



Universidade Federal do Rio de Janeiro  
Centro de Ciências Matemáticas e da Natureza  
Observatório do Valongo



Uso de *Machine Learning* para  
classificação espectral estelar

Júlia Camões Alves

Rio de Janeiro  
2022

Uso de *Machine Learning* para  
classificação espectral estelar

Júlia Camões Alves

Trabalho de conclusão de curso / Projeto final submetido ao Curso de Graduação em Astronomia, Observatório do Valongo, da Universidade Federal do Rio de Janeiro (OV/UFRJ), como parte dos requisitos necessários à obtenção do título de Astrônomo.

Orientador: Wagner Luiz Ferreira Marcolino

Rio de Janeiro  
2022

## CIP - Catalogação na Publicação

C185u Camões Alves, Júlia  
Uso de Machine Learning para Classificação  
Espectral Estelar / Júlia Camões Alves. -- Rio de  
Janeiro, 2022.  
153 f.

Orientador: Wagner Luiz Ferreira Marcolino.  
Trabalho de conclusão de curso (graduação) -  
Universidade Federal do Rio de Janeiro, Observatório  
do Valongo, Bacharel em Astronomia, 2022.

1. Estrelas de alta massa. 2. Estrelas tipo O.  
3. Classificação espectral. 4. Machine Learning. I.  
Luiz Ferreira Marcolino, Wagner, orient. II. Título.



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
CCMN - OBSERVATÓRIO DO VALONGO  
DEPARTAMENTO DE ASTRONOMIA



## PROJETO FINAL

### RELATÓRIO DA COMISSÃO JULGADORA

**ALUNA:** Júlia Camões Alves (DRE 117097263)

**TÍTULO DO TRABALHO:** “USO DE MACHINE LEARNING PARA CLASSIFICAÇÃO ESPECTRAL ESTELAR”

**DATA DA DEFESA:** 19 de agosto de 2022 às 10:30 h

#### MEMBROS DA COMISSÃO JULGADORA:

Prof. Wagner Luiz Ferreira Marcolino – Presidente/Orientador - (OV/UFRJ)

Prof. Carlos Roberto Rabaça – (OV/UFRJ)

MSc. Aline Novais – (OV/UFRJ)

**CANDIDATO:**

  
Júlia Camões Alves

Rio de Janeiro, 19 de agosto de 2022.

Prof. Paulo Afrânio Augusto Lopes  
Coord. de Grad. do Curso de Astronomia

# *Agradecimentos*

Em primeiro lugar, gostaria de agradecer a minha família. Ao meu pai, à minha mãe e ao meu irmão, que são os meus maiores exemplos de generosidade e sabedoria, agradeço todo o suporte, incentivo e amor que recebi desde o dia em que nasci. Muito obrigada por sempre estarem comigo, sem o apoio de vocês esse sonho não seria possível.

Aos meus avós agradeço pelo amor e carinho de sempre. Em especial, gostaria de agradecer à minha avó Altaci por ter me acolhido quando comecei a graduação e por ter construído a melhor parceria que poderia pedir.

À Luciana, agradeço por sempre cuidar e me fazer sorrir. Obrigada por estar sempre presente.

Ao meu namorado, que esteve comigo em todos os momentos mais felizes e difíceis que essa graduação me proporcionou nos últimos anos, me incentivando a continuar e a mostrar o potencial que ele sempre soube que eu tinha. Muito obrigada por sempre acreditar em mim e também por me ajudar a sanar as minhas dúvidas de português.

Às minhas melhores amigas Stella e Yanna. Vocês sempre me inspiraram a ser a minha melhor versão e sempre cuidaram de mim quando precisei. Obrigada por sempre estarem comigo e serem as irmãs que sempre quis ter.

Às melhores pessoas que poderia conhecer ao longo da graduação: Augusto, Érica, Paulo, Cadu, João, Gabi, Marco, Jéssica e Mylenna. Obrigada por dividirem comigo essa trajetória, sem vocês essa jornada não teria sido tão gratificante e bonita.

Ao meu orientador, Dr. Wagner Marcolino, a quem admiro e em quem me espelho desde 2017 na minha trajetória acadêmica. Obrigada por ser sempre compreensivo e paciente, e por ser o exemplo de professor e cientista que buscarei me tornar um dia.

À todos os professores que estiveram presentes na graduação e que impactaram em quem sou hoje. Em especial gostaria de agradecer ao professor Luan Ghezzi, por sempre ter se colocado disponível não só como professor, mas também como um amigo. Às professoras Diana Andrade e Denise Gonçalves, por terem se disponibilizado a me ajudar com o projeto Ad Astra, levando a Astronomia para alunos da Cidade de Deus e encantando a todos com competência e gentileza.

À toda equipe administrativa, biblioteca, secretaria e funcionários do Observatório do Valongo e da UFRJ, por sempre estarem prontos para auxiliar em quaisquer dúvidas ou problemas que apareceram ao longo desse caminho, e também por todo esforço e dedicação em fazer essa experiência o mais agradável possível para todos os alunos.

À FAPERJ e ao CNPQ pelo apoio financeiro através das bolsas de iniciação científica que recebi ao longo da graduação.

*“Young people, especially young women, often ask me for advice. Here it is, valeat quantum. Do not undertake a scientific career in quest of fame or money. There are easier and better ways to reach them. Undertake it only if nothing else will satisfy you; for nothing else is probably what you will receive. Your reward will be the widening of the horizon as you climb. And if you achieve that reward you will ask no other.”*

— Cecilia Payne-Gaposchkin

# Resumo

## Uso de *Machine Learning* para Classificação Espectral Estelar

Júlia Camões Alves

Orientador: Wagner Luiz Ferreira Marcolino

O presente projeto tem como objetivo analisar a performance de diferentes algoritmos de *Machine Learning* (ML ou Aprendizado de Máquina), aplicando-os à classificação espectral, um problema clássico em astrofísica estelar. Em particular, o objeto de estudo é a classificação espectral de estrelas de alta massa ( $> 10M_{\odot}$ ). Estrelas de alta massa são muito raras se comparadas às estrelas do tipo solar e suas trajetórias evolutivas, especialmente pós-Sequência Principal, ainda não são bem compreendidas até hoje (Martins & Palacios, 2013). São objetos de extrema importância, uma vez que são progenitores de supernovas, estrelas de nêutrons, buracos negros e *gamma-ray bursts*, sendo possivelmente fontes de ondas gravitacionais, quando em sistemas binários. A classificação correta desses objetos é, portanto, uma questão fundamental. Exploramos os algoritmos de ML denominados: *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), *Decision Tree Classifier* (DTC) e *Random Forest Classifier* (RFC). Eles foram testados quanto à significância nos resultados finais em quantidades probabilísticas e também na forma como realizam o processo classificatório — caso haja a necessidade de saber como a separação de dados foi feita, por exemplo. Utilizamos espectros óticos de alta resolução de 113 estrelas, de Martins (2018), e também de baixa resolução de 495 estrelas, do Catálogo GOS (*Galactic O-Star Catalog*) (Apellániz et al., 2016). Para o processo classificatório de tipos espectrais, foram utilizadas as linhas de absorção He I 4471 e He II 4542, e, especialmente, a linha de absorção He II 4686 para a classificação de luminosidade. Encontramos que, de modo geral, os modelos performaram de forma parecida, por mais que possuam mecanismos diferentes para realizar a tarefa de classificação. Quando utilizamos a amostra com espectros de alta e de baixa resolução, vemos claramente uma melhora na capacidade de classificação de tipo espectral, na qual atingimos mais de 85% de classificações corretas para os modelos SVM (86%), DTC (87%) e RFC (87%), e, para o KNN, obtivemos 82%. No entanto, para a classificação de tipo espectral em conjunto a de luminosidade, a capacidade de classificação manteve-se baixa, para os quais o SVM teve o melhor resultado com 66%. Algumas dificuldades encontradas foram a pequena quantidade de dados de alta resolução disponíveis, o que resultou em um número insuficiente de estrelas por classe prejudicando o aprendizado de máquina. A qualidade inferior dos dados de baixa resolução também foi um problema por influenciar diretamente a precisão dos parâmetros utilizados para classificação, aumentando a chance de erro. Com isso, concluímos que por mais que a performance dos



modelos não tenha sido ruim no processo de classificação de tipo espectral, a classificação manual ainda é mais segura e precisa e retorna resultados mais acurados devido a análise ser feita em cada espectro. No entanto, acreditamos que caso tivéssemos disponíveis milhares de espectros de alta resolução, os resultados dos modelos seriam ainda melhores.

**palavras chave:** Astrofísica Estelar, estrelas de alta massa, espectroscopia, *Machine Learning*

# *Abstract*

## **Use of Machine Learning for Stellar Spectral Classification**

Júlia Camões Alves

Advisor: Wagner Luiz Ferreira Marcolino

The following project has the purpose of analyzing the performance of different Machine Learning (ML) algorithms when applied to spectral classification, a classic problem in stellar astrophysics. In particular, our subject of study is the spectral classification of massive stars ( $> 10M_{\odot}$ ). Massive stars are very rare when compared to solar type stars and their evolutionary paths, especially post main-sequence, are still not well understood (Martins & Palacios, 2013). They are extremely important objects, since they are the progenitors of supernovae, neutron stars, black holes and gamma-ray bursts, being possibly sources of gravitational waves, when in binary systems. The correct classification of these objects is, therefore, a fundamental matter. We explored the following ML algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree Classifier (DTC) and Random Forest Classifier (RFC). They were tested according to their final results in probabilistic quantities and also in how they performed the classification process — in case there is the need to understand how the separation of the data occurred. We used 113 optical high-resolution spectra, from Martins (2018), and also 495 low-resolution spectra, from the GOS Catalog (Galactic O-Star Catalog) (Apellániz et al., 2016). For the spectral type classification process, we mainly used the absorption lines of He I 4471 and He II 4542, and, especially, the absorption line of He II 4686 for the luminosity classification. We found that, in general, the models had similar performances, even though they use different mechanisms to get the classification task done. When we used the sample with both high and low resolution spectra, we clearly saw an improvement in the capability of spectral type classification, in which we reached more than 85% of correct classifications for the models SVM (86%), DTC (87%) and RFC (87%), and for KNN, we reached 82%. However, when we tried the classification of spectral type with luminosity, the classification capacity did not improve as much as expected, and the best result was for SVM with 66%. Some of the difficulties faced were the small quantity of high-resolution data available, which led to an insufficient number of stars per class and turned the learning process of the machine much more difficult. The inferior quality of the low-resolution data was also a problem as it interfered directly on the precision of the parameters used for classification, increasing the chance of error. Therefore, we conclude that, even though the performance of the models were reasonable for the spectral type classification, the manual inspection of the spectra is still a

more safe and precise procedure which returns more accurate results. Nevertheless, our results indicate that if we had thousands of high-resolution spectra, the performances of the models would be vastly improved.

**keywords:** Stellar Astrophysics, Massive stars, spectroscopy, *Machine Learning*

# Lista de Figuras

1.1	Comparação entre espectros de tipos espectrais O, B, A e F, na região do UV. . . . .	19
1.2	Comparação entre espectros do tipo O ao M. . . . .	20
1.3	Exemplos de fenômenos causados por estrelas de tipo espectral O. . . . .	21
1.4	A dependência da intensidade das linhas espectrais com a temperatura. . . . .	23
1.5	Diagrama HR. . . . .	24
2.1	Exemplos de espectros de alta resolução. . . . .	31
2.2	Exemplos de espectros de baixa resolução. . . . .	32
3.1	Representação da largura equivalente. . . . .	34
3.2	Representação do somatório para encontrar a largura equivalente. . . . .	35
3.3	No eixo $x$ vemos os tipos espectrais e no eixo $y$ os respectivos valores de $W'$ dos espectros. . . . .	37
3.4	Espectros de alta resolução classificados entre O8 e O9.7. . . . .	39
3.5	Razões entre as larguras equivalentes das linhas utilizadas para classificação de tipo espectral de objetos já classificados entre O8 e O9.7. . . . .	40
3.6	Linhas do N III, C III e He II no espectro da estrela HD 155806, para a qual obtivemos múltiplas classificações. . . . .	43
3.7	No eixo $x$ vemos as classes de luminosidade e no eixo $y$ os valores correspondentes de $W_\lambda(\text{He II } \lambda 4686)$ . . . . .	44
3.8	Espectros de alta resolução de estrelas supergigantes classificadas entre O8 e O9.7. . . . .	45
3.9	Intervalos de integração para espectros de alta resolução. . . . .	48
3.10	Intervalos de integração para espectros de baixa resolução. . . . .	49
3.11	Linha do He II $\lambda 4542$ do espectro de alta resolução HD 42088. . . . .	50
3.12	Linha do He II $\lambda 4542$ do espectro de baixa resolução CPD -59 2554. . . . .	51
3.13	Imagens da base de dados <i>MNIST</i> . . . . .	52
3.14	Esquema que demonstra a escolha do modelo de ML de melhor performance. . . . .	53
4.1	Ilustrações que indicam os casos de <i>Underfitting</i> , de <i>overfitting</i> e um cenário ideal. . . . .	58
4.2	Relação entre erro e complexidade do modelo. . . . .	59
4.3	<i>Cross Validation</i> . . . . .	60
4.4	Matriz de Confusão para uma dada classe. . . . .	61
4.5	Passo-a-passo do funcionamento do modelo KNN utilizando 3 vizinhos que estão representados dentro da região circular na segunda imagem. . . . .	64
4.6	Plot resultante do KNN para classificação do Conjunto de Dados Iris. . . . .	64
4.7	Exemplo de separação de dados de duas classes: verde e azul (SVM). . . . .	66

4.8	Diferenças entre os dados não-escalonados e escalonados no SVM. . . . .	67
4.9	Separação suave entre os dados de duas classes distintas (azul e verde) a partir da desconsideração dos <i>outliers</i> . . . . .	68
4.10	Comportamento das margens do SVM a partir de diferentes valores do parâmetro $C$ . . . . .	68
4.11	Exemplo de árvore de decisão para classificação do Conjunto de dados Flor Iris. . . . .	70
4.12	Alpha versus acurácia para os conjuntos teste e treino. . . . .	72
4.13	Exemplo de RFC utilizando quatro árvores de decisão. . . . .	73
5.1	Valores encontrados para as larguras equivalentes de He I 4542 e He I 4471 para os espectros de alta resolução. . . . .	76
5.2	Valores encontrados para as larguras equivalentes de He I 4542 e He I 4471 para os espectros de baixa resolução. . . . .	76
5.3	Histograma que mostra a quantidade de espectros de alta resolução em cada classe de tipo espectral. . . . .	77
5.4	Histograma que mostra a quantidade de espectros de alta resolução em cada classe de tipo espectral e de luminosidade. . . . .	78
5.5	Histograma que mostra a quantidade de espectros de alta e de baixa resolução em cada classe de tipo espectral. . . . .	78
5.6	Histograma que mostra a quantidade de espectros de alta e de baixa resolução em cada classe de tipo espectral e de luminosidade. . . . .	78
5.7	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta resolução por KNN. A <i>colorbar</i> indica a quantidade de espectros nas posições da matriz. . . . .	81
5.8	Gráfico de KNN resultante para classificação de tipo espectral para espectros de alta resolução. . . . .	82
5.9	No eixo $x$ temos diferentes valores de $C$ e no eixo $y$ os respectivos resultados de $F1-Score$ . . . . .	83
5.10	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta resolução por SVM. . . . .	84
5.11	No eixo $x$ temos diferentes valores de $\alpha$ e no eixo $y$ temos os valores de acurácia resultantes. . . . .	85
5.12	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta resolução por DTC. . . . .	85
5.13	. . . . .	85
5.14	Árvore de decisão para classificação de tipo espectral para espectros de alta resolução . . . . .	86
5.15	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta resolução por RFC. . . . .	87
5.16	. . . . .	87
5.17	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral e de luminosidade com espectros de alta resolução por KNN. . . . .	89
5.18	No eixo $x$ temos diferentes valores de $C$ e no eixo $y$ os respectivos resultados de $F1-Score$ . . . . .	90
5.19	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral e de luminosidade com espectros de alta resolução por SVM. . . . .	90

5.20	No eixo $x$ temos diferentes valores de $\alpha$ e no eixo $y$ temos os respectivos valores de acurácia. . . . .	91
5.21	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral e de luminosidade com espectros de alta resolução por DTC. . . . .	92
5.22	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral e de luminosidade com espectros de alta resolução por RFC. . . . .	93
5.23	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por KNN. . . . .	94
5.24	Gráfico resultante da classificação de tipo espectral utilizando KNN para espectros de alta e de baixa resolução. . . . .	95
5.25	No eixo $x$ temos diferentes valores de $C$ e no eixo $y$ os respectivos resultados de $F1-Score$ . . . . .	96
5.26	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por SVM. . . . .	97
5.27	No eixo $x$ temos diferentes valores de $\alpha$ e no eixo $y$ temos os valores de acurácia resultantes. . . . .	98
5.28	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por DTC. . . . .	98
5.29	árvore de decisão para classificação de tipo espectral para espectros de alta e de baixa resolução. . . . .	99
5.30	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por RFC. . . . .	100
5.31	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por KNN. . . . .	102
5.32	No eixo $x$ temos diferentes valores de $C$ e no eixo $y$ os respectivos resultados de $F1-Score$ . . . . .	103
5.33	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por SVM. . . . .	103
5.34	No eixo $x$ temos diferentes valores de $\alpha$ e no eixo $y$ temos os valores de acurácia resultantes. . . . .	104
5.35	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por DTC. . . . .	105
5.36	Árvore de decisão para classificação de tipo espectral e de luminosidade, para espectros de alta e de baixa resolução. . . . .	106
5.37	Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por RFC. . . . .	107

# Lista de Tabelas

2.1	Fontes e propriedades do espectros de alta resolução. . . . .	28
3.1	Intervalos de $W'$ para a classificação de tipo espectral das estrelas (Mathys, 1988). . . . .	36
3.2	Tabela 1 em Martins (2018). Intervalos da razão entre linhas de absorção usadas para diagnóstico de tipo espectral entre O8 e O9.7. . . . .	38
3.3	Intervalos de $W_\lambda(\text{HeII}\lambda 4686)$ para a classificação de luminosidade dos espectros entre O3-O7.5 (Martins, 2018). . . . .	41
3.4	Intervalos de $W_\lambda(\text{He II } \lambda 4686)$ para o fenômeno $f$ de Martins (2018). . . . .	42
3.5	Relação entre tipos espectrais e o fenômeno $f$ de Sota et al. (2014). Esta tabela foi adaptada utilizando as classes de luminosidade deste projeto e contendo apenas as relações para espectros entre O3 e O7.5. . . . .	42
3.6	Sistema utilizado para classificação de luminosidade de espectros classificados entre O3-O7.5. Desenvolvido a partir da Figura 3.7 de Martins (2018). . . . .	43
3.7	Intervalos de $W^+$ para a classificação de luminosidade dos espectros. . . . .	45
3.8	Intervalos de integração para encontrar o valor de largura equivalente de cada linha espectral. . . . .	46
5.1	Resultados dos modelos de ML para classificação dos espectros de alta resolução. . . . .	109
5.2	Resultados dos modelos de ML para classificação dos espectros de alta e baixa resolução. . . . .	109
A.1	Lista de estrelas com as respectivas classificações e larguras equivalentes para os espectros de alta resolução. . . . .	117
A.2	Lista de estrelas com as respectivas classificações e larguras equivalentes para os espectros de baixa resolução. . . . .	124

# Abreviações

<b>ML</b>	<b>M</b> achine <b>L</b> earning (Aprendizado de Máquina)
<b>KNN</b>	<b>K</b> - <b>N</b> earest <b>N</b> eighbours (K-Vizinhos Mais Próximos)
<b>SVM</b>	<b>S</b> upport <b>V</b> ector <b>M</b> achines (Máquinas de Vetores de Suporte)
<b>DTC</b>	<b>D</b> ecision <b>T</b> ree <b>C</b> lassifier (Árvores de Decisão)
<b>RFC</b>	<b>R</b> andom <b>F</b> orest <b>C</b> lassifier (Florestas Aleatórias)
<b>GOSC</b>	<b>G</b> alactic <b>O</b> - <b>S</b> tars <b>C</b> atalog (Catálogo de Estrelas do tipo O da Galáxia)



# Sumário

<b>1</b>	<b>Introdução</b>	<b>18</b>
1.1	Estrelas tipo-O . . . . .	18
1.2	Busca pela classificação espectral . . . . .	21
1.3	Vantagens da classificação via aprendizado de máquina . . . . .	25
1.4	Objetivos . . . . .	25
<b>2</b>	<b>Dados Utilizados</b>	<b>27</b>
2.1	Origem dos espectros . . . . .	28
2.2	Amostra de espectros . . . . .	29
<b>3</b>	<b>Metodologia</b>	<b>33</b>
3.1	Largura Equivalente . . . . .	34
3.2	Critérios de classificação . . . . .	35
3.2.1	Classificação de tipo espectral . . . . .	35
3.2.2	Classificação de luminosidade . . . . .	40
3.2.2.1	Classificação de estrelas de tipo O3-O7.5 . . . . .	41
3.2.2.2	Classificação de espectros O8-O9.7 . . . . .	44
3.3	Intervalos de integração . . . . .	46
3.3.1	Medidas dos erros das larguras equivalentes . . . . .	47
3.4	Aplicação dos modelos de <i>Machine Learning</i> . . . . .	51
<b>4</b>	<b>Modelos de Machine Learning</b>	<b>55</b>
4.1	Conceitos Fundamentais . . . . .	56
4.1.1	<i>Labels</i> e <i>Features</i> . . . . .	56
4.1.2	Aprendizados supervisionado e não-supervisionado . . . . .	56
4.1.3	Conjuntos Teste e Treino . . . . .	57
4.1.4	<i>Underfitting</i> e <i>Overfitting</i> . . . . .	57
4.1.5	<i>Cross Validation</i> . . . . .	58
4.1.6	Matriz de Confusão . . . . .	60
4.1.7	<i>Precision</i> , <i>Recall</i> ou <i>F1-Score</i> ? . . . . .	61
4.2	Modelos utilizados . . . . .	62
4.2.1	<i>K-Nearest Neighbors</i> . . . . .	63
4.2.2	<i>Support Vector Machines</i> . . . . .	66
4.2.3	<i>Decision Tree Classifier</i> . . . . .	69
4.2.4	<i>Random Forest Classifier</i> . . . . .	72
<b>5</b>	<b>Resultados</b>	<b>75</b>
5.1	Resultados da classificação pré-ML . . . . .	75
5.2	Resultados de ML utilizando espectros de alta resolução . . . . .	79

---

5.2.1	Classificação de tipo espectral . . . . .	80
5.2.1.1	<i>K-Nearest Neighbors</i> . . . . .	80
5.2.1.2	<i>Support Vector Machine</i> . . . . .	83
5.2.1.3	<i>Decision Tree Classifier</i> . . . . .	84
5.2.1.4	<i>Random Forest Classifier</i> . . . . .	87
5.2.2	Classificação de tipo espectral e de luminosidade . . . . .	88
5.2.2.1	<i>K-Nearest Neighbors</i> . . . . .	88
5.2.2.2	<i>Support Vector Machine</i> . . . . .	89
5.2.2.3	<i>Decision Tree Classifier</i> . . . . .	90
5.2.2.4	<i>Random Forest Classifier</i> . . . . .	92
5.3	Resultados utilizando espectros de alta e de baixa resolução . . . . .	93
5.3.1	Classificação de tipo espectral . . . . .	93
5.3.1.1	<i>K-Nearest Neighbors</i> . . . . .	93
5.3.1.2	<i>Support Vector Machine</i> . . . . .	96
5.3.1.3	<i>Decision Tree Classifier</i> . . . . .	97
5.3.1.4	<i>Random Forest Classifier</i> . . . . .	100
5.3.2	Classificação de tipo espectral e de luminosidade . . . . .	101
5.3.2.1	<i>K-Nearest Neighbors</i> . . . . .	101
5.3.2.2	<i>Support Vector Machines</i> . . . . .	102
5.3.2.3	<i>Decision Tree Classifier</i> . . . . .	103
5.3.2.4	<i>Random Forest Classifier</i> . . . . .	105
5.4	Resultados finais de ML . . . . .	107
<b>6</b>	<b>Conclusão</b>	<b>110</b>
	<b>Referências Bibliográficas</b>	<b>113</b>
<b>A</b>	<b>Classificação de tipo espectral e de luminosidade</b>	<b>116</b>

# Capítulo 1

## Introdução

Neste trabalho de conclusão de curso (TCC), investigamos a performance de diferentes algoritmos de aprendizado de máquina para classificação espectral de estrelas de alta massa. Em particular, consideramos as estrelas de alta massa do tipo O. Sendo assim, iniciaremos apresentando a definição e as características básicas desses objetos, discutiremos brevemente classificação espectral e, em seguida, apresentaremos a motivação e a estrutura do restante do TCC.

### 1.1 Estrelas tipo-O

Estrelas do tipo O são estrelas massivas ( $M > 10M_{\odot}$ ). Elas são progenitoras de objetos fascinantes, como por exemplo LBVs (variáveis luminosas azuis), Wolf-Rayet's e supergigantes vermelhas. No final de suas vidas, dão origem à estrelas de nêutrons, supernovas, “gamma-ray bursts” (do tipo longo) e buracos negros. Se evoluem em pares, podem dar origem à ondas gravitacionais, como as recém detectadas pela colaboração LIGO (Abbott et al., 2016).

Devido às suas altas temperaturas ( $T$  maior que  $\sim 30.000$  K), a maior parte da energia desses objetos é emitida na parte ultravioleta do espectro e, se comparados aos das estrelas de mais baixa massa (p.e., do tipo G), vemos que são objetos que apresentam poucas linhas no ótico, mas muitas na região do ultravioleta como mostrado na Figura 1.1. Em geral vemos absorções de H, He I, He II na região do visível e algumas poucas linhas metálicas (p.e., de C II-III, O III). Os perfis são relativamente alargados,

devido às suas altas taxas de rotação (Howarth et al., 1997). Na figura 1.2, para fins de comparação, mostramos alguns espectros óticos de estrelas do tipo O até M.

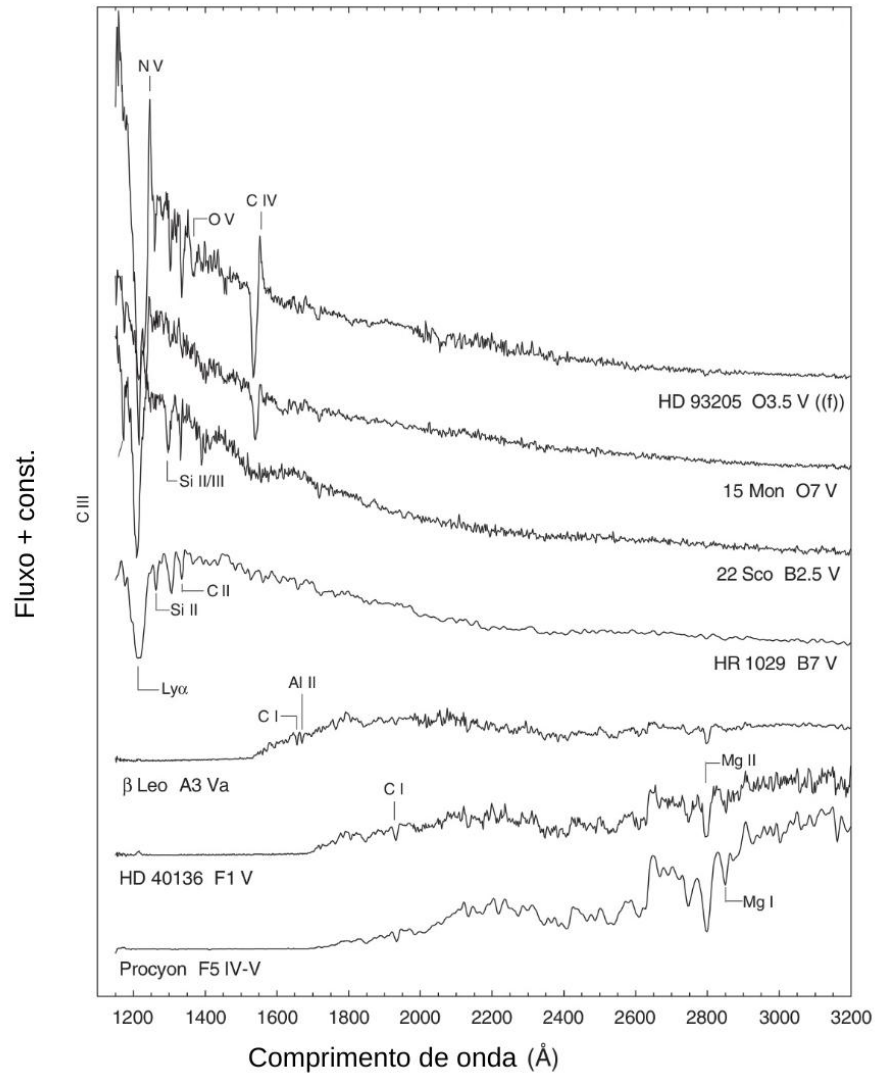


FIGURA 1.1. Comparação entre espectros de tipos espectrais O, B, A e F, na região do UV. Figura extraída de Gray & Corbally (2021). Detalhes sobre classificação espectral serão discutidos mais adiante.

Uma característica fundamental das estrelas O é a presença dos ventos estelares. A atmosfera desses objetos não se encontra em equilíbrio hidrostático, mas sim em contínua expansão. Isso é revelado pelos perfis de linha P-Cygni — absorção azul seguida de emissão vermelha (Lamers et al., 1999) — geralmente encontrados no ultravioleta. No ótico, algumas linhas podem ser parcialmente ou inteiramente formadas no vento (p.e., da série de Balmer), apresentando uma absorção relativamente rasa ou pura emissão (ver p.e., Bouret et al., 2012).

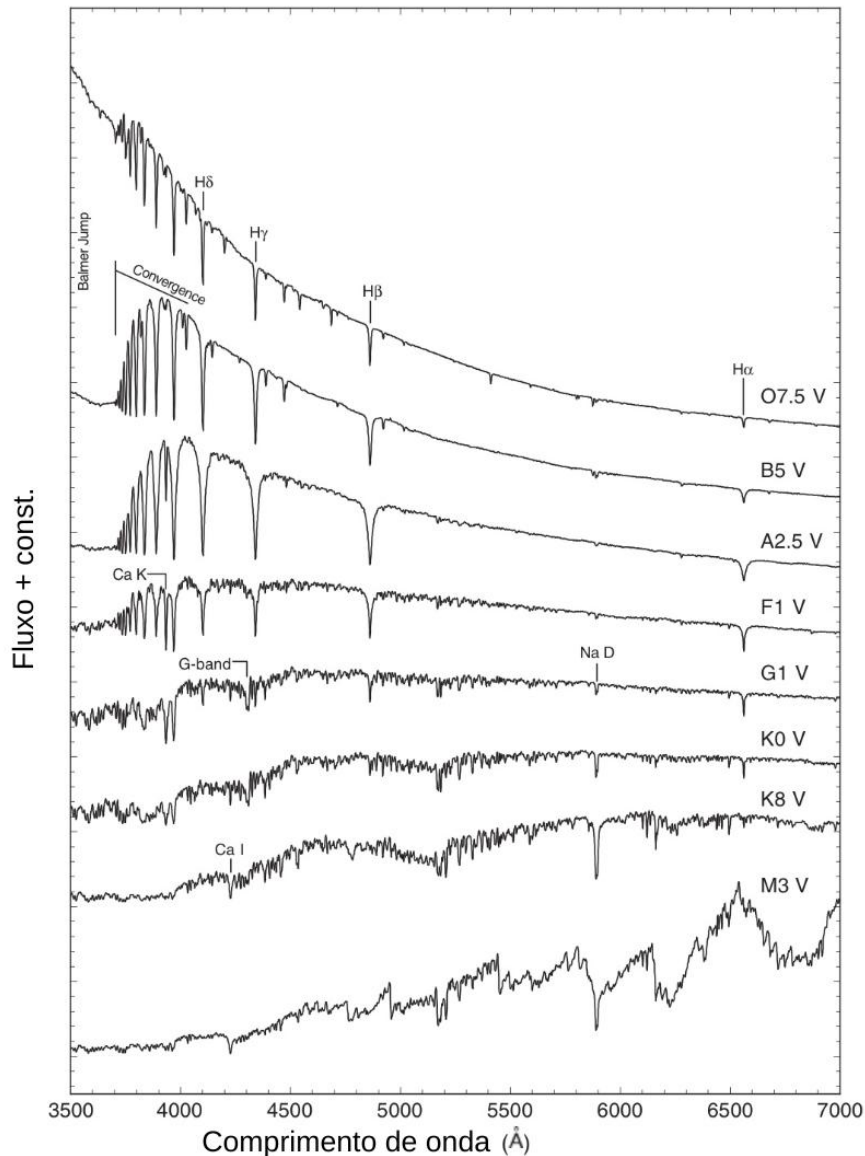


FIGURA 1.2. Comparação entre espectros do tipo O ao M. Figura extraída de Gray & Corbally (2021).

Os ventos representam perdas de massa importantes, com impactos significativos nas trajetórias evolutivas das estrelas massivas (Ekström et al., 2012). Para darmos um exemplo, uma estrela de  $40 M_{\odot}$  pode finalizar a sua vida com apenas  $12 M_{\odot}$ , tendo perdido 70% de sua massa inicial (Ekström et al., 2012). Por isso, os ventos são extremamente importantes para as respectivas galáxias hospedeiras como um todo, por depositar energia e novos elementos químicos, provenientes da perda de massa, no meio interestelar. De fato, o espectro de galáxias *starbursts* em altos *redshifts* revelam perfis de vento de estrelas de alta massa (p.e., Steidel et al. 1996). Para interpretarmos populações estelares adequadamente, precisamos entender bem a dependência da intensidade desses ventos com a metalicidade. Atualmente, isso é um assunto de fronteira na literatura

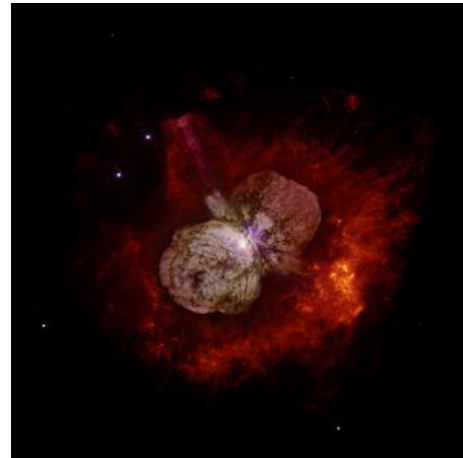
(Marcolino et al., 2022).

Esses objetos são consideradas motores cósmicos (Maeder et al., 2007) devido ao seu potencial de influência na evolução da galáxia na qual estão contidas (Kyritsis et al., 2022), sendo também responsáveis pela existência de regiões ativas de formação estelar, contribuindo fortemente para a nucleossíntese estelar (Irrgang et al., 2014).

De forma a ilustrar o resultado do impacto desses objetos no meio interestelar, podemos citar a Nebulosa de Caranguejo (Figura 1.3a), uma remanescente de supernova que se encontra próxima à constelação de Touro a aproximadamente 6,5 anos-luz de distância da Terra, e  $\eta$  Carinae (Figura 1.3b), uma LBV que sofreu uma grande erupção no século 19 se tornando uma das estrelas mais brilhantes no céu.



(A) Nebulosa do Caranguejo. Crédito da imagem: NASA, ESA, J. Hester, A. Loll (ASU).



(B)  $\eta$  Carinae. Crédito da imagem: Nathan Smith (University of California, Berkeley), and NASA.

FIGURA 1.3. Exemplos de fenômenos causados por estrelas de tipo espectral O. Imagens como essa mostram a influência que as estrelas de alta massa têm em suas vizinhanças.

## 1.2 Busca pela classificação espectral

Como trataremos de classificação espectral estelar via algoritmos de aprendizado de máquina, iremos discutir este tema aqui de forma breve e, mais tarde, em mais detalhes para o caso de estrelas O.

A classificação estelar remonta ao nascimento da espectroscopia, que passou a ser uma importante área de estudo na astronomia após a identificação da linha de absorção do sódio no espectro solar por Joseph Fraunhofer, em 1814. Em 1860, Robert Wilhelm Bunsen e Gustav Kirchoff, observaram que cada elemento químico produzia um conjunto

único de linhas espectrais. A partir disso, a análise espectral passou a ser feita para outras estrelas tendo como objetivo inferir as suas respectivas composições químicas.

Em 1901, Annie Jump Cannon e Williamina P. Fleming, desenvolverem a Classificação Espectral de Harvard baseada na similaridade aparente entre os espectros: "O B A F G K M"<sup>1</sup>, onde atualmente sabemos que as mais quentes são as de tipo espectral O e as mais frias são as de tipo espectral M. Como uma complementação à classificação anterior, também desenvolveram a subclassificação numérica de 0 (estrelas mais quentes) a 9 (estrelas mais frias) — por exemplo, estrelas B9 são menos quentes que estrelas B0.<sup>2</sup> Com esses estudos, posteriormente, foi levantado o questionamento: porque há tantas diferenças entre os espectros?

Essa questão foi respondida por Cecilia Payne em sua tese de doutorado *Stellar Atmospheres; a Contribution to the Observational Study of High Temperature in the Reversing Layers of Stars* (Payne, 1925), que é considerada uma das mais brilhantes teses de doutorado já escritas na Astronomia. Payne percebeu que as diferenças observadas no perfil dos espectros não era dada por uma diferença na sua composição química, mas sim pela diferença de temperatura dos objetos. A temperatura influencia diretamente na quantidade de átomos ionizados e no estado de excitação dos elétrons, de forma que determinados elementos estarão presentes em maior abundância apenas nos espectros de estrelas mais quentes. De forma a ilustrar esse comportamento, a Figura 1.2 (seção 1.1) mostra o perfil de espectros com diferentes temperaturas e a Figura 1.4 mostra a dependência da intensidade das linhas espectrais em relação a temperatura.

A Classificação Espectral de Harvard é utilizada até hoje, porém, outras duas classes L e T, com temperaturas ainda mais baixas que as estrelas de tipo espectral M, foram adicionadas a essa sequência, resultando na sequência: "O B A F G K M L T"<sup>3</sup>.

Para que esse esquema de classificação fosse desenvolvido, Cannon classificou aproximadamente 200.000 espectros entre 1911 e 1914, de forma inteiramente manual, que foram disponibilizadas no catálogo HD (Henry Draper).

<sup>1</sup>Na astronomia, é comum utilizar a frase *Oh Be A Fine Guy/Girl Kiss Me* como uma forma de mnemônico para a ordem das classes de tipo espectral.

<sup>2</sup>Atualmente sabemos a relação direta entre a temperatura e o perfil dos espectros, mas neste momento da história, a classificação é baseada na morfologia dos espectros.

<sup>3</sup>*Oh Be A Fine Guy/Girl Kiss Me — Less Talk!*

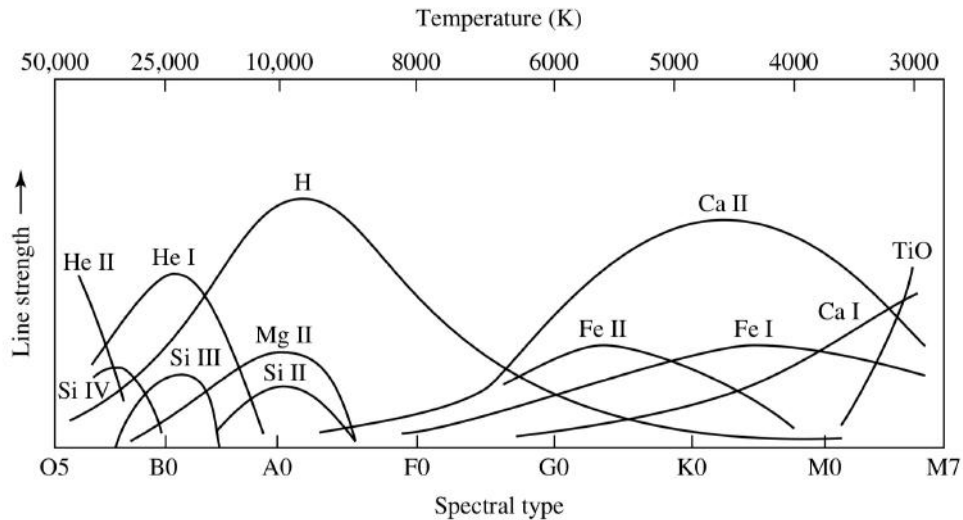


FIGURA 1.4. Figura 8.11 de Carroll & Ostlie (2017): A dependência da intensidade das linhas espectrais com a temperatura.

Posteriormente, Ejnar Hertzsprung e Henry Norris Russel, notaram, ao mesmo tempo, que estrelas de mesmo tipo espectral poderiam ter magnitudes absolutas distintas. Assim, dentro de uma mesma classe de tipo espectral, ambos passaram a chamar as estrelas mais luminosas de gigantes, e Russel começou a chamar as menos luminosas de anãs. Esse conhecimento deu origem, posteriormente, ao Diagrama HR (Hertzsprung-Russel), apresentado na Figura 1.5, no qual observamos uma relação entre a classificação de luminosidade (indicada no eixo  $y$ ), o raio e a temperatura (indicada no eixo  $x$ ) dos objetos.

Outra forma de notar a relação entre temperatura, luminosidade e raio da estrela, é a partir da Lei de Stefan-Boltzmann, indicada na Equação 1.1. Nesta, vemos que para uma temperatura fixa, quanto maior o raio, maior precisa ser a luminosidade da estrela.

$$R = \frac{1}{T_e^2} \sqrt{\frac{L}{4\pi\sigma}} \quad (1.1)$$

Hertzsprung, então, passou a se questionar se a variação da luminosidade poderia ser observada no espectro das estrelas, o que foi respondido a partir da análise dos espectros catalogados por Antonia Maury, junto a Edward Charles Pickering, em Maury & Pickering (1897), onde uma variação, que ela chamou de característica  $c$ , na largura das linhas de absorção, era vista para diferentes magnitudes absolutas.



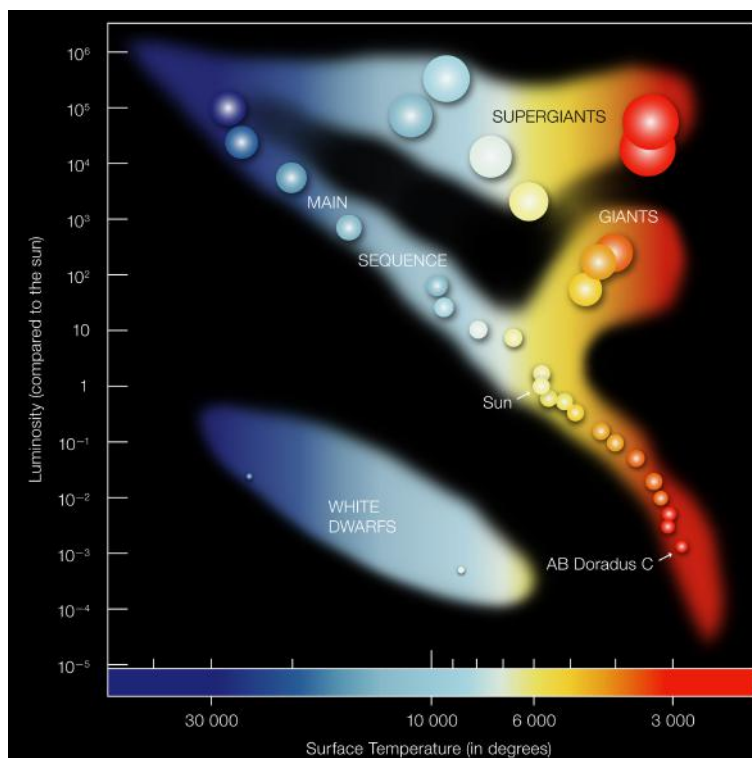


FIGURA 1.5. Diagrama HR. A luminosidade, no eixo  $y$ , está indicada em termos da luminosidade solar, de forma que o Sol possui luminosidade igual a 1. Nesta figura vemos indicado as regiões de classificação de luminosidade para as anãs brancas, para as estrelas da sequência principal (ou anãs), para as gigantes e para as supergigantes. Créditos da imagem: ESO.

