

UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO

Escola de Ciências e Tecnologia – Departamento de Engenharias

Classificação de castas em vinhos do Douro usando Redes Neurais Artificiais

JOAQUIM MARIA LOPES VELHO



Vila Real

2012



UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO

Escola de Ciências e Tecnologia – Departamento de Engenharias

Classificação de castas em vinhos do Douro usando Redes Neurais Artificiais

JOAQUIM MARIA LOPES VELHO

Dissertação de Mestrado, submetida à Universidade de Trás-os-Montes e Alto Douro, para cumprimento dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Informática, sob a orientação do Prof. Doutor Vítor Filipe e coorientação do Prof. Doutor João Barroso

Vila Real

2012

**Dado que o homem é o único animal
que bebe sem sede, convém
que o faça com discernimento.**

(L. Farnoux-Reynaud)

Agradecimentos

Muito especialmente aos orientadores da dissertação, Prof. Doutor Vítor Filipe e Prof. Doutor João Barroso. Ao colega David Luna que, no âmbito do seu Doutoramento trabalha na mesma área e que foi uma ajuda importante principalmente no pontapé de saída.

Ao Prof. Doutor Fernando Nunes do Departamento de Química que foi quem forneceu os dados estudados nesta dissertação, sempre interessado, mostrou-se constantemente disponível para me acompanhar até ao fim.

A todos os professores que, com maior ou menor influência, me permitiram obter as bases necessárias para chegar até este ponto que culmina com a dissertação, a última etapa deste percurso. Professores esses que pertencem maioritariamente ao Departamento de Engenharias da Escola de Ciência e Tecnologia da Universidade de Trás-os-Montes e Alto Douro.

Aos amigos e familiares, que sempre me ajudaram e incentivaram. A alguns colegas de curso que foram fundamentais neste longo caminho através da partilha de experiências e de sabedoria, em especial aos que fizeram parte da caminhada desde o primeiro minuto (ano), e àqueles que ao longo da caminhada me fui cruzando, tendo eles um papel igualmente importante.

Ao pai, mãe e irmão que foram e são a base de tudo.

Índice

<i>Agradecimentos</i>	<i>i</i>
<i>Índice de figuras</i>	<i>iv</i>
<i>Índice de tabelas</i>	<i>v</i>
<i>Índice de abreviaturas</i>	<i>vi</i>
<i>Resumo</i>	<i>vii</i>
<i>Abstract</i>	<i>viii</i>
1 <i>Introdução</i>	1
1.1 <i>Objetivos</i>	2
1.2 <i>Organização do documento</i>	3
2 <i>Revisão bibliográfica</i>	4
2.1 <i>Benefícios do vinho para a saúde</i>	4
2.2 <i>A química do vinho</i>	4
2.2.1 <i>Não Flavonóides</i>	4
2.2.2 <i>Flavonóides</i>	5
2.2.3 <i>Compostos fenólicos</i>	6
2.3 <i>Técnicas de análise química ao vinho</i>	6
2.3.1 <i>Espectroscopia de absorção atômica</i>	6
2.3.2 <i>Análise por espectrometria de infravermelho</i>	8
2.3.3 <i>Análise cromatográfica</i>	11
2.4 <i>Vinhos Monocasta e Multicasta</i>	12
2.4.1 <i>Castas do Douro</i>	12
2.5 <i>Trabalhos relacionados</i>	14
3 <i>Métodos desenvolvidos para a classificação de vinhos</i>	18
3.1 <i>Redes neuronais</i>	18
3.1.1 <i>História</i>	18
3.1.2 <i>Definição</i>	18
3.1.3 <i>Fases</i>	21

3.1.3.1	Fase de treino ou aprendizagem.....	21
3.1.3.2	Fase de Validação	22
3.1.3.3	Fase de teste	22
3.1.4	Classificação	22
3.2	Caracterização do conjunto de amostras usadas no estudo	23
3.3	Implementação	29
3.3.1	Configuração e parametrização da rede neuronal artificial.....	29
3.4	Resultados	31
3.4.1	Objetivo 1 – Classificação dos vinhos por castas	31
3.4.1.1	Ensaio 1 – Classificação das castas nos 53 vinhos com todos os resultados de antocianinas	32
3.4.1.2	Ensaio 2 - Classificação das castas em todos os 53 vinhos com resultados de antocianinas selecionadas.....	33
3.4.2	Objetivo 2 – Classificação dos vinhos por idade	34
3.4.2.1	Ensaio 1-Classificaçãodos vinhos das duas castas por idade	34
3.4.2.2	Ensaio 2 - Classificação das idades nos vinhos