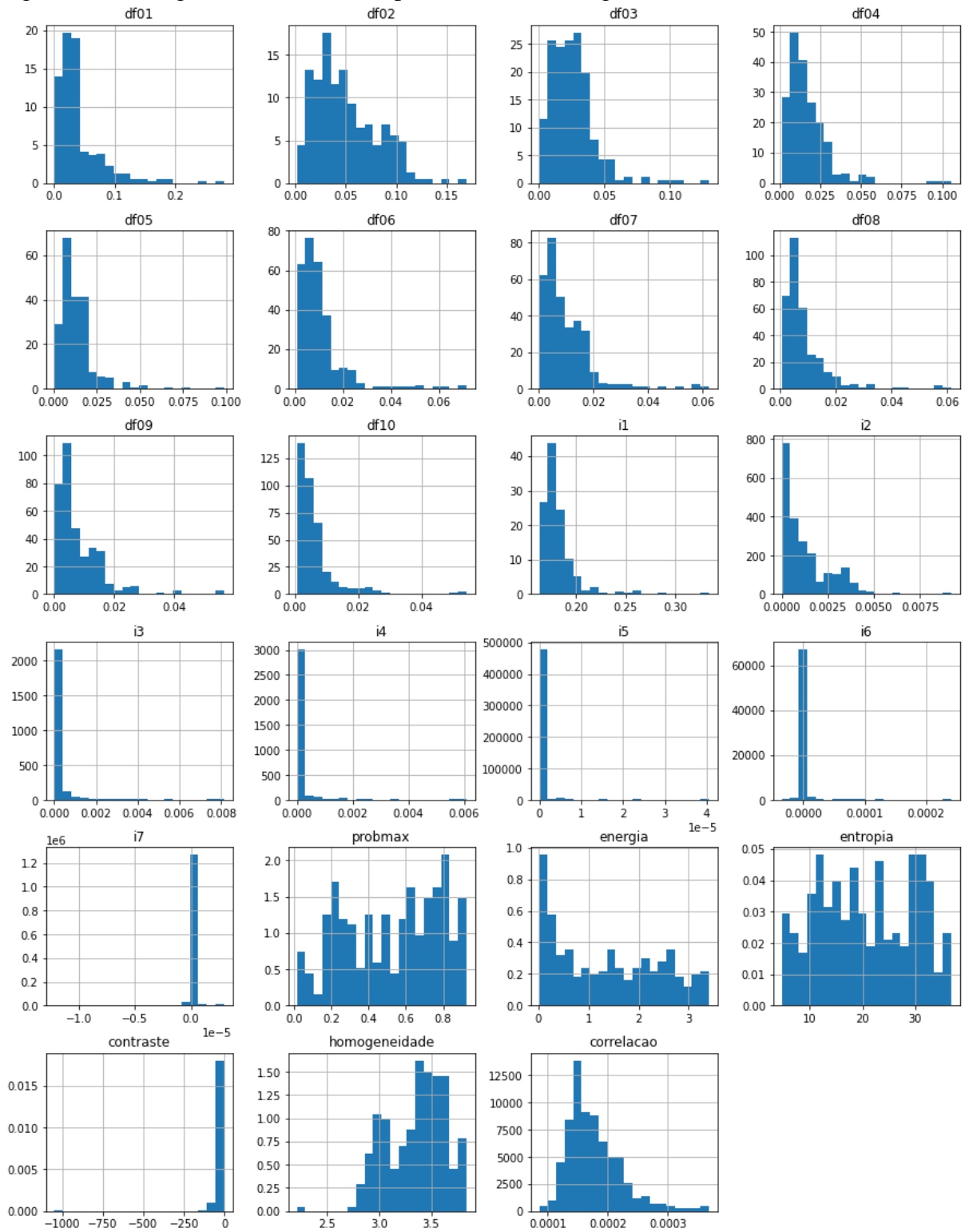
Figura 30 - Histogramas das variáveis provenientes dos sensores.



Fonte: Elaborada pelo autor (2022).

APÊNDICE H – HISTOGRAMAS DAS VARIÁVEIS PROVENIENTES DE IMAGENS

Figura 31 - Histogramas das variáveis provenientes de imagens.



Fonte: Elaborada pelo autor (2022).

APÊNDICE I - RESULTADOS DOS TESTES T COM CLASSIFICADORES DE TIPOS DE LÂMPADA

Quadro 7 - Resultados do teste t com classificadores de tipos de lâmpada.

Modelo	st_knn	st_dt	st_svm	st_rf	st_xgb	st_mlp	it_knn	it_dt	it_svm	it_rf	it_xgb	it_mlp	ct_knn	ct_dt	ct_svm	ct_rf	ct_xgb	ct_mlp	ft_knn	ft_dt	ft_svm	ft_rf	ft_xgb	ft_mlp
st_knn	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
st_dt	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
st_svm	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
st_rf	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
st_xgb	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
st_mlp	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
it_knn	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
it_dt	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
it_svm	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
it_rf	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
it_xgb	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
it_mlp	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
ct_knn	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
ct_dt	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
ct_svm	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
ct_rf	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
ct_xgb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
ct_mlp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
ft_knn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ft_dt	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1
ft_svm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ft_rf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ft_xgb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ft_mlp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fonte: Elaborada pelo autor (2021).