Além disso, em seu trabalho intitulado *Radiation from Stars* (Radiação das estrelas), no qual cita o trabalho de Maury, Hertzsprung explora a existência de estrelas de mesma temperatura que são pouco luminosas e menores (anãs), e outras que eram maiores e mais luminosas (gigantes).

Assim, William W. Morgan e Phillip C. Keenan, do Observatório de Yerkes, em 1943, utilizaram os estudos iniciais de Hertzsprung e Maury para desenvolver o Atlas Espectral Estelar, contendo o sistema de classificação MK (Morgan et al., 1943; Morgan & Keenan, 1973) que consiste em uma classificação 2D de temperatura e luminosidade, ainda utilizada nos dias atuais. Diante disso, uma estrela também passou a ser classificada como: D (Anã Branca), VI (Sub-anã), V (Anã), IV (Subgigante), III (Gigante), II (Gigante Luminosa), Ib (Supergigante pouco luminosa) e Ia (Supergigante luminosa) (Carroll & Ostlie, 2017).

### 1.3 Vantagens da classificação via aprendizado de máquina

É importante destacar que a classificação espectral teve início de forma manual a partir da análise visual pelos astrônomos e que, por mais que esteja sujeita a erros, tem sido bem sucedida até mesmo nos tempos atuais. No entanto, diante da grande quantidade de dados sendo obtidos atualmente por telescópios, tanto terrestres quanto espaciais, e pelo crescente número de *surveys* (Kyritsis et al., 2022), o processo de classificação inteiramente manual pode se mostrar inviável.

Atualmente, estamos vivendo a era do *Big Data* na Astronomia (Pesenson et al., 2010) — podemos citar catálogos espectroscópios de estrelas de tipo espectral O e B como IACOB<sup>4</sup> (Simón-Díaz et al., 2011), NoMaDS<sup>5</sup> (Pellerin et al., 2012), e GOSC (Apellániz et al., 2016) — de forma que a implementação de novas formas de tratamento de dados a partir de uma automatização, visando um processo mais acurado e eficiente, é essencial. Diante disso, estudos acerca da aplicação de modelos da inteligência artificial (ou *machine learning*; aprendizado de máquina) na classificação de objetos na Astronomia têm crescido por serem capazes de realizar tarefas sem a necessidade de instruções explícitas. Isso se deve a sua capacidade de analisar centenas e até milhares de dados de uma só vez. Como exemplo, podemos citar os projetos Kuntzer et al. (2016), que buscou realizar a classificação de tipo espectral, em 13 classes entre O5 e M5, a partir do formato do padrão de difração das estrelas utilizando técnicas de PCA e de redes neurais; e Kyritsis et al. (2022), que utilizou três diferentes modelos de ML de florestas aleatórias para realizar a classificação de tipo espectral de estrelas do tipo OB.

### 1.4 Objetivos

Motivados pelos fatos descritos acima e pelo fato de termos acesso à dados de centenas de estrelas O (que serão descritos no Cap. 2), que são objetos fundamentais na Astronomia, o presente TCC teve como objetivo analisar a eficiência de diferentes métodos de aprendizado de máquina na classificação espectral desses objetos. Desta forma, nossos objetivos naturalmente incluíram o domínio de vários algoritmos e métodos, assim como da linguagem Python, escolhida para programação.

---

<sup>4</sup>Projeto observacional que visa a criação de uma base de dados de espectros óticos alta resolução para estrelas de alta massa observadas no hemisfério Norte.

<sup>5</sup>Northern Massive Dim Stars

---

A estrutura da primeira parte desse trabalho foi construída de forma que, inicialmente, o leitor compreenda a qualidade e a fonte dos espectros utilizados, assim como suas principais características (Capítulo 2). Então, apresentaremos na metodologia o processo de derivação dos parâmetros utilizados para classificação — as larguras equivalentes das linhas de diagnóstico — bem como os critérios utilizados para definição das classes de tipo espectral e luminosidade (Capítulo 3). Abordaremos também os conceitos básicos utilizados em ML e uma breve explicação sobre as particularidades e o funcionamento dos modelos aplicados (Capítulo 4). Por fim, teremos uma discussão sobre os resultados obtidos (Capítulo 5) e, por último, a conclusão (Capítulo 6).

## Capítulo 2

# Dados Utilizados

Os conjuntos de espectros de estrelas do tipo O utilizados neste projeto são:

- 113 espectros de alta resolução do artigo *Quantitative Spectral Classification of Galactic O Stars* (Martins, 2018).
- 500 espectros de baixa resolução do GOSC (*Galactic O-Star Catalog*) (Apellániz et al., 2016).

Todos os espectros foram previamente normalizados, de forma que esta não é uma etapa que faz parte deste projeto. Os espectros de alta resolução foram cedidos gentilmente por Fabrice Martins (U. Montpellier/França). Já os de resolução menor são públicos.

Como citado, a nossa amostra inicial de espectros contém 113 espectros de alta resolução e 500 espectros de baixa resolução. No entanto, achamos importante mencionar que, ao final da classificação espectral, a amostra total passará a ter 110 espectros de alta resolução e 483 espectros de baixa resolução. Os motivos pelos quais isso ocorre será explicado em detalhes no Capítulo 3, na seção 3.2.

Abaixo descrevemos os detalhes mais importantes acerca desses dados.

## 2.1 Origem dos espectros

As fontes, o poder resolutor e a faixa de comprimento de onda dos espectros de alta resolução estão dispostos na Tabela 2.1.

Fonte do espectro	Poder Resolutor	Faixa de $\lambda$
<i>SOPHIE Archive</i> <sup>1</sup>	$R = 40.000$	3872Å - 6943Å
<i>ELODIE archive</i> <sup>2</sup>	$R = 42.000$	3850Å - 6800Å
<i>CFHT Science Archive</i> <sup>3</sup>	$R = 68.000$	3700Å - 10500Å
Arquivo PolarBase do espectro NARVAL <sup>4</sup>	$R = 68.000$	3700Å - 10500Å
Arquivo ESO dos espectros FEROS <sup>5</sup>	$R=48.000$	3500Å - 9200Å
HARSPol	$R = 115.000$	3780Å - 6910Å

TABELA 2.1. Fontes e propriedades dos espectros de alta resolução.  $\lambda$  indica os comprimentos de onda.

Os espectros de baixa resolução do GOSC são todos da Via Láctea e foram obtidos pelos espectrógrafos Albireo/OSN (Observatorio de Sierra Nevada, telescópio de 1,5m), TWIN/CAHA (Calar Alto, telescópio 3,5m), Boller & Chivens/LCO (Las Campanas, telescópio du Pont 2,5m), ISIS/WHT-OHM (Telescópio William Herschel, Observatório del Roque de los Muchachos, La Palma - Espanha), e OSIRIS/GTC (Gran Telescopio Canarias, 10,4m em ORM). O poder resolutor de todos os espectros é  $R \sim 2.500$ , de forma que mesmo que algum espectro tivesse uma resolução maior, este seria degradado com o objetivo de se formar um catálogo uniforme (Apellániz et al., 2016). Os intervalos de comprimento de onda desses espectros sofrem variação mas, de modo geral, tem início em 3900Å e terminam em 5100Å. Mais detalhes podem ser encontrados no site do projeto GOSC - <https://gosc.cab.inta-csic.es/galactic-o-star-catalog>. Por exemplo, são fornecidos dados fotométricos, coordenadas e notas sobre determinados objetos (p.e., pertinência à aglomerados, binaridade). Espectros individuais (arquivos no formato FITS) podem ser obtidos diretamente da tabela online.

<sup>1</sup>(Moultaka et al., 2004a)

<sup>2</sup>(Moultaka et al., 2004b)

<sup>3</sup>Centro de Dados Astronômicos Canadense (*Canadian Astronomical Data Center*).

<sup>4</sup>(Petit et al., 2014)

<sup>5</sup>Coletado pelo *OWN survey* (Barbá et al., 2010)

## 2.2 Amostra de espectros

Antes de darmos início às análises deste projeto com os dados citados anteriormente, notamos alguns problemas que exigiram a modificação do conjunto total de espectros, tanto de alta quanto de baixa resolução.

Inicialmente, haviam 508 espectros no conjunto de baixa resolução, mas alguns precisaram ser descartados por não conterem a faixa de comprimentos de onda necessária para observarmos as linhas de absorção que serão utilizadas no processo de classificação. São estes: *Bajamar\_Star\_B2500.fits* (4432 - 5100Å), *Cyg\_OB2-7\_B2500.fits* (4441 - 5100Å), *Sh\_2-158\_1\_B2500.fits* (4441 - 5100Å), *ALS\_7833\_B2500.fits* (4441 - 5100Å), *BD\_+66\_1674\_B2500.fits* (4441 - 5100Å), *Cyg\_OB2-A11\_B2500.fits* (4432 - 5100Å), *ALS\_-15\_108\_AB\_B2500.fits* (4428 - 5100Å) e *BD\_-15\_1909\_B2500.fits* (3900.72 - 3903.39Å).

Além disso, há estrelas que estão presentes tanto no conjunto de espectros de alta quanto no de baixa resolução. Com isso, optamos por excluir 14 espectros do catálogo de alta resolução, que também estavam presentes no catálogo GOSC, visando a uniformidade do conjunto total (devido a maioria ser de baixa resolução). Isso causa uma perda de espectros de qualidade mais alta, porém como a quantidade não é significativa, a diferença nos resultados será desprezível.

Assim, chegamos na amostra de 113 espectros de alta resolução e de 500 espectros de baixa resolução.

Nas Figuras 2.1 e 2.2, temos exemplos de espectros de alta e baixa resolução com diferentes tipos espectrais que foram utilizados nesse projeto. As linhas indicadas são exemplos de linhas de diagnóstico utilizadas para classificação de tipo espectral (He I  $\lambda 4471$  e He II  $\lambda 4542$ ) e de luminosidade (He II  $\lambda 4686$  e He I  $\lambda 4388$ ) em estrelas de tipo espectral O. As estrelas de tipo espectral O9.7 são as mais frias enquanto que as de tipo espectral O3 são as mais quentes.

Podemos notar pelas Figuras 2.1 e 2.2 que a profundidade da linha He I  $\lambda 4471$  diminui conforme a temperatura da estrela aumenta e, de forma contrária, a profundidade da linha He II  $\lambda 4542$  aumenta junto à temperatura. Isso ocorre porque, quanto maior a temperatura, maior a abundância de átomos ionizados no meio.

No caso da linha He II  $\lambda 4686$ , o seu comportamento é influenciado pelo vento estelar, o que está diretamente relacionado à classe de luminosidade desses objetos (Conti & Leep, 1974). Assim, de acordo com Martins (2018), o perfil da linha de absorção He II  $\lambda 4686$  em estrelas supergigantes (classe I) será em emissão, em estrelas gigantes (classe III) será uma forte absorção e em estrelas anãs (classe V) será uma absorção muito forte, o que é visível nas Figuras 2.1 e 2.2.

No entanto, para estrelas com classificação espectral entre O8 e O9.7, a intensidade da linha He II  $\lambda 4686$  diminui consideravelmente devido a sensibilidade a temperatura da linha (Mathys, 1988). De forma que utilizar apenas a sua influência para definir a classe de luminosidade do objeto não será o suficiente. Diante disso, utiliza-se também a linha de absorção He I  $\lambda 4388$ , que assim como He II  $\lambda 4686$ , é mais forte em anãs que em supergigantes, mas não sofre tanta influência da temperatura. Os detalhes quantitativos acerca da classificação espectral serão discutidos no próximo capítulo.

É importante notar que, por mais que a maior parte das linhas em espectros de estrelas quentes esteja na forma de absorção, somos capazes de enxergar picos estreitos na Figura 2.1 — diferentemente das linhas de emissão formadas em  $\lambda 4640$ , aproximadamente, e em He II  $\lambda 4686$ , que são mais largas. Estes picos podem ter sido causadas por raios cósmicos ou por defeitos no próprio CCD no momento de captura do espectro. Assim, estas linhas não fazem parte do espectro e podemos simplesmente desconsiderá-las na nossa análise.

A discussão de como se procedeu a classificação e quais critérios foram utilizados, será feito no próximo capítulo.

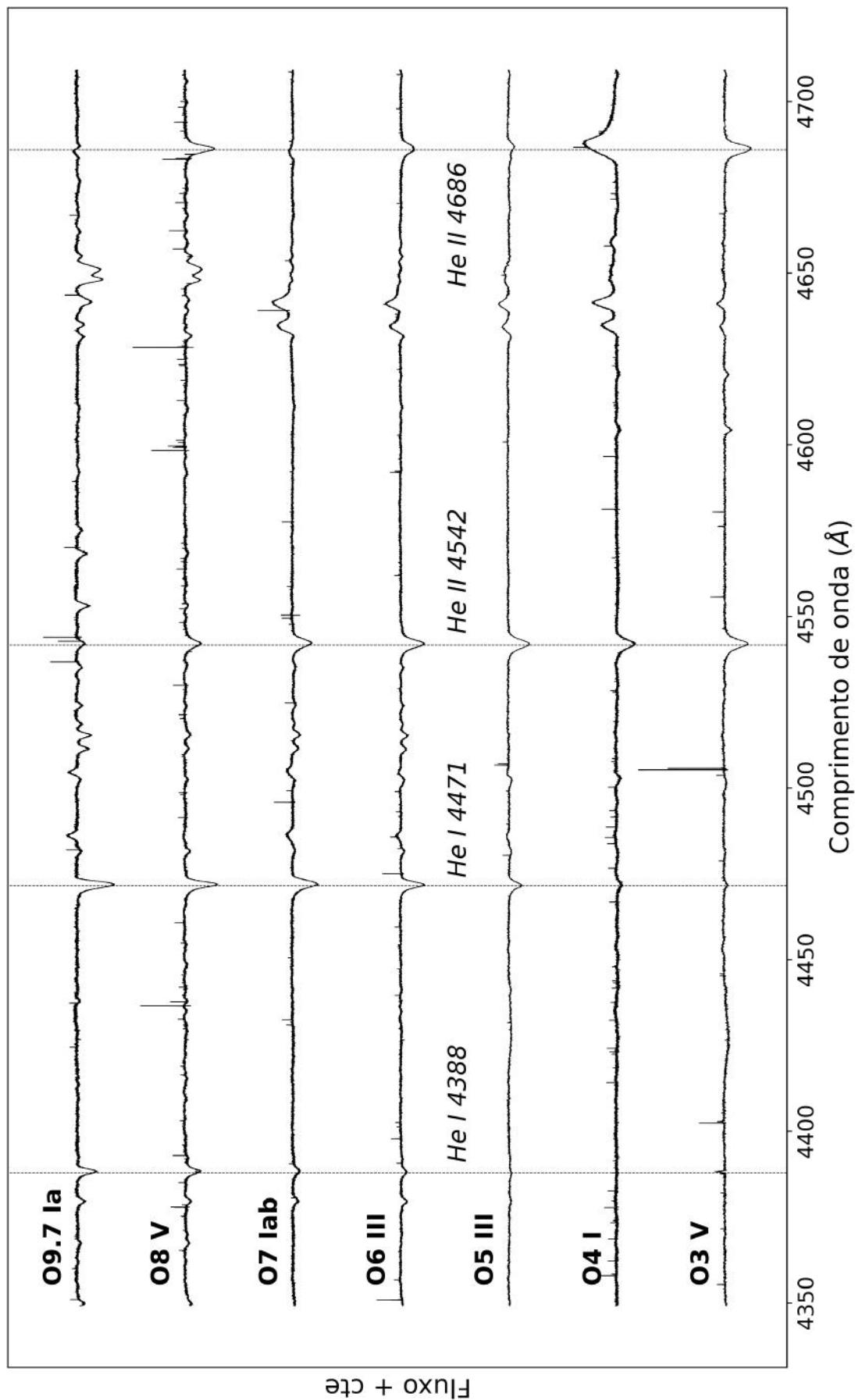


FIGURA 2.1. Exemplos de espectros de alta resolução. A profundidade da linha He I  $\lambda 4471$  diminui conforme a temperatura da estrela aumenta e, de forma contrária, a profundidade da linha He II  $\lambda 4542$  aumenta junto à temperatura. Já o perfil das linhas de absorção He II  $\lambda 4686$  e He I  $\lambda 4388$  em estrelas supergigantes (classe I) será em emissão, em estrelas gigantes (classe III) será uma forte absorção e em estrelas anãs (classe V) será uma absorção muito forte.



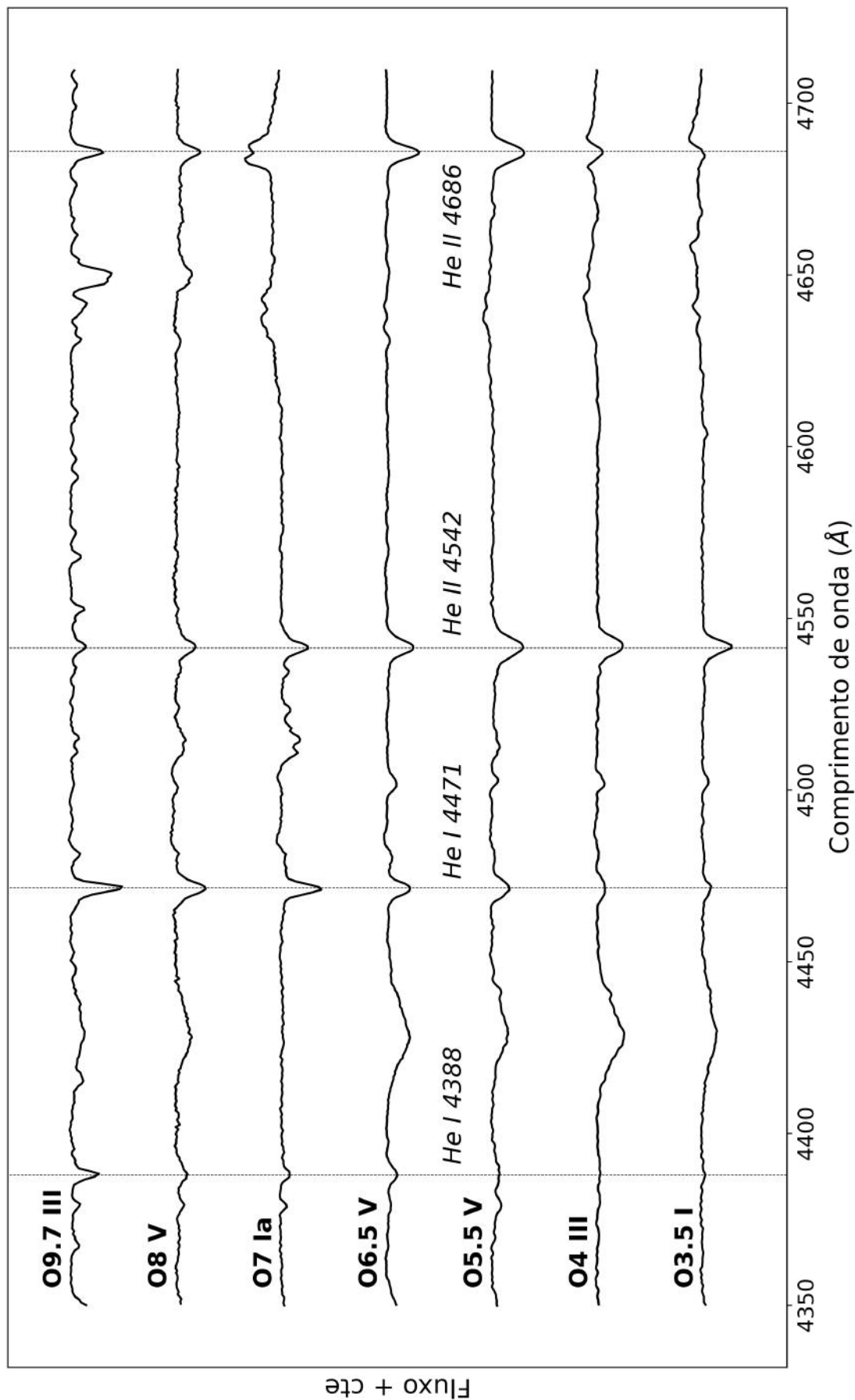


FIGURA 2.2. Exemplos de espectros de baixa resolução. A profundidade da linha He I  $\lambda 4471$  diminui conforme a temperatura da estrela aumenta e, de forma contrária, a profundidade da linha He II  $\lambda 4542$  aumenta junto à temperatura. Já o perfil das linhas de absorção He II  $\lambda 4686$  e He I  $\lambda 4388$  em estrelas supergigantes (classe I) será em emissão, em estrelas gigantes (classe III) será uma forte absorção e em estrelas anãs (classe V) será uma absorção muito forte.

## Capítulo 3

# Metodologia

Neste capítulo iremos explorar os critérios utilizados na literatura para a classificação de tipo espectral e de luminosidade de estrelas de tipo O. Todos foram definidos a partir da largura equivalente: medida que nos indica a profundidade das linhas de absorção utilizadas para classificação — ver seção 3.1. Por fim, iniciaremos a discussão sobre o processo de escolha dos modelos de *Machine Learning*. Porém, as suas particularidades e modos de operação serão vistos em maior profundidade no Capítulo 4.

A classificação moderna dos objetos com tipo espectral O é derivada dos artigos publicados por N. R. Walborn (Walborn, 1971, 1972, 1973), P. S. Conti (Conti & Alschuler, 1971; Conti & Leep, 1974; Conti & Frost, 1977) e G. Mathys (Mathys, 1988) e, neste projeto, iremos nos basear nesses estudos para realizar a tarefa de classificação.

É importante ressaltar que, inicialmente, o objetivo deste trabalho era a classificação de tipo espectral e, posteriormente, a classificação de tipo espectral e luminosidade para os 113 espectros de alta resolução (Martins, 2018). Porém, os resultados utilizando somente essas amostras não se mostraram satisfatórios — ver Capítulo 5. Assim, adicionamos ao conjunto inicial 500 espectros de baixa resolução (Apellániz et al., 2016), de forma a verificar se a aumentaríamos a capacidade de aprendizado de máquina utilizando uma quantidade maior de espectros, mesmo que estes tenham menor qualidade.

### 3.1 Largura Equivalente

Como vimos no Capítulo 2, o perfil das linhas de absorção varia com a temperatura do objeto e, com isso, podemos utilizar a medida da profundidade dessas linhas como um parâmetro de classificação de tipo espectral e de luminosidade. Essa medida é dada pela Largura Equivalente  $W_\lambda$ , ilustrada na Figura 3.1, que é definida pela largura de um retângulo de altura 1 que possua a mesma área da região formada entre o perfil de absorção e o fluxo do contínuo ( $F_\lambda/F_c = 1$ ), sendo  $F_c$  o fluxo do contínuo e  $F_\lambda$  o fluxo para um determinado comprimento de onda  $\lambda$ .

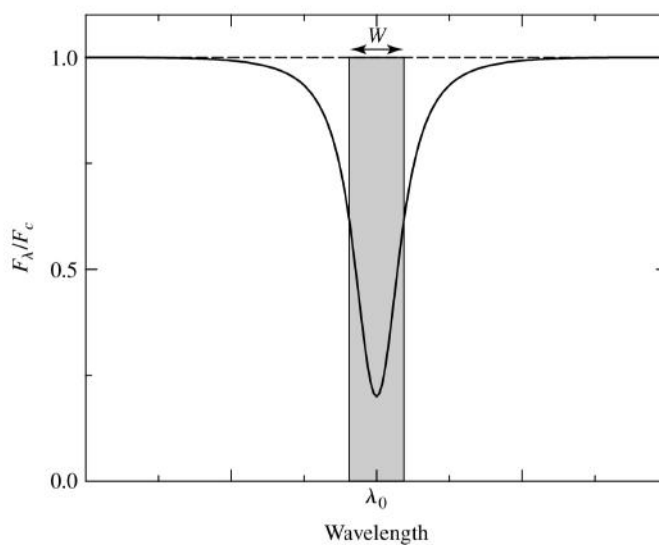


FIGURA 3.1. Representação da largura equivalente. Extraído de Carroll & Ostlie (2017).

Como o termo  $(F_c - F_\lambda)/F_c$  é definido como a profundidade da linha, podemos escrever a expressão 3.1 para a largura equivalente.

$$W_\lambda = \int \frac{F_c - F_\lambda}{F_c} d\lambda \quad (3.1)$$

Neste projeto, as larguras equivalentes foram medidas de forma automatizada, a partir de um código desenvolvido por nós em *Python 3* que funciona da seguinte forma: dados os intervalos de comprimento de onda entre  $\lambda_i$  e  $\lambda_f$  (inicial e final, respectivamente), soma-se a contribuição de fluxo para cada  $d\lambda$ , encontrando, por fim, o valor total da largura equivalente, da forma como está indicada na Figura 3.2. Além disso,

nosso código é capaz de medir  $W_\lambda$  para todos os objetos e para quantas linhas forem necessárias de uma só vez.

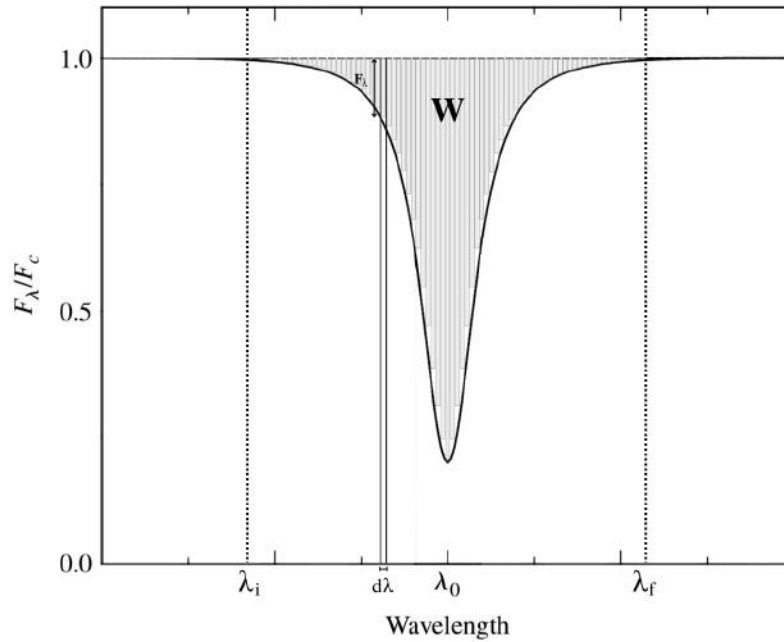


FIGURA 3.2. Representação do somatório para encontrar a largura equivalente. Figura adaptada de Carroll & Ostlie (2017).

Assim, podemos definir a expressão 3.2 que representa esse somatório, na qual  $W_\lambda$ , agora, corresponde a área total da linha de absorção abaixo do contínuo.

$$W_\lambda = \sum_{\lambda_i}^{\lambda_f} \frac{F_c - F_{\lambda_i}}{F_c} * (\lambda_{i+1} - \lambda_i) = \sum_{\lambda_i}^{\lambda_f} \left(1 - \frac{F_{\lambda_i}}{F_c}\right) d\lambda \quad (3.2)$$

Diante disso, utilizaremos a medida de  $W_\lambda$  das linhas de diagnóstico para tipo espectral e luminosidade para realizar a classificação e, na próxima seção, veremos quais foram os critérios utilizados.

## 3.2 Critérios de classificação

### 3.2.1 Classificação de tipo espectral

Para a classificação de tipo espectral, as linhas de diagnóstico comumente utilizadas na literatura são He II  $\lambda 4542$  e He I  $\lambda 4471$  e utilizamos o critério de Conti &

Leep (1974); Conti & Frost (1977); Mathys (1988). Esses autores definem a grandeza  $W'$  como sendo o seguinte logaritmo:

$$W' \equiv \log[W_\lambda(\text{HeI}4471)/W_\lambda(\text{HeII}4542)] \quad (3.3)$$

Os intervalos utilizados para a classificação de tipo espectral estão indicados na Tabela 3.1 (Mathys, 1988). Por exemplo, se na prática medimos a largura de He I  $\lambda 4471$  e a de He II  $\lambda 4542$ , e vemos que o logaritmo da razão dessas larguras (eq. 3.3) é 0.15, a classificação da estrela será então O8.

Tipo Espectral	Intervalos $W'$
O9.7	$W' \geq +0,65$
O9.5	$+0,65 > W' \geq +0,45$
O9	$+0,45 > W' \geq +0,30$
O8.5	$+0,30 > W' \geq +0,20$
O8	$+0,20 > W' \geq +0,10$
O7.5	$+0,10 > W' \geq 0$
O7	$0 > W' \geq -0,10$
O6.5	$-0,10 > W' \geq -0,20$
O6	$-0,20 > W' \geq -0,30$
O5.5	$-0,30 > W' \geq -0,45$
O5	$-0,45 > W' \geq -0,60$
O4	$-0,60 > W' \geq -0,90$
O3	$-0,90 > W'$

TABELA 3.1. Intervalos de  $W'$  para a classificação de tipo espectral das estrelas (Mathys, 1988).

No entanto, para os espectros classificados entre O8 e O9.7, um ajuste no critério de classificação se fez necessário devido a uma sobreposição das medidas de  $W'$  para essas classes (p.e., uma estrela com  $W' = 0.45$ , poderia ser uma O8.5, O9, O9.5 ou O9.7) — ver Figura 3.3, de Martins (2018).

Por consequência, precisamos utilizar a Tabela 3.2 de Martins (2018), na qual temos novos critérios de diagnóstico para essas classes espectrais utilizando a razão entre as larguras equivalentes de linhas de absorção do hélio e também do silício. Além

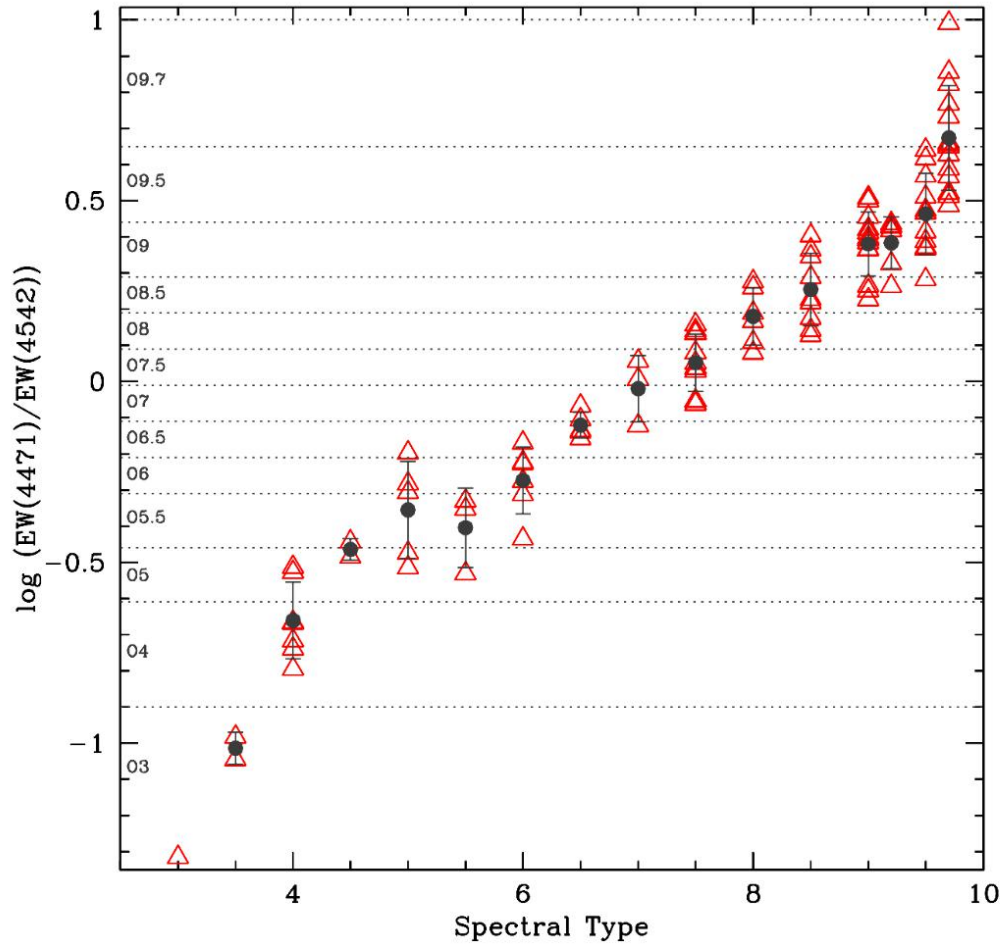


FIGURA 3.3. No eixo  $x$  vemos os tipos espectrais e no eixo  $y$  os respectivos valores de  $W'$  dos espectros. Os pontos em cinza escuro indicam a média dos valores de  $W'$  para cada tipo espectral e, as barras de erro, a dispersão. Figura extraída de Martins (2018).

disso, notamos agora a presença de uma nova classe: O9.2, de forma que passamos a ter 14 classes de tipo espectral distintas.

De maneira a compreendermos a origem dos intervalos presentes na Tabela 3.2, podemos fazer uma análise mais detalhada do perfil das linhas de absorção nela contidas — He I  $\lambda 4144$ , He II  $\lambda 4200$ , He I  $\lambda 4388$ , He II  $\lambda 4542$  e Si III  $\lambda 4552$  — a partir da Figura 3.4, na qual utilizamos espectros de alta resolução de objetos classificados entre O8 e O9.7. Na figura, vemos que as linhas de hélio neutro (He I  $\lambda 4144$  e He I  $\lambda 4388$ ) e de Si III  $\lambda 4552$  são mais profundas para estrelas mais frias e as linhas de hélio ionizado (He II  $\lambda 4200$  e He II  $\lambda 4542$ ) são mais profundas para as mais quentes <sup>1</sup>. Desta forma,

<sup>1</sup>Por mais que o silício também esteja ionizado, e energia para o seu segundo potencial de ionização é  $\sim 1577, 1kJ/mol$ , enquanto que para o primeiro potencial de ionização do He é  $\sim 2372, 3kJ/mol$ . Assim, o hélio ionizado ocorre em temperaturas mais altas do que o silício duas vezes ionizado.

como as razões das larguras equivalentes da tabela são formadas entre as linhas do He I e do Si III sobre as linhas de He II, espera-se que para estrelas mais frias o resultado dessa razão seja maior do que para estrelas quentes, devido a maior influência da largura equivalente do He I e do Si III. Isso é exatamente o que observamos.

Tipo Espectral	$\frac{W_{\lambda}(HeI4144)}{W_{\lambda}(HeII4200)}$	$\frac{W_{\lambda}(HeI4388)}{W_{\lambda}(HeII4542)}$	$\frac{W_{\lambda}(SiIII4552)}{W_{\lambda}(HeII4542)}$
O8	$0,38 \pm 0,14$	$0,55 \pm 0,25$	$0,07 \pm 0,04$
O8.5	$0,67 \pm 0,15$	$0,96 \pm 0,17$	$0,13 \pm 0,04$
O9	$0,85 \pm 0,28$	$1,08 \pm 0,20$	$0,28 \pm 0,09$
O9.2	$1,11 \pm 0,27$	$1,14 \pm 0,13$	$0,31 \pm 0,09$
O9.5	$1,37 \pm 0,32$	$1,53 \pm 0,40$	$0,39 \pm 0,17$
O9.7	$1,98 \pm 1,00$	$2,45 \pm 0,99$	$0,96 \pm 0,57$

TABELA 3.2. Tabela 1 em Martins (2018). Intervalos da razão entre linhas de absorção usadas para diagnóstico de tipo espectral entre O8 e O9.7.

É importante ressaltar que os intervalos da Tabela 3.2, de Martins (2018), baseiam-se somente nos espectros de alta resolução, de forma que não há uma relação desenvolvida especificamente para os espectros de baixa resolução. Assim, mantivemos os mesmos intervalos para todos os espectros utilizados nesse projeto de forma a garantir a uniformidade da classificação.

A exigência de que um espectro passe a acordar com três intervalos de razões de linhas distintas resultará em uma classificação mais precisa, por mais que ainda exista uma sobreposição entre as classes quando analisamos os gráficos individualmente. Por exemplo, ao realizarmos a classificação do espectro de alta resolução da nossa amostra HD14633, chegamos em  $W' = 0,34$ , o que indicaria que é uma estrela de tipo O9. Porém, ao utilizarmos os intervalos da tabela 3.2 chegamos em:  $\frac{W_{\lambda}(HeI4144)}{W_{\lambda}(HeII4200)} \approx 0,59$ ,  $\frac{W_{\lambda}(HeI4388)}{W_{\lambda}(HeII4542)} \approx 0,99$  e  $\frac{W_{\lambda}(SiIII4552)}{W_{\lambda}(HeII4542)} \approx 0,13$ , o que indica que seu tipo espectral é O8.5, que é a sua classificação real de acordo com Martins (2018).

É importante notar que cinco espectros de baixa resolução obtiveram valores de  $W_{\lambda}(HeI4471)$  negativos, o que enxergamos como sendo um reflexo da resolução desses espectros. Assim, precisamos retirá-los da amostra porque não conseguimos realizar

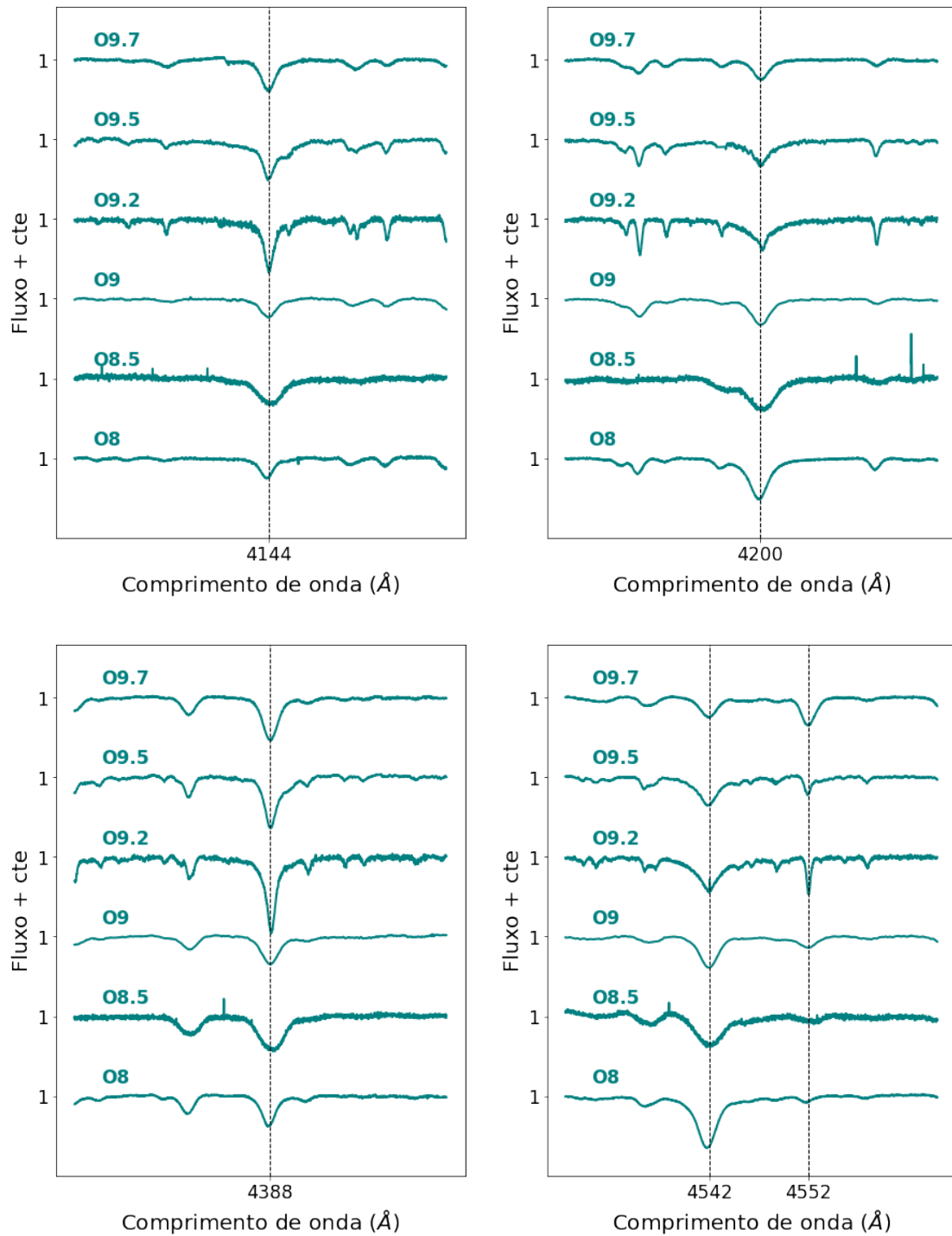


FIGURA 3.4. Espectros de alta resolução classificados entre O8 e O9.7.

a operação de log para definir os seus tipos espectrais. Diante disso, finalizamos a classificação espectral com 113 espectros de alta resolução e 495 espectros de baixa resolução.

Com isso, finalizamos os critérios para a classificação de tipo espectral e podemos explorar a classificação também em termos das classes de luminosidade.



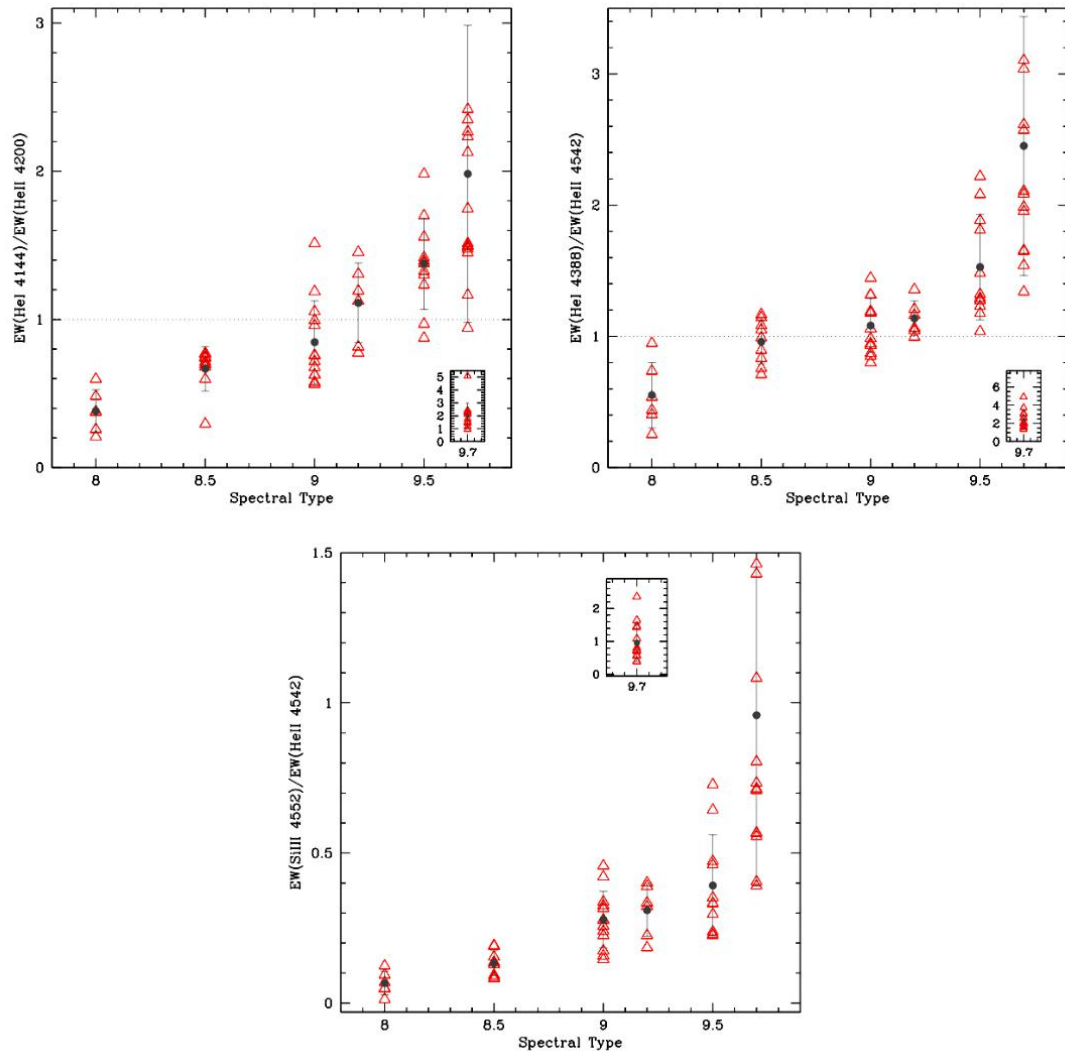


FIGURA 3.5. Razões entre as larguras equivalentes das linhas utilizadas para classificação de tipo espectral de objetos já classificados entre O8 e O9.7. Figura extraída de Martins (2018).

### 3.2.2 Classificação de luminosidade

As estrelas tipo O podem ser classificadas nas seguintes luminosidades: V (Anã), IV (Subgigante), III (Gigante), II (Gigante Luminosa), Ib (Supergigante pouco luminosa) e Ia (Supergigante luminosa) (Carroll & Ostlie, 2017).

Assim, ao realizarmos uma classificação de tipo espectral e de luminosidade, teríamos um conjunto de 84 classes distintas. Isso será um problema quando atingirmos a etapa de aplicação dos dados nos modelos de *Machine Learning* pois dificultará o processo de aprendizado — o que será visto em mais detalhes no Capítulo 5.

Desta forma, optamos por utilizar apenas três classes: V (Anãs), III (juntando as

Subgigantes, Gigantes e Gigantes luminosas) e I (juntando Supergigantes pouco luminosas e Supergigante luminosa). Com isso, teremos 42 classes possíveis, metade do que seria caso utilizássemos todas as classes de luminosidade. Desta forma, a partir desse momento, ao mencionarmos a classe das Gigantes III, estaremos nos referindo à união das classes das Subgigantes, Gigantes e Gigantes luminosas.

Nas próximas duas subseções apresentaremos os critérios utilizados para classificação de luminosidade dos objetos com tipos espectrais entre O3 e O7.5, e entre O8 e O9.7, respectivamente.

### 3.2.2.1 Classificação de estrelas de tipo O3-O7.5

Para as estrelas com tipos espectrais entre O3 e O7.5, utilizamos o critério presente em Martins (2018), que, inicialmente, separa os espectros em anãs e supergigantes a partir de  $W_\lambda(\text{HeII}\lambda 4686)$  — os intervalos são apresentados na Tabela 3.3.

Classe de Luminosidade	Intervalos de $W_\lambda(\text{He II } \lambda 4686)$
V	$W_\lambda(\text{He II } \lambda 4686) > +0,60$
Diversas possibilidades	$+0,60 \geq W_\lambda(\text{He II } \lambda 4686) \geq -0,20$
I	$-0,20 > W_\lambda(\text{He II } \lambda 4686)$

TABELA 3.3. Intervalos de  $W_\lambda(\text{HeII}\lambda 4686)$  para a classificação de luminosidade dos espectros entre O3-O7.5 (Martins, 2018).

No entanto, observamos que não existe uma relação direta para a classificação de estrelas gigantes, dado que para  $-0,20 \leq W_\lambda(\text{HeII}\lambda 4686) \leq 0,60$ , podemos ter anãs, gigantes e supergigantes. Diante disso, utilizamos o fenômeno  $f$ , que indica a relação entre a intensidade de absorção, ou de emissão, das linhas do N III  $\lambda 4634\text{-}40\text{-}42$  e da linha do He II  $\lambda 4686$ . Este é definido nas seguintes formas:

- $((f))$ : fraca emissão das linhas de N III  $\lambda 4634\text{-}40\text{-}42$  e forte absorção da linha de He II  $\lambda 4686$ ;
- $(f)$ : emissão mediana das linhas de N III  $\lambda 4634\text{-}40\text{-}42$  e fraca ou neutra absorção da linha de He II  $\lambda 4686$ ;
- $f$ : forte emissão da linha He II  $\lambda 4686$  e das linhas N III  $\lambda 4634\text{-}40\text{-}42$ .

Definimos um espectro como sendo  $((f))$ ,  $(f)$  ou  $f$  a partir de intervalos de  $W_\lambda(\text{He II } \lambda 4686)$ , indicados na Tabela 3.4 (Martins, 2018).

Classificador $f$	Intervalos de $W_\lambda(\text{He II } \lambda 4686)$
$((f))$	$W_\lambda(\text{He II } \lambda 4686) > +0,60$
$(f)$ ou $((f))$	$+0,60 \geq W_\lambda(\text{He II } \lambda 4686) > +0,40$
$(f)$	$+0,40 \geq W_\lambda(\text{He II } \lambda 4686) \geq 0$
$f$ ou $(f)$	$0 > W_\lambda(\text{He II } \lambda 4686) \geq -0,20$
$f$	$-0,20 > W_\lambda(\text{He II } \lambda 4686)$

TABELA 3.4. Intervalos de  $W_\lambda(\text{He II } \lambda 4686)$  para o fenômeno  $f$  de Martins (2018).

Após definirmos o fenômeno  $f$ , realizamos a sua correlação com o tipo espectral de cada objeto, utilizando a Tabela 3.5, adaptada de Sota et al. (2014). É possível notar que nas Tabelas 3.4 e 3.5, há intervalos que não definem somente um fenômeno  $f$  ou uma classe de luminosidade em específico (p.e., uma estrela de tipo espectral O4 com  $0,40 < W_\lambda(\text{He II } \lambda 4686) < 0,60$ , poderia pertencer tanto à classe das anãs quanto à das gigantes). Diante disso, para os 13 espectros de alta resolução com múltiplas classificações que encontramos, fizemos uma análise visual direta comparando o perfil das linhas de N III  $\lambda 4634\text{-}40\text{-}42$  e de He II  $\lambda 4686$ , diminuindo, assim, as possibilidades de classificação de luminosidade.

Tipo Espectral	$((f))$	$(f)$	$f$
O3-O5.5	V	III	I
O6-O6.5	V-III	III-I	I
O7-O7.5	V-III	III-I	I

TABELA 3.5. Relação entre tipos espectrais e o fenômeno  $f$  de Sota et al. (2014). Esta tabela foi adaptada utilizando as classes de luminosidade deste projeto e contendo apenas as relações para espectros entre O3 e O7.5.

De forma a ilustrar essa análise visual direta dos espectros que obtiveram múltiplas classificações, mostramos na Figura 3.6 o espectro da estrela HD 155806, no qual podemos comparar a profundidade das linhas do N III e do He II. Assim, vemos que N III está em fraca emissão e He II está em forte absorção, chegando a conclusão de podemos classificar esse objeto como  $((f))$ , o que está de acordo com Martins (2018).

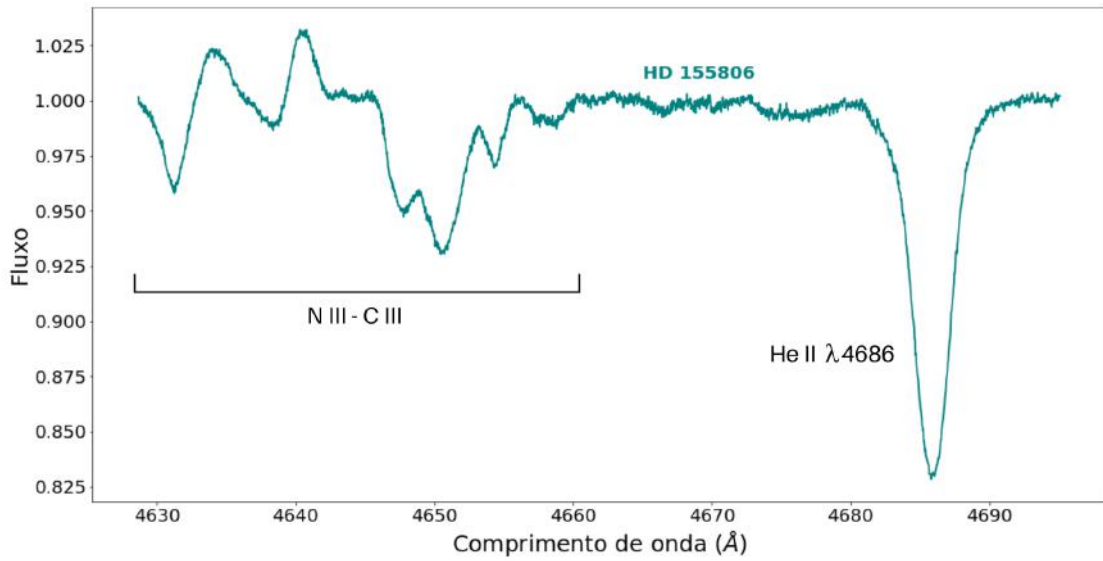


FIGURA 3.6. Linhas do N III, C III e He II no espectro da estrela HD 155806, para a qual obtivemos múltiplas classificações.

Porém, é importante notar que todos os espectros com tipos espectrais entre O6 e O7.5, tendo o fenômeno  $f$  definido como ( $f$ ) ou ( $(f)$ ), ainda possuem mais de uma possibilidade de classe de luminosidade, podendo ser classificados como V ou III e III ou I, respectivamente. Assim, como haviam poucos objetos nessa situação, assumimos estes como sendo gigantes para evitar uma análise visual ainda mais profunda de cada espectro.

No caso dos espectros de baixa resolução, a grande quantidade de objetos encontrados com múltiplas classificações (85 objetos), impediu a análise visual direta. Diante disso, criamos uma relação utilizando  $W_\lambda(\text{He II } \lambda 4686)$  para classificação de luminosidade, a partir da Figura 3.7, de Martins (2018).

Os intervalos de  $W_\lambda(\text{He II } \lambda 4686)$  definidos para os diferentes tipos espectrais desses objetos podem ser vistos na Tabela 3.6.

	I	III	V
O3-4	$W_\lambda(\lambda 4686) < -0.2$	$-0.2 < W_\lambda(\lambda 4686) < 0.6$	$W_\lambda(\lambda 4686) > 0.6$
O5-6	$W_\lambda(\lambda 4686) < -0.2$	$-0.2 < W_\lambda(\lambda 4686) < 0.3$	$W_\lambda(\lambda 4686) > 0.3$
O6.5-7.5	$W_\lambda(\lambda 4686) < 0.25$	$0.25 < W_\lambda(\lambda 4686) < 0.6$	$W_\lambda(\lambda 4686) > 0.6$

TABELA 3.6. Sistema utilizado para classificação de luminosidade de espectros classificados entre O3-O7.5. Desenvolvido a partir da Figura 3.7 de Martins (2018).

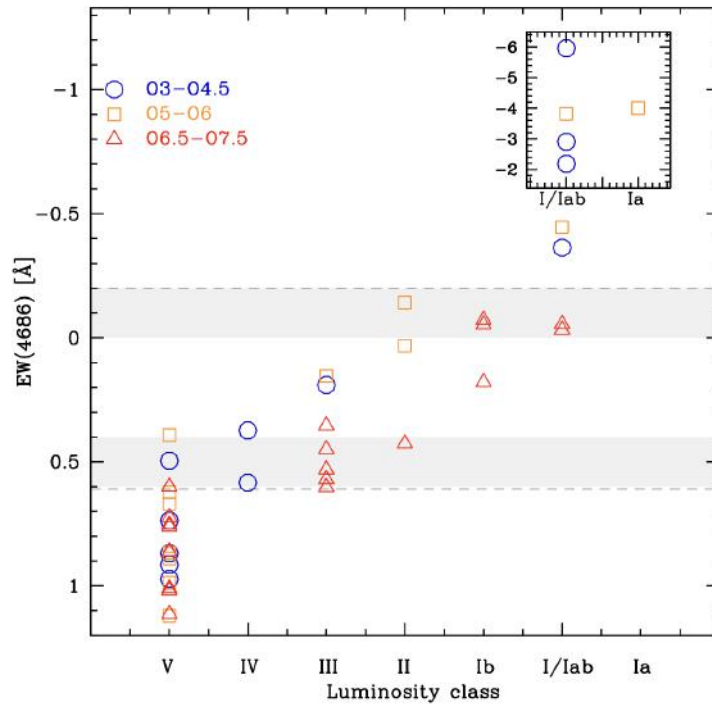


FIGURA 3.7. Figura de Martins (2018). No eixo  $x$  vemos as classes de luminosidade e no eixo  $y$  os valores correspondentes de  $W_{\lambda}(\text{He II } \lambda 4686)$ .

E, com isso, finalizamos a classificação de luminosidade de espectros entre O3 e O7.5.

### 3.2.2.2 Classificação de espectros O8-O9.7

Para as estrelas com tipos espectrais entre O8 e O9.7, utilizar apenas a linha He II  $\lambda 4686$  não será o suficiente, devido a sua sensibilidade à temperatura, tornando-se fraca em temperaturas mais baixas. Porém, em Mathys (1988), é definido uma nova grandeza  $W^+$ , que utiliza a linha He I  $\lambda 4388$ , em conjunto com He II  $\lambda 4686$ , por continuar forte em temperaturas mais baixas e se manter sensível à luminosidade do objeto. Essa grandeza é definida abaixo:

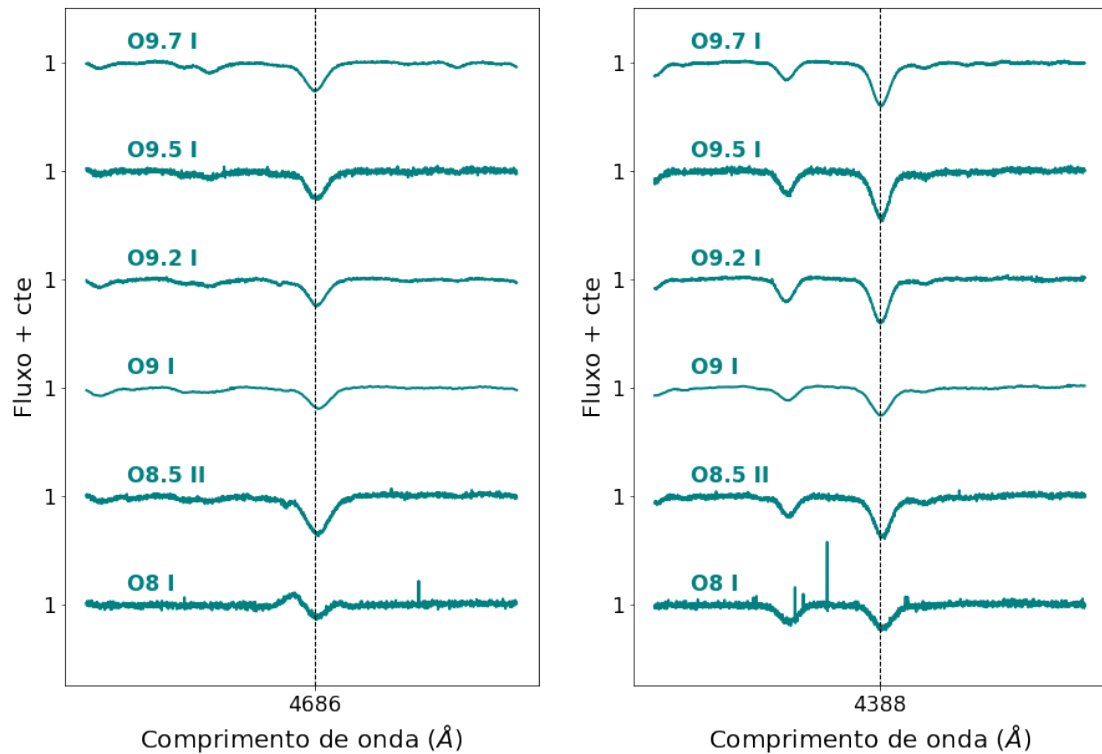
$$\log W^+ \equiv \log[W_{\lambda}(\text{HeI}4388)] + \log[W_{\lambda}(\text{HeII}4686)] \quad (3.4)$$

Os intervalos para a classificação são apresentados na Tabela 3.7, de Mathys (1988).

Classe de Luminosidade	Limite $\log W^+$
V	$\log W^+ \geq 5,35$
III	$5,35 > \log W^+ \geq 5,05$
I	$5,05 > \log W^+$

TABELA 3.7. Intervalos de  $W^+$  para a classificação de luminosidade dos espectros.

A relação entre a profundidade das linhas He I  $\lambda 4388$  e He II  $\lambda 4686$  e a temperatura se torna mais nítida ao observarmos a Figura 3.8, na qual temos diferentes espectros de estrelas supergigantes, e gigante luminosa<sup>2</sup>, de alta resolução classificadas entre O8 e O9.7. Note que a profundidade da linha He II  $\lambda 4686$  se torna cada vez mais fraca para objetos mais frios, enquanto que He I  $\lambda 4388$  se mantém forte.

FIGURA 3.8. Espectros de alta resolução de estrelas supergigantes classificadas entre O8 e O9.7. Nesta figura comparamos a profundidade das linhas de absorção He II  $\lambda 4686$  e He I  $\lambda 4388$ .

É importante notar que, para realizar a classificação de luminosidade, precisamos que tanto as larguras equivalentes da linha He I  $\lambda 4388$  e He II  $\lambda 4686$  sejam positivas para ser possível realizar a operação de  $\log$  proposta por Mathys (1988). No entanto, tivemos três espectros de alta resolução e 12 espectros de baixa resolução que obtiveram

<sup>2</sup>Não há estrelas supergigantes de tipo espectral O8.5 em Martins (2018).

valores negativos e, conseqüentemente, precisamos retirá-los da amostra total. Assim, a classificação de tipo espectral e de luminosidade foi feita para 110 espectros de alta resolução e 483 espectros de baixa resolução.

E, com isso, finalizamos a classificação de tipo espectral e luminosidade.

### 3.3 Intervalos de integração

Os intervalos de comprimento de onda utilizados para definir as larguras equivalentes das linhas de absorção citadas na seção anterior, foram definidos a partir de uma análise visual direta de diferentes espectros, verificando onde seria o início  $\lambda_i$  e o fim da linha  $\lambda_f$ .

Os valores definidos de  $\lambda_i$  e  $\lambda_f$ , tanto para espectros de alta quanto de baixa resolução, estão listados na Tabela 3.8.

Linha Espectral	Alta Resolução	Baixa Resolução
He I 4471	(4467.80, 4473.80)	(4465.20, 4476.20)
He II 4542	(4538.10, 4545.10)	(4536.00, 4548.00)
He II 4686	(4681.60, 4690.10)	(4680.00, 4691.50)
He I 4388	(4385.80, 4390.30)	(4383.60, 4392.10)
Si III 4552	(4550.15, 4554.15)	(4549.00, 4555.00)
He I 4144	(4142.10, 4146.10)	(4140.00, 4148.00)
He I 4200	(4196.60, 4203.60)	(4194.50, 4205.50)

TABELA 3.8. Intervalos de integração para encontrar o valor de largura equivalente de cada linha espectral.

É possível notar diferenças entre os intervalos de alta e baixa resolução, o que é uma consequência direta da qualidade desses espectros: precisamos definir intervalos mais largos para os de baixa resolução devido a dificuldade de definir o início e fim da linha. Além disso, a definição dos intervalos para as linhas Si III  $\lambda 4552$ , He I  $\lambda 4144$ , e He I  $\lambda 4200$ , foi feita baseando-se apenas nos espectros com tipos espectrais entre O8 e O9.7, porque se relacionam somente à classificação desses objetos. Então, os valores encontrados na Tabela 3.8 são referentes a esse grupo de objetos.

Nas Figuras 3.9 e 3.10, vemos os intervalos de  $\lambda_i$  e  $\lambda_f$  (indicados na Tabela 3.8) para 30 espectros selecionados de maneira aleatória, contendo diferentes tipos espectrais e luminosidades, de alta e de baixa resolução, respectivamente. Podemos ver que, por mais que tenhamos uma alta variação no formato dessas linhas tanto nas de alta quanto na de baixa resolução, provocada pela amostra aleatória de espectros utilizada para formar as figuras, os intervalos definidos para  $\lambda_i$  e  $\lambda_f$  estão de acordo com o início e o final das linhas de absorção, sendo coerente utilizá-los para medir as larguras equivalentes.

### 3.3.1 Medidas dos erros das larguras equivalentes

Como a determinação dos intervalos de comprimento de onda foi feita de forma manual, é importante estimar a incerteza nas medidas das larguras equivalentes.

Para realizar essa tarefa, utilizamos o espectro de menor resolução de cada conjunto: no caso de alta resolução  $R \sim 40.000$  e de baixa resolução  $R \sim 2.500$ . Como não há variação na resolução dos espectros de baixa resolução, a escolha pode ser feita de forma arbitrária.

Após a escolha do espectro, variamos o intervalo  $(\lambda_i, \lambda_f)$  utilizado para medir a largura equivalente em uma quantidade  $\Delta\lambda$ , de forma que realizaríamos a medida a partir de um novo intervalo  $(\lambda_i - \Delta\lambda, \lambda_f + \Delta\lambda)$  — que pode ser maior ou menor que o intervalo original.

- Para os espectros de alta resolução, os valores utilizados de  $\Delta\lambda$  foram  $0\text{\AA}, \pm 0.2\text{\AA}, \pm 0.4\text{\AA}, \pm 0.6\text{\AA}, \pm 0.8\text{\AA}, \pm 1.0\text{\AA}, \pm 1.2\text{\AA}$ .
- Para espectros de baixa resolução, os valores utilizados de  $\Delta\lambda$  foram  $0\text{\AA}, \pm 0.4\text{\AA}, \pm 0.8\text{\AA}, \pm 1.2\text{\AA}, \pm 1.6\text{\AA}, \pm 2.0\text{\AA}, \pm 2.4\text{\AA}$ . Os intervalos nesse caso foram maiores por haver uma maior dificuldade em definir o início e final da linha para resoluções mais baixas.

Essa variação da largura equivalente foi feita para todas as linhas de absorção citadas na seção de critérios de classificação: He I  $\lambda 4471$ , He II  $\lambda 4542$ , He I  $\lambda 4200$ , He I  $\lambda 4144$ , He I  $\lambda 4388$ , Si III  $\lambda 4552$  e He II  $\lambda 4686$ .



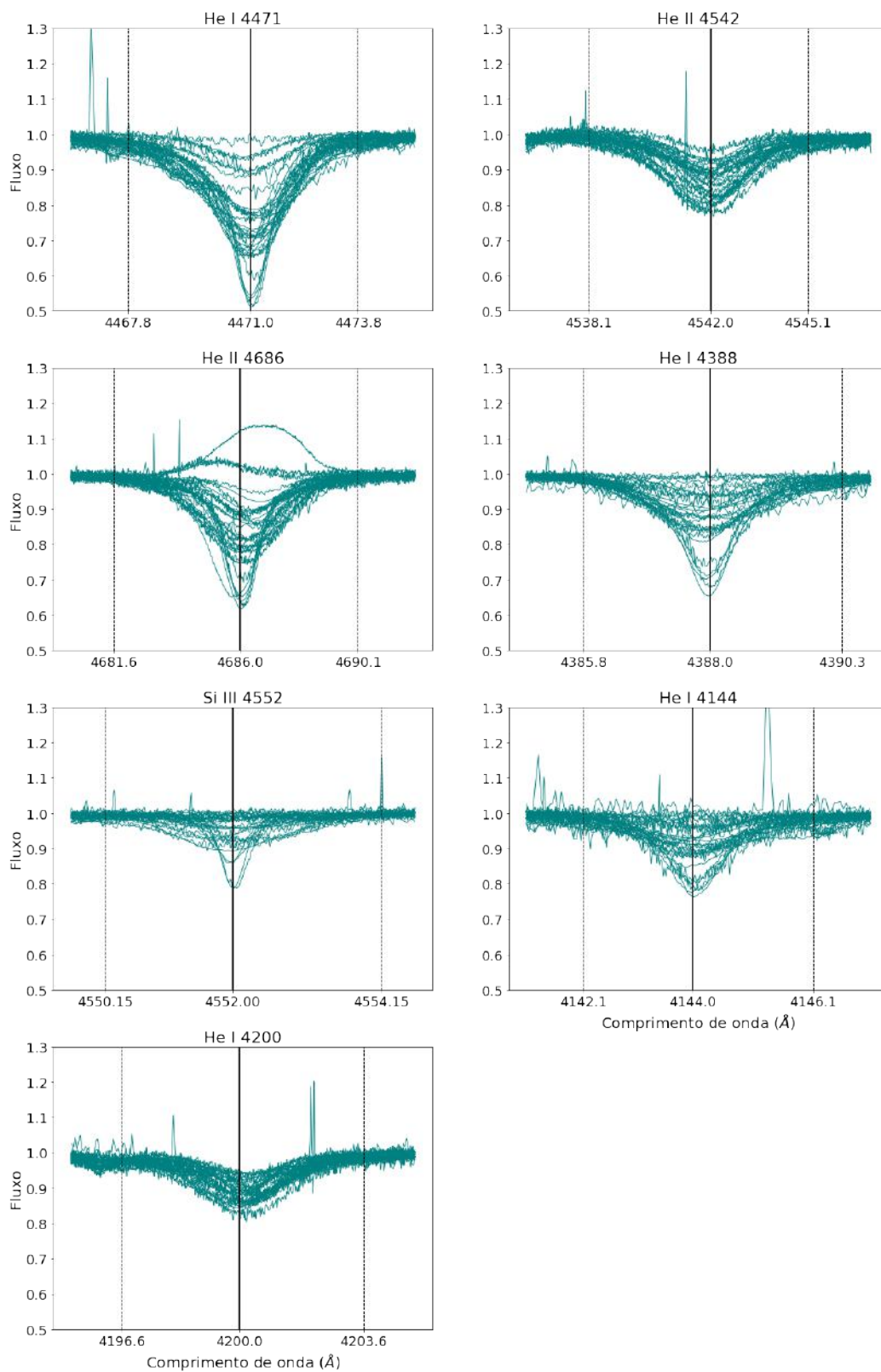


FIGURA 3.9. Espectros de alta resolução. No eixo  $x$  temos os comprimentos de onda em angstroms e no eixo  $y$  o fluxo normalizado. As retas tracejadas indicam  $\lambda_i$  e  $\lambda_f$ , e a reta mais larga e contínua, o centro da linha. Podemos verificar que os intervalos escolhidos são razoáveis e pegam inteiramente os perfis das estrelas selecionadas aleatoriamente.

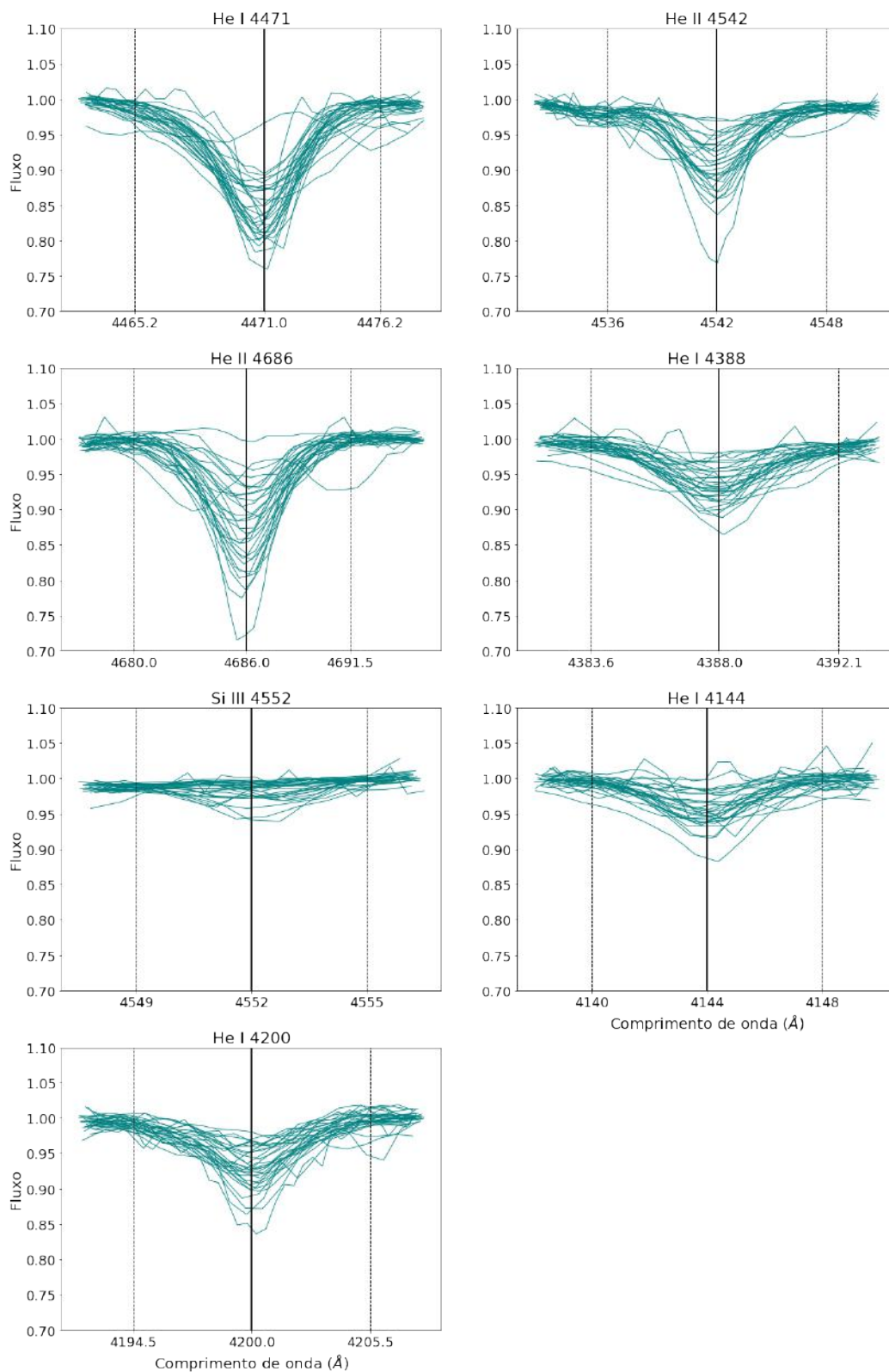


FIGURA 3.10. Espectros de baixa resolução. No eixo  $x$  temos os comprimentos de onda em angstroms e no eixo  $y$  o fluxo normalizado. As retas tracejadas indicam  $\lambda_i$  e  $\lambda_f$ , e a reta mais larga e contínua, o centro da linha. Podemos verificar que os intervalos escolhidos são razoáveis e pegam inteiramente os perfis das estrelas selecionadas aleatoriamente.

Depois, realizamos o cálculo da largura equivalente para cada um desses novos intervalos (indicados por  $x_i$  em 3.5) e definimos a média aritmética simples  $\mu$  desses valores para encontrar a incerteza da nossa medida: o desvio padrão  $\sigma$ .

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \quad (3.5)$$

Onde  $N$  é a quantidade total de intervalos utilizados.

Para os espectros de alta resolução, utilizamos a estrela HD 42088. Na Figura 3.11, ilustramos o caso da linha de absorção He II  $\lambda$  4542, na qual vemos como são os novos intervalos ( $\lambda_i - \Delta\lambda$ ,  $\lambda_f + \Delta\lambda$ ) indicados pelas linhas tracejadas mais finas e de mesma cor.

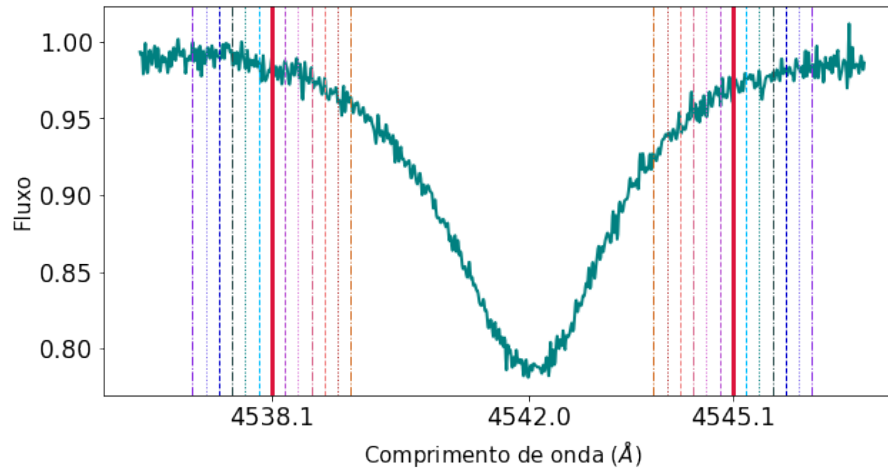


FIGURA 3.11. Linha do He II  $\lambda$ 4542 do espectro de alta resolução HD 42088. As linhas vermelhas mais grossas indicam o intervalo utilizado originalmente para obter a medida de largura equivalente e, as linhas coloridas, testam novos intervalos a partir da variação do  $\Delta\lambda = 0\text{\AA}, \pm 0.2\text{\AA}$ , etc.

Para os espectros de baixa resolução, utilizamos a estrela CPD -59 2554 e, na Figura 3.12, vemos os intervalos resultantes para o caso da linha de absorção He II  $\lambda$  4542.

Em ambas as Figuras 3.11 e 3.12, as linhas mais grossas em vermelho indicam o intervalo originalmente utilizado para realizar a medida da largura equivalente, mostrado nas Figuras 3.9 e 3.10, e também na Tabela 3.8.

Por fim, encontramos que o erro nas medidas de largura equivalente é de  $\pm 0,025\text{\AA}$  para os espectros de alta resolução e de  $\pm 0,046\text{\AA}$  para os espectros de baixa resolução.

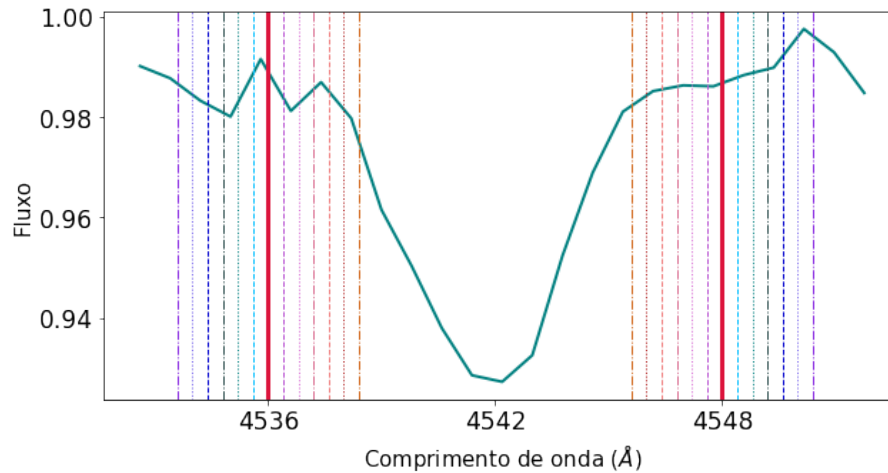


FIGURA 3.12. Linha do He II  $\lambda 4542$  do espectro de baixa resolução CPD -59 2554. As linhas vermelhas mais grossas indicam o intervalo utilizado originalmente para obter a medida de largura equivalente e, as linhas coloridas, testam novos intervalos a partir da variação do  $\Delta\lambda = 0\text{\AA}, \pm 0.4\text{\AA}$ , etc.

Esses valores foram obtidos a partir da média dos desvios padrões de todas as linhas presentes na Tabela 3.8 e, assim, assumimos que todas as linhas possuem o mesmo erro.

Idealmente, as incertezas deveriam ser medidas para cada objeto. Porém, não é nosso intuito analisar em detalhes os erros para cada objeto de nossa amostra, e sim apenas estimar a ordem de grandeza dos erros, para as resoluções e SNR típicas utilizadas. Os resultados mostrados indicam uma precisão bastante razoável das medidas e, com isso, podemos iniciar a aplicação do aprendizado de máquina no nosso problema.

### 3.4 Aplicação dos modelos de *Machine Learning*

Nesta seção, iremos definir o que é *Machine Learning* e introduzir alguns conceitos fundamentais para compreensão de seu funcionamento. No entanto, uma análise mais detalhada de suas principais ferramentas e sobre o que foi utilizado neste projeto, será visto no Capítulo 4.

*Machine Learning*, ou Aprendizado de Máquina, é um ramo da Inteligência Artificial que visa a construção de programas que sejam capazes de aprender sobre os dados que lhe são fornecidos, de forma a identificar similaridades e diferenças entre eles para, no fim, realizar a análise pretendida (p.e., classificação ou regressão dos dados) sem a necessidade de instruções explícitas. Como exemplo, podemos citar a base de dados

*MNIST*<sup>3</sup>, constituída por dezenas de milhares de números escritos à mão (exemplificados na Figura 3.13), utilizada para o desenvolvimento de programas de reconhecimento de imagem com o objetivo de classificar o que está escrito.

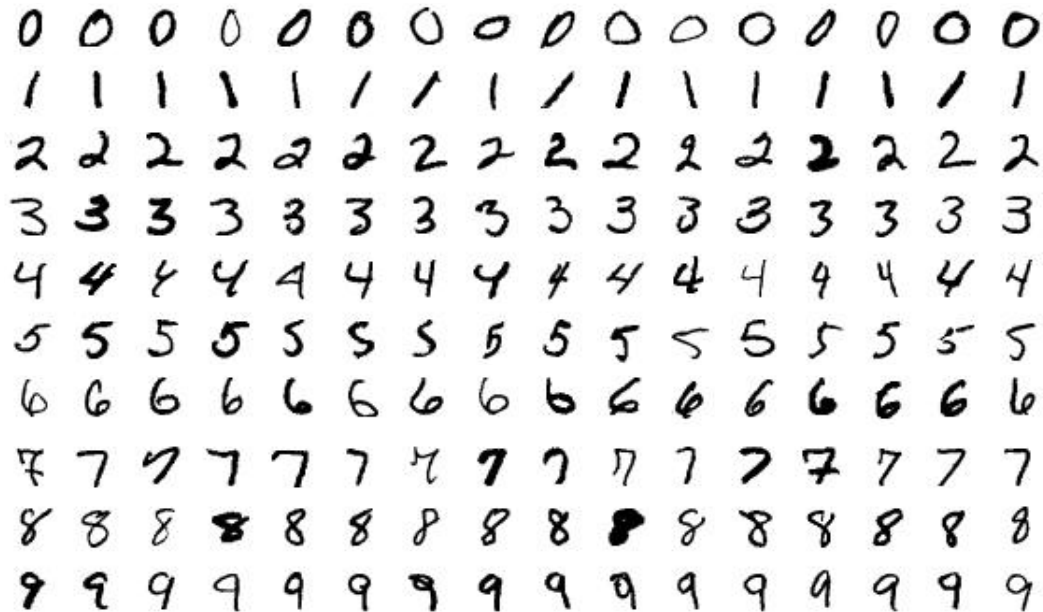


FIGURA 3.13. Imagens da base de dados *MNIST*.

O programa utiliza um conjunto treino, formado por uma parte das imagens já classificadas da base de dados (p.e., todos os números da primeira fileira são iguais a zero, os da última fileira são iguais a nove, etc.), para aprender quais características definem uma determinada classe. Por fim, utiliza um conjunto teste, formado pelo restante das imagens da base de dados, para classificá-lo a partir da lógica desenvolvida no aprendizado e verificar a acurácia da classificação. De maneira geral, os conjuntos treino e teste constituem 70% e 30%, respectivamente, de todos os dados.

Como aprendizado de máquina pode ser utilizado para resolução de problemas estatísticos variados, existem diversos programas — os modelos de ML — com diferentes funcionalidades e capacidades de adaptação aos dados (p.e., há modelos de classificação que realizam a separação dos dados linearmente, outros a partir de uma função polinomial).

<sup>3</sup>Para mais informações acesse: <http://yann.lecun.com/exdb/mnist/>

Assim, vários modelos podem parecer úteis na resolução de um problema, o que pode provocar o seguinte questionamento: como escolher o melhor modelo? Ilustramos esse processo no esquema indicado na Figura 3.14.

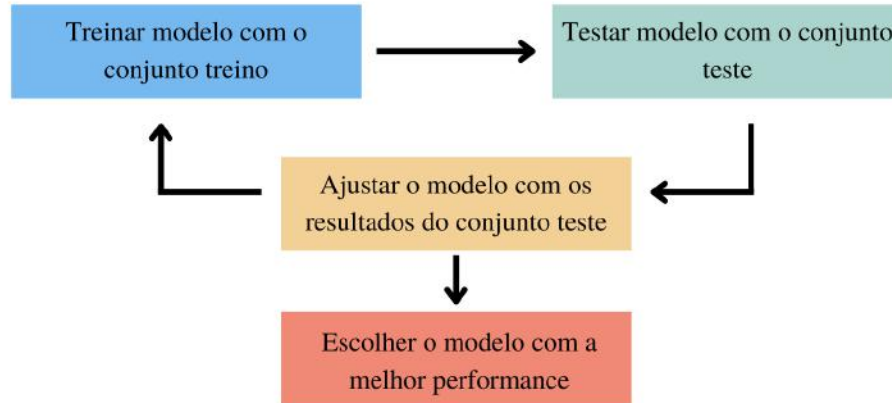


FIGURA 3.14. Esquema que demonstra a escolha do modelo de ML de melhor performance.

No problema proposto por este projeto, ao invés de imagens, os modelos irão aprender o que caracteriza uma determinada classe de tipo espectral e de luminosidade a partir das larguras equivalentes das linhas de diagnóstico, ou, em alguns casos, a razão entre as larguras equivalentes das linhas, que serão os nossos parâmetros de classificação (ou *features*).

No entanto, é importante notar que, em ML, não necessariamente alimentaremos os modelos com todas as larguras equivalentes obtidas nas seções anteriores: é necessário haver um equilíbrio com relação à quantidade de informação. Se alimentarmos o modelo com um excesso de informação, podemos causar uma desordem no momento da correlação entre as *features*, diminuindo a capacidade de aprendizado. Desta forma, muitas vezes podemos ter poucos parâmetros, mas um bom resultado de acurácia da classificação. Esse equilíbrio só é atingido a partir de testes e análise da performance em cada caso.

Além disso, neste projeto optamos por utilizar apenas modelos de ML supervisionados, que utilizam a classificação dos espectros na fase de treinamento para correlacioná-la às respectivas *features*.

Nós utilizamos os seguintes algoritmos: *K-Nearest Neighbors* (KNN), *Support Vector Machines* (SVM), *Decision Tree Classifier* (DTC) e *Random Forest Classifier* (RFC). Estes serão elaborados em mais detalhes na seção 4.2 do Capítulo 4.

Para realizar essa parte do projeto, utilizamos os pacotes *Scikit-Learn*<sup>4</sup> do Python 3 — além dos pacotes padrão como *Numpy*<sup>5</sup>, *Matplotlib*<sup>6</sup>, *Pandas*<sup>7</sup> — que contém todas as ferramentas necessárias para utilizar os modelos e visualizá-los posteriormente.

Caso o leitor esteja familiarizado com as ferramentas utilizadas em ML, pode seguir diretamente para o Capítulo 5, de resultados, onde mostraremos a performance de cada modelo e os ajustes utilizados em cada um deles. Caso contrário, sugerimos que faça a leitura do Capítulo 4, onde explicamos em detalhes as ferramentas e as nomenclaturas presentes em ML e também o funcionamento dos modelos utilizados — que será necessário para compreensão dos Capítulos 5 e 6.

---

<sup>4</sup>Endereço: <https://scikit-learn.org/stable/>

<sup>5</sup>Endereço: <https://numpy.org/>

<sup>6</sup>Endereço: <https://matplotlib.org/stable/index.html>

<sup>7</sup>Endereço: <https://pandas.pydata.org/>

## Capítulo 4

# Modelos de Machine Learning

*Um programa de computador é dito aprender com experiência  $E$  com respeito a uma tarefa  $T$  e alguma medida de performance  $P$ , se a sua performance em  $T$ , medida por  $P$ , é aprimorada com  $E$ .*

— Tom Mitchell, 1997

Introduzimos este capítulo a partir de uma citação de Tom Mitchell, cientista da computação e professor da Universidade de Carnegie Mellon, com fortes contribuições no ramo da inteligência artificial. Esta busca explicar de maneira simples como funciona o processo de Aprendizado de Máquina, o que será explorado mais profundamente nas próximas seções.

Neste capítulo, temos como objetivo realizar uma breve introdução sobre os principais conceitos que envolvem o *Machine Learning*, assim como elaborar sobre o funcionamento dos modelos que foram utilizados neste projeto, visando a classificação de tipo espectral e luminosidade — o que será fundamental para a compreensão dos Capítulos 5 e 6. Além disso, enxergamos esta como uma oportunidade de desenvolver sobre uma ferramenta em ascensão na Astronomia por sua capacidade de capazes de realizar tarefas sem a necessidade de instruções explícitas e de tratar milhares de dados de uma só vez de maneira automatizada, o que tem sido cada vez mais necessário.