Legenda: O valor 0 representa que a hipótese nula foi aceita, que a média de acurácia do modelo A (identificado pela primeira linha/cabeçalho do quadro) é menor ou igual à média de acurácia do modelo B (identificado pela primeira coluna do quadro). O valor 1 representa que a hipótese nula foi rejeitada, ou seja, que a média de acurácia do modelo A foi maior que a média de acurácia do modelo B.

APÊNDICE J - RESULTADOS DOS TESTES T COM CLASSIFICADORES DE POTÊNCIA DA LÂMPADA

Quadro 8 - Resultados do teste t com classificadores de potência da lâmpada.

Modelo	sp_knn	sp_dt	sp_svm	sp_rf	sp_xgb	sp_nhp	ip_knn	ip_dt	ip_svm	ip_rf	ip_xgb	ip_nhp	cp_knn	cp_dt	cp_svm	cp_rf	cp_xgb	cp_nhp	fp_knn	fp_dt	fp_svm	fp_rf	fp_xgb	fp_nhp
sp_knn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sp_dt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sp_svm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sp_rf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sp_xgb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sp_nhp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ip_knn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ip_dt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ip_svm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ip_rf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ip_xgb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ip_nhp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cp_knn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cp_dt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cp_svm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cp_rf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cp_xgb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cp_nhp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fp_knn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fp_dt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fp_svm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fp_rf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fp_xgb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fp_nhp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fonte: Elaborada pelo autor (2021).

Legenda: O valor 0 representa que a hipótese nula foi aceita, que a média de acurácia do modelo A (identificado pela primeira linha/cabeçalho do quadro) é menor ou igual à média de acurácia do modelo B (identificado pela primeira coluna do quadro). O valor 1 representa que a hipótese nula foi rejeitada, ou seja, que a média de acurácia do modelo A foi maior que a média de acurácia do modelo B.

APÊNDICE K - RESULTADOS DOS TESTES T COM CLASSIFICADORES DE CLASSIFICAÇÃO ÚNICA

Quadro 9 - Resultados do teste t com classificadores únicos.

Modelo	su kam	su dt	su svtn	su rf	su xgb	su nnp	iu kam	iu dt	iu svtn	iu rf	iu xgb	iu nnp	cu kam	cu dt	cu svtn	cu rf	cu xgb	cu nnp	fu kam	fu dt	fu svtn	fu rf	fu xgb	fu nnp
su kam	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
su dt	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
su svtn	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
su rf	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
su xgb	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
su nnp	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
iu kam	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
iu dt	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
iu svtn	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
iu rf	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
iu xgb	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
iu nnp	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
cu kam	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
cu dt	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
cu svtn	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
cu rf	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
cu xgb	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
cu nnp	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
fu kam	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
fu dt	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
fu svtn	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
fu rf	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
fu xgb	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
fu nnp	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0

Fonte: Elaborada pelo autor (2021).

Legenda: O valor 0 representa que a hipótese nula foi aceita, que a média de acurácia do modelo A (identificado pela primeira linha/cabeçalho do quadro) é menor ou igual à média de acurácia do modelo B (identificado pela primeira coluna do quadro). O valor 1 representa que a hipótese nula foi rejeitada, ou seja, que a média de acurácia do modelo A foi maior que a média de acurácia do modelo B.

APÊNDICE L - RESULTADOS DOS TESTES T COM CLASSIFICADORES DE MELHOR PERFORMANCE

Quadro 10 - Resultados do teste t com os classificadores de melhor performance.

Modelo	sp_svm	fp_svm	fp_rf	fp_xgb	fp_mlp	su_svm	cu_rf	cu_xgb	fu_svm	fu_rf	fu_xgb	fu_mlp
sp_svm	-	0	0	0	0	0	0	0	0	0	0	0
fp_svm	0	-	0	0	0	0	0	0	0	0	0	0
fp_rf	1	0	-	0	0	0	0	0	1	0	0	0
fp_xgb	1	0	0	-	0	0	0	0	1	0	0	0
fp_mlp	0	0	0	0	-	0	0	0	0	0	0	0
su_svm	0	0	0	0	0	-	0	0	0	0	0	0
cu_rf	0	0	0	0	0	0	-	0	0	0	0	0
cu_xgb	1	0	0	0	0	0	0	-	0	0	0	0
fu_svm	0	0	0	0	0	0	0	0	-	0	0	0
fu_rf	0	0	0	0	0	0	0	0	0	-	0	0
fu_xgb	0	0	0	0	0	0	0	0	0	0	-	0
fu_mlp	0	0	0	0	0	0	0	0	0	0	0	-

Fonte: Elaborada pelo autor (2021).

Legenda: O valor 0 representa que a hipótese nula foi aceita, que a média de acurácia do modelo A (identificado pela primeira linha/cabeçalho do quadro) é menor ou igual à média de acurácia do modelo B (identificado pela primeira coluna do quadro). O valor 1 representa que a hipótese nula foi rejeitada, ou seja, que a média de acurácia do modelo A foi maior que a média de acurácia do modelo B.