## 4.1 Conceitos Fundamentais

### 4.1.1 Labels e Features

*Labels* e *features* são os conjuntos fundamentais de informação utilizados em *Machine Learning*. Estes são definidos por:

- *Labels*: são as classificações de cada espectro, o que estamos querendo que o modelo de *Machine Learning* seja capaz de prever. Por exemplo: a *label* de um espectro pode ser *O3 I*.
- *Features*: são as informações atribuídas a cada espectro, o que o modelo utilizará para aprender sobre o problema. Por exemplo: uma das *features* utilizadas é a largura equivalente da linha espectral He I  $\lambda 4471$ .

### 4.1.2 Aprendizados supervisionado e não-supervisionado

Uma diferença que podemos evidenciar entre os diversos modelos de ML é a forma como realizam o aprendizado. De modo geral, a compreensão acerca dos dados é feita de maneira supervisionada ou não-supervisionada, no qual o primeiro exige que os dados já estejam classificados antes de aplicá-los aos modelos de ML, e o segundo não.

Em ambos os aprendizados, as correlações entre as *features* são feitas a partir da identificação de suas similaridades e diferenças. No entanto, no caso de modelos supervisionados, essa correlação é associada às *labels*, de forma a se compreender quais parâmetros pertencem a qual classe. Então, utiliza-se um novo conjunto de dados para validar e obter a acurácia definida a partir da comparação direta entre a classificação feita pelo modelo e as *labels* já existentes.

No caso do aprendizado não-supervisionado, os espectros não precisam ser previamente classificados para serem aplicados ao modelo de ML, de forma que a separação dos dados será feita a partir de agrupamentos por similaridade entre as *features*.

Neste projeto, utilizaremos apenas modelos de ML supervisionados voltados para classificação, de forma que a explicação das próximas seções será direcionada para esse

tipo de modelo. Caso o leitor tenha interesse em saber mais informações sobre modelos não-supervisionados, recomendamos a leitura do livro Celebi & Aydin (2016), que contém uma extensa análise sobre esse tipo de aprendizado.

### 4.1.3 Conjuntos Teste e Treino

Ao utilizarmos ML, precisamos separar o conjunto total de dados em dois grupos, de forma que um será utilizado para o aprendizado da máquina, e o outro será utilizado para testar se o aprendizado foi feito corretamente. Estes são, respectivamente, os conjuntos treino e teste.

O objetivo do conjunto treino é realizar correlações entre as *features* dos espectros, buscando compreender as suas similaridades e diferenças, comparando diretamente com as suas respectivas *labels* (p.e., uma *feature* com valor X, está relacionada à classe Y).

Após o aprendizado ser feito, o conjunto teste irá validar se a etapa anterior foi feita de forma correta, a partir da sua classificação de acordo com o aprendizado desenvolvido pelo conjunto treino. Isso nos retornará o valor de acurácia, que dirá a eficiência do modelo em resolver o nosso problema.

Usualmente, 30% da amostras serão utilizadas para teste e 70% para treino. No entanto, isso depende da quantidade de dados e do problema sendo proposto, e pode ser facilmente ajustado na configuração do modelo.

### 4.1.4 *Underfitting* e *Overfitting*

Os termos *underfitting* e *overfitting* são comumente utilizados em *Machine Learning*, pois descrevem quando há uma falta e um excesso de aprendizado, respectivamente.

No primeiro caso, o modelo não é capaz de compreender o problema e realizar uma correlação entre os dados, de forma que não é capaz de realizar a classificação.

No segundo caso, o modelo é capaz de compreender o conjunto treino, que é o que buscamos, indicando que o aprendizado foi bem sucedido. Porém, quando utilizamos o conjunto teste para verificar a sua eficiência, a acurácia da classificação é baixa. Isso é um indicativo de que o modelo não aprendeu sobre o problema de forma a ser capaz de

realizar uma generalização (p.e., ignorando a presença de *outliers*), mas que se ajustou exclusivamente a nossa amostra treino, não sendo capaz de classificar novos dados.

Ilustramos, na Figura 4.1, o comportamento de um modelo de ML nos cenários de *Underfitting*, *overfitting* e no caso ideal, onde temos uma separação equilibrada entre as classes.

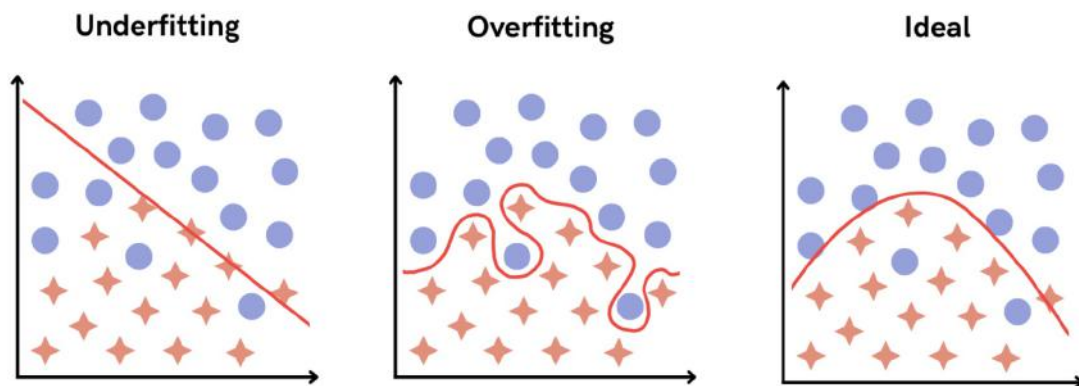


FIGURA 4.1. Ilustrações que indicam os casos de *Underfitting*, de *overfitting* e um cenário ideal.

Com isso, podemos dizer que o que buscamos é um aprendizado que seja capaz de executar a análise com diferentes grupos de dados e, ainda assim, obter resultados de acurácia satisfatórios.

Outra maneira de mostrar o equilíbrio desejado na fase de aprendizado é a partir da relação entre erro e complexidade do modelo, ilustrada na Figura 4.2. Quanto maior a complexidade do modelo, maior a probabilidade de aprendizado, por termos mais ferramentas de ajuste e adaptação, mas também maior o risco disso ocorrer de forma excessiva e termos um caso de *overfitting*.

Por isso, é importante observarmos e questionarmos quando o erro no treinamento está muito baixo.

#### 4.1.5 *Cross Validation*

*Cross validation* (ou validação cruzada) é uma forma de realizar o treinamento e a testagem  $n$  vezes, variando os espectros dentro dos conjuntos treino e teste em cada iteração, o que resultará em valores distintos de acurácia. Assim, o resultado final será dado pela média dessas acurácias.

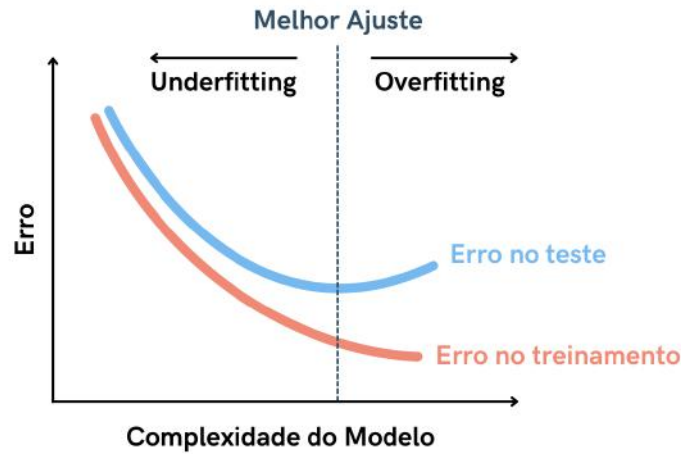


FIGURA 4.2. Relação entre erro e complexidade do modelo.

Utilizar a validação cruzada é fundamental quando não temos todas as classes sendo representadas no conjunto treino, de forma que o aprendizado é realizado de forma incompleta. Isso ocorre quando o número de classes é alto com relação à quantidade total de espectros a serem classificados e/ou quando não há uma uniformidade nos dados, havendo expressivamente mais itens em uma classe do que nas outras. Por exemplo, caso tenhamos 20 espectros de tipo O9V e dois espectros de tipo O3I em toda a amostra, será mais fácil garantirmos a presença de espectros O9V nos conjuntos teste e treino do que os de tipo O3I.

Na ilustração indicada na Figura 4.3 temos um esquema que mostra o *Cross Validation* ocorrendo para 5 iterações distintas, onde o conjunto teste (azul claro) corresponde a 20% da amostra, enquanto o restante é designado para treinamento. A amostra total indica, a partir de diferentes cores, as classes existentes no conjunto.

É possível observar na figura que, caso realizássemos somente uma iteração (iteração 1), o aprendizado seria feito sem considerar as classes vermelhas (estrelas), de forma que não haveriam correlações entre as *features* para essas cores e, conseqüentemente, o modelo não seria capaz de classificá-las. No entanto, utilizando *Cross Validation*, o aprendizado é garantido para todas as classes na fase de treinamento, mesmo que em iterações distintas, de forma que a acurácia final levará em conta o aprendizado completo.

O valor de  $n$  é arbitrário, e a única maneira de defini-lo é testando-o dentro do problema. Porém, usualmente são utilizadas 5 ou 10 iterações.

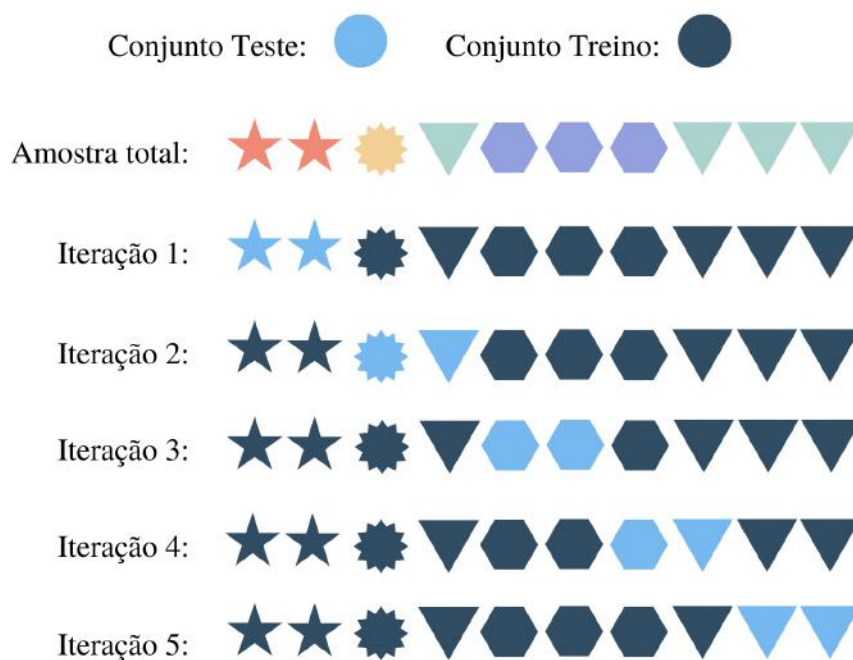


FIGURA 4.3. Cross Validation

#### 4.1.6 Matriz de Confusão

A matriz de confusão é formada após a aplicação dos dados em ML e mostra como espectros de uma determinada classe foram classificadas pelo modelo. A sua estrutura é dada por uma matriz quadrada que contém todas as classes possíveis em ambos os eixos, onde as colunas indicam a *label* real e as linhas indicam as classificações realizadas pelo modelo.

De forma a ilustrarmos uma matriz de confusão, podemos utilizar uma matriz quadrada  $2 \times 2$  com as decisões genéricas *positivo* e *negativo* — ver Figura 4.4. Nesta, as posições fora da diagonal principal (em vermelho) são *Falso Positivo* e *Falso Negativo*, e na diagonal principal (em verde) por *Verdadeiro Positivo* e *Verdadeiro Negativo*, que ocorre quando a previsão é igual a realidade.

Diante da análise da Figura 4.4, percebemos que a matriz de confusão ideal terá apenas as posições da diagonal principal (Verdadeiro Positivo e Verdadeiro Negativo) preenchidas — qualquer outra posição da matriz indicará uma classificação incorreta.

Na seção de Resultados (Capítulo 5) mostraremos as Matrizes de confusão resultantes do nosso problema.

Previsão	Real	Positivo	Negativo
Positivo	Verdadeiro Positivo	Falso Positivo	
Negativo	Falso Negativo	Verdadeiro Negativo	

FIGURA 4.4. Matriz de Confusão para uma dada classe.

#### 4.1.7 *Precision, Recall* ou *F1-Score*?

Quando vamos avaliar a capacidade de um modelo de realizar a classificação dos dados, é comum utilizarmos a acurácia, dada pela Fórmula 4.1, que indica o número de classificações corretas sobre o total de classificações para uma dada classe.

$$Acuracia = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.1)$$

Sendo VP, VN, FP e FN a quantidade de verdadeiros positivos, de verdadeiros negativos, de falsos positivos e de falsos negativos, respectivamente — ver Figura 4.4.

No entanto, ao construirmos um modelo, podemos priorizar saber a razão de verdadeiros positivos dentre os objetos classificados da mesma forma, saber a razão de classificações corretas dentre os objetos que pertencem a uma mesma classe, ou ambas as questões anteriores ao mesmo tempo. Estes são a precisão, o *recall* e o F1-Score, respectivamente.

★ *Precisão*: indica, dentro de um conjunto de espectros que foram igualmente classificados, o número de classificações corretas positivas (verdadeiros positivos) sobre o número de classificações positivas (verdadeiros positivos e falsos positivos). Assim, a precisão prioriza a quantidade de classificações corretas dentro da previsão feita pelo modelo.

$$Precisao = \frac{VP}{VP + FP} \quad (4.2)$$

★ *Recall*: indica, dentro de um conjunto de espectros que pertencem a uma determinada classe, o número de classificações corretas positivas (verdadeiros positivos) sobre

o número real total de objetos na classe (verdadeiros positivos e falsos negativos). Este também é chamado de sensibilidade ou a taxa de verdadeiros positivos.

$$Recall = \frac{VP}{VP + FN} \quad (4.3)$$

★ *F1-Score*: combina a influência da precisão e do *recall* a partir de uma média harmônica entre esses parâmetros. Terá valores altos apenas quando tanto a precisão quanto o *recall* tiverem resultados altos.

$$F_1 = \frac{2}{\frac{1}{precisao} + \frac{1}{recall}} = 2 \times \frac{precisao \times recall}{precisao + recall} = \frac{VP}{VP + \frac{FN+FP}{2}} \quad (4.4)$$

No livro de Géron (2019), o autor dá um exemplo interessante para entendermos quando é interessante utilizar o critério de precisão e de *recall*:

”Por exemplo, se você treinar um modelo para detectar vídeos que são seguros para crianças, você provavelmente iria preferir um modelo que rejeita muitos vídeos bons (baixo *recall*) mas que mantém apenas os que são seguros (alta precisão), do que um modelo que possui um *recall* alto mas que permite que alguns vídeos ruins apareçam. [...]”

Neste projeto, utilizamos o *F1-Score*, porque o nosso objetivo é evitar tanto os falsos negativos quanto os falsos positivos, de forma a obter a melhor classificação possível.

Com isso, finalizamos os fundamentos do Aprendizado de Máquina e exploraremos, na próxima seção, o funcionamento dos modelos de ML utilizados neste trabalho.

## 4.2 Modelos utilizados

Os algoritmos de *Machine Learning* utilizados nesse trabalho foram:

- ★ KNN (*K-Nearest Neighbors*);
- ★ SVM (*Support Vector Machines*);
- ★ DTC (*Decision Tree Classifier*);
- ★ RFC (*Random Forest Classifier*).

Todos são modelos supervisionados de classificação, o que significa que cada espectro precisa ser previamente classificado antes da sua aplicação nos modelos. Essa classificação prévia foi feita por nós a partir dos critérios utilizados no Capítulo 3 de Metodologia. Assim, o modelo utiliza as larguras equivalentes obtidas como *features* e as classificações como as *labels*.

#### 4.2.1 *K-Nearest Neighbors*

O KNN (*K-Nearest Neighbors*) é um dos modelos de ML mais simples que se baseia na distância euclidiana entre um dado não-classificado e os outros  $K$  vizinhos mais próximos com classificações conhecidas.

O funcionamento do código está ilustrado na Figura 4.5 e é dado da seguinte forma:

1. Inicialmente definimos um valor de  $K$  arbitrário, que corresponde à quantidade de vizinhos que iremos analisar.
2. A partir desses  $K$  vizinhos mais próximos são medidas as distâncias euclidianas entre eles e a amostra não-classificada (indicada na Figura 4.5 como um ponto de interrogação).
3. A amostra mais próxima corresponderá a classificação final.

A definição do valor ideal de  $K$  vizinhos é dada a partir de testes com o próprio conjunto de dados, avaliando qual retorna a melhor taxa de classificação.

Além disso, após utilizarmos o modelo em nosso problema, teremos também a possibilidade de construir um gráfico que mostra as regiões de classificação correspondentes para cada classe, definidas na fase de treinamento. Por exemplo, um espectro novo em uma posição  $(x, y)$  desse gráfico terá a mesma classificação da região na qual está contido. Um exemplo disso pode ser visto na Figura 4.6 do site *scikit-learn*, onde utilizou-se o Conjunto de dados flor Iris, que consistem em 50 amostras de cada tipo de flor Iris: Iris setosa, Iris virginica e Iris versicolor, desenvolvido por Ronald Fisher (Fisher, 1936). As informações disponíveis para a classificação foram o comprimento e a largura das sépalas e das pétalas.



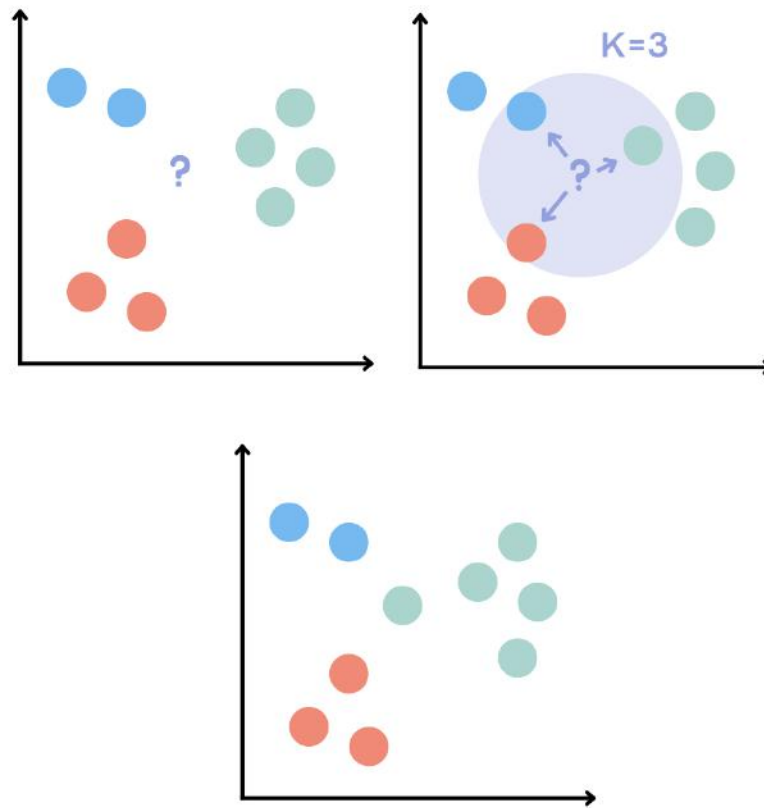


FIGURA 4.5. Passo-a-passo do funcionamento do modelo KNN utilizando 3 vizinhos que estão representados dentro da região circular na segunda imagem.

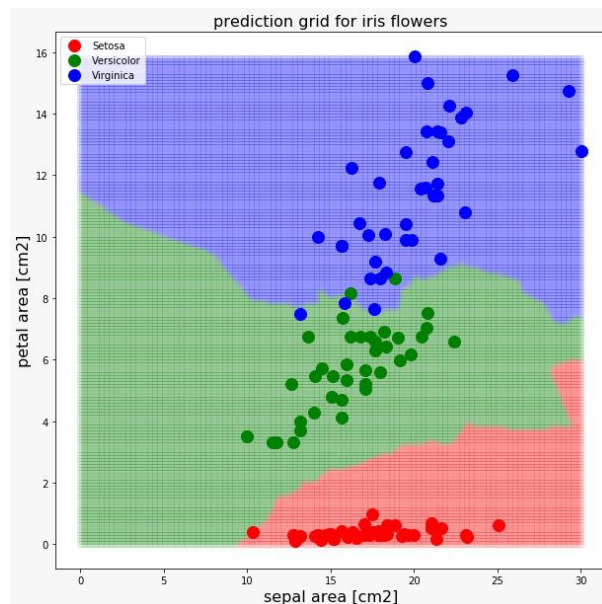


FIGURA 4.6. Plot resultante do KNN para classificação do Conjunto de Dados Iris cedido por Dr. Wagner Marcolino e feito em seus estudos pessoais de *Machine Learning*. No eixo  $x$  temos a área da sépala em  $cm^2$  e no eixo  $y$  a área da pétala em  $cm^2$ .

Com isso, qualquer nova flor que se encontre na região vermelha, será classificada como Iris Setosa, na região em verde como Iris Versicolor e na região em azul como Iris Virginica.

Abaixo, listamos parte de um código simples, em Python, para construção de um modelo KNN onde não definimos valores para os parâmetros de ajuste, porque isso dependerá completamente do problema sendo analisado. Não entraremos em detalhes sobre o que cada instrução faz, mas podemos ressaltar alguns pontos principais: (i) Inicialmente, separamos o conjunto teste e de treino, com a função `train_test_split()`; (ii) a instrução `knn.fit()` é a parte do aprendizado com o conjunto treino; (iii) `knn.predict()` é a previsão do modelo, com o conjunto separado da amostra (teste); (iv) a acurácia é averiguada utilizando `metrics.accuracy_score()`.

---

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn import preprocessing
from sklearn.model_selection import cross_val_score
lab_enc = preprocessing.LabelEncoder()
y_encoded = lab_enc.fit_transform(y)
X_train, X_test, y_train, y_test = train_test_split(X,y_encoded, test_size)
knn = KNeighborsClassifier(n_neighbors=K)
knn.fit(X_train, y_train)
pred = knn.predict(X_test)
metrics.accuracy_score(y_test,pred)
kfold = model_selection.KFold(n_splits)
scores = cross_val_score(knn, X, y, cv, scoring)
```

---

### 4.2.2 Support Vector Machines

O SVM (*Support Vector Machine*) é um modelo de ML amplamente utilizado devido a sua alta capacidade de adaptação aos dados, e também por ser capaz de realizar separações lineares e não-lineares (basta definir a função desejada para o comportamento da separação) entre as diferentes classes. Neste projeto, utilizamos a separação linear porque foi a que melhor se ajustou aos nossos dados.

Em um caso simples de separação linear entre duas classes distintas, o modelo utiliza a maior separação possível entre os dados. Na Figura 4.7, o gráfico à esquerda mostra diferentes possibilidades de separação linear, mas que são construídas de maneira não otimizada por estarem muito próximas dos itens de cada classe, de forma que podem provocar classificações equivocadas com a chegada de novos dados.

No entanto, o SVM utiliza os chamados *support vectors* ou os vetores de suporte: vetores formados pelos itens mais próximos de classificações opostas (circuladas no gráfico à direita na Figura 4.7) indicando a maior distância possível entre as classes.

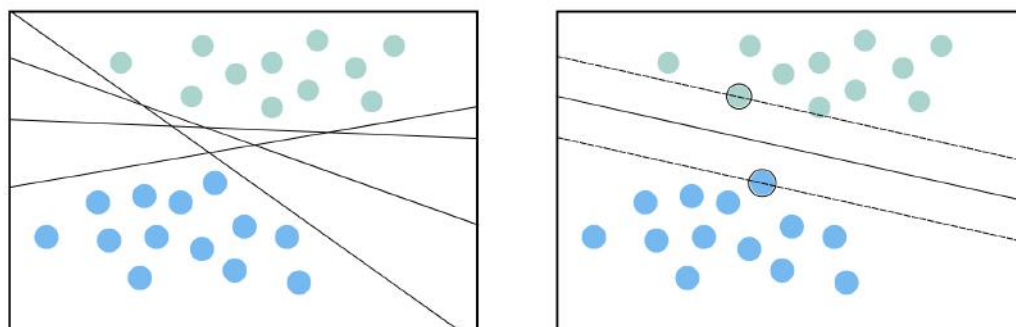


FIGURA 4.7. Exemplo de separação de dados de duas classes: verde e azul (SVM).

Com isso, por mais que classificações erradas possam ocorrer, essa separação otimizada garante uma região mais segura para cada classe e evita classificações incorretas.

Outra funcionalidade deste modelo é a possibilidade de escalonar os dados para aumentarmos ainda mais a separação, como mostrado na Figura 4.8, de Géron (2019). Isso é feito utilizando a função *preprocessing* do pacote *scikit-learn*.

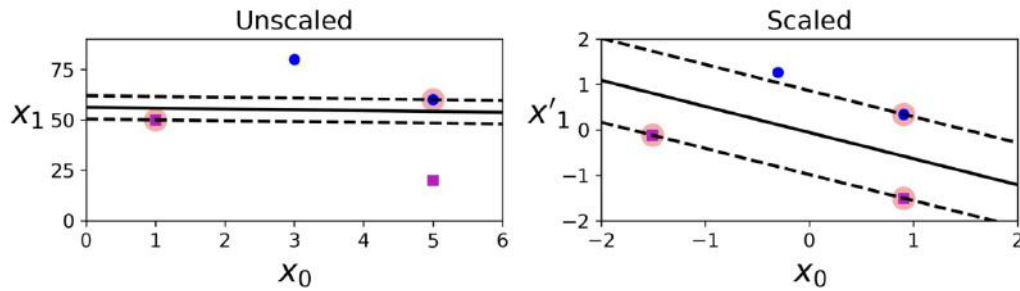


FIGURA 4.8. Diferenças entre os dados não-escalados e escalados no SVM. As linhas tracejadas indicam a posição dos vetores de suporte utilizados para realizar a separação final, que é indicada pela linha sólida em preto.  $X_0$  e  $X_1$  indicam as *features* utilizadas para realizar a classificação, e  $X'_1$  seria a *feature*  $X_1$  escalonada. Figura 5-2 de Géron (2019).

Além disso, o SVM é capaz de identificar a presença de *outliers* da forma como está indicado na Figura 4.9 e está descrita a seguir:

1. Ainda na fase de treinamento, o modelo é capaz de identificar quando algum dado não possui o mesmo comportamento que o restante de mesma classificação. Com isso, caso não houvéssimos o entendimento de que estes são *outliers* (circulados em preto no painel 1 da figura), a separação seria a indicada nos painéis 1 e 2 da figura, chamada de *hard margin classification*.
2. Assim, no momento de classificar os novos dados, indicados pelos círculos preenchidos em preto (painel 2 da Figura 4.9), estes seriam classificados erroneamente.
3. Porém, como é possível identificar os *outliers*, esses dados são desconsiderados, e a separação ótima é feita de acordo com o restante do conjunto (painel 3 da figura 4.9). Essa separação é chamada de *soft margin classification*.

Uma maneira que temos de definir o quanto iremos permitir de violações na nossa margem suave de classificação é utilizando o parâmetro de ajuste do modelo  $C$ . Este funciona da seguinte forma: quanto maior (menor) for o seu valor, menor (maior) será a quantidade de itens que serão desconsiderados ao se definir as distâncias das margens. Esse comportamento é visto na Figura 4.10, que indica os dados utilizados para a definição da margem sombreados de vermelho.

Para definirmos  $C$ , precisamos testar a capacidade de classificação utilizando diferentes valores para encontrar um equilíbrio que optimize a performance do modelo. Uma margem maior irá permitir mais classificações erradas, porém resultará em uma melhor generalização para os dados novos (Géron, 2019).

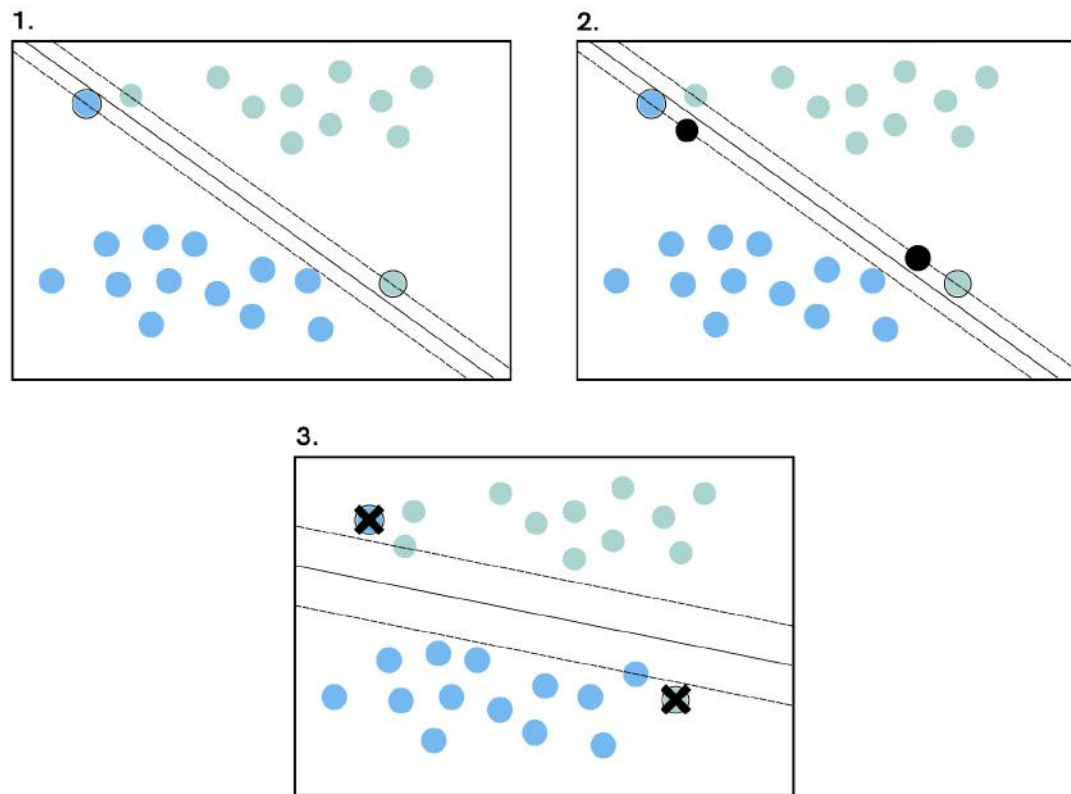


FIGURA 4.9. Separação suave entre os dados de duas classes distintas (azul e verde) a partir da desconsideração dos outliers.

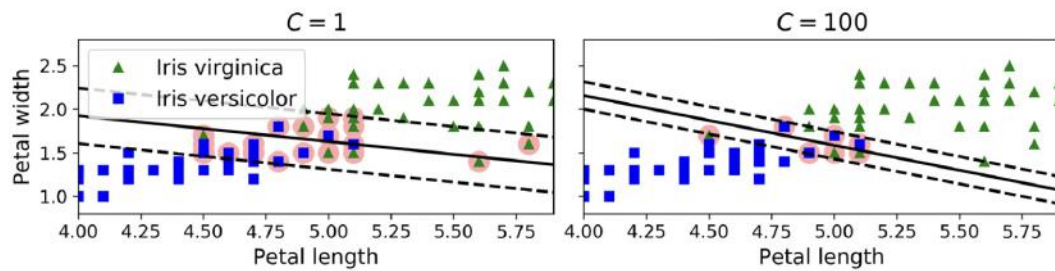


FIGURA 4.10. Comportamento das margens do SVM a partir de diferentes valores do parâmetro  $C$ . Figura extraída de Géron (2019).

Abaixo, deixamos como exemplo um código simples, em *Python*, para construção de um modelo SVM. Note que a estrutura é bem similar ao caso do KNN apresentado anteriormente.

```

from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix
    
```

```
from sklearn import preprocessing
Scaled_X = preprocessing.scale(X)
y=data['Spectral Type']
lab_enc = preprocessing.LabelEncoder()
y_encoded = lab_enc.fit_transform(y)
X_train, X_test, y_train, y_test = train_test_split(Scaled_X,y_encoded, test_
size)
svclassifier = SVC(kernel)
svclassifier.fit(X_train, y_train)
sv_pred = svclassifier.predict(X_test)
metrics.accuracy_score(y_test,sv_pred)
kfold = model_selection.KFold(n_splits)
scores = cross_val_score(svclassifier, standardized_X, y_encoded, cv=kfold)
```

---

### 4.2.3 Decision Tree Classifier

O DTC (*Decision Tree Classifier*) é um modelo de ML que busca dividir ao máximo os dados ao longo da ramificação de uma árvore, com o objetivo de separá-los em suas devidas classificações. Isto é feito utilizando critérios de separação definidos pelas *features* do problema (p.e., para  $W_\lambda(\text{He I } \lambda 4471) \geq X$ , a classe será O3). O objetivo, ao final, é ter um modelo que seja capaz de formar subconjuntos que contenham apenas uma classe, através da lógica desenvolvida pelo conjunto treino. Além disso, esse modelo não requer o tratamento dos dados antes de ser aplicado.

De forma a ilustrar esse processo, observamos um exemplo de árvore de decisão para o o Conjunto de dados Flor Iris, presente na Figura 4.11, que tem como objetivo classificar as flores nas espécies *Setosa* (Marrom), *Versicolor* (Verde) e *Virginica* (Roxo).

A classificação dessa árvore se procedeu da seguinte forma:

1. A árvore tem início na raiz, onde temos 150 flores para classificação (*samples=150*). Nesta, temos o conjunto total que contém as três espécies de Iris distintas com 50 itens cada. O critério utilizado para separação é Comprimento da Pétala (cm)  $\leq 2.45$ , formando à esquerda um conjunto de 50 flores todas classificadas como

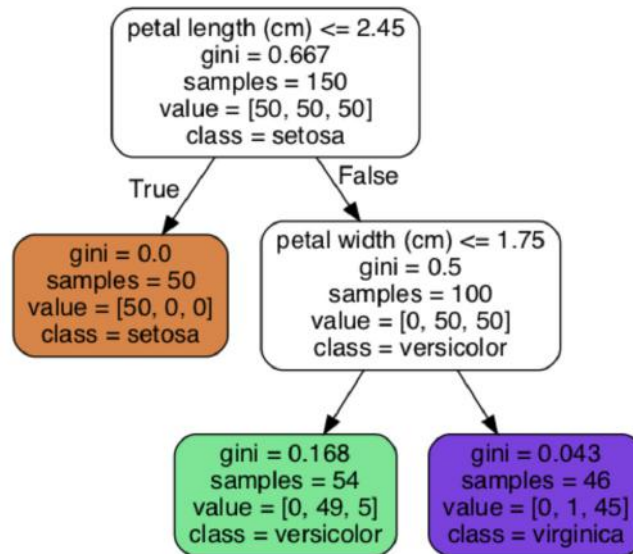


FIGURA 4.11. Exemplo de árvore de decisão para classificação do Conjunto de dados Flor Iris. Figura extraída de Géron (2019).

*Iris Setosa* (indicada por *True*) e à direita com 100 espécies de *Iris Versicolor* e *Virginica* (indicada por *False*).

- Em cada uma das folhas acima é possível observar o termo *Gini*, que é o índice de impureza, e que indica o quão desorganizada uma folha é. Quanto maior for esse valor, maior é a quantidade de itens de classes diferentes e maior a desorganização no conjunto. Assim, idealmente, a ramificação ocorrerá até que esse índice seja igual a zero, como vemos no caso da *Iris Setosa* (em marrom).
- Por possuir mais de uma classe no total ( $gini \neq 0$ ), o conjunto com classificação *Versicolor*, irá se dividir novamente formando a última ramificação usando Comprimento da Pétala (cm)  $\leq 1.75$ . Os conjuntos resultantes foram classificados como *Versicolor* e *Virginica*, mas com índices de *gini* diferentes de zero por haver classificações erradas (49 *Versicolor* e 5 *Virginica* sendo classificadas como *Versicolor*, e 1 *Versicolor* (em verde) e 45 *Virginica* (em roxo) sendo classificadas como *Virginica*).
- Nem sempre o resultado final terá todas as folhas com índice de impureza igual a zero. Caso contrário, poderíamos ter uma árvore que se ajusta bem demais aos dados e teríamos um caso de *overfitting*.

Nas árvores de decisão existem duas possibilidades de Índice de Impureza: o de *Gini* (que aparece na Figura 4.11), e o de Entropia. Ambos indicam a quantidade de

desorganização no conjunto de dados definidos abaixo em 4.5.

$$GiniIndex = 1 - \sum_j p_j^2 \quad Entropy = - \sum_j p_j \cdot \log_2(p_j) \quad (4.5)$$

Dado que  $p_j$  é a probabilidade de obtermos a classe  $j$  em ambos os casos.

O Índice de Gini mede o quanto um item escolhido de forma aleatória será classificado incorretamente, e o índice de Entropia indica a desordem do conjunto de acordo com a classe pretendida. No entanto, é importante testar o comportamento de ambos para descobrir qual melhor se adaptará ao problema.

Para evitar a ocorrência de *overfitting* da árvore de decisão, podemos utilizar artifícios pré-definidos no código para limitar o crescimento da árvore como: *max\_depth* (profundidade máxima da árvore), *max\_leaf\_nodes* (quantidade máxima de nós ou divisões permitidas), *min\_samples\_split* (quantidade mínima de amostras que uma folha precisa ter para se dividir), *min\_samples\_leaf* (quantidade mínima de amostras que uma folha precisa ter), *min\_weight\_fraction\_leaf* (o mesmo propósito que *min\_samples\_leaf*, mas considerando a porcentagem ao invés de quantidade), *max\_features* (quantidade máxima de parâmetros que precisam ser avaliados antes de realizar a separação em cada nó) (Géron, 2019). Com isso, impedimos que a árvore utilize critérios de separação específicos para o nosso conjunto amostral (*overfitting*), tendo um comportamento generalizado que se adaptará bem à chegada de novos dados.

Além das funções anteriores para evitar o *overfitting*, também podemos realizar a poda da árvore de decisão depois de pronta. Isso é feito utilizando o parâmetro *alpha*, que nos indica o quanto precisaremos cortar da árvore para que a acurácia tanto para o conjunto teste quanto para o conjunto treino sejam iguais ou as mais próximas possíveis.

A Figura 4.12 nos mostra, para diferentes valores de *alpha* no eixo  $x$ , os valores de acurácia tanto para o conjunto treino quanto para o teste para um problema genérico.

Nesse caso, o valor de *alpha* estaria em torno de  $\alpha \approx 0.05$  e a acurácia final do modelo seria de aproximadamente 55%.

Abaixo, deixamos como exemplo um código simples, em Python, para construção de um modelo DTC.



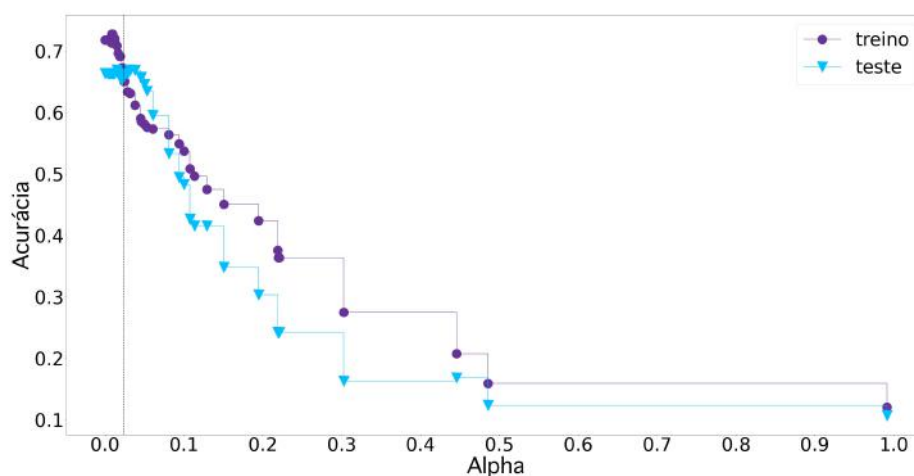


FIGURA 4.12. Alpha versus acurácia para os conjuntos teste e treino.

---

```

from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size=0.3)
dtc = DecisionTreeClassifier(criteriom, splitter, max_depth, ccp.alpha)
dtc = dtc.fit(X_train,y_train)
y_pred = dtc.predict(X_test)
metrics.accuracy_score(y_test,y_pred)
kfold = model_selection.KFold(n_splits, random_state=42)
scores = cross_val_score(dtc, X, y, cv=kfold, scoring)

```

---

#### 4.2.4 *Random Forest Classifier*

O RFC (*Random Forest Classifier*) é um modelo que utiliza  $n$  árvores de decisão para realizar a classificação do conjunto de dados, de forma que é mais seguro na hora de evitar o *overfitting* pela classificação ser realizada  $n$  vezes — por mais que uma árvore tenha tido um problema na classificação, as outras poderão corrigir esse erro. Além disso, utiliza diferentes conjuntos treino e de *features*, o que forma diferentes critérios no momento de gerar a maior separação possível, fazendo com que as árvores tenham baixa correlação umas com as outras. Com isso, aumentamos a variedade do aprendizado das

árvores e também a possibilidade de classificações distintas para um mesmo item, mas sempre buscamos a que será mais recorrente dentre as  $n$  árvores.

Ilustramos na Figura 4.13 um exemplo de RFC para quatro árvores de decisão buscando classificar um mesmo item, que terá classe final B, previsto pela maioria árvores.

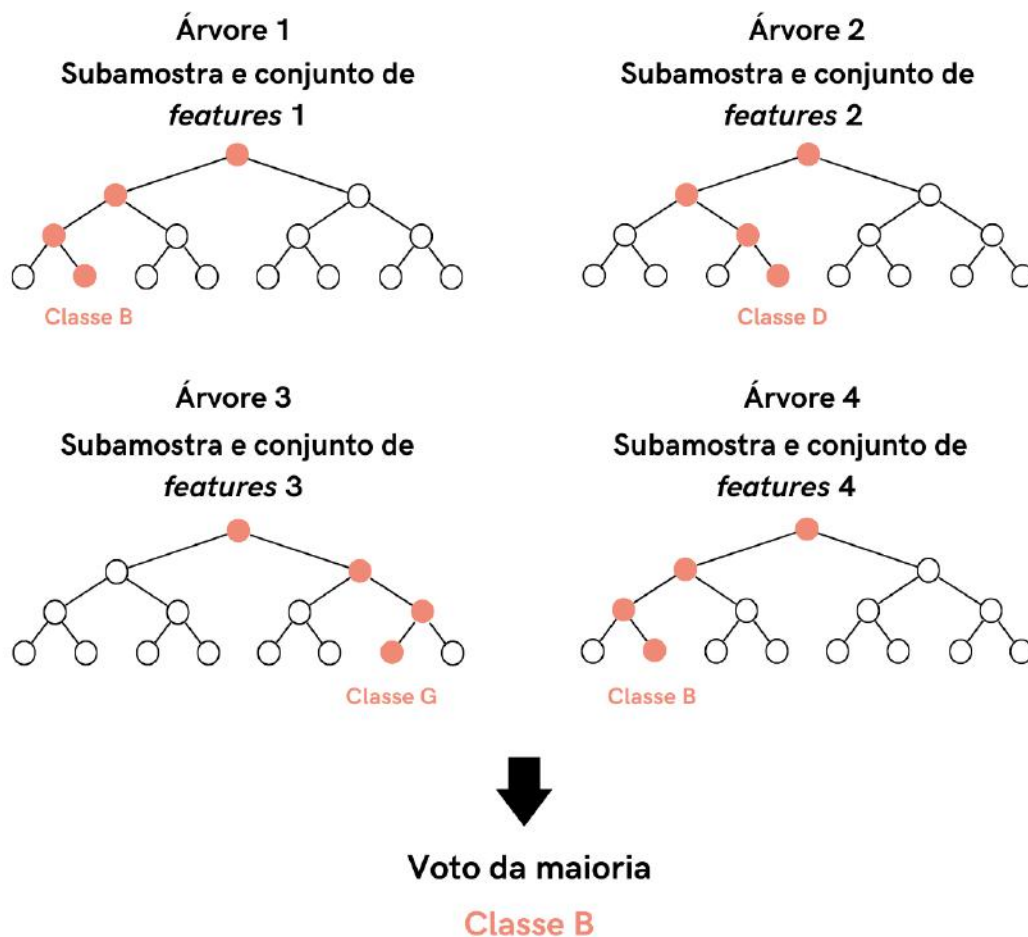


FIGURA 4.13. Exemplo de RFC utilizando quatro árvores de decisão.

A quantidade ideal de árvores de decisão é dada a partir de testes e verificação da acurácia do modelo, e depende diretamente do problema buscando solução. No entanto, normalmente são utilizadas 5 ou 10 árvores. Além disso, os mesmos critérios de delimitação do tamanho da árvore e de poda, vistos na subseção do DTC, podem ser utilizados nesse caso.

Abaixo, deixamos como exemplo um código simples, em Python, para construção de um modelo RFC.

---

```
from sklearn.model_selection import cross_val_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size=0.3)
RFC = RandomForestClassifier(n_estimators, criterion, max_depth, ccp_alpha)
RFC = RFC.fit(X_train,y_train)
rfc_pred = classifier.predict(X_test)
kfold = model_selection.KFold(n_splits=5, shuffle=True, random_state=42)
scores = cross_val_score(RFC, X, y, cv=kfold, scoring='f1_weighted')
```

---

Caso o leitor tenha interesse em se aprofundar mais no funcionamento dos modelos de ML utilizados nesse projeto, recomendamos a leitura do livro Géron (2019) e do artigo Baron (2019). Em Géron (2019), você conhecerá outras propriedades fundamentais sobre o aprendizado de máquina, assim como outros modelos com outros objetivos além da classificação supervisionada, e terá a oportunidade de se aprofundar ainda mais nas ferramentas que são disponíveis. Por sua vez, Baron (2019), tem como foco a análise feita por modelos não-supervisionados, por mais que também explore o funcionamento de modelos supervisionados como o *Random Forest Classifier*, o *Support Vector Machines* e as Redes Neurais.

No próximo capítulo abordaremos os resultados obtidos utilizando os dados definidos no Capítulo 3 de Metodologia aplicados nos modelos apresentados neste capítulo.

# Capítulo 5

## Resultados

Neste capítulo, temos como objetivo explorar, inicialmente, os resultados obtidos na classificação espectral feita antes da aplicação dos dados nos modelos de *Machine Learning* (classificação pré-ML), discutida em detalhes no Capítulo 3, e, por fim, apresentaremos os resultados obtidos pelos modelos de KNN, SVM, DTC e RFC, assim como as suas respectivas configurações (p.e., número de vizinhos no KNN, quantidade destinada aos conjuntos treino e teste, etc.).

### 5.1 Resultados da classificação pré-ML

Podemos começar a discussão sobre classificação a partir dos valores encontrados para as larguras equivalentes das linhas He I  $\lambda 4471$  e He II  $\lambda 4542$  dos espectros de alta e de baixa resolução, que são mostradas nas Figuras 5.1 e 5.2, respectivamente. Nessas figuras, enxergamos uma tendência na qual os espectros que possuem maiores valores de  $W_\lambda(\text{HeI}\lambda 4471)$ , possuem menores valores de  $W_\lambda(\text{HeII}\lambda 4542)$ , o que acontece exatamente para os objetos com menores temperaturas, indicados pelos X vermelhos (O9-O9.7). O contrário acontece para os objetos mais quentes, indicados pelos círculos azuis (O3 até O7.5).

O comportamento observado nas Figuras 5.1 e 5.2 é esperado, uma vez que estrelas mais quentes terão uma abundância maior de elementos ionizados que as estrelas mais frias, o que nos traz ainda mais segurança nas medidas obtidas pelo cálculo de largura

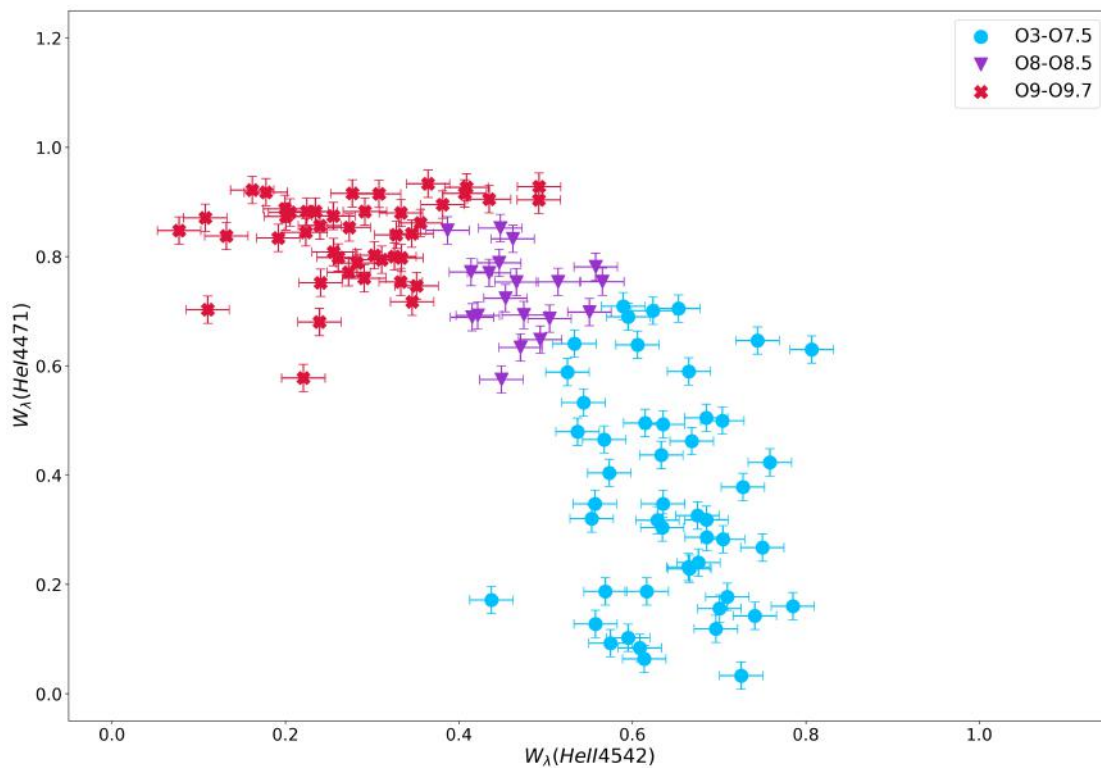


FIGURA 5.1. Valores encontrados para as larguras equivalentes de He I 4542 e He I 4471 para os espectros de alta resolução.

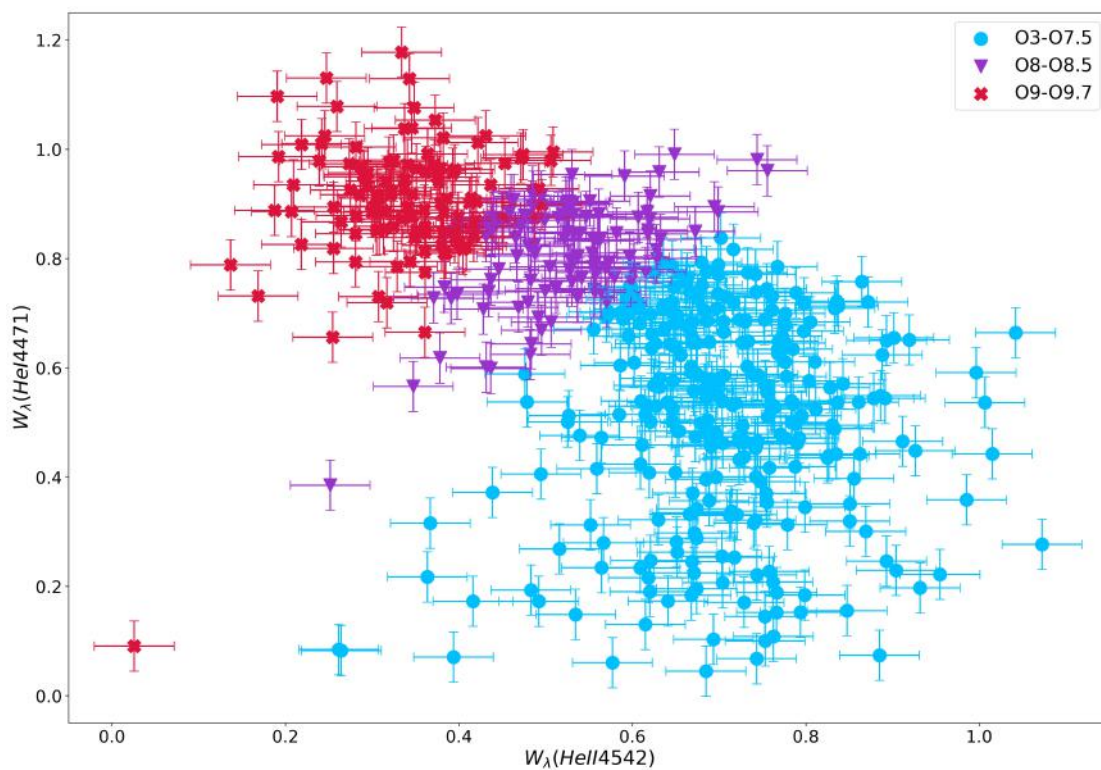


FIGURA 5.2. Valores encontrados para as larguras equivalentes de He I 4542 e He I 4471 para os espectros de baixa resolução.

equivalente descrito no Capítulo 3. No entanto, ainda existem alguns objetos que possuem um comportamento fora do padrão, como o espectro classificado como O9-O9.7 na região inferior esquerda da Figura 5.2). Este objeto é a estrela PZ Gem (ou HD 45314) — classificado como O9 na plataforma Simbad, e como O9.5 por nós, o que não é muito distante — considerada uma estrela Oe (de tipo O com linhas em emissão), e que, nos últimos 10 anos, aproximadamente, tem sofrido variações que alteram a morfologia e a intensidade de suas linhas em emissão na região do ótico (Rauw et al., 2018), o que pode ser o motivo pelo qual ela está distante do seu grupo na Figura 5.2. Essas variações de comportamento de alguns dos objetos, assim como a baixa qualidade dos espectros do GOSC, são os motivos, de modo geral, que atrapalham a classificação de forma automatizada.

Assim, após termos encontrado todos os valores das larguras equivalentes necessárias para classificação de tipo espectral e de luminosidade, conseguimos definir a classe de cada um dos objetos a partir dos critérios discutidos no Capítulo 3. Desta forma, vemos na Figura 5.3 a quantidade de espectros de alta resolução em cada classe de tipo espectral e, na Figura 5.4, em cada classe de tipo espectral e de luminosidade.

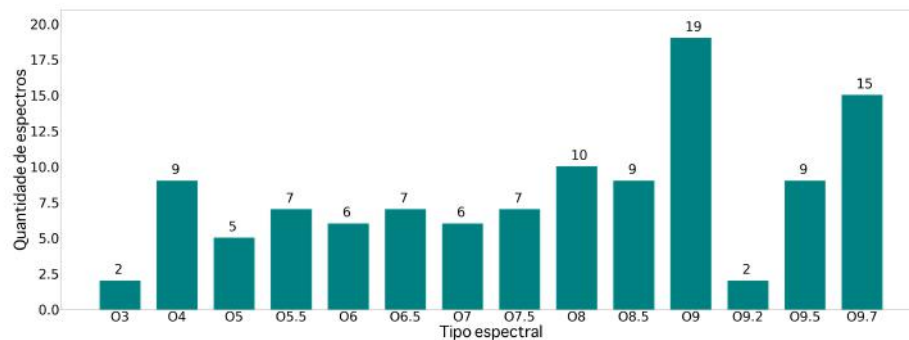


FIGURA 5.3. Histograma que mostra a quantidade de espectros de alta resolução em cada classe de tipo espectral.

No caso do conjunto de espectros de alta e de baixa resolução, vemos na Figura 5.5, a quantidade de objetos em cada classe de tipo espectral e, na Figura 5.6, em cada classe de tipo espectral e de luminosidade.

Podemos ver que as estrelas de tipo O3 e O9.2 aparecem menor quantidade nos histogramas presentes nas Figuras 5.3 e 5.5. As estrelas O3 são objetos mais raros no meio interestelar por serem as estrelas mais quentes dentre as estrelas de tipo O, e, com relação as estrelas de tipo O9.2, acreditamos que tenhamos poucos objetos por essa

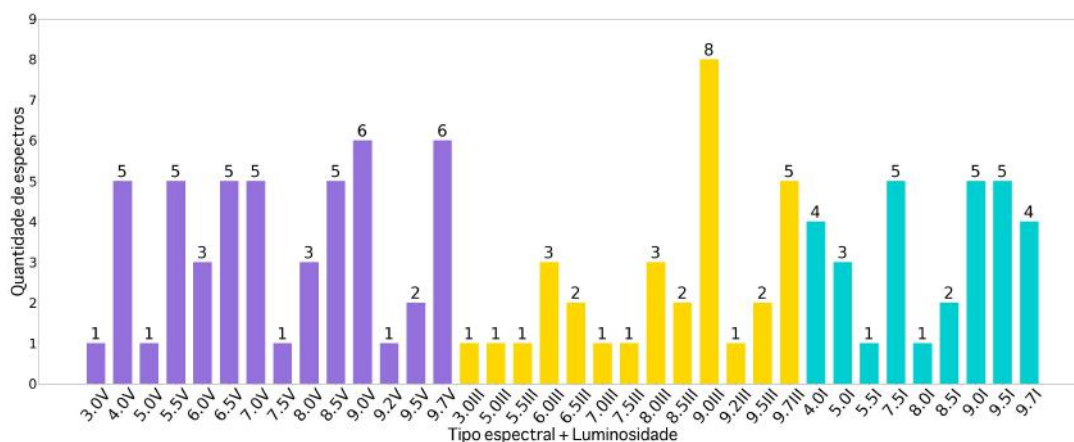


FIGURA 5.4. Histograma que mostra a quantidade de espectros de alta resolução em cada classe de tipo espectral e de luminosidade. As barras em roxo indicam estrelas V (anãs), em amarelo as estrelas III (gigantes), e em azul as estrelas I (supergigantes).

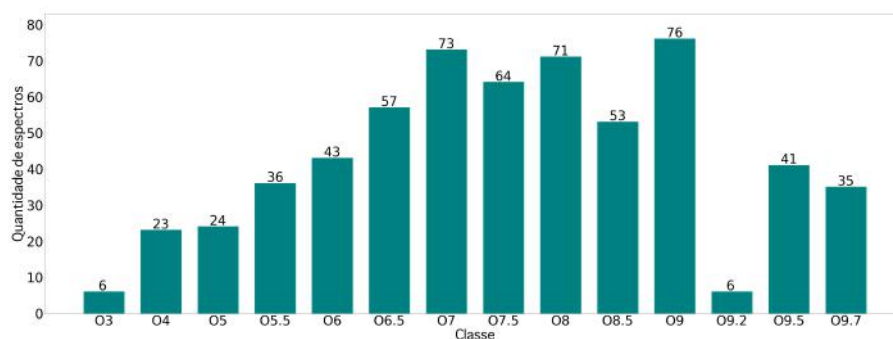


FIGURA 5.5. Histograma que mostra a quantidade de espectros de alta e de baixa resolução em cada classe de tipo espectral.

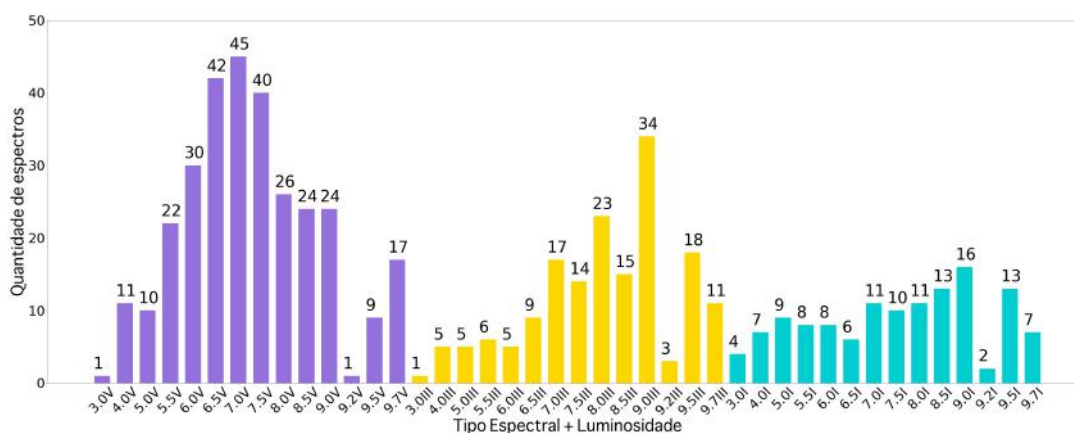


FIGURA 5.6. Histograma que mostra a quantidade de espectros de alta e de baixa resolução em cada classe de tipo espectral e de luminosidade. As barras em roxo indicam estrelas V (anãs), em amarelo as estrelas III (gigantes), e em azul as estrelas I (supergigantes).

classificação aparecer apenas em Martins (2018) e pelo problema da sobreposição entre as classificações O8 e O9.7, exposto no Capítulo 3.

Caso seja de interesse do leitor, as classificações encontradas, além dos valores das larguras equivalentes das linhas utilizadas para classificação de tipo espectral e de luminosidade, para cada um dos espectros de alta e de baixa resolução, se encontram disponibilizadas nas Tabelas A.1 e A.2 presentes no Apêndice A.

Com a classificação concluída, podemos finalmente discutir os resultados de classificação obtidos pelos modelos de ML, que estarão organizados da seguinte forma: na seção 5.2 apresentaremos os resultados de classificação para os espectros de alta resolução, inicialmente, apenas para classificação de tipo espectral e, posteriormente, para tipo espectral e luminosidade em conjunto; na seção 5.3, apresentaremos os resultados de classificação para o conjunto de espectros formado tanto pelos de alta quanto os de baixa resolução. Da mesma forma que na seção 5.2, primeiramente abordaremos a classificação de tipo espectral e posteriormente de tipo espectral e de luminosidade.

Assim, antes de iniciarmos a discussão, é importante notar que os conjuntos de *features* que obtiveram os melhores resultados, tanto para os espectros de alta resolução, quanto para o conjunto de espectros de alta e de baixa resolução, foram:

- Classificação de tipo espectral:  $\log[W_\lambda(HeI\lambda4471)/W_\lambda(HeII\lambda4542)]$  e  $W_\lambda(HeI\lambda4471)$ ;
- Classificação de tipo espectral e de luminosidade:  $W_\lambda(HeI\lambda4471)$ ,  $\log[W_\lambda(HeI\lambda4388)] + \log[W_\lambda(HeII\lambda4686)]$  e  $\log[W_\lambda(HeI\lambda4471)/W_\lambda(HeII\lambda4542)]$ .

Caso o leitor não tenha interesse na discussão acerca da configuração de cada um dos modelos aplicados, além da visualização dos seus respectivos gráficos, pode se dirigir à seção 5.4, onde apresentamos as Tabelas 5.1 e 5.2 que contém os resultados finais obtidos de *F1-Score* utilizando *Cross Validation*, além das configurações utilizadas para cada modelo.

## 5.2 Resultados de ML utilizando espectros de alta resolução

Nesta seção apresentaremos os resultados de classificação para os espectros de alta resolução utilizando os modelos KNN, SVM, DTC e RFC. É importante lembrar que, nesse caso, foram utilizados 113 espectros para classificação de tipo espectral e 110 para classificação de tipo espectral e de luminosidade, como comentado no Capítulo 3.



### 5.2.1 Classificação de tipo espectral

Para a classificação de tipo espectral temos 14 classes disponíveis para classificação: O3, O4, O5, O5.5, O6, O6.5, O7, O7.5, O8, O8.5, O9, O9.2, O9.5 e O9.7.

#### 5.2.1.1 *K-Nearest Neighbors*

O melhor resultado obtido utilizando o modelo KNN foi a partir da seguinte configuração:

- 30% da amostra sendo conjunto teste e 70% conjunto treino;
- $K=3$ ;
- *Cross validation* com 5 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.71 \pm 0.14$ .

As figuras presentes em 5.7 são, respectivamente, o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes da classificação. O Relatório de Classificação nos mostra, na primeira coluna, as classes disponíveis para classificação e, nas três colunas seguintes, os resultados de precisão, *recall* e *F1-Score*, para cada classe. Na última coluna, chamada de *support*, temos a quantidade de espectros de cada classe presente no conjunto teste.

É possível notar que algumas classes tiveram resultado zero nas colunas de precisão, *recall* e *F1-Score*, o que indica que todos os espectros destas classes presentes no conjunto teste foram classificados incorretamente e, conseqüentemente, podemos concluir que o aprendizado não foi realizado. Além disso, algumas não continham nenhum representante no conjunto teste (p.e., como vemos para a classe O3, na qual *support*=0). O que causa esses resultados é a sub-amostragem de determinadas classes, como vimos no histograma da Figura 5.3, de forma que a baixa quantidade de espectros nos impede de realizar correlações entre as *features* e, muitas vezes, nos impedem de realizar a testagem do pouco aprendizado que foi capaz de ser feito.

Na Figura 5.7, temos a Matriz de Confusão e vemos todas as classes disponíveis para classificação tanto no eixo  $x$  (previsto) quanto no eixo  $y$  (real). No cenário ideal, gostaríamos que todos espectros estivessem presentes apenas na diagonal principal, como comentado no Capítulo 4. No entanto, por mais que esse não seja o caso, vemos que as classificações previstas não se distanciam muito das reais, de forma que estão logo acima ou logo abaixo da classe que é considerada como o verdadeiro positivo (p.e., os espectros de tipo O7, quando incorretamente classificados, foram classificados como O6.5). A exceção seria o caso dos objetos da classe O9.5, que foram classificados como O9, o que também não se distancia muito da realidade.

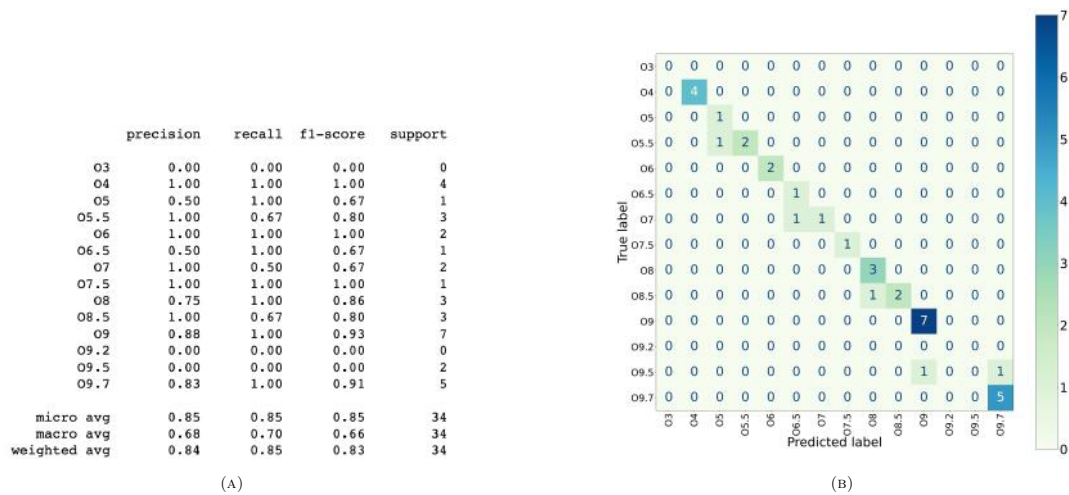


FIGURA 5.7. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta resolução por KNN. A *colorbar* indica a quantidade de espectros nas posições da matriz.

É importante lembrar que, como o *Cross Validation* utiliza  $n$  conjuntos de grupos teste e treino para realizar a sua classificação final, os resultados mostrados na Figura 5.7 são relativos a um dos conjuntos utilizados para se obter o resultado de *F1-Score* final. Lembrando que isso é importante porque classes que não obtiveram classificações corretas nesse caso, podem ter tido uma melhor performance em outros testes.

Por fim, podemos gerar o gráfico resultante da classificação por KNN, que é mostrado na Figura 5.8. Nesta podemos ver as regiões de classificação indicadas pelas diferentes cores — estas regiões indicam que qualquer novo objeto que esteja inserido na região branca, por exemplo, terá classificação O3.

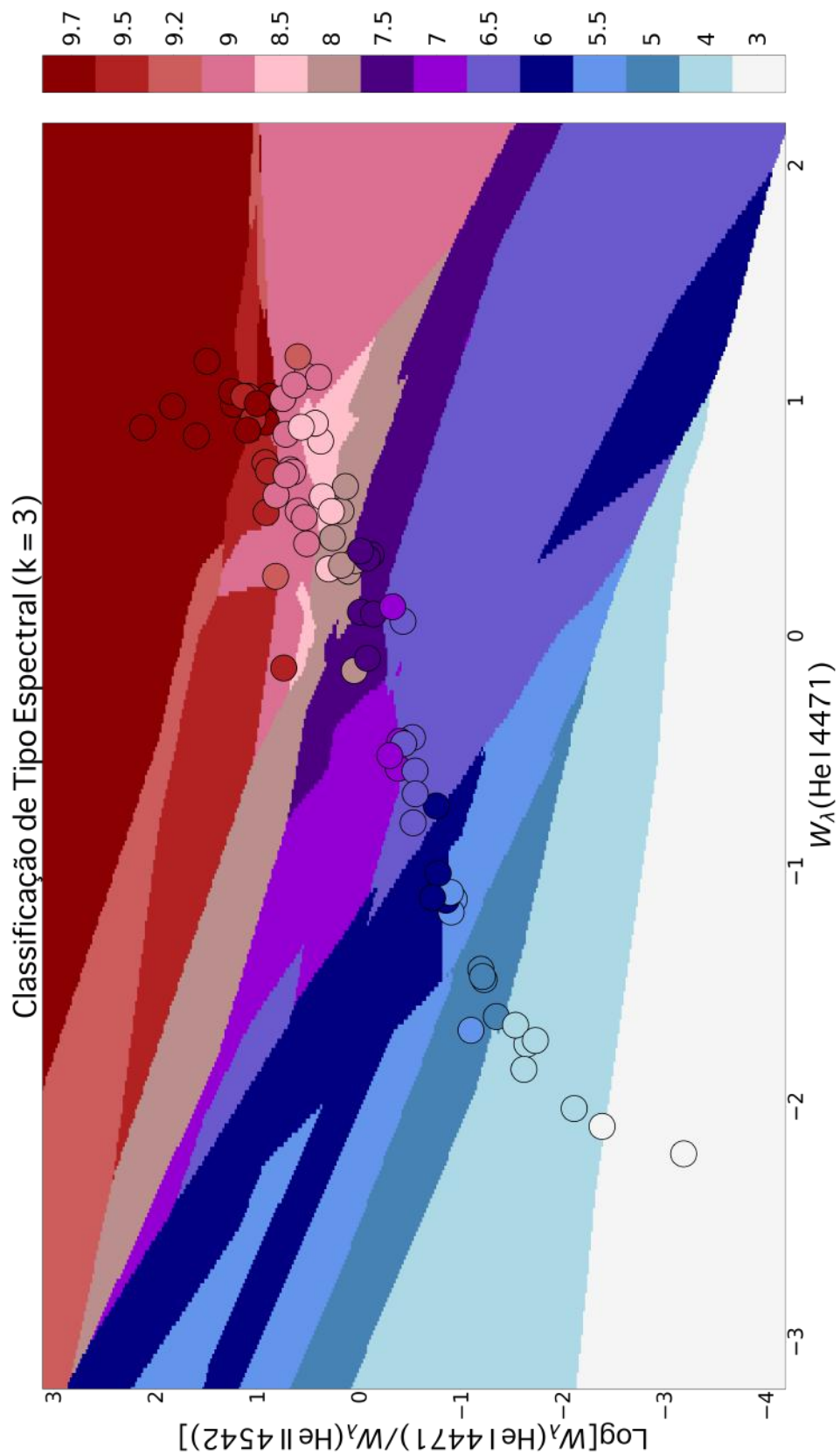


FIGURA 5.8. Gráfico de KNN resultante para classificação de tipo espectral para espectros de alta resolução.

### 5.2.1.2 Support Vector Machine

O melhor resultado obtido utilizando o modelo SVM foi a partir da seguinte configuração:

- 30% da amostra sendo conjunto teste e 70% conjunto treino;
- $C=27$ ;
- *Cross validation* com 10 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.73 \pm 0.25$ .

O valor ideal de  $C$  foi encontrado a partir de um gráfico no qual o eixo  $x$  indica diferentes valores de  $C$  e o eixo  $y$  os respectivos valores de *F1-Score*, como podemos ver na Figura 5.9. Assim, buscamos um valor de  $C$  que nos retornasse bons resultados de *F1-Score* e também a menor incerteza possível.

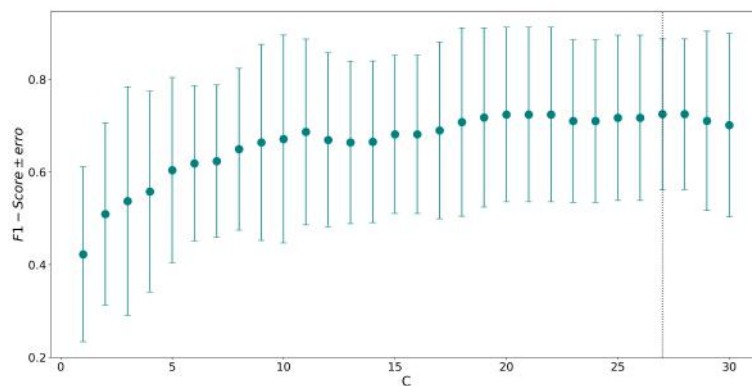


FIGURA 5.9. No eixo  $x$  temos diferentes valores de  $C$  e no eixo  $y$  os respectivos resultados de *F1-Score*.

Na Figura 5.10 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes da classificação. Note que neste caso tivemos representantes para todas as classes no conjunto teste, mas ainda tivemos casos onde precisão, *recall* e *F1-Score* são iguais a zero, mostrando uma falta de aprendizado para essas classes (p.e., a classe O5.5).

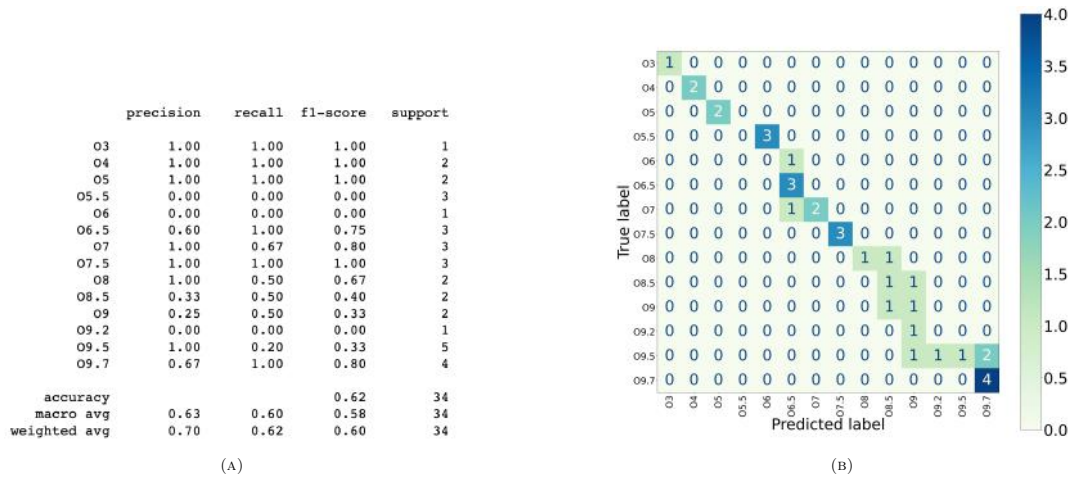


FIGURA 5.10. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta resolução por SVM.

### 5.2.1.3 Decision Tree Classifier

O melhor resultado obtido utilizando o modelo DTC foi a partir da seguinte configuração:

- 30% da amostra sendo conjunto teste e 70% conjunto treino;
- $max\_depth = 5$ ;
- Critério de impureza: Entropia;
- *Cross validation* com 5 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.72 \pm 0.21$ .

Além disso, tivemos que o valor de  $\alpha$ , ou seja, o valor para o qual a diferença entre a performance do conjunto teste e a do conjunto treino é a menor possível, equivale a  $\alpha = 0.070$ , como representado pela linha tracejada preta na Figura 5.11. Lembrando que  $\alpha$  é o parâmetro de poda da árvore de decisão, visto em detalhes no Capítulo 4.

Na Figura 5.13 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. Assim como no caso do KNN, tivemos, para algumas classes,  $support=0$ , e também os resultados de precisão, *recall* e *F1-Score* iguais a zero (p.e., classe O6).

Por fim, temos a árvore de decisão resultante deste problema, ilustrada na figura 5.14. Nesta árvore temos indicadas as *features* e seus respectivos critérios de separação,

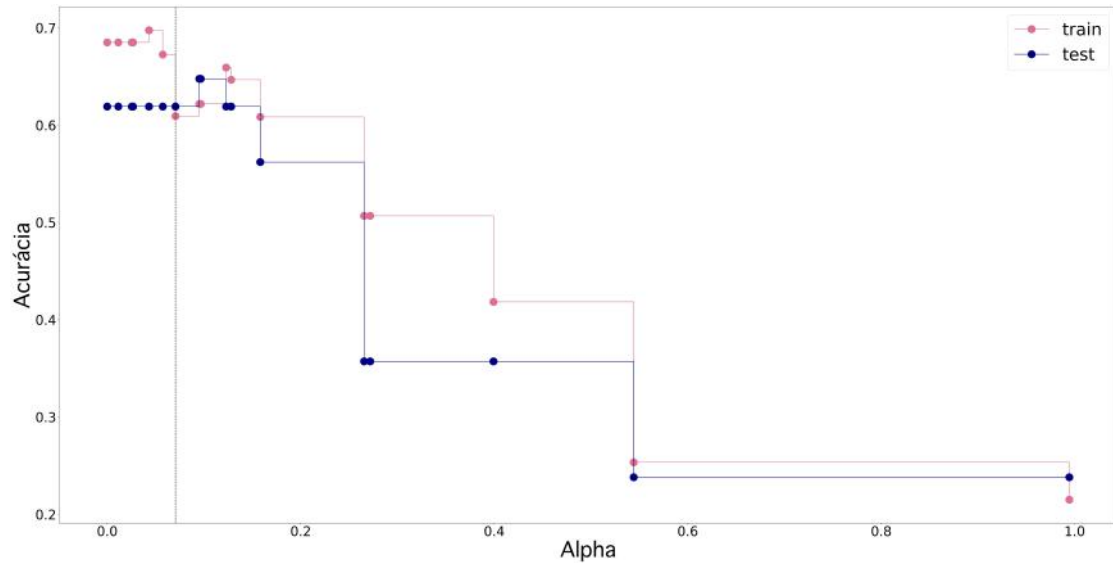


FIGURA 5.11. No eixo  $x$  temos diferentes valores de  $\alpha$  e no eixo  $y$  temos os valores de acurácia resultantes.

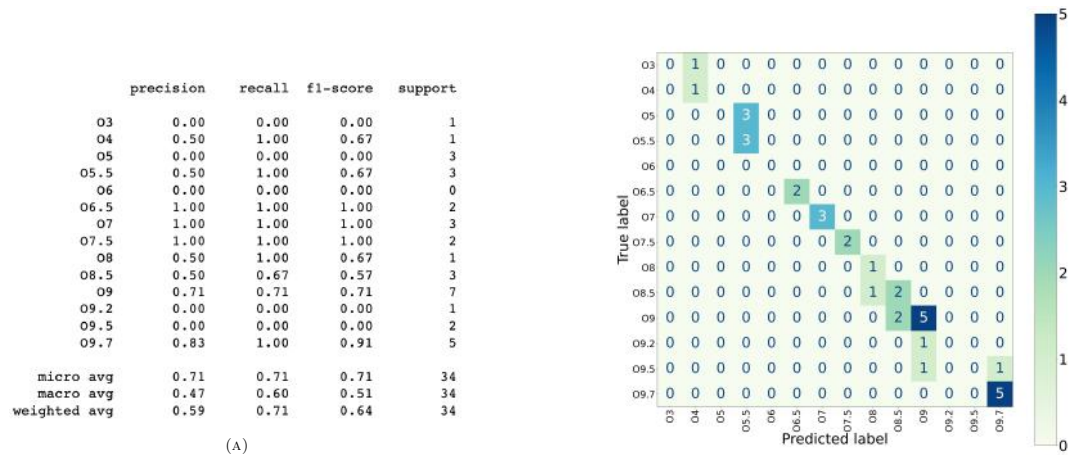


FIGURA 5.12. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta resolução por DTC.

FIGURA 5.13

assim como foi mostrado no Capítulo 4. Veja que ela se divide no máximo cinco vezes, por termos definido  $max\_depth=5$ . Além disso, vemos que nem sempre o indicador de impureza será igual a zero.

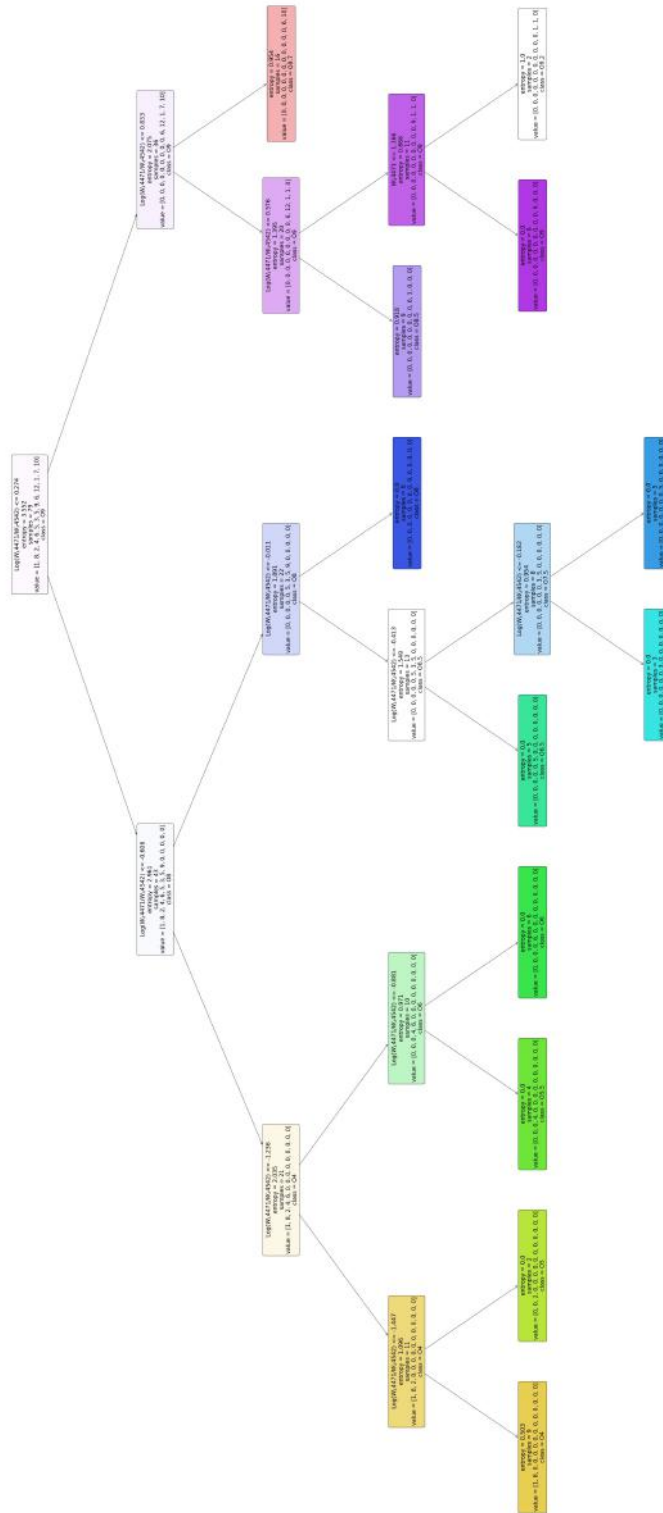


FIGURA 5.14. Essa figura é apenas para ilustrar o resultado da árvore de decisão para classificação de tipo espectral para espectros de alta resolução. Uma versão de alta resolução, que permite a leitura de cada folha da árvore, está disponível eletronicamente clicando aqui. Em cada retângulo informamos o critério de separação utilizado para separar as folhas seguintes (caso haja separação), o valor de entropia, a quantidade de espectros contidos e a classificação da folha até aquele momento.

### 5.2.1.4 Random Forest Classifier

No caso do modelo RFC, utilizamos 10 árvores iguais a descrita para o modelo DTC.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.69 \pm 0.19$ .

Abaixo, na Figura 5.16, temos o Relatório de Classificação (A) e a Matriz de Confusão (B) para esse problema. Assim como para o DTC, temos a classe O6 com resultados de precisão, *recall* e *F1-Score* iguais a zero e sem representantes dessa classe no conjunto teste (*support*=0).

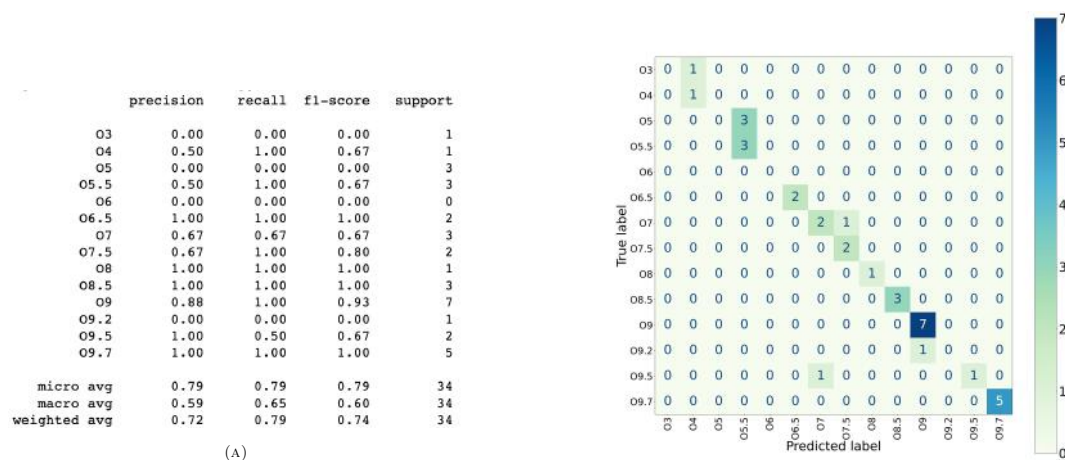


FIGURA 5.15. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta resolução por RFC.

FIGURA 5.16

Neste caso não fizemos o gráfico da árvore, já que seriam 10 árvores de decisão parecidas com a da Figura 5.14.



## 5.2.2 Classificação de tipo espectral e de luminosidade

Para a classificação de tipo espectral e de luminosidade, passamos a ter 36 classes disponíveis para classificação: O3V, O4V, O5V, O5.5V, O6V, O6.5V, O7V, O7.5V, O8V, O8.5V, O9V, O9.2V, O9.5V, O9.7V, O3III, O5III, O5.5III, O6III, O6.5III, O7III, O7.5III, O8III, O8.5III, O9III, O9.2III, O9.5III, O9.7III, O4I, O5I, O5.5I, O7.5I, O8I, O8.5I, O9I, O9.5I e O9.7I.

### 5.2.2.1 *K-Nearest Neighbors*

O melhor resultado obtido utilizando o modelo KNN foi a partir da seguinte configuração:

- 30% da amostra sendo conjunto teste e 70% conjunto treino;
- $K=4$ ;
- *Cross validation* com 10 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.48 \pm 0.17$ .

Na Figura 5.17 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes da classificação. Note que agora passamos a ter mais classes disponíveis para classificação, nas quais diversas obtiveram resultados de precisão, *recall* e *F1-Score* iguais a zero, além de não terem representantes no conjunto teste (p.e., O3III, O3V, etc.). Isso é um indicador do porquê o resultado final de *F1-Score* ser baixo: a falta de aprendizado para a maioria das classes causada pela subamostragem, como vimos na Figura 5.4.

Devido a grande quantidade de classes e o fato de termos utilizado três *features* diferentes para classificação, o gráfico de KNN não nos ajudaria na compreensão do problema neste caso, por termos 36 regiões de classificação em três dimensões, além da performance de classificação ter sido baixa.

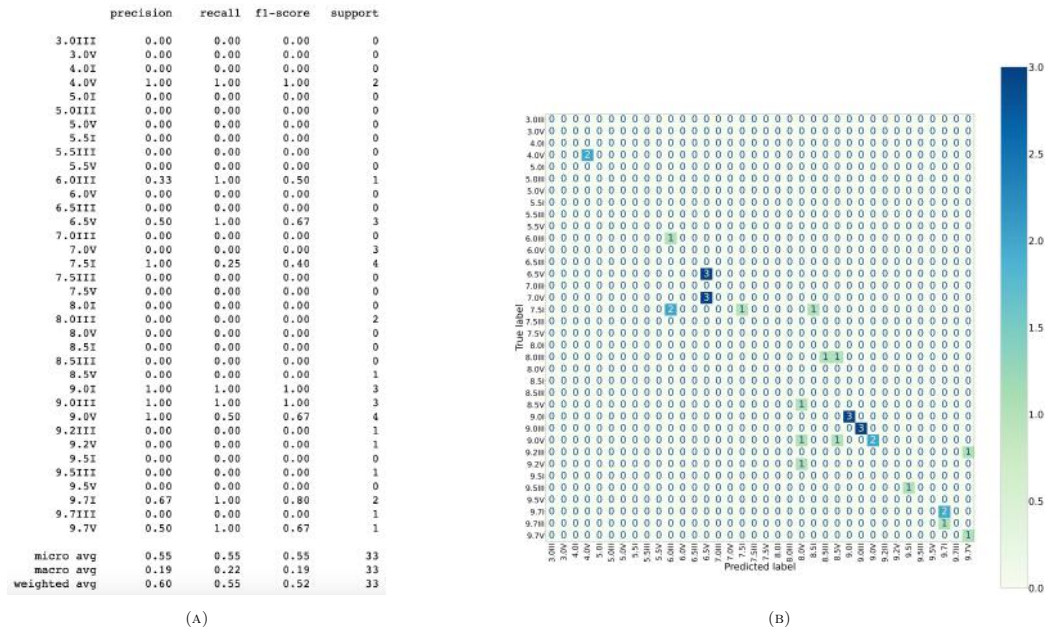


FIGURA 5.17. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral e de luminosidade com espectros de alta resolução por KNN.

5.2.2.2 Support Vector Machine

O melhor resultado obtido utilizando o modelo SVM foi a partir da seguinte configuração:

- 30% da amostra sendo conjunto teste e 70% conjunto treino;
- $C=10$ ;
- *Cross validation* com 5 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.52 \pm 0.18$ .

Na Figura 5.18 temos indicado o valor ideal de  $C$ , explicado anteriormente na subseção 5.2.1.2 de classificação de tipo espectral para objetos de alta resolução.

Na Figura 5.19 temos o Relatório de Classificação e a Matriz de Confusão resultantes da classificação. Aqui encontramos o mesmo problema da subamostragem nas classes indicado na seção anterior do KNN e, com isso, o resultado final de classificação também é baixo.

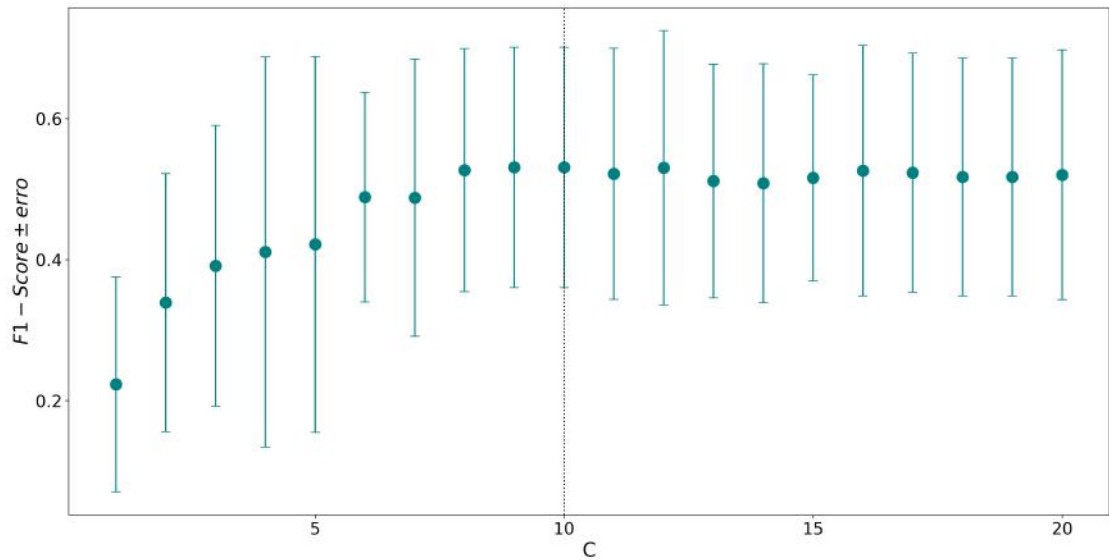
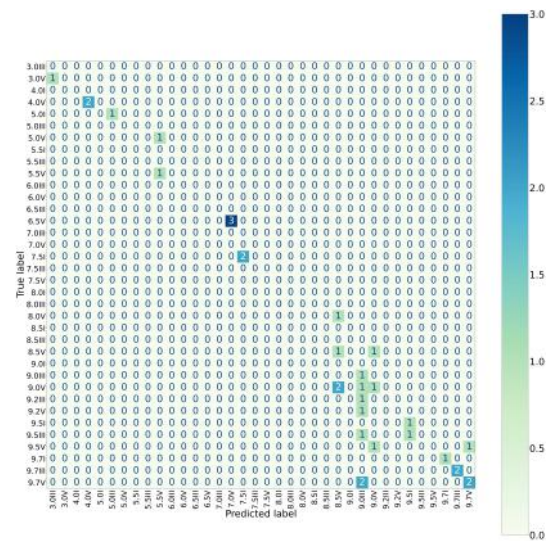


FIGURA 5.18. No eixo  $x$  temos diferentes valores de  $C$  e no eixo  $y$  os respectivos resultados de  $F1-Score$ .

	precision	recall	f1-score	support
3.0III	0.00	0.00	0.00	0
3.0V	0.00	0.00	0.00	1
4.0I	0.00	0.00	0.00	0
4.0V	1.00	1.00	1.00	2
5.0I	0.00	0.00	0.00	1
5.0III	0.00	0.00	0.00	0
5.0V	0.00	0.00	0.00	1
5.5I	0.00	0.00	0.00	0
5.5III	0.00	0.00	0.00	0
5.5V	0.50	1.00	0.67	1
6.0III	0.00	0.00	0.00	0
6.0V	0.00	0.00	0.00	0
6.5III	0.00	0.00	0.00	0
6.5V	0.00	0.00	0.00	3
7.0III	0.00	0.00	0.00	0
7.0V	0.00	0.00	0.00	0
7.5I	1.00	1.00	1.00	2
7.5III	0.00	0.00	0.00	0
7.5V	0.00	0.00	0.00	0
8.0I	0.00	0.00	0.00	0
8.0III	0.00	0.00	0.00	0
8.0V	0.00	0.00	0.00	1
8.5I	0.00	0.00	0.00	0
8.5III	0.00	0.00	0.00	0
8.5V	0.25	0.50	0.33	2
9.0I	0.00	0.00	0.00	0
9.0III	0.14	1.00	0.25	1
9.0V	0.33	0.25	0.29	4
9.2III	0.00	0.00	0.00	1
9.2V	0.00	0.00	0.00	1
9.5I	0.50	1.00	0.67	1
9.5III	0.00	0.00	0.00	2
9.5V	0.00	0.00	0.00	2
9.7I	1.00	1.00	1.00	1
9.7III	1.00	1.00	1.00	2
9.7V	0.67	0.50	0.57	4
micro avg	0.42	0.42	0.42	33
macro avg	0.18	0.23	0.19	33
weighted avg	0.38	0.42	0.38	33

(A)



(B)

FIGURA 5.19. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral e de luminosidade com espectros de alta resolução por SVM.

### 5.2.2.3 Decision Tree Classifier

O melhor resultado obtido utilizando o modelo DTC foi a partir da seguinte configuração:

- 35% da amostra sendo conjunto teste e 65% conjunto treino;
- $max\_depth = 10$ ;
- Critério de impureza: Entropia;

- *Cross validation* com 5 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.47 \pm 0.22$ .

Além disso, tivemos que  $\alpha = 0.039$ , o parâmetro de poda da árvore de decisão, indicado na Figura 5.20 pela linha tracejada em preto.

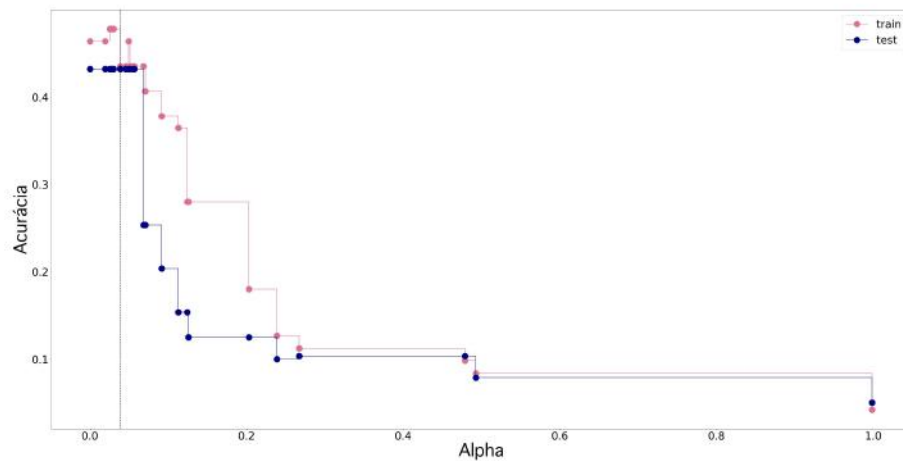


FIGURA 5.20. No eixo  $x$  temos diferentes valores de  $\alpha$  e no eixo  $y$  temos os respectivos valores de acurácia.

Na Figura 5.21 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. Assim como para os modelos anteriores, a capacidade de aprendizado foi baixa e, por esse motivo, não geramos uma árvore de decisão para este caso já que a classificação não foi bem sucedida.

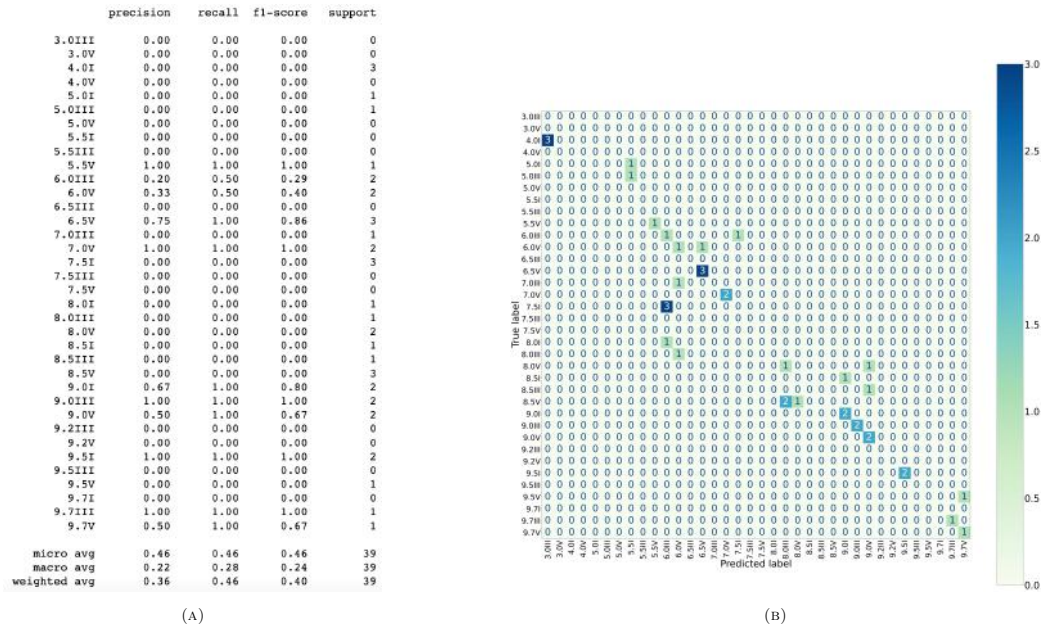


FIGURA 5.21. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral e de luminosidade com espectros de alta resolução por DTC.

### 5.2.2.4 Random Forest Classifier

No caso do modelo RFC, utilizamos 10 árvores iguais à descrita no DTC.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.45 \pm 0.24$ .

Na Figura 5.22 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. Assim como para os modelos anteriores, a capacidade de aprendizado foi baixa.

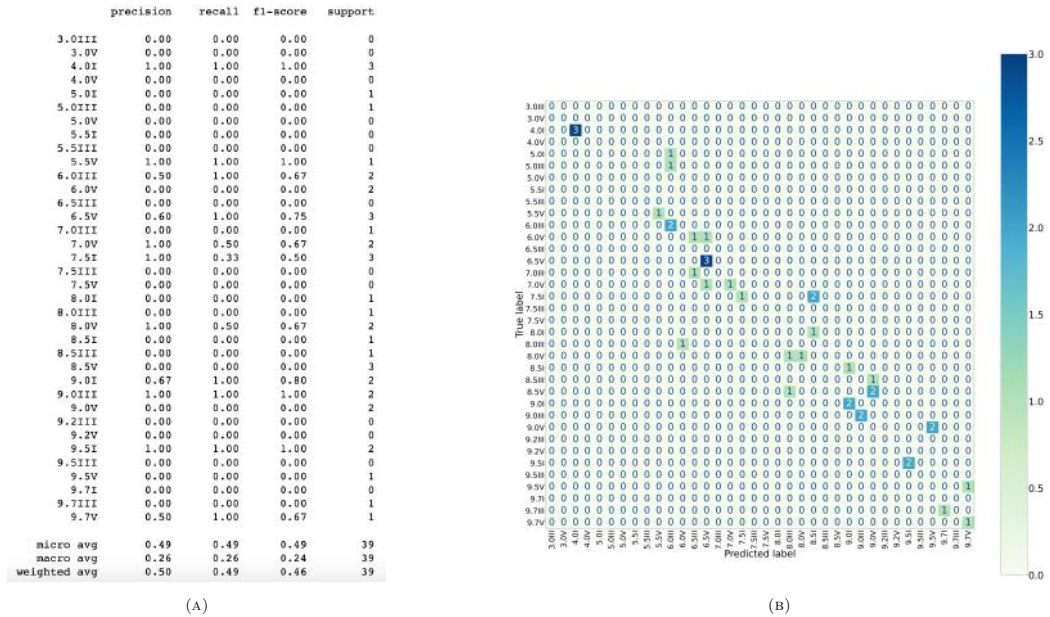


FIGURA 5.22. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral e de luminosidade com espectros de alta resolução por RFC.

### 5.3 Resultados utilizando espectros de alta e de baixa resolução

Nesta seção apresentaremos os resultados de classificação para os espectros de alta e de baixa resolução utilizando os modelos KNN, SVM, DTC e RFC. É importante lembrar que, nesse caso, foram 113 espectros de alta resolução junto a 495 espectros de baixa resolução para classificação de tipo espectral (formando uma amostra total de 608 espectros), e 110 de alta resolução junto a 483 de baixa resolução para classificação de tipo espectral e de luminosidade (formando uma amostra total de 593 espectros), como comentado no Capítulo 3.

#### 5.3.1 Classificação de tipo espectral

Para a classificação de tipo espectral temos 14 classes disponíveis para classificação: O3, O4, O5, O5.5, O6, O6.5, O7, O7.5, O8, O8.5, O9, O9.2, O9.5 e O9.7.

##### 5.3.1.1 K-Nearest Neighbors

O melhor resultado obtido utilizando o modelo KNN foi a partir da seguinte configuração:



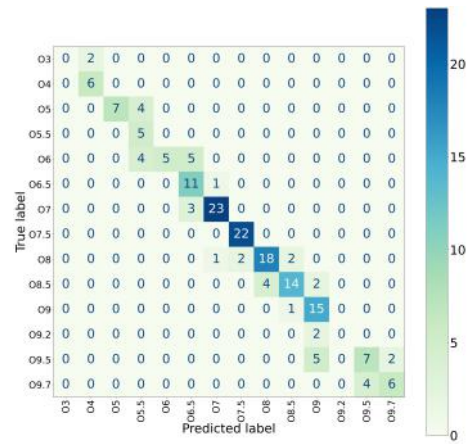
- 30% da amostra sendo conjunto teste e 70% conjunto treino;
- $K=3$ ;
- *Cross validation* com 10 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.82 \pm 0.07$ .

Na Figura 5.23 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. Neste caso, como a quantidade de espectros no conjunto teste aumentou em comparação ao que vimos na seção anterior, passamos a ter representantes para todas as classes na coluna *support*. No entanto, ainda temos classes que possuem pouca representatividade na amostra total, para as quais os resultados de precisão, *recall* e *F1-Score* são iguais a zero (p.e., classe O9.2). Porém, basta olharmos o histograma na Figura 5.5, que observamos como houve uma melhora expressiva na quantidade de objetos por classe ao compararmos com o histograma na Figura 5.3.

	precision	recall	f1-score	support
O3	1.00	1.00	1.00	2
O4	1.00	1.00	1.00	6
O5	1.00	0.73	0.84	11
O5.5	0.50	1.00	0.67	5
O6	1.00	0.50	0.67	14
O6.5	0.57	1.00	0.73	12
O7	0.95	0.81	0.88	26
O7.5	0.91	0.91	0.91	22
O8	0.74	0.87	0.80	23
O8.5	0.76	0.65	0.70	20
O9	0.63	0.75	0.69	16
O9.2	0.00	0.00	0.00	2
O9.5	0.77	0.71	0.74	14
O9.7	0.78	0.70	0.74	10
accuracy			0.78	183
macro avg	0.76	0.76	0.74	183
weighted avg	0.81	0.78	0.78	183

(A)



(B)

FIGURA 5.23. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por KNN.

Por fim, podemos gerar o gráfico resultante da classificação utilizando KNN para espectros de alta e de baixa resolução, que é mostrado na Figura 5.24. Note a similaridade com o gráfico gerado para classificação de tipo espectral para espectros de alta resolução (Figura 5.8).

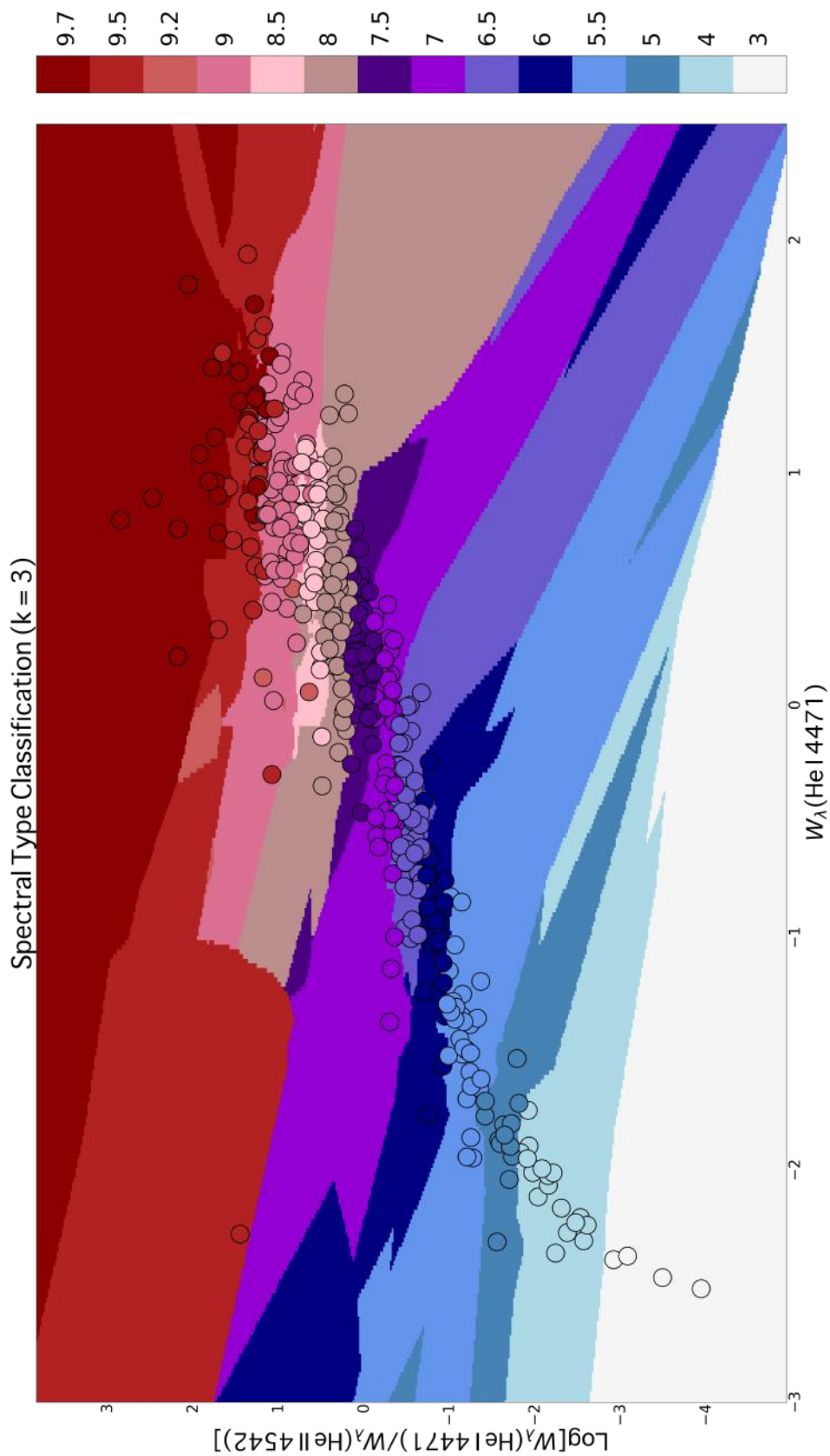


FIGURA 5.24. Gráfico resultante da classificação de tipo espectral utilizando KNN para espectros de alta e de baixa resolução.



### 5.3.1.2 Support Vector Machine

O melhor resultado obtido utilizando o modelo SVM foi a partir da seguinte configuração:

- 30% da amostra sendo conjunto teste e 70% conjunto treino;
- $C=27$ ;
- *Cross validation* com 10 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.86 \pm 0.03$ .

O valor ideal de  $C$  é mostrado pela linha pontilhada preta na Figura 5.25, assim como para os conjuntos de espectros anteriores.

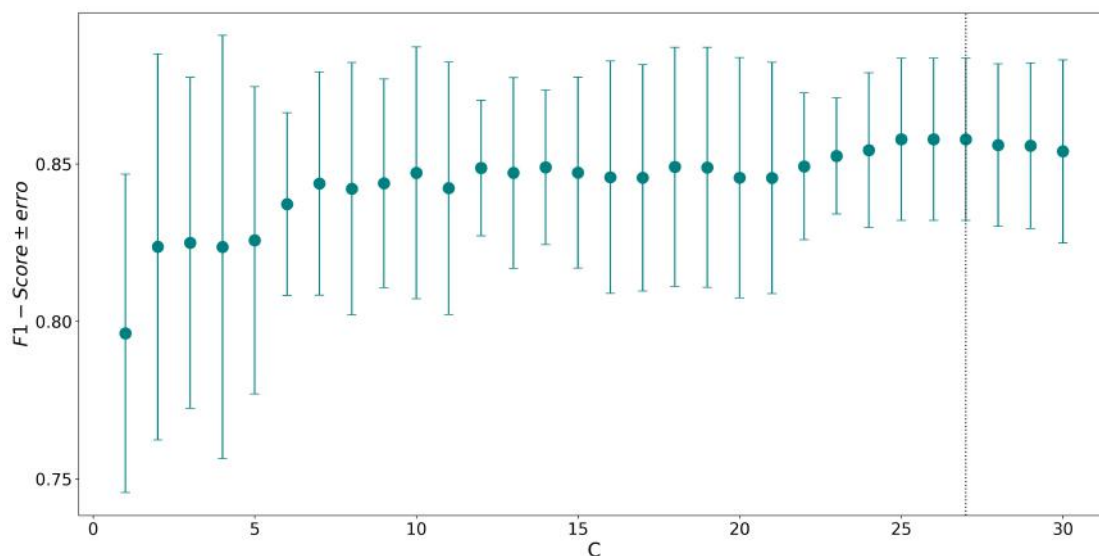


FIGURA 5.25. No eixo  $x$  temos diferentes valores de  $C$  e no eixo  $y$  os respectivos resultados de *F1-Score*.

Na Figura 5.26 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. Neste caso tivemos os resultados de precisão, *recall* e *F1-Score* iguais a zero, assim como a quantidade de itens no conjunto teste, para as duas classes com menor representatividade na amostra total: O3 e O9.2.

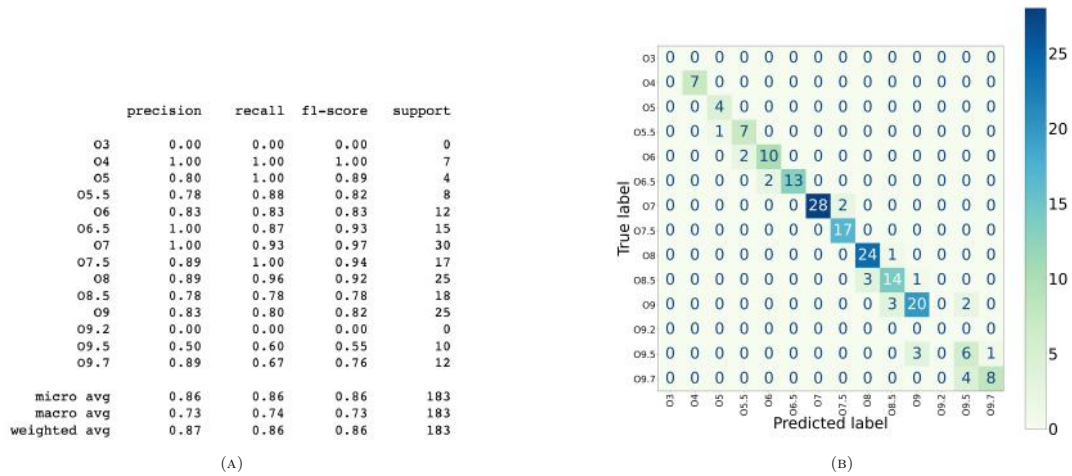


FIGURA 5.26. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por SVM.

### 5.3.1.3 Decision Tree Classifier

O melhor resultado obtido utilizando o modelo DTC foi a partir da seguinte configuração:

- 30% da amostra sendo conjunto teste e 70% conjunto treino;
- $max\_depth = 7$ ;
- Critério de impureza: Entropia;
- *Cross validation* com 10 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.87 \pm 0.11$ .

Além disso, tivemos que  $\alpha = 0.006$ , como podemos ver na Figura 5.27, na qual a linha traceja indica o valor ideal de  $\alpha$ .

Na Figura 5.28 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. Neste caso, vemos que os resultados para precisão, *recall* e *F1-Score* foram iguais a zero exatamente para as classes com a menor quantidade de espectros: 03 e 09.2. Além disso, estas classes não estão presentes no conjunto teste que vemos em *support*=0, assim como no modelo SVM.

Por fim, temos a árvore de decisão resultante deste problema, indicada na Figura 5.29. Vemos que, neste caso, ela se divide até no máximo 7 vezes, por termos definido

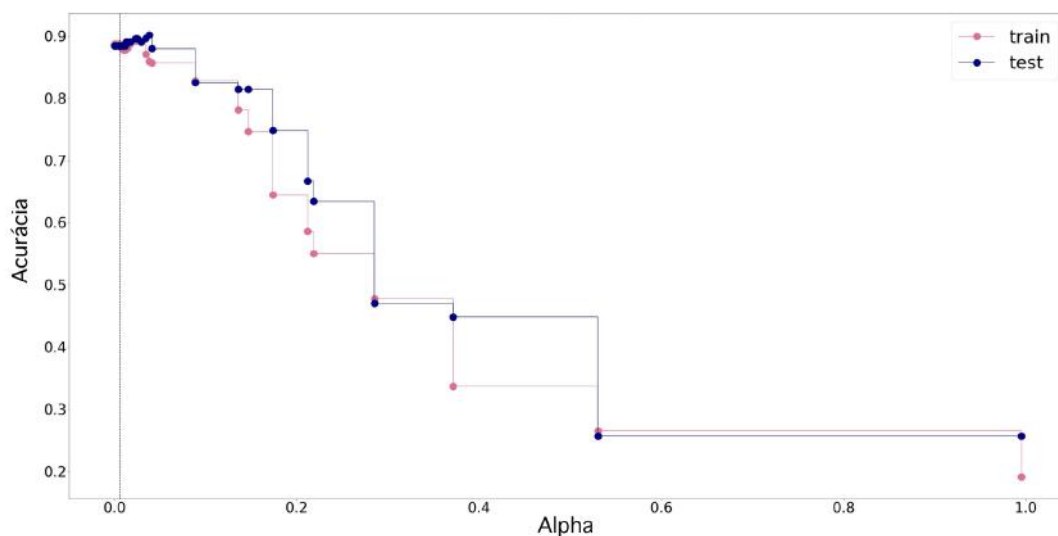


FIGURA 5.27. No eixo  $x$  temos diferentes valores de  $\alpha$  e no eixo  $y$  temos os valores de acurácia resultantes.

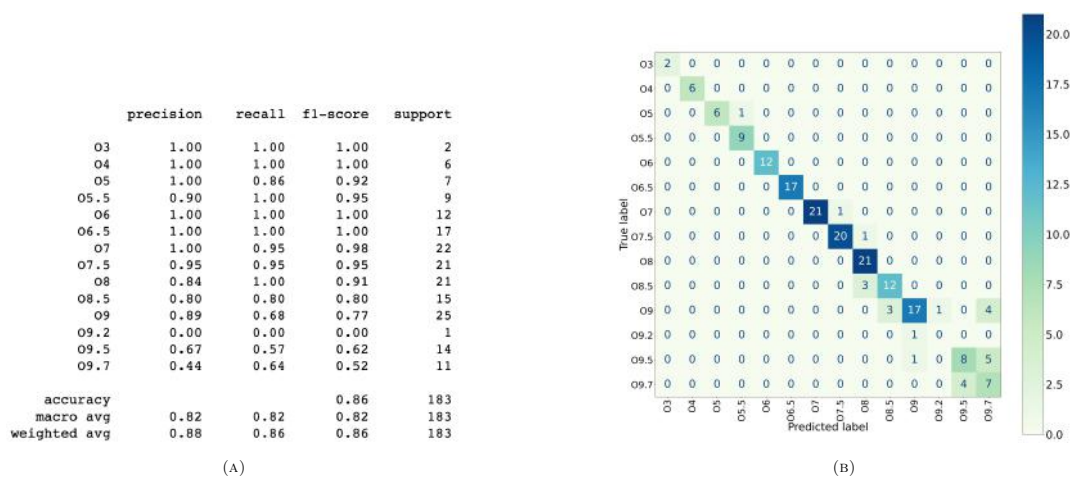


FIGURA 5.28. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por DTC.

$max\_depth=7$ , e nem todas as classes chegam no índice de impureza igual a zero (p.e., classe O8.5 tem Entropia = 0,954).

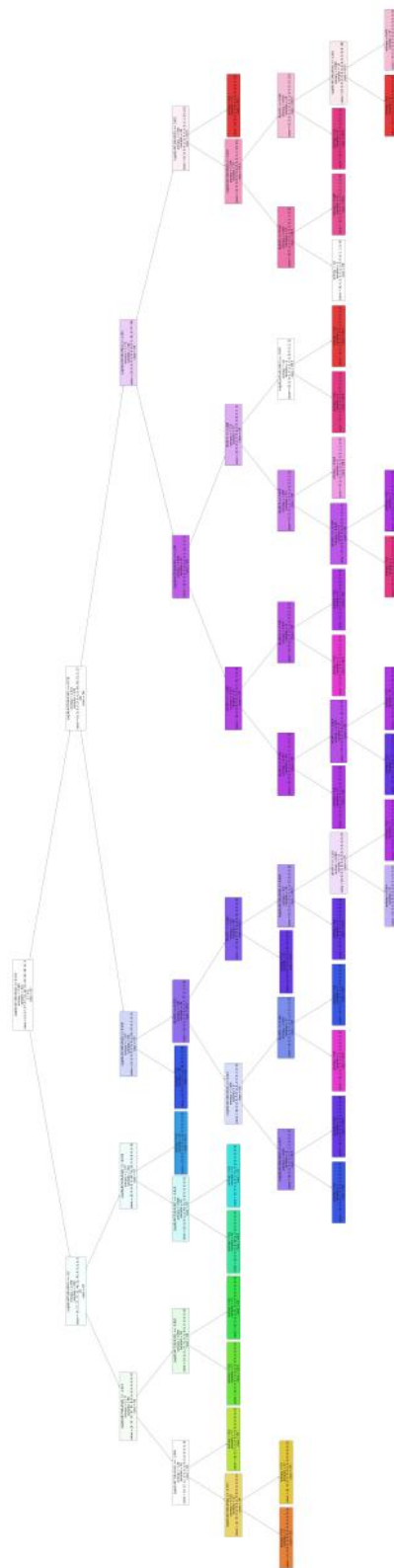


FIGURA 5.29. Essa figura é apenas para ilustrar o resultado da árvore de decisão para classificação de tipo espectral para espectros de alta e de baixa resolução. Uma versão de alta resolução, que permite a leitura de cada folha da árvore, está disponível eletronicamente clicando aqui. Em cada retângulo informamos o critério de separação utilizado para separar as folhas seguintes (caso haja separação), o valor de entropia, a quantidade de espectros contidos e a classificação da folha até aquele momento.

### 5.3.1.4 Random Forest Classifier

No caso do modelo RFC, utilizamos 10 árvores iguais a descrita no DTC.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.87 \pm 0.06$ .

Na Figura 5.28 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. Neste caso tivemos apenas a classe O9.2 com resultados *precision*, *recall* e *F1-Score* iguais a zero, por mais que tenhamos tido 1 representante no conjunto teste, o que indica que a quantidade presente no conjunto treino não foi o suficiente para realizar o aprendizado.

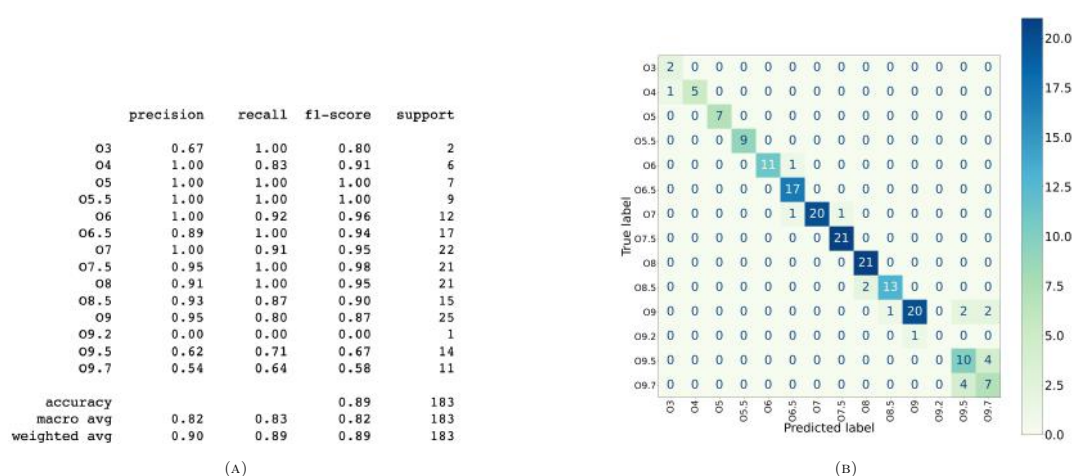


FIGURA 5.30. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por RFC.

### 5.3.2 Classificação de tipo espectral e de luminosidade

Para a classificação de tipo espectral e de luminosidade, passamos a ter 42 classes disponíveis para classificação: O3V, O4V, O5V, O5.5V, O6V, O6.5V, O7V, O7.5V, O8V, O8.5V, O9V, O9.2V, O9.5V, O9.7V, O3III, O4III, O5III, O5.5III, O6III, O6.5III, O7III, O7.5III, O8III, O8.5III, O9III, O9.2III, O9.5III, O9.7III, O3I, O4I, O5I, O5.5I, O6I, O6.5I, O7I, O7.5I, O8I, O8.5I, O9I, O9.2I, O9.5I e O9.7I.

#### 5.3.2.1 *K-Nearest Neighbors*

O melhor resultado obtido utilizando o modelo KNN foi a partir da seguinte configuração:

- 30% da amostra sendo conjunto teste e 70% conjunto treino;
- $K=6$ ;
- *Cross validation* com 5 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.51 \pm 0.06$ .

Na Figura 5.31 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. Vemos que, por mais que tenhamos mais espectros nessa amostra, algumas classes ainda aparecem com resultados de precisão, *recall* e *F1-Score* iguais a zero, além de não terem representatividade no conjunto teste (*support=0*). No entanto, isso já não acontece com tanta frequência quando comparamos à amostra com apenas espectros de alta resolução, como vimos na Figura 5.31.

Por estarmos utilizando três *features*, além da alta quantidade de classes disponíveis para classificação, realizar um gráfico de KNN nesse caso não nos ajudaria na visualização do problema por termos 42 regiões de classificação em três dimensões.

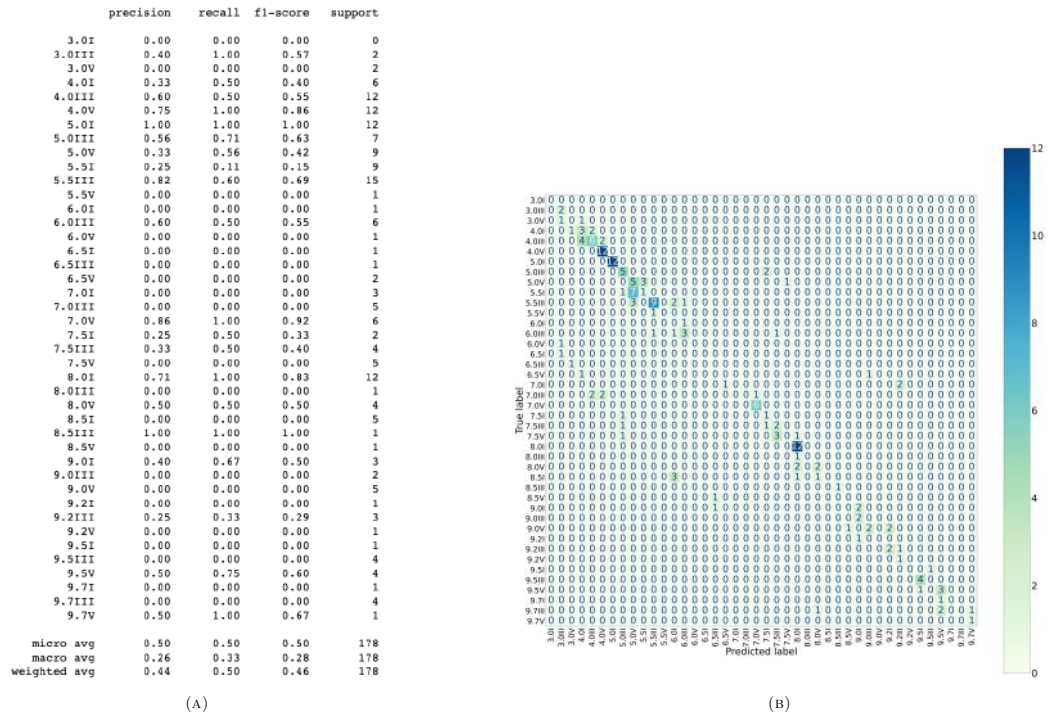


FIGURA 5.31. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por KNN.

### 5.3.2.2 Support Vector Machines

O melhor resultado obtido utilizando o modelo SVM foi a partir da seguinte configuração:

- 35% da amostra sendo conjunto teste e 65% conjunto treino;
- $C=47$ ;
- *Cross validation* com 5 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*:  $0.66 \pm 0.08$ .

O valor ideal de  $C$  é mostrado pela linha pontilhada preta na Figura 5.32.

Na Figura 5.33 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. O mesmo que comentamos para o caso do KNN pode ser observado no Relatório de Classificação e na Matriz de Confusão para o caso do SVM: houve uma melhora no resultado final, mas a falta de itens em determinadas classes ainda atrapalha o aprendizado para classificação.



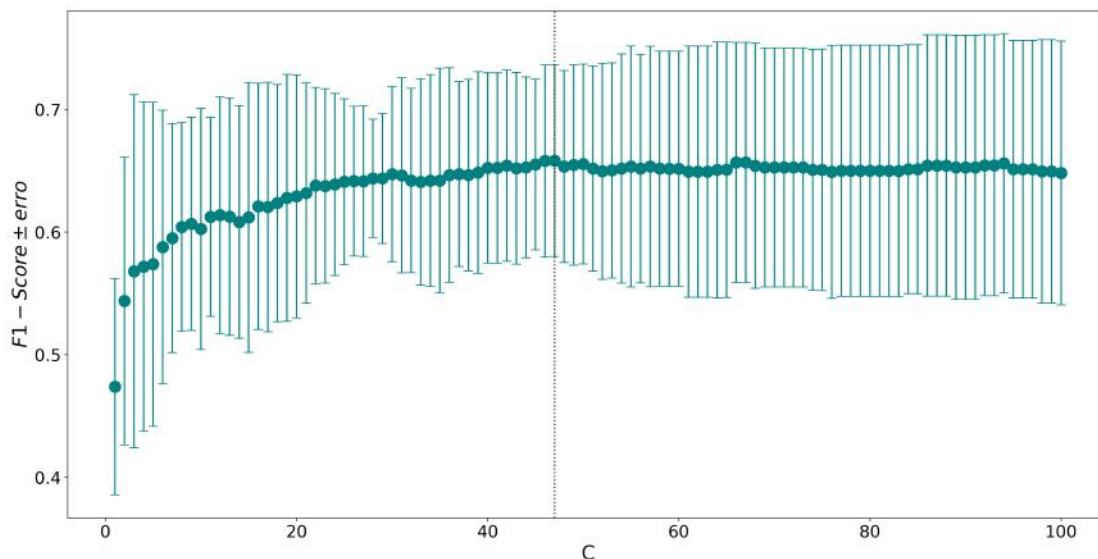
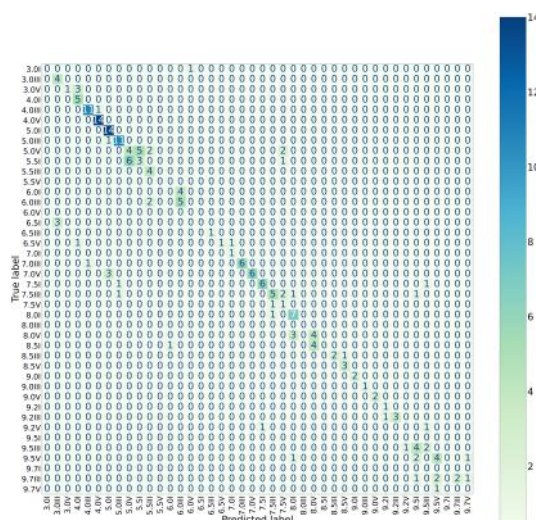


FIGURA 5.32. No eixo  $x$  temos diferentes valores de  $C$  e no eixo  $y$  os respectivos resultados de  $F1-Score$ .

	precision	recall	f1-score	support
3.0I	0.00	0.00	0.00	1
3.0III	0.57	1.00	0.73	4
3.0V	1.00	0.25	0.40	4
4.0I	0.56	1.00	0.71	5
4.0III	0.92	0.92	0.92	12
4.0V	0.93	1.00	0.97	14
5.0I	0.78	1.00	0.88	14
5.0III	0.85	0.92	0.88	12
5.0V	0.40	0.31	0.35	13
5.5I	0.38	0.30	0.33	10
5.5III	0.50	1.00	0.67	4
5.5V	0.00	0.00	0.00	0
6.0I	0.00	0.00	0.00	4
6.0III	0.56	0.71	0.63	7
6.0V	0.00	0.00	0.00	0
6.5I	0.00	0.00	0.00	3
6.5III	1.00	1.00	1.00	1
6.5V	1.00	0.33	0.50	3
7.0I	0.50	1.00	0.67	1
7.0III	1.00	0.86	0.92	7
7.0V	1.00	0.67	0.80	9
7.5I	0.86	0.75	0.80	8
7.5III	0.71	0.50	0.59	10
7.5V	0.17	0.50	0.25	2
8.0I	0.58	0.88	0.70	8
8.0III	0.00	0.00	0.00	0
8.0V	0.50	0.57	0.53	7
8.5I	0.00	0.00	0.00	5
8.5III	1.00	0.67	0.80	3
8.5V	0.75	1.00	0.86	3
9.0I	1.00	1.00	1.00	2
9.0III	1.00	1.00	1.00	1
9.0V	1.00	1.00	1.00	2
9.2I	0.50	1.00	0.67	1
9.2III	1.00	0.75	0.86	4
9.2V	0.00	0.00	0.00	2
9.5I	0.00	0.00	0.00	0
9.5III	0.50	0.29	0.36	7
9.5V	0.57	0.50	0.53	8
9.7I	0.00	0.00	0.00	0
9.7III	1.00	0.33	0.50	6
9.7V	0.00	0.00	0.00	1
micro avg	0.65	0.65	0.65	208
macro avg	0.55	0.55	0.52	208
weighted avg	0.68	0.65	0.64	208

(A)



(B)

FIGURA 5.33. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por SVM.

### 5.3.2.3 Decision Tree Classifier

O melhor resultado obtido utilizando o modelo DTC foi a partir da seguinte configuração:

- 30% da amostra sendo conjunto teste e 70% conjunto treino;



- $max\_depth = 7$ ;
- Critério de impureza: Entropia;
- *Cross validation* com 5 conjuntos distintos de grupos teste e treino.

Resultado final da porcentagem de acertos do modelo utilizando  $F1-Score$ :  $0.63 \pm 0.05$ .

Além disso, temos que  $\alpha = 0.018$ , como podemos ver na Figura 5.34, na qual a linha tracejada indica o valor ideal de  $\alpha$ .

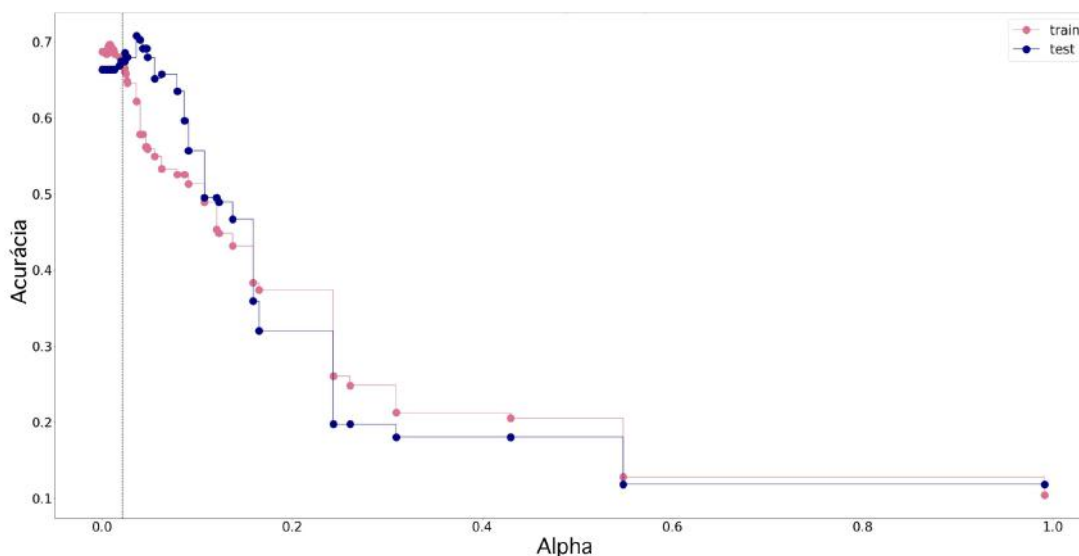


FIGURA 5.34. No eixo  $x$  temos diferentes valores de  $\alpha$  e no eixo  $y$  temos os valores de acurácia resultantes.

Na Figura 5.35 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes. O mesmo que comentamos para o caso do KNN pode ser observado no Relatório de Classificação e na Matriz de Confusão para o caso do DTC: houve uma melhora no resultado final, mas a falta de itens em determinadas classes ainda atrapalha o aprendizado para classificação.

Por fim, temos a árvore de decisão resultante deste problema, indicada na Figura 5.36. Vemos que esta também se divide no máximo 7 vezes, dado que definimos  $max\_depth=7$ .

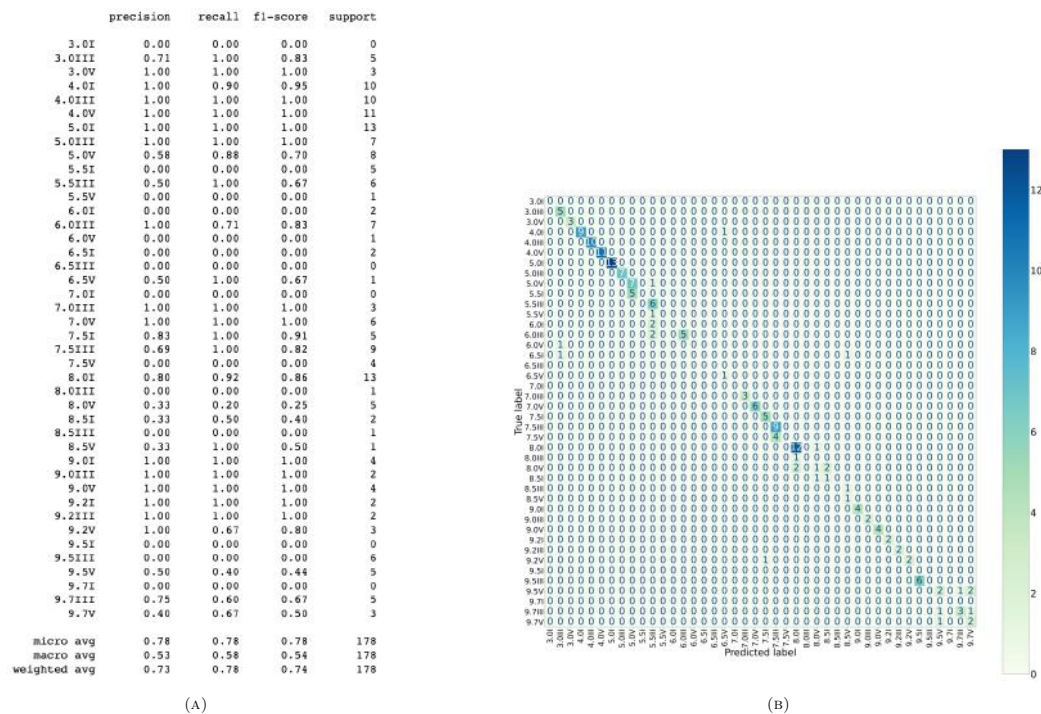


FIGURA 5.35. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por DTC.

### 5.3.2.4 Random Forest Classifier

No caso do modelo RFC, utilizamos 10 árvores iguais a descrita para o modelo DTC.

Resultado final da porcentagem de acertos do modelo utilizando *F1-Score*: 0.61 ± 0.19.

Na Figura 5.37 temos o Relatório de Classificação (A) e a Matriz de Confusão (B) resultantes.

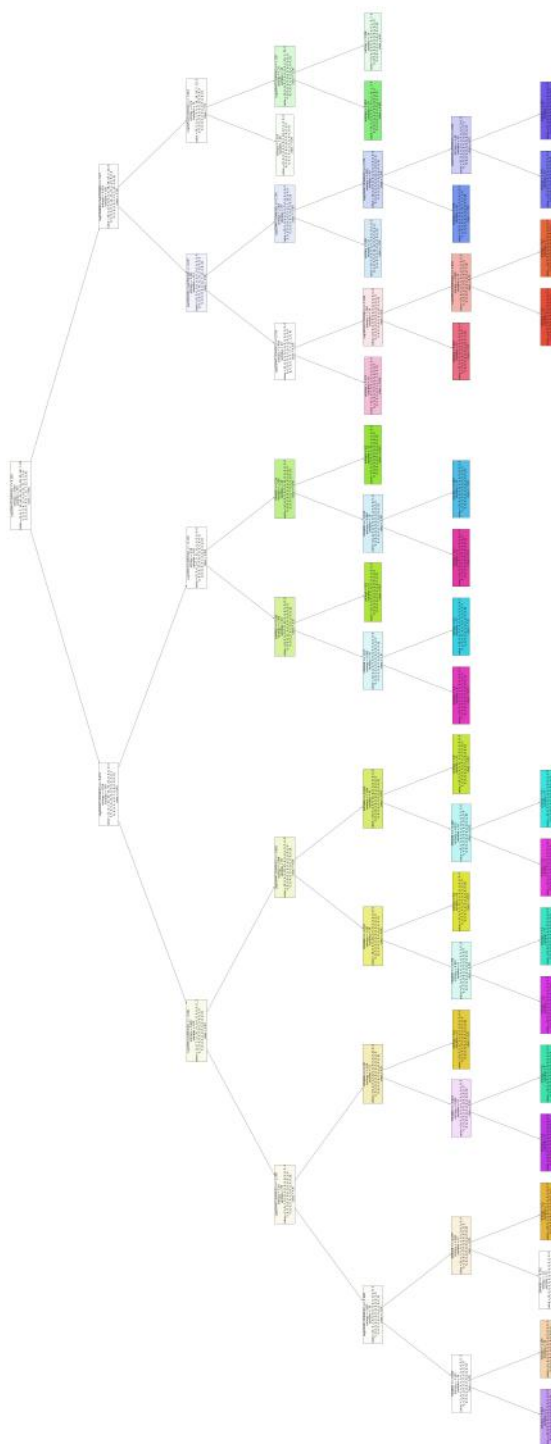


FIGURA 5.36. Essa figura é apenas para ilustrar o resultado da árvore de decisão para classificação de tipo espectral e de luminosidade, para espectros de alta e de baixa resolução. Uma versão de alta resolução, que permite a leitura de cada folha da árvore, está disponível eletronicamente clicando aqui. Em cada retângulo informamos o critério de separação utilizado para separar as folhas seguintes (caso haja separação), o valor de entropia, a quantidade de espectros contidos e a classificação da folha até aquele momento.

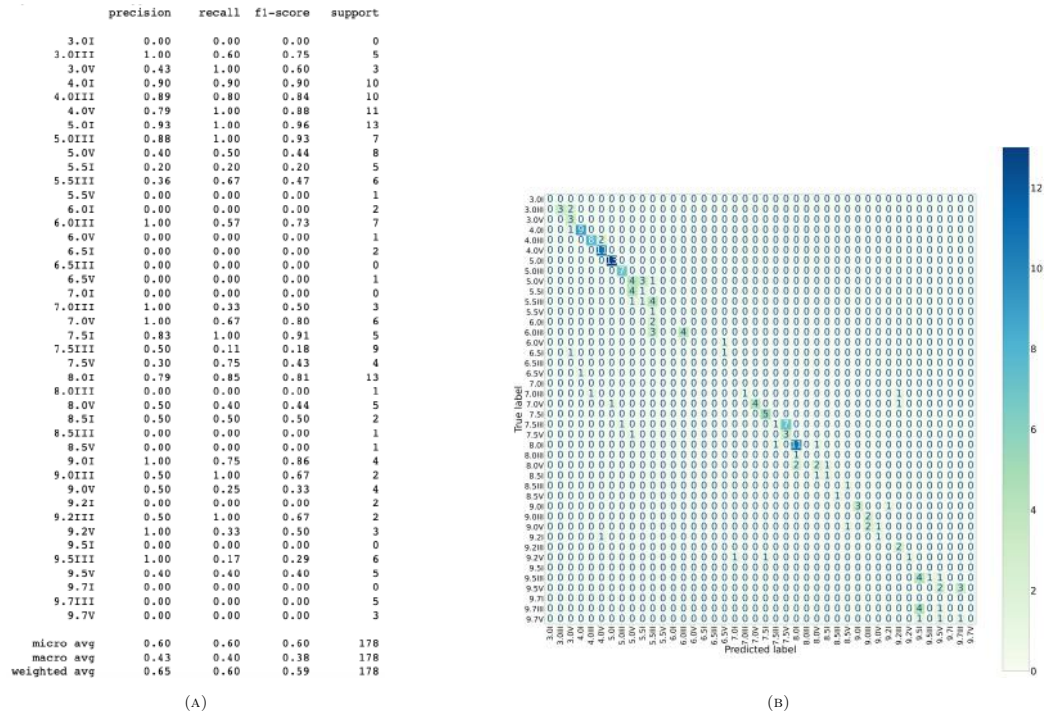


FIGURA 5.37. Relatório de Classificação (A) e Matriz de Confusão (B) para classificação de tipo espectral com espectros de alta e baixa resolução por RFC.

## 5.4 Resultados finais de ML

Abaixo, nas Tabelas 5.1 e 5.2, apresentamos as configurações utilizadas e os resultados de *F1-Score* obtidos pelos modelos de *Machine Learning* utilizando *Cross Validation*. Estes dados são apresentados para o conjunto de espectros de alta resolução e para o conjunto de espectros de alta e de baixa resolução, respectivamente.

É importante lembrar, neste momento, que um alto resultado de *F1-Score* também indica altos resultados de *Recall* e de *Precisão*, o que significa uma predominância de classificações corretas (Verdadeiros Positivos), para uma dada classe, tanto no total de espectros pertencentes a ela, quanto no total de previsões feitas para essa classe, respectivamente. Esses conceitos são elaborados em mais detalhes na subseção 4.1.7, do Capítulo 4.

Observamos que houve uma melhora expressiva nos valores de *F1-Score* ao utilizarmos uma amostra de 495 (classificação de tipo espectral) e 483 (classificação de tipo espectral em conjunto a de luminosidade) espectros. Isso é algo esperado quando falamos de *Machine Learning*, já que os modelos precisam ter espectros o suficiente no conjunto treino para que o aprendizado seja possível, e também no conjunto teste, para

---

que haja a validação da capacidade de classificação do modelo. Quando temos poucos dados por classe, o aprendizado não é feito de maneira correta, de forma que a capacidade de classificação não é satisfatória e os resultados de *F1-Score* traduzem isso com valores baixos. Com isso, podemos dizer que o aumento na quantidade de espectros foi o principal fator para o aumento dos resultados de *F1-Score*.

TABELA 5.1. Resultados dos modelos de ML para classificação dos espectros de alta resolução.

Conjunto de espectros de alta resolução				
-	Tipo Espectral		Tipo Espectral e Luminosidade	
	<i>F1-Score</i> <sup>a</sup>	Configuração	<i>F1-Score</i>	Configuração
KNN	(0.71 ± 0.14)	K=3	(0.48 ± 0.17)*	K=4
SVM	(0.73 ± 0.25)*	C=27	(0.52 ± 0.18)	C=10
DTC	(0.72 ± 0.21)	<i>max_depth</i> =5, $\alpha = 0,07$	(0.47 ± 0.22)	<i>max_depth</i> =10, $\alpha = 0,04$
RFC	(0.69 ± 0.19)	10 árvores	(0.45 ± 0.24)	10 árvores

TABELA 5.2. Resultados dos modelos de ML para classificação dos espectros de alta e baixa resolução.

Conjunto de espectros de alta e baixa resolução				
-	Tipo Espectral		Tipo Espectral e Luminosidade	
	<i>F1-Score</i> <sup>b</sup>	Configuração	<i>F1-Score</i>	Configuração
KNN	(0.82 ± 0.07)	K=3	(0.51 ± 0.06)	K=6
SVM	(0.86 ± 0.03)	C=27	(0.66 ± 0.08)	C=47
DTC	(0.87 ± 0.11)	<i>max_depth</i> =7, $\alpha = 0,006$	(0.63 ± 0.05)	<i>max_depth</i> =7, $\alpha = 0,018$
RFC	(0.87 ± 0.06)	10 árvores	(0.61 ± 0.19)	10 árvores

<sup>a</sup>Todos os resultados foram obtidos a partir de 5 conjuntos de *Cross Validation*, a não ser os marcados com asterisco que foram obtidos com 10.<sup>b</sup>Todos os resultados para tipo espectral foram obtidos a partir de 5 conjuntos de *Cross Validation*, a os de tipo espectral e luminosidade foram obtidos com 10.

# Capítulo 6

## Conclusão

A utilização de modelos de ML supervisionados na classificação espectral é um desafio, principalmente quando busca-se realizar a classificação anterior ao *Machine Learning* (pré-ML) de forma automatizada, como o proposto por este projeto. Diante disso, utilizamos diversos critérios presentes na literatura (p.e., Martins (2018) e Mathys (1988)) para atingirmos esse objetivo. No entanto, ainda não existem critérios o suficiente para separarmos todos os nossos espectros em classificação de luminosidade diretamente — como foi feito para a classificação de tipo espectral — de forma que precisamos realizar uma análise visual direta para alguns dos espectros de alta resolução, com tipos espectrais entre O3 e O7.5, no momento da determinação das classes de luminosidade (ver Subseção 3.2.2.1).

É verdade que poderíamos ter utilizado as classificações de tipo espectral e de luminosidade disponíveis na literatura, e esse projeto se basearia apenas na performance dos modelos de *Machine Learning* aplicados ao problema. No entanto, consideramos importante realizar a classificação espectral e, para isso, desenvolvemos um código para o cálculo das larguras equivalentes em *Python*, como explicado em detalhes na seção 3.1 do capítulo de Metodologia. Consequentemente, também realizamos o estudo acerca do comportamento das linhas de diagnóstico para classificação de temperatura e luminosidade, com o objetivo de aplicar esses resultados aos modelos — consideramos esses desenvolvimentos fundamentais para a formação de um astrônomo.

Com relação aos resultados obtidos pelos modelos de ML, notamos uma melhora ao adicionarmos uma nova amostra de espectros à amostra inicial de alta resolução, por

mais que tenham sido espectros de baixa resolução — ver seção 5.3. Como os modelos de ML utilizam os próprios dados para realizar o aprendizado acerca do problema sendo proposto, esperamos que quanto maior o número de dados, melhor sejam as suas performances. Isso ocorre porque a baixa quantidade de espectros em uma dada classe diminui a capacidade de correlação entre as *features*, além de atrapalhar o processo de aprendizado — o que é visto claramente nos baixos resultados de *F1-Score* quando realizamos a classificação de tipo espectral e de luminosidade para os espectros de alta resolução, explorado na Subseção 5.3.2, por possuírem diversas classes com apenas um representante — ver Histograma 5.4. Assim, podemos concluir que a quantidade foi o fator dominante na melhora nos resultados de *F1-Score*, por mais que tenham sido espectros de menor qualidade.

Além disso, notamos, a partir das Tabelas 5.1 e 5.2, que as incertezas também diminuem ao utilizarmos uma amostra maior, o que também é um reflexo da quantidade de dados. Quando utilizamos uma menor quantidade de espectros, temos uma variação muito grande das classes presentes nos conjuntos teste e treino cada vez que aplicamos o modelo ao problema, o que resulta em capacidades de aprendizado muito diferentes a cada iteração do nosso modelo (p.e., na amostra de espectros de alta resolução, temos apenas dois espectros de tipo espectral O3 e, com isso, podemos ter todos no conjunto teste, todos no conjunto treino, ou um em cada conjunto, em diferentes tentativas).

Assim, quando temos uma grande quantidade de espectros de cada classe na amostra, podemos ter conjuntos teste e treino mais equilibrados, e o aprendizado ocorre de forma similar em cada iteração do modelo ao problema, por mais que ainda possa haver uma variação na quantidade de itens por classe. Com isso, as performances serão parecidas em todas as tentativas utilizadas pelo *Cross Validation*, resultando em uma baixa incerteza.

É importante notar que, em um cenário ideal, teríamos milhares de espectros de altíssima resolução, e acreditamos que isso nos retornaria performances ainda melhores desses modelos. No entanto, isso está longe da realidade das estrelas de tipo O, que são objetos mais raros de serem observados na natureza, e, por isso, temos poucos espectros disponíveis para análise — principalmente quando buscamos espectros de alta resolução.

Todos os modelos utilizados tiveram bons resultados de *F1-Score* na classificação de tipo espectral em ambos os conjuntos amostrais — ver Tabelas 5.1 e 5.2. No entanto,



a classificação de tipo espectral junto à de luminosidade retornaram resultados baixos em ambos os casos. Isso está relacionado à grande quantidade de classes disponíveis para classificação (36 classes para alta resolução, e 42 classes para alta e baixa resolução), além da baixa quantidade de espectros por classe em alguns casos, como mostrado nos Histogramas 5.4 e 5.6.

Caso o leitor desejasse realizar esse mesmo processo de classificação com seus próprios dados, a partir da análise feita nesse projeto, recomendamos dois modelos a serem utilizados: o *Support Vector Machines* (Subseção 4.2.2) e o *Random Forest Classifier* (Subseção 4.2.4). De modo geral, a classificação feita pelo modelo do SVM obteve resultados melhores (ver Tabelas 5.1 e 5.2), mas a classificação por RFC possui dois diferenciais importantes: a capacidade de observarmos como a separação dos dados foi feita (por utilizar a mesma estrutura das árvores de decisão) e também pela sua capacidade de evitar o *overfitting*, como também foi explicado na Subseção 4.2.4, do Capítulo 4.

Além disso, podemos observar que os resultados obtidos pelos modelos, de modo geral, são próximos, o que nos traz segurança com relação aos resultados obtidos. As diferenças nos seus resultados finais são uma consequência do seu modo de operação. No entanto, podemos dizer que o KNN (Subseção 4.2.1) tem uma performance inferior aos outros ao realizar a classificação de tipo espectral e de luminosidade para o conjunto de espectros de alta e baixa resolução, o que acreditamos ser devido a sua sensibilidade a quantidade de classes presentes, por utilizar a distância euclidiana entre os vizinhos mais próximos para obter a classificação final.

Podemos concluir com esse projeto que, os modelos obtidos aqui, por mais que tenham obtido resultados satisfatórios, ainda não nos oferecem tanta segurança e precisão quanto a análise visual humana, principalmente devido ao tamanho amostral. Porém, é uma excelente ferramenta para uma primeira análise, na qual somos capazes de separar os objetos com mais de 80% de certeza em classes de tipos espectrais (ver Tabelas 5.1 e 5.2) e, posteriormente, ter uma análise visual para garantir a acurácia da classificação.

Futuramente, obtendo um maior conjunto de espectros de alta resolução, podemos retornar a este projeto e realizar uma nova classificação dos objetos, de forma a analisar a performance desses modelos para conjuntos de dados ainda maiores — de preferência, em alta resolução.

# Referências Bibliográficas

Abbott B. P., et al., 2016, Physical review letters, 116, 131103

Apellániz J. M., et al., 2016, The Astrophysical Journal Supplement Series, 224, 4

Barbá R., Gamen R., Arias J., Morrell N., Apellániz J. M., Alfaro E., Walborn N., Sota A., 2010, Revista Mexicana de Astronomia y Astrofisica, 38, 30

Baron D., 2019, arXiv preprint arXiv:1904.07248

Bouret J.-C., Hillier D., Lanz T., Fullerton A., 2012, Astronomy & Astrophysics, 544, A67

Carroll B. W., Ostlie D. A., 2017, An introduction to modern astrophysics. Cambridge University Press

Celebi M. E., Aydin K., 2016, Unsupervised learning algorithms. Springer

Conti P. S., Alschuler W. R., 1971, The Astrophysical Journal, 170, 325

Conti P., Frost S., 1977, The Astrophysical Journal, 212, 728

Conti P., Leep E., 1974, The Astrophysical Journal, 193, 113

Ekström S., et al., 2012, Astronomy & Astrophysics, 537, A146

Fisher R. A., 1936, Annals of eugenics, 7, 179

Géron A., 2019, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. "O'Reilly Media, Inc."

Gray R. O., Corbally C. J., 2021, in , Stellar Spectral Classification. Princeton university press

- Howarth I. D., Siebert K. W., Hussain G. A., Prinja R. K., 1997, *Monthly Notices of the Royal Astronomical Society*, 284, 265
- Irrgang A., Przybilla N., Heber U., Böck M., Hanke M., Nieva M.-F., Butler K., 2014, *Astronomy & Astrophysics*, 565, A63
- Kuntzer T., Tewes M., Courbin F., 2016, *Astronomy & Astrophysics*, 591, A54
- Kyritsis E., Maravelias G., Zezas A., Bonfini P., Kovelakas K., Reig P., 2022, *Astronomy and Astrophysics*, 657
- Lamers H. J., Cassinelli J. P., Cassinelli J., 1999, *Introduction to stellar winds*. Cambridge university press
- Maeder A., Meynet G., Ekström S., Hirschi R., Georgy C., 2007, *Proceedings of the International Astronomical Union*, 3, 3
- Marcolino W., Bouret J.-C., Rocha-Pinto H., Bernini-Peron M., Vink J., 2022, *Monthly Notices of the Royal Astronomical Society*, 511, 5104
- Martins F., 2018, *AA*, 616
- Martins F., Palacios A., 2013, *Astronomy & Astrophysics*, 560, A16
- Mathys G., 1988, *Astronomy and Astrophysics Supplement Series*, 76, 427
- Maury A. C., Pickering E. C., 1897, *Annals of Harvard College Observatory*, 28, 1
- Morgan W. W., Keenan P. C., 1973, *ARA&A*, 11, 29
- Morgan W. W., Keenan P. C., Kellman E., 1943, *An atlas of stellar spectra, with an outline of spectral classification*
- Moultaka J., Ilovaisky S., Prugniel P., Soubiran C., 2004a, *SF2A-2004: Semaine de l'Astrophysique Francaise*, p. 547
- Moultaka J., Ilovaisky S., Prugniel P., Soubiran C., 2004b, *Publications of the Astronomical Society of the Pacific*, 116, 693
- Payne C. H., 1925
- Pellerin A., Maíz Apellániz J., Simón-Díaz S., Barbá R. H., 2012, in *American Astronomical Society Meeting Abstracts #219*. p. 224.03

- 
- Pesenson M. Z., Pesenson I. Z., McCollum B., 2010, *Advances in Astronomy*, 2010, 1
- Petit P., Louge T., Théado S., Paletou F., Manset N., Morin J., Marsden S., Jeffers S., 2014, *Publications of the Astronomical Society of the Pacific*, 126, 469
- Rauw G., et al., 2018, *Astronomy & Astrophysics*, 615, A44
- Simón-Díaz S., Castro N., Garcia M., Herrero A., Markova N., 2011, *Bulletin de la Societe Royale des Sciences de Liege*, 80, 514
- Sota A., Apellániz J. M., Morrell N. I., Barba R. H., Walborn N. R., Gamen R. C., Arias J. I., Alfaro E. J., 2014, *The Astrophysical Journal Supplement Series*, 211, 10
- Walborn N. R., 1971, *The Astrophysical Journal Supplement Series*, 23, 257
- Walborn N. R., 1972, *The Astronomical Journal*, 77, 312
- Walborn N., 1973, *The Astronomical Journal*, 78, 1067

## Apêndice A

# Classificação de tipo espectral e de luminosidade

Neste apêndice serão apresentados as classificações de tipo espectral e de luminosidade, além dos valores de largura equivalente para as linhas.

TABELA A.1. Lista de estrelas com as respectivas classificações e larguras equivalentes para os espectros de alta resolução.

<b>Estrela</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 64568	3	V	0.033	0.726	0.502	-0.004	-0.021	0.008	0.874
HD 319718	3	III	0.064	0.614	0.396	0.099	0.079	0.035	0.086
HD 46223	4	V	0.156	0.7	0.485	0.005	0.035	0.023	0.705
HD 164794	4	V	0.142	0.741	0.56	0.031	0.029	0.004	0.814
HD 93205	4	V	0.084	0.608	0.427	0.028	0.04	0.024	0.479
HD 93250	4	V	0.118	0.696	0.513	-0.022	0.018	-0.013	0.462
HD 96715	4	V	0.16	0.785	0.563	-0.016	0.048	-0.009	0.914
HD 66811	4	I	0.177	0.709	0.611	0.014	0.026	-0.011	-1.265
HD 93129	4	I	0.102	0.595	0.337	-0.014	0.054	0.033	-0.887
HD 15570	4	I	0.128	0.558	0.347	-0.016	0.004	0.035	-1.51
HD 93128	4	I	0.092	0.575	0.342	0.009	0.036	0.002	-0.975
HD 153919	5	I	0.232	0.664	0.446	-0.02	0.017	0.01	-2.912
CPD -472963	5	I	0.228	0.666	0.522	-0.003	0.09	0.013	-0.286
HD 15558	5	III	0.187	0.569	0.27	-0.035	0.017	0.008	0.195
HD 319699	5	V	0.24	0.676	0.497	0.003	0.101	0.02	0.485
HD 16691	5	I	0.187	0.616	0.412	0.016	0.004	-0.019	-3.273
Continua na próxima página									

**Tabela A.1— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 5005	5.5	V	0.326	0.675	0.347	-0.055	0.227	0.043	0.66
HD 46150	5.5	V	0.304	0.634	0.501	0.051	0.12	0.036	0.846
CPD -592673	5.5	V	0.318	0.686	0.6	0.021	0.082	-0.007	0.798
BD -164826	5.5	V	0.283	0.705	0.513	0.049	0.125	0.031	0.781
HD 168112	5.5	III	0.286	0.686	0.597	-0.017	0.06	0.007	0.412
HD 192281	5.5	V	0.267	0.75	0.599	-0.005	0.036	-0.001	0.622
ALS 2063	5.5	I	0.171	0.437	0.081	-0.175	0.005	-0.008	-2.85
HD 93843	6	III	0.317	0.629	0.52	0.05	0.075	-0.014	0.181
HD 93204	6	V	0.378	0.727	0.637	0.01	0.043	0.01	0.622
HD e229196	6	III	0.32	0.553	0.519	0.061	0.016	-0.0	0.079
HD 303311	6	V	0.423	0.759	0.625	0.074	0.142	-0.002	0.96
CPD -592600	6	V	0.347	0.557	0.49	0.052	0.133	0.006	0.608
HD 152233	6	III	0.347	0.635	0.496	0.061	0.096	0.009	-0.017
HD 12993	6.5	V	0.63	0.807	0.651	0.106	0.21	0.028	0.938
HD 190864	6.5	III	0.505	0.685	0.612	0.067	0.141	0.009	0.491
HD 199579	6.5	V	0.462	0.668	0.531	0.088	0.121	0.016	0.825
HD e242926	6.5	V	0.493	0.635	0.435	0.11	0.184	0.039	0.899
Continua na próxima página									

**Tabela A.1— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 42088	6.5	V	0.5	0.704	0.579	0.058	0.159	0.016	1.03
HD 64315	6.5	V	0.437	0.633	0.516	0.13	0.156	0.021	0.81
HD 17505	6.5	III	0.404	0.573	0.367	-0.029	0.049	0.022	0.51
HD 164492	7	V	0.646	0.744	0.598	0.134	0.259	0.033	1.013
HD 36879	7	V	0.59	0.665	0.489	0.126	0.178	0.029	0.685
HD 228841	7	V	0.465	0.567	0.492	0.028	0.152	0.022	0.773
HD 155806	7	III	0.479	0.537	0.404	0.134	0.169	0.032	0.593
HD 47839	7	V	0.532	0.544	0.396	0.055	0.231	0.023	0.706
HD 305525	7	V	0.496	0.615	0.501	0.116	0.128	-0.01	0.831
HD 36861	7.5	V	0.709	0.589	0.459	0.149	0.219	0.042	0.717
HD 192639	7.5	I	0.64	0.533	0.443	0.099	0.161	-0.014	-0.069
HD 193514	7.5	I	0.589	0.525	0.43	0.086	0.148	0.037	-0.068
HD 156154	7.5	I	0.701	0.624	0.505	0.108	0.191	0.022	0.154
HD 34656	7.5	I	0.705	0.653	0.557	0.113	0.23	-0.002	0.37
HD 186980	7.5	III	0.638	0.606	0.462	0.109	0.242	0.03	0.542
HD 24912	7.5	I	0.69	0.595	0.512	0.101	0.176	0.009	0.4
HD 13268	8	III	0.633	0.471	0.4	0.083	0.244	0.043	0.591
Continua na próxima página									



**Tabela A.1— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 207898	8	V	0.77	0.435	0.252	0.254	0.397	0.07	0.732
HD 193322	8	III	0.692	0.422	0.305	0.184	0.296	0.063	0.668
HD 66788	8	V	0.772	0.414	0.355	0.204	0.34	0.042	0.674
HD 35619	8	III	0.648	0.494	0.368	0.071	0.274	0.015	0.742
HD 162978	8	I	0.698	0.55	0.446	0.135	0.215	0.027	0.399
HD 48279	8	V	0.754	0.566	0.446	0.23	0.454	0.049	0.946
HD 151804	8	-	0.754	0.514	0.303	0.071	0.123	0.025	-0.441
HD 188001	8	-	0.781	0.558	0.392	0.058	0.218	0.03	-0.085
HD 17603	8	-	0.686	0.505	0.381	0.064	0.166	0.005	-0.074
HD 14633	8.5	V	0.848	0.387	0.473	0.278	0.383	0.051	0.783
HD 46966	8.5	III	0.693	0.475	0.436	0.241	0.325	0.091	0.683
HD 216532	8.5	V	0.689	0.415	0.336	0.232	0.363	0.066	0.725
HD 218195	8.5	III	0.852	0.448	0.441	0.231	0.343	0.042	0.652
HD 214680	8.5	V	0.832	0.462	0.416	0.245	0.4	0.082	0.862
HD 46149	8.5	V	0.753	0.466	0.398	0.223	0.397	0.061	0.902
HD 191978	8.5	V	0.724	0.454	0.405	0.178	0.33	0.043	0.809
HD 225160	8.5	I	0.788	0.446	0.502	0.105	0.213	0.006	0.052
Continua na próxima página									

**Tabela A.1— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 203064	8.5	I	0.575	0.449	0.4	0.095	0.162	0.017	0.404
HD 155889	9	V	0.797	0.334	0.312	0.347	0.384	0.077	0.647
HD 152249	9	I	0.794	0.311	0.286	0.154	0.244	0.099	0.23
HD 167263	9	III	0.771	0.273	0.241	0.212	0.272	0.094	0.491
HD 34078	9	V	0.916	0.407	0.366	0.31	0.473	0.146	0.765
HD 46202	9	V	0.905	0.435	0.348	0.332	0.469	0.102	0.75
HD 207198	9	III	0.842	0.346	0.38	0.222	0.34	0.064	0.454
BD -60499	9	V	0.928	0.492	0.374	0.39	0.471	0.122	0.866
HD 152247	9	III	0.84	0.328	0.291	0.243	0.331	0.112	0.451
HD 24431	9	III	0.754	0.333	0.308	0.236	0.334	0.086	0.554
HD 154368	9	I	0.88	0.333	0.297	0.216	0.323	0.134	0.268
HD 16429	9	III	0.8	0.326	0.27	0.158	0.311	0.109	0.378
HD 193443	9	III	0.717	0.346	0.285	0.243	0.368	0.127	0.538
HD 201345	9	V	0.904	0.492	0.405	0.4	0.447	0.165	0.801
HD 209481	9	III	0.76	0.291	0.273	0.198	0.286	0.084	0.463
HD 153426	9	III	0.746	0.351	0.339	0.18	0.256	0.048	0.53
HD 218915	9	I	0.802	0.302	0.304	0.198	0.296	0.119	0.237
Continua na próxima página									

**Tabela A.1— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 210809	9	I	0.862	0.356	0.352	0.212	0.301	0.086	0.119
HD 202124	9	I	0.788	0.283	0.287	0.159	0.232	0.095	0.193
HD 38666	9	V	0.895	0.381	0.314	0.341	0.435	0.132	0.743
HD 258691	9.2	V	0.927	0.409	0.308	0.396	0.48	0.116	0.67
HD 36483	9.2	III	0.68	0.239	0.185	0.214	0.257	0.055	0.506
HD 206183	9.5	V	0.933	0.365	0.288	0.335	0.511	0.112	0.753
HD 328856	9.5	I	0.916	0.277	0.251	0.231	0.336	0.205	0.328
HD 188209	9.5	I	0.882	0.224	0.295	0.228	0.368	0.14	0.288
HD 30614	9.5	I	0.808	0.255	0.261	0.173	0.296	0.115	0.108
HD 192001	9.5	V	0.857	0.24	0.281	0.304	0.432	0.111	0.637
HD 149757	9.5	III	0.578	0.221	0.152	0.161	0.295	0.085	0.46
HD 15137	9.5	I	0.752	0.24	0.138	0.165	0.317	0.12	0.346
HD 209975	9.5	I	0.798	0.261	0.314	0.174	0.322	0.108	0.33
HD 13745	9.5	III	0.882	0.235	0.27	0.207	0.41	0.154	0.37
HD 149038	9.7	I	0.834	0.192	0.194	0.24	0.345	0.207	0.296
HD 189957	9.7	V	0.844	0.223	0.16	0.233	0.408	0.159	0.568
HD 225146	9.7	III	0.847	0.077	0.21	0.271	0.4	0.21	0.38
Continua na próxima página									

**Tabela A.1— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 15642	9.7	I	0.703	0.11	0.134	0.239	0.23	0.084	0.369
HD 36486	9.7	III	0.881	0.205	0.167	0.187	0.348	0.149	0.341
HD 36512	9.7	V	0.915	0.308	0.116	0.299	0.544	0.141	0.617
HD 101131	9.7	V	0.874	0.255	0.249	0.32	0.499	0.156	0.499
HD 195592	9.7	I	0.838	0.132	0.213	0.188	0.363	0.187	0.037
HD 55879	9.7	V	0.882	0.292	0.274	0.352	0.449	0.165	0.572
HD 46106	9.7	III	0.917	0.177	0.105	0.371	0.45	0.103	0.372
HD 167264	9.7	I	0.922	0.162	0.164	0.299	0.387	0.267	0.267
HD 154643	9.7	III	0.887	0.2	0.196	0.319	0.38	0.162	0.504
HD 37468	9.7	V	0.853	0.273	0.187	0.401	0.53	0.131	0.492
HD 207538	9.7	V	0.873	0.201	0.221	0.4	0.485	0.146	0.528
HD 47432	9.7	III	0.871	0.108	0.217	0.301	0.404	0.179	0.309

TABELA A.2. Lista de estrelas com as respectivas classificações e larguras equivalentes para os espectros de baixa resolução.

Estrela	TE	CL	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
SS 215	3	I	0.06	0.577	0.151	-0.152	-0.014	0.035	-2.87
HD 93 129 AaAb	3	I	0.045	0.685	0.341	0.009	0.007	-0.004	-1.051
HD 190 429 A	3	I	0.074	0.885	0.51	0.025	0.03	0.014	-3.582
BD +43 3654	3	I	0.068	0.743	0.4	-0.006	0.049	0.019	-2.196
BD +50 886	4	V	0.1	0.753	0.512	-0.016	0.089	0.017	0.756
NGC 3603 HST-24	4	V	0.145	0.753	0.493	-0.026	-0.007	0.019	0.613
ALS 4067 AB	4	I	0.071	0.394	0.118	0.018	0.019	-0.017	-2.722
Pismis 24-1 AB	4	III	0.108	0.763	0.332	-0.002	0.032	0.026	-0.048
LS III +46 11	4	I	0.103	0.693	0.367	-0.056	0.067	-0.003	-0.202
HD 168 076 AB	4	III	0.171	0.729	0.427	0.059	0.085	0.035	0.468
9 Sgr AB	4	V	0.153	0.794	0.555	0.007	-0.013	0.012	0.824
HD 193 682	4	III	0.222	0.955	0.584	-0.009	0.056	0.025	0.509
ALS 19 618	4	V	0.152	0.766	0.474	-0.034	0.063	-0.007	0.608
HD 150 136 AaAb	4	III	0.13	0.615	0.421	0.018	0.03	0.002	0.284
HD 5005 A	4	V	0.156	0.848	0.612	-0.009	0.054	0.018	0.785
zeta Pup	4	I	0.189	0.767	0.56	-0.009	0.023	-0.004	-1.609
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 15 629	4	III	0.184	0.799	0.52	-0.027	0.083	0.03	0.518
HDE 229 232 AB	4	V	0.197	0.932	0.577	0.034	0.054	0.015	0.601
Pismis 24-2	5	V	0.301	0.869	0.678	0.023	0.092	0.007	0.788
Cyg OB2-8 C	5	I	0.184	0.668	0.416	-0.012	0.049	-0.005	-0.287
ALS 12 050	5	V	0.253	0.718	0.429	0.033	0.09	0.011	0.55
ALS 17 591	5	III	0.229	0.904	0.653	0.024	0.106	-0.023	0.166
ALS 4962	5	III	0.207	0.704	0.564	-0.008	0.09	-0.019	-0.158
HD 120 678	5	V	0.149	0.535	0.349	0.344	0.359	0.143	0.652
ALS 18 551	5	V	0.173	0.641	0.309	-0.019	0.025	0.018	0.679
HD 152 386	5	I	0.084	0.261	0.11	0.123	0.063	-0.027	-5.076
HD 155 913	5	V	0.246	0.893	0.599	0.029	0.071	0.014	0.721
ALS 15 210	5	I	0.196	0.674	0.47	0.008	0.02	0.015	-0.201
HDE 228 854	5	III	0.172	0.492	0.206	0.022	0.11	0.073	0.223
Cyg OB2-9	5	I	0.19	0.62	0.457	0.003	0.061	-0.003	-1.462
HD 93 632	5	I	0.216	0.618	0.43	0.055	0.042	0.002	-0.511
LS III +46 12	5	V	0.221	0.744	0.518	-0.003	0.072	-0.006	0.452
ALS 11 244	5	III	0.227	0.758	0.443	0.035	0.077	0.007	-0.171
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 93 162	5	I	0.082	0.264	0.233	0.032	0.022	-0.053	-6.067
HDE 303 308 AB	5	V	0.225	0.671	0.417	0.043	0.062	0.015	0.551
ALS 19 265	5	V	0.277	1.072	0.857	-0.034	0.085	0.06	1.182
HDE 242 908	5	V	0.208	0.762	0.57	-0.023	0.048	0.006	0.857
ARV2008 206	5.5	V	0.37	0.754	0.615	0.048	0.076	0.035	0.705
ALS 19 692	5.5	V	0.331	0.721	0.288	0.117	0.224	0.043	0.571
HDE 256 725 A	5.5	V	0.32	0.851	0.639	0.027	0.09	0.013	0.874
HDE 319 699	5.5	V	0.255	0.703	0.447	-0.025	0.094	-0.007	0.585
BD -14 5040	5.5	V	0.354	0.755	0.588	0.022	0.103	0.004	0.801
BD +45 3216 A	5.5	V	0.358	0.985	0.78	0.004	0.066	0.023	1.046
DH Cep	5.5	V	0.28	0.567	0.423	0.007	0.107	0.03	0.563
NGC 3603 MTT 25	5.5	V	0.245	0.669	0.564	-0.042	0.096	-0.011	0.462
HDE 303 311	5.5	V	0.397	0.856	0.708	0.053	0.164	0.022	1.002
Cyg OB2-8 A	5.5	I	0.281	0.651	0.52	0.005	0.073	-0.0	-0.358
V441 Pup	5.5	V	0.173	0.416	0.38	0.169	0.108	-0.031	0.532
HD 97 253 AaAb	5.5	III	0.288	0.674	0.45	0.057	0.06	0.008	-0.051
Cyg OB2-22 A	5.5	III	0.332	0.666	0.431	-0.123	0.251	-0.047	-0.109
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
Cyg OB2-11	5.5	I	0.247	0.621	0.404	0.031	0.112	0.007	-1.361
HD 169 582	5.5	I	0.351	0.851	0.575	0.085	0.117	-0.006	-0.678
V747 Cep	5.5	V	0.33	0.712	0.511	0.025	0.154	0.012	0.655
HD 39 680	5.5	V	0.193	0.483	0.335	0.19	0.203	0.012	0.595
ALS 18 769	5.5	I	0.313	0.779	0.527	0.068	0.11	0.099	-0.257
HD 14 442	5.5	I	0.345	0.799	0.59	0.02	0.063	0.022	-0.308
BD +60 134	5.5	V	0.443	1.015	0.673	0.072	0.231	0.022	0.913
V662 Car	5.5	V	0.316	0.74	0.445	0.018	0.06	0.083	0.729
2MASS J10583238-6110565	5.5	V	0.263	0.652	0.375	0.047	0.098	0.016	0.62
HD 93 403	5.5	III	0.334	0.669	0.429	0.057	0.096	0.019	0.001
Cyg OB2-B17	5.5	I	0.298	0.672	0.406	0.096	0.086	-0.08	-2.958
HD 101 190 AaAb	5.5	V	0.337	0.714	0.496	0.082	0.124	0.026	0.534
NGC 3603 HST-51	5.5	III	0.234	0.609	0.529	0.085	0.138	-0.015	0.364
HD 14 947	5.5	I	0.321	0.742	0.472	0.003	0.034	0.004	-2.159
BD +60 2635	5.5	V	0.448	0.926	0.65	0.062	0.189	0.031	0.924
ALS 18 748	5.5	III	0.234	0.564	0.395	-0.09	0.111	0.016	0.099
BD -16 4826	6	V	0.391	0.748	0.56	-0.014	0.189	0.014	0.861
Continua na próxima página									



**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
ALS 3386	6	I	0.313	0.552	0.48	0.022	0.096	-0.019	-0.808
HD 99 897	6	V	0.453	0.742	0.491	0.059	0.121	0.015	0.592
Pismis 24-13	6	V	0.442	0.835	0.609	0.058	0.241	0.059	0.859
CPD -58 2611	6	V	0.44	0.827	0.619	0.077	0.194	0.012	0.91
HD 96 946	6	V	0.472	0.792	0.544	0.114	0.16	0.051	0.415
HD 157 857	6	V	0.537	0.861	0.607	0.09	0.145	-0.005	0.325
HD 150 958 AB	6	I	0.432	0.723	0.481	0.061	0.063	-0.007	-1.687
HD 92 206 B	6	V	0.489	0.831	0.46	-0.046	0.075	0.052	0.825
CPD -59 2641	6	V	0.4	0.697	0.508	0.031	0.127	0.02	0.474
HD 165 052	6	V	0.544	0.878	0.564	0.055	0.156	0.045	0.959
HD 148 937 AaAb	6	I	0.342	0.674	0.396	0.053	0.097	0.009	-0.735
HD 130 298	6	V	0.465	0.912	0.602	0.083	0.1	0.026	0.322
HD 124 314 AaAb	6	V	0.478	0.771	0.371	0.042	0.099	0.063	0.371
HDE 322 417	6	V	0.437	0.729	0.427	0.014	0.141	0.003	0.524
HDE 305 532	6	V	0.471	0.776	0.534	0.085	0.187	0.021	0.923
V479 Sct	6	V	0.536	1.007	0.662	0.12	0.062	0.019	1.056
HD 76 556	6	III	0.396	0.685	0.469	0.036	0.088	0.025	0.221
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
ALS 12 688	6	V	0.322	0.63	0.465	-0.009	0.11	-0.028	0.499
BD +62 424	6	V	0.419	0.788	0.466	0.02	0.092	0.042	0.743
HD 14 434	6	V	0.488	0.833	0.518	0.049	0.101	0.023	0.412
ALS 4880	6	V	0.435	0.825	0.53	0.082	0.175	0.033	0.823
ALS 18 747	6	I	0.269	0.516	0.418	-0.02	0.055	-0.028	-1.403
ALS 4626	6	V	0.591	0.996	0.845	0.079	0.179	0.006	1.151
HD 92 206 A	6	V	0.462	0.79	0.63	0.013	0.12	0.021	0.813
HD 191 612	6	I	0.371	0.669	0.437	0.117	0.132	0.003	-0.347
Cyg OB2-8 B	6	III	0.409	0.65	0.433	0.013	0.09	-0.015	0.002
HD 172 175	6	I	0.442	0.862	0.527	0.054	0.123	0.023	-0.429
HDE 305 525	6	V	0.401	0.742	0.538	-0.006	0.081	0.006	0.833
ALS 19 693	6	V	0.357	0.688	0.54	0.086	0.099	-0.043	0.664
ALS 15 360	6	V	0.544	0.891	0.467	0.065	0.234	0.083	0.917
CPD -28 2561	6	I	0.46	0.741	0.494	0.114	0.205	0.036	-1.629
HDE 227 018	6	V	0.465	0.745	0.553	0.069	0.151	0.036	0.831
HDE 242 935 A	6	V	0.495	0.83	0.626	0.092	0.192	0.017	0.973
NGC 3603 HST-48	6	I	0.217	0.364	0.352	-0.014	0.002	-0.032	-2.549
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 167 633	6	V	0.549	0.886	0.649	0.072	0.15	0.016	0.741
HD 101 298	6	V	0.417	0.758	0.468	0.037	0.11	0.023	0.539
HDE 227 245	6.5	V	0.532	0.76	0.506	0.084	0.195	0.016	0.872
HDE 338 931	6.5	III	0.498	0.788	0.546	0.044	0.126	0.01	0.251
HD 163 758	6.5	I	0.571	0.843	0.54	0.069	0.111	0.01	-0.883
ALS 12 370	6.5	V	0.5	0.783	0.515	0.042	0.177	0.042	0.82
ALS 6351	6.5	V	0.65	0.893	0.707	0.191	0.314	0.014	1.035
HDE 344 784 A	6.5	V	0.494	0.727	0.53	0.133	0.249	0.035	0.709
HDE 305 524	6.5	V	0.532	0.715	0.548	0.07	0.164	0.018	0.793
HD 206 267 AaAb	6.5	III	0.459	0.611	0.423	0.026	0.106	0.018	0.372
BD -10 4682	6.5	V	0.54	0.783	0.426	0.033	0.195	0.06	0.713
BD +60 501	6.5	V	0.568	0.753	0.512	0.087	0.21	0.032	0.927
ALS 12 320	6.5	III	0.505	0.687	0.543	0.082	0.213	0.022	0.541
HD 91 824	6.5	V	0.564	0.828	0.543	0.166	0.236	0.048	0.868
HD 90 273	6.5	V	0.651	0.919	0.707	0.132	0.275	0.015	0.854
HDE 303 316 A	6.5	V	0.528	0.761	0.632	0.051	0.13	0.02	0.886
HD 194 649 AB	6.5	V	0.481	0.685	0.474	0.054	0.149	0.034	0.634
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
ALS 12 619	6.5	V	0.611	0.81	0.413	0.171	0.277	0.034	1.064
Tyc 7370-00460-1	6.5	V	0.408	0.619	0.436	-0.006	0.141	0.059	0.638
HD 152 723 AaAb	6.5	III	0.474	0.673	0.419	0.06	0.154	0.041	0.381
HD 101 436	6.5	V	0.51	0.756	0.512	0.055	0.167	0.023	0.725
HD 150 135 AaAb	6.5	V	0.534	0.785	0.584	0.051	0.152	0.013	0.948
HD 54 662 AB	6.5	V	0.524	0.811	0.615	0.075	0.181	0.017	0.937
BD +55 2722 C	6.5	V	0.539	0.703	0.503	0.084	0.216	0.006	0.748
ALS 18 553	6.5	I	0.531	0.763	0.458	-0.085	0.068	0.042	0.116
2MASS J20315961+4114504	6.5	V	0.587	0.745	0.455	0.093	0.402	0.083	0.903
HD 168 075	6.5	V	0.498	0.678	0.44	0.125	0.208	0.063	0.711
Cyg OB2-A24	6.5	III	0.611	0.776	0.522	0.106	0.438	0.022	0.466
ALS 207	6.5	V	0.512	0.795	0.582	0.072	0.143	-0.018	0.914
HD 93 161 B	6.5	III	0.424	0.609	0.518	0.021	0.051	-0.015	0.41
HD 48 099	6.5	V	0.456	0.691	0.478	0.111	0.163	0.028	0.791
HD 93 146 A	6.5	V	0.552	0.698	0.513	0.062	0.19	0.016	0.749
HD 91 572	6.5	V	0.481	0.696	0.48	0.105	0.207	-0.006	0.752
HDE 319 703 BaBb	6.5	V	0.473	0.722	0.542	0.108	0.169	0.049	0.754
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
V467 Vel	6.5	III	0.484	0.654	0.553	0.033	0.149	0.016	0.551
HD 217 086	6.5	V	0.502	0.646	0.46	0.029	0.15	0.02	0.869
CPD -47 2962	6.5	V	0.624	0.888	0.5	0.036	0.248	0.066	0.85
CPD -58 2620	6.5	V	0.584	0.777	0.542	0.104	0.243	0.021	0.972
CPD -26 2716	6.5	I	0.555	0.755	0.464	0.08	0.119	0.027	-0.785
Tyc 0737-01170-1	6.5	V	0.576	0.804	0.594	0.159	0.249	0.026	1.012
HDE 228 766	6.5	I	0.415	0.559	0.352	0.093	0.098	-0.043	-2.64
ALS 458	6.5	V	0.527	0.764	0.651	0.074	0.18	0.004	0.852
HD 193 595	6.5	V	0.655	0.901	0.575	0.125	0.204	0.032	0.906
BD +61 411 A	6.5	V	0.466	0.724	0.576	0.122	0.261	0.01	0.934
lambda Cep	6.5	I	0.462	0.704	0.514	0.039	0.073	-0.007	-0.886
HDE 326 775	6.5	V	0.464	0.73	0.589	-0.001	0.153	-0.001	0.773
HDE 227 465	6.5	V	0.499	0.683	0.433	0.142	0.225	0.049	0.626
Tyc 4026-00424-1	6.5	V	0.535	0.719	0.621	0.112	0.225	0.026	0.836
HD 110 360	6.5	V	0.664	1.042	0.698	0.292	0.301	0.066	1.001
BD +60 2522	6.5	I	0.484	0.652	0.442	0.022	0.09	-0.011	-0.678
HDE 228 759	6.5	V	0.459	0.692	0.469	0.082	0.195	0.024	0.768
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HDE 228 841	6.5	V	0.537	0.836	0.617	0.046	0.161	0.041	0.825
HD 152 248 AaAb	7	I	0.472	0.564	0.385	0.005	0.111	0.016	-0.046
BD -04 4503	7	V	0.758	0.865	0.515	0.272	0.379	0.073	0.884
MY Ser AaAb	7	I	0.372	0.439	0.262	0.084	0.136	0.023	-0.932
HD 117 797	7	I	0.647	0.727	0.459	0.115	0.156	0.006	-0.644
HD 156 738 AB	7	III	0.515	0.585	0.401	0.148	0.211	0.053	0.35
Cyg OB2-4 A	7	III	0.576	0.647	0.462	0.054	0.201	0.014	0.342
HD 93 222 AB	7	V	0.603	0.732	0.455	0.115	0.186	0.025	0.713
BN Gem	7	III	0.316	0.367	0.362	0.251	0.255	0.07	0.503
HDE 319 703 A	7	V	0.586	0.706	0.514	0.069	0.214	0.062	0.624
BD +66 1675	7	V	0.649	0.733	0.537	0.112	0.292	0.028	0.859
HD 94 370 A	7	I	0.558	0.671	0.348	-0.013	0.134	0.026	0.039
CPD -26 2704	7	V	0.719	0.798	0.474	0.076	0.212	0.047	0.833
2MASS J10224096-5930305	7	V	0.537	0.64	0.559	0.126	0.189	0.049	0.733
HD 168 137 AaAb	7	V	0.745	0.755	0.487	0.141	0.301	0.088	0.952
Cyg OB2-29	7	V	0.53	0.627	0.585	0.08	0.137	-0.01	0.844
HD 18 326	7	V	0.591	0.668	0.407	0.116	0.25	0.049	0.821
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
ALS 18 707	7	V	0.545	0.685	0.563	0.189	0.246	0.012	0.715
HD 168 504	7	V	0.647	0.668	0.406	0.084	0.236	0.031	0.856
HD 97 319	7	V	0.646	0.674	0.48	0.105	0.154	0.027	0.625
HD 93 160 AB	7	I	0.53	0.619	0.374	0.13	0.175	0.047	0.134
HDE 305 438	7	V	0.709	0.833	0.602	0.171	0.277	0.05	1.017
CPD -49 2322	7	V	0.637	0.65	0.51	0.066	0.189	0.042	0.717
HD 99 546	7	V	0.667	0.797	0.565	0.156	0.286	0.044	0.805
HD 120 521	7	I	0.688	0.723	0.49	0.192	0.258	0.056	0.037
ALS 8294	7	V	0.688	0.752	0.586	0.129	0.268	0.039	0.906
HD 53 975	7	V	0.659	0.668	0.423	0.149	0.261	0.028	0.798
HD 152 590	7	V	0.684	0.805	0.597	0.122	0.302	0.048	0.891
HD 168 461	7	V	0.706	0.761	0.621	0.2	0.293	0.068	0.795
HD 5689	7	V	0.634	0.785	0.506	0.14	0.203	0.022	0.697
NGC 1624-2	7	I	0.713	0.838	0.504	0.157	0.298	0.05	-0.75
BD -13 4927	7	III	0.501	0.621	0.477	0.09	0.157	0.024	0.258
HD 152 408	7	I	0.406	0.495	0.076	-0.004	-0.08	-0.089	-4.839
HD 46 485	7	V	0.647	0.689	0.478	0.053	0.187	0.036	0.766
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
BD +62 2078	7	V	0.658	0.773	0.502	0.067	0.19	0.044	0.928
HD 175 514	7	III	0.476	0.539	0.445	0.047	0.157	-0.009	0.593
HD 74 920	7	III	0.7	0.776	0.426	0.096	0.24	0.04	0.589
HDE 344 777	7	V	0.623	0.656	0.432	0.119	0.303	0.077	0.794
EM Car	7	III	0.501	0.526	0.353	0.057	0.146	0.034	0.459
HD 94 963	7	III	0.581	0.685	0.487	0.074	0.133	0.018	0.277
BD +33 1025 A	7	V	0.649	0.768	0.586	0.124	0.295	0.035	0.924
HD 163 800	7	III	0.602	0.667	0.458	0.107	0.121	0.032	0.51
theta <sup>1</sup> <i>OriCaCb</i>	7	III	0.568	0.625	0.425	0.155	0.187	0.045	0.262
ALS 18 770	7	V	0.526	0.646	0.558	0.186	0.176	0.076	0.753
Pismis 24-3	7	V	0.72	0.872	0.302	0.162	0.324	0.133	0.759
HD 213 023 A	7	V	0.684	0.73	0.616	0.194	0.324	0.046	0.861
HD 152 623 AaAbB	7	V	0.566	0.692	0.414	0.104	0.149	0.049	0.67
BD +55 2840	7	V	0.685	0.705	0.528	0.153	0.256	0.053	0.665
HD 151 515	7	III	0.574	0.698	0.469	0.121	0.18	0.034	0.259
HDE 338 916	7	V	0.687	0.73	0.582	0.135	0.294	0.049	0.993
BD +60 497	7	V	0.565	0.631	0.338	0.095	0.199	0.04	0.698
Continua na próxima página									



**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 167 659	7	III	0.561	0.683	0.441	0.064	0.104	0.036	0.342
Pismis 24-15	7	V	0.739	0.743	0.571	0.042	0.278	0.017	0.909
HD 93 130	7	III	0.539	0.61	0.423	0.084	0.198	0.048	0.313
ALS 85	7	V	0.678	0.776	0.411	0.217	0.364	0.081	0.767
Cyg OB2-5 B	7	I	0.588	0.739	0.45	0.119	0.075	-0.024	-0.398
BD +60 513	7	V	0.581	0.631	0.441	0.102	0.167	0.061	0.677
HD 159 176	7	III	0.515	0.615	0.366	0.097	0.152	0.023	0.512
HD 46 573	7	V	0.737	0.789	0.559	0.177	0.271	0.031	0.874
HD 44 811	7	V	0.645	0.777	0.595	0.169	0.259	0.051	0.967
HD 171 589	7	III	0.722	0.836	0.572	0.148	0.257	0.044	0.446
BD +60 586 A	7	V	0.634	0.767	0.549	0.16	0.244	0.048	0.887
HD 97 966	7	V	0.531	0.644	0.427	0.095	0.216	0.028	0.827
HD 41 997	7	V	0.665	0.705	0.39	0.136	0.23	0.02	0.616
HD 101 205 AB	7	I	0.512	0.527	0.343	0.127	0.191	0.032	0.058
HD 69 464	7	I	0.588	0.675	0.432	0.114	0.124	0.018	0.131
CPD -61 3973	7	V	0.73	0.761	0.576	0.181	0.249	0.026	0.809
HD 175 876	7	III	0.626	0.768	0.485	0.104	0.176	0.021	0.463
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
Cyg OB2-16	7.5	V	0.644	0.643	0.381	0.082	0.34	0.038	0.608
HD 167 771	7.5	III	0.609	0.602	0.418	0.078	0.215	0.05	0.467
ALS 15 111	7.5	V	0.774	0.626	0.383	0.183	0.386	0.028	0.767
68 Cyg	7.5	III	0.672	0.637	0.476	0.136	0.216	0.051	0.393
BD +52 805	7.5	V	0.777	0.685	0.398	0.31	0.361	0.074	0.85
BD +55 2722 A	7.5	V	0.709	0.697	0.489	0.15	0.316	0.047	0.873
HD 173 783	7.5	I	0.701	0.586	0.443	0.244	0.359	0.144	0.214
xi Per	7.5	III	0.785	0.767	0.532	0.1	0.188	0.057	0.398
HD 161 853	7.5	V	0.704	0.597	0.473	0.236	0.272	0.024	0.78
HD 93 343	7.5	V	0.736	0.699	0.471	0.173	0.357	0.058	0.859
ALS 15 114	7.5	V	0.644	0.627	0.323	0.015	0.224	0.021	0.798
Cyg OB2-24	7.5	V	0.671	0.555	0.462	0.158	0.305	0.041	0.666
HD 17 505 B	7.5	V	0.786	0.646	0.502	0.253	0.314	0.061	0.708
HDE 326 331	7.5	III	0.74	0.596	0.354	0.114	0.193	0.048	0.445
CPD -59 2636 AB	7.5	V	0.756	0.662	0.41	0.169	0.308	0.053	0.765
HD 124 979	7.5	III	0.684	0.669	0.47	0.121	0.206	0.066	0.476
15 Mon AaAb	7.5	V	0.662	0.622	0.529	0.141	0.233	0.03	0.732
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
ALS 15 204	7.5	V	0.698	0.621	0.436	0.146	0.222	0.019	0.774
63 Oph	7.5	III	0.712	0.635	0.435	0.185	0.203	0.043	0.412
CPD -34 2135	7.5	I	0.73	0.707	0.514	0.094	0.183	0.028	0.002
HD 164 536	7.5	V	0.751	0.68	0.493	0.195	0.278	0.035	0.801
ALS 8272	7.5	V	0.636	0.622	0.341	0.094	0.32	0.024	0.667
BD +60 261	7.5	III	0.677	0.662	0.5	0.117	0.171	0.017	0.432
Cyg OB2-25 A	7.5	V	0.682	0.649	0.535	0.21	0.321	0.031	0.804
HDE 305 612	7.5	V	0.734	0.601	0.441	0.158	0.286	0.058	0.806
Cyg OB2-73	7.5	V	0.72	0.717	0.453	0.128	0.334	0.022	0.757
HD 95 589	7.5	III	0.788	0.699	0.484	0.17	0.279	0.052	0.541
CPD -59 2626 AB	7.5	V	0.681	0.608	0.535	0.179	0.279	0.066	0.681
HD 93 161 A	7.5	V	0.708	0.601	0.303	0.189	0.333	0.052	0.707
HD 97 166	7.5	III	0.658	0.596	0.443	0.127	0.227	0.043	0.59
HD 47 129	7.5	I	0.793	0.636	0.504	0.238	0.298	0.019	-0.552
HD 135 591	7.5	V	0.702	0.573	0.4	0.15	0.196	0.051	0.645
BD +40 4179	7.5	V	0.774	0.733	0.562	0.197	0.352	0.054	0.845
HD 166 734	7.5	I	0.537	0.478	0.262	0.081	0.149	0.012	-0.85
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 123 590 AB	7.5	V	0.705	0.683	0.499	0.166	0.315	0.058	0.817
Trumpler 14-9	7.5	V	0.838	0.702	0.436	0.271	0.379	0.093	0.859
HDE 329 100 A	7.5	III	0.769	0.656	0.418	0.114	0.293	0.144	0.591
Cyg OB2-22 B	7.5	III	0.699	0.647	0.545	-0.034	0.334	0.007	0.57
ALS 15 115	7.5	V	0.817	0.716	0.407	0.242	0.335	0.104	0.801
ALS 11 355	7.5	V	0.696	0.652	0.465	0.17	0.292	0.053	0.681
Cyg OB2-17	7.5	V	0.771	0.742	0.482	0.244	0.399	0.116	0.776
HDE 305 539	7.5	V	0.77	0.699	0.584	0.219	0.352	0.078	0.869
HD 101 191	7.5	V	0.739	0.666	0.426	0.216	0.29	0.076	0.747
lambda Ori A	7.5	V	0.716	0.609	0.52	0.168	0.252	0.035	0.63
HD 46 056	7.5	V	0.693	0.57	0.312	0.121	0.282	0.045	0.668
CPD -59 2610	7.5	V	0.764	0.663	0.519	0.232	0.379	0.088	0.829
HDE 319 702	7.5	V	0.725	0.593	0.431	0.16	0.337	0.074	0.61
HDE 229 202	7.5	V	0.604	0.586	0.306	0.123	0.212	0.052	0.635
Cyg OB2-6	7.5	V	0.689	0.589	0.453	0.196	0.361	0.062	0.611
HD 94 024	7.5	V	0.793	0.68	0.469	0.221	0.333	0.071	0.757
HD 101 223	7.5	V	0.749	0.597	0.453	0.207	0.295	0.035	0.719
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 97 434	7.5	III	0.686	0.686	0.416	0.192	0.239	0.043	0.349
HD 17 520 A	7.5	V	0.726	0.675	0.396	0.29	0.225	-0.015	0.924
HD 175 754	7.5	I	0.747	0.599	0.44	0.177	0.232	0.033	0.167
HD 97 848	7.5	V	0.748	0.648	0.428	0.194	0.295	0.055	0.744
HD 165 921	7.5	III	0.589	0.476	0.371	0.087	0.199	0.065	0.536
ALS 15 134	7.5	V	0.72	0.646	0.46	0.112	0.354	0.013	0.853
CPD -59 2591	8	V	0.848	0.525	0.317	0.216	0.357	0.076	0.684
Herschel 36	8	III	0.775	0.559	0.509	0.274	0.318	0.033	0.651
V572 Car	8	III	0.669	0.495	0.281	0.152	0.24	0.034	0.589
HD 216 898	8	V	0.905	0.551	0.371	0.287	0.433	0.09	0.719
CPD -59 2635	8	III	0.693	0.491	0.32	0.191	0.309	0.034	0.662
HD 102 415	8	V	0.872	0.62	0.4	0.356	0.406	0.133	0.664
BD -11 4586	8	I	0.785	0.505	0.453	0.174	0.291	0.037	0.154
HDE 319 703 D	8	III	0.875	0.514	0.207	0.188	0.364	0.159	0.546
CPD -59 2629	8	V	0.778	0.528	0.499	0.246	0.403	0.069	0.657
HD 124 314 BaBb	8	V	0.897	0.524	0.2	0.236	0.421	0.127	0.585
HD 75 211	8	I	0.78	0.602	0.41	0.184	0.294	0.043	0.178
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
CPD -59 2551	8	V	0.881	0.561	0.437	0.244	0.432	0.099	0.812
HD 122 313	8	V	0.835	0.56	0.489	0.271	0.405	0.056	0.819
HDE 228 779	8	I	0.874	0.44	0.181	0.189	0.308	0.116	0.202
ALS 18 768	8	V	0.99	0.649	0.614	0.188	0.305	0.123	0.85
HDE 298 429	8	V	0.852	0.618	0.396	0.271	0.332	0.089	0.714
HD 151 018	8	I	0.793	0.573	0.383	0.263	0.392	0.153	0.286
BD -08 4617	8	III	0.683	0.506	0.385	0.155	0.273	0.02	0.555
NX Vel AB	8	III	0.745	0.512	0.368	0.247	0.329	0.042	0.548
HD 116 282	8	III	0.785	0.555	0.291	0.154	0.303	0.107	0.637
HD 101 413 AB	8	V	0.791	0.557	0.375	0.244	0.34	0.056	0.673
HD 89 625	8	V	0.914	0.495	0.379	0.364	0.419	0.13	0.577
HDE 305 523	8	III	0.884	0.458	0.306	0.275	0.378	0.1	0.412
2MASS J10224377-5930182	8	III	0.729	0.536	0.49	0.21	0.271	0.061	0.611
HD 12 323	8	V	0.958	0.631	0.45	0.372	0.472	0.124	0.839
ALS 18 767	8	V	0.914	0.486	0.419	0.313	0.395	0.192	0.589
HD 155 756	8	I	0.761	0.437	0.306	0.241	0.323	0.182	0.307
Cyg OB2-8 D	8	V	0.835	0.563	0.444	0.285	0.439	0.046	0.765
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 114 886 AaAb	8	III	0.84	0.431	0.249	0.212	0.343	0.072	0.478
LS III +41 14	8	V	0.924	0.485	0.24	0.37	0.493	0.1	0.696
HD 152 246 AaAb	8	V	0.822	0.48	0.327	0.245	0.355	0.086	0.633
HD 93 028	8	V	0.903	0.523	0.419	0.293	0.419	0.092	0.684
BD -14 5014	8	III	0.624	0.482	0.294	0.11	0.243	-0.002	0.535
HD 164 438	8	III	0.865	0.443	0.276	0.294	0.373	0.146	0.58
HD 191 423	8	III	0.879	0.476	0.405	0.229	0.371	0.111	0.427
HD 96 917	8	I	0.847	0.582	0.42	0.153	0.29	0.045	0.033
HD 191 201 A	8	III	0.781	0.446	0.234	0.204	0.366	0.132	0.36
HD 123 008	8	I	0.877	0.547	0.408	0.324	0.443	0.155	0.248
BD +39 1328	8	I	0.805	0.467	0.503	0.154	0.259	0.059	0.019
HD 5005 D	8	V	0.902	0.528	0.35	0.369	0.499	0.084	0.76
BD +36 4063	8	I	0.618	0.378	0.178	0.308	0.363	0.155	0.294
ALS 6967	8	V	0.758	0.53	0.419	0.207	0.385	0.037	0.666
CPD -59 2627	8	V	0.904	0.527	0.249	0.375	0.428	0.088	0.824
LY Aur A	8	I	0.707	0.428	0.257	0.067	0.341	0.087	0.232
HD 92 504	8	V	0.803	0.597	0.371	0.262	0.368	0.094	0.791
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 152 314 AaAb	8	III	0.908	0.46	0.322	0.269	0.368	0.141	0.568
HD 105 627	8	V	0.9	0.531	0.339	0.286	0.403	0.084	0.575
HD 92 206 C	8	III	0.711	0.469	0.332	0.185	0.293	0.045	0.507
delta Cir AaAbAc	8	III	0.644	0.483	0.35	0.17	0.264	0.035	0.454
BD +36 4145	8	III	0.758	0.599	0.411	0.167	0.267	0.092	0.668
HD 92 607	8	III	0.76	0.483	0.324	0.213	0.285	0.054	0.646
HD 41 161	8	III	0.885	0.699	0.413	0.223	0.325	0.066	0.609
Pismis 24-10	8	V	0.981	0.744	0.482	0.227	0.445	0.186	0.749
Cyg OB2-5 A	8	-	0.72	0.479	0.306	0.081	0.08	-0.036	-3.416
HD 96 670 AB	8	-	0.748	0.383	0.302	0.124	0.277	0.037	-0.396
Muzzio III-9	8	-	0.783	0.569	0.447	0.055	0.2	0.039	-0.408
HD 108	8	-	0.96	0.756	0.488	0.045	0.111	-0.031	-0.397
9 Sge	8	-	0.85	0.672	0.443	0.116	0.177	0.001	-0.436
29 CMa	8	-	0.765	0.579	0.461	0.031	0.138	-0.001	-1.438
2MASS J10584671-6105512	8	-	0.599	0.437	0.444	0.083	0.127	0.009	-0.419
HD 149 404	8	-	0.601	0.431	0.262	0.168	0.237	0.068	-0.029
HD 112 244	8	-	0.739	0.434	0.352	0.155	0.222	0.039	-0.164
Continua na próxima página									



**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
GS Mus	8.5	I	0.728	0.371	0.188	0.283	0.409	0.19	0.069
HD 149 452	8.5	III	0.858	0.471	0.279	0.154	0.314	0.064	0.51
ALS 18 771	8.5	V	0.879	0.531	0.559	0.3	0.396	0.142	0.708
HD 144 647	8.5	V	0.81	0.553	0.392	0.235	0.372	0.052	0.735
ALS 4923	8.5	III	0.72	0.571	0.37	0.165	0.306	0.064	0.612
HDE 328 209 AB	8.5	III	0.861	0.432	0.47	0.231	0.375	0.143	0.337
HDE 344 758	8.5	V	0.738	0.556	0.325	0.178	0.344	0.063	0.659
ALS 12 749	8.5	V	0.846	0.532	0.37	0.308	0.433	0.117	0.673
LM Vel	8.5	I	0.816	0.628	0.355	0.11	0.261	0.088	0.084
CPD -35 2105 AaAbB	8.5	III	0.854	0.439	0.309	0.226	0.37	0.117	0.417
HD 75 759 AB	8.5	V	0.838	0.465	0.392	0.246	0.38	0.073	0.712
HD 57 236	8.5	V	0.775	0.615	0.48	0.237	0.387	0.059	0.789
HD 148 546	8.5	I	0.837	0.522	0.358	0.243	0.261	0.125	0.204
AB Cru	8.5	I	0.743	0.556	0.365	0.178	0.265	0.075	0.389
CC Cas	8.5	I	0.727	0.391	0.33	0.142	0.238	0.025	0.378
HD 91 651	8.5	V	0.881	0.519	0.29	0.375	0.44	0.155	0.59
ALS 18 049	8.5	V	0.895	0.695	0.547	0.252	0.433	0.18	0.869
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
CPD -41 7733	8.5	V	0.832	0.556	0.311	0.259	0.375	0.104	0.628
CPD -54 6791 AB	8.5	V	0.954	0.53	0.401	0.376	0.455	0.145	0.686
HD 150 574	8.5	III	0.914	0.621	0.357	0.298	0.389	0.129	0.521
HD 52 266	8.5	V	0.908	0.463	0.292	0.344	0.446	0.124	0.504
HD 76 341	8.5	III	0.849	0.438	0.275	0.299	0.367	0.141	0.593
HD 114 737 AB	8.5	III	0.805	0.537	0.43	0.211	0.342	0.066	0.557
HD 125 241	8.5	I	0.835	0.633	0.419	0.172	0.341	0.082	0.12
HD 91 837	8.5	III	0.795	0.53	0.389	0.192	0.274	0.065	0.714
CPD -59 2644	8.5	V	0.875	0.607	0.487	0.273	0.414	0.092	0.869
TU Mus	8.5	I	0.566	0.347	0.302	0.095	0.16	0.037	0.504
HD 167 411	8.5	III	0.872	0.45	0.157	0.262	0.428	0.285	0.395
HD 52 533 A	8.5	III	0.741	0.499	0.288	0.201	0.33	0.057	0.542
Tyc 8626-02506-1	8.5	V	0.885	0.617	0.331	0.249	0.418	0.077	0.799
HD 165 246	8.5	V	0.764	0.541	0.434	0.188	0.336	0.06	0.718
ALS 15 207	8.5	V	0.951	0.591	0.467	0.342	0.464	0.096	0.823
HD 93 190	8.5	I	0.385	0.252	0.074	0.279	0.207	0.039	0.373
HD 113 659	8.5	III	0.873	0.45	0.375	0.249	0.369	0.096	0.587
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 117 490	8.5	III	0.87	0.499	0.344	0.369	0.408	0.128	0.535
HD 116 852	8.5	I	0.809	0.488	0.32	0.153	0.268	0.087	0.391
HD 145 217	8.5	V	0.8	0.628	0.382	0.201	0.368	0.097	0.699
AO Cas	8.5	I	0.734	0.398	0.36	0.191	0.219	0.077	0.268
HD 71 304	8.5	I	0.847	0.54	0.262	0.213	0.364	0.137	0.304
10 Lac	8.5	V	0.845	0.618	0.499	0.27	0.413	0.088	0.782
HD 115 455	8.5	III	0.738	0.564	0.325	0.187	0.293	0.037	0.579
BD -12 4979	8.5	V	0.896	0.485	0.188	0.352	0.447	0.141	0.584
ALS 15 125	8.5	V	0.826	0.488	0.258	0.143	0.443	0.083	0.703
HD 17 520 B	9	I	0.665	0.361	0.213	0.222	0.418	0.14	0.211
HD 69 106	9	III	0.922	0.322	0.177	0.238	0.419	0.15	0.409
ALS 18 083	9	V	1.053	0.373	0.217	0.424	0.533	0.118	0.511
HD 93 521	9	III	0.942	0.377	0.228	0.255	0.346	0.101	0.428
HD 93 027	9	V	0.965	0.379	0.295	0.386	0.488	0.093	0.603
V573 Car	9	I	0.788	0.137	0.098	0.278	0.32	0.026	0.278
HD 37 737	9	III	0.917	0.286	0.185	0.296	0.409	0.141	0.39
AE Aur	9	V	0.99	0.474	0.36	0.358	0.456	0.105	0.711
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 13 022	9	III	0.896	0.302	0.182	0.272	0.396	0.168	0.413
HD 154 811	9	I	0.886	0.313	0.19	0.246	0.341	0.21	0.254
Cyg OB2-22 C	9	V	0.894	0.256	0.28	0.195	0.554	0.063	0.579
HD 125 206 AB	9	III	0.953	0.29	0.183	0.327	0.453	0.115	0.471
HDE 344 783	9	V	0.96	0.362	0.324	0.318	0.478	0.146	0.56
HD 152 147	9	I	0.886	0.207	0.161	0.284	0.369	0.161	0.221
HD 190 429 B	9	III	0.945	0.338	0.17	0.319	0.417	0.115	0.451
HD 93 576	9	III	0.949	0.375	0.187	0.361	0.459	0.099	0.475
HD 96 622	9	V	0.911	0.414	0.332	0.319	0.425	0.137	0.581
HD 118 198	9	V	0.917	0.321	0.276	0.299	0.438	0.18	0.522
Sh 2-158 2	9	III	0.889	0.36	0.521	0.208	0.494	0.081	0.407
HD 16 832	9	III	0.85	0.392	0.233	0.244	0.365	0.121	0.488
HD 202 214 AaAb	9	V	0.995	0.509	0.47	0.405	0.455	0.098	0.645
HD 93 249 A	9	III	0.87	0.42	0.232	0.266	0.325	0.099	0.575
HDE 328 856	9	III	0.91	0.386	0.226	0.266	0.357	0.227	0.315
14 Cep	9	III	0.824	0.38	0.371	0.258	0.334	0.082	0.444
ALS 11 761	9	III	0.935	0.437	0.326	0.328	0.381	0.137	0.428
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 101 545 AaAb	9	III	0.814	0.361	0.294	0.257	0.34	0.15	0.416
HD 158 186	9	III	0.859	0.361	0.298	0.333	0.387	0.086	0.539
HD 18 409	9	III	0.883	0.383	0.257	0.275	0.382	0.164	0.328
CPD -58 2627	9	V	0.898	0.495	0.303	0.28	0.454	0.109	0.707
HD 57 682	9	V	0.875	0.487	0.432	0.306	0.44	0.095	0.714
HDE 305 619	9	III	0.901	0.379	0.284	0.266	0.385	0.181	0.346
QZ Car AaAc	9	I	0.795	0.344	0.18	0.165	0.292	0.163	0.099
CPD -59 5634	9	III	0.846	0.382	0.408	0.259	0.374	0.151	0.321
HD 163 892	9	V	0.86	0.417	0.302	0.246	0.409	0.134	0.604
HDE 237 211	9	I	0.91	0.308	0.251	0.213	0.36	0.088	0.228
BD +55 2722 B	9	V	0.839	0.349	0.111	0.346	0.454	0.08	0.545
Cyg OB2-10	9	I	0.853	0.313	0.204	0.268	0.349	0.215	0.161
HD 166 546	9	V	0.902	0.412	0.298	0.319	0.451	0.168	0.594
16 Sgr AaAb	9	III	0.847	0.362	0.247	0.247	0.342	0.121	0.498
Y Cyg	9	III	0.875	0.345	0.155	0.202	0.384	0.071	0.551
HD 168 941	9	III	0.963	0.395	0.33	0.294	0.351	0.171	0.599
HD 173 010	9	I	0.656	0.255	0.218	0.144	0.303	0.233	0.039
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 165 319	9	III	0.872	0.334	0.215	0.219	0.414	0.208	0.274
HDE 323 110	9	I	0.83	0.4	0.353	0.146	0.266	0.092	0.156
theta Mus B	9	III	0.817	0.405	0.325	0.271	0.331	0.098	0.544
mu Col	9	V	0.98	0.506	0.339	0.38	0.494	0.137	0.716
HD 164 816	9	V	1.012	0.422	0.183	0.369	0.485	0.146	0.583
delta Ori Aa	9	I	0.853	0.308	0.325	0.249	0.308	0.132	0.329
CPD -46 8221	9	III	0.93	0.337	0.209	0.317	0.433	0.206	0.432
19 Cep	9	I	0.852	0.351	0.253	0.19	0.282	0.113	0.254
HD 152 218	9	III	0.719	0.317	0.227	0.185	0.267	0.069	0.492
HDE 305 536	9	V	0.974	0.453	0.354	0.356	0.467	0.123	0.593
HD 151 003 AaAb	9	III	0.84	0.417	0.323	0.157	0.293	0.085	0.429
HDE 256 725 B	9	V	0.875	0.451	0.318	0.35	0.521	0.094	0.524
CPD -58 2625	9	V	0.928	0.492	0.341	0.336	0.452	0.113	0.763
HDE 303 492	9	-	0.809	0.385	0.197	0.172	0.263	0.059	-0.311
alpha Cam	9	-	0.81	0.383	0.305	0.216	0.302	0.101	-0.08
HD 90 087	9.2	III	0.984	0.472	0.329	0.315	0.418	0.175	0.464
ALS 18 556	9.2	I	0.73	0.307	0.125	0.123	0.273	0.142	0.145
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
zeta Ori AaAb	9.2	I	0.859	0.325	0.228	0.245	0.338	0.124	0.186
HD 96 264	9.2	III	0.895	0.341	0.261	0.281	0.385	0.134	0.526
delta Ori Ab	9.5	III	1.097	0.19	0.218	0.412	0.386	0.152	0.335
zeta Ori B	9.5	III	0.925	0.275	0.234	0.315	0.439	0.132	0.312
HD 152 424	9.5	I	0.826	0.218	0.246	0.186	0.189	0.115	0.035
zeta Oph	9.5	III	0.838	0.395	0.361	0.281	0.32	0.098	0.503
HDE 308 813	9.5	V	0.971	0.372	0.246	0.36	0.485	0.196	0.538
HD 117 856	9.5	III	0.95	0.291	0.241	0.316	0.415	0.178	0.373
HD 5005 B	9.5	V	1.024	0.245	0.086	0.508	0.579	0.118	0.4
theta <sup>2</sup> OriA	9.5	V	0.991	0.364	0.219	0.399	0.512	0.088	0.574
HD 10 125	9.5	III	0.973	0.274	0.247	0.329	0.452	0.154	0.267
1 Cam A	9.5	I	1.009	0.219	0.168	0.348	0.36	0.096	0.301
V1034 Sco	9.5	III	0.823	0.409	0.284	0.19	0.34	0.112	0.513
HD 68 450	9.5	III	0.923	0.322	0.228	0.287	0.351	0.176	0.452
tau CMa AaAb	9.5	I	0.785	0.329	0.264	0.225	0.251	0.043	0.264
HD 155 775	9.5	III	0.986	0.192	0.143	0.431	0.531	0.138	0.361
HD 93 146 B	9.5	V	1.131	0.247	0.054	0.456	0.548	0.156	0.419
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
HD 76 968	9.5	I	0.775	0.361	0.287	0.274	0.379	0.138	0.228
HD 165 174	9.5	III	0.956	0.393	0.267	0.277	0.424	0.165	0.4
HD 191 781	9.5	III	0.972	0.341	0.212	0.384	0.44	0.258	0.278
HD 152 405	9.5	III	0.933	0.294	0.232	0.332	0.406	0.22	0.364
HD 156 292	9.5	III	0.867	0.264	0.184	0.277	0.356	0.163	0.393
HD 191 201 B	9.5	V	1.177	0.334	0.164	0.584	0.603	0.096	0.569
HD 123 056	9.5	V	1.039	0.346	0.338	0.42	0.533	0.1	0.524
HD 152 003	9.5	I	0.819	0.255	0.161	0.279	0.362	0.195	0.168
HD 161 807	9.5	III	0.916	0.306	0.221	0.259	0.388	0.135	0.319
sigma Ori AaAb	9.5	V	1.129	0.343	0.257	0.428	0.573	0.1	0.675
HD 75 222	9.5	I	0.845	0.281	0.272	0.298	0.388	0.145	0.244
Cyg OB2-27 AB	9.5	III	0.731	0.169	0.164	0.168	0.359	0.029	0.378
HD 209 339	9.5	III	0.957	0.295	0.232	0.364	0.43	0.147	0.472
HDE 305 518	9.5	III	0.942	0.316	0.206	0.31	0.423	0.162	0.474
PZ Gem	9.5	I	0.091	0.026	0.029	0.288	0.075	-0.159	0.264
HD 104 565	9.5	I	0.878	0.281	0.134	0.241	0.379	0.239	0.218
Cyg X-1	9.5	-	0.794	0.281	0.27	0.249	0.438	0.159	-0.222
Continua na próxima página									



**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
CPD -59 2554	9.7	V	1.025	0.431	0.309	0.4	0.517	0.071	0.635
HD 154 368	9.7	I	0.857	0.398	0.29	0.285	0.343	0.133	0.21
HD 164 019	9.7	V	0.876	0.399	0.279	0.295	0.427	0.191	0.54
BD +66 1661	9.7	V	0.904	0.42	0.239	0.325	0.399	0.108	0.576
CPD -59 2624	9.7	V	1.078	0.259	0.11	0.507	0.56	0.122	0.435
Cyg OB2-22 D	9.7	III	0.852	0.305	0.041	0.143	0.325	0.059	0.505
HD 115 071	9.7	III	0.918	0.346	0.196	0.315	0.415	0.185	0.294
HD 152 219	9.7	III	0.885	0.299	0.28	0.252	0.39	0.11	0.474
HD 152 200	9.7	III	0.979	0.239	0.19	0.409	0.473	0.154	0.434
upsilon Ori	9.7	V	1.021	0.382	0.21	0.418	0.565	0.119	0.6
CPD -41 7721 A	9.7	V	0.98	0.325	0.151	0.378	0.474	0.157	0.482
mu Nor	9.7	I	0.888	0.188	0.182	0.262	0.366	0.189	0.25
BD +60 498	9.7	III	1.01	0.242	0.153	0.428	0.514	0.095	0.415
HD 204 827 AaAb	9.7	V	0.964	0.351	0.238	0.412	0.531	0.145	0.544
HDE 326 329	9.7	V	1.038	0.337	0.203	0.409	0.509	0.18	0.525
HD 89 137	9.7	III	0.968	0.287	0.238	0.446	0.515	0.195	0.395
HDE 303 312	9.7	V	1.076	0.348	0.23	0.429	0.523	0.149	0.496
Continua na próxima página									

**Tabela A.2— Continuação da página anterior**

<b>Estrelas</b>	<b>TE</b>	<b>CL</b>	$W_{\lambda}(4471)$	$W_{\lambda}(4542)$	$W_{\lambda}(4200)$	$W_{\lambda}(4144)$	$W_{\lambda}(4388)$	$W_{\lambda}(4552)$	$W_{\lambda}(4686)$
15 Sgr	9.7	I	0.935	0.209	0.165	0.286	0.391	0.255	0.251
HD 54 879	9.7	V	1.004	0.281	0.217	0.436	0.53	0.091	0.527
HDE 344 782	9.7	V	0.977	0.319	0.378	0.381	0.489	0.128	0.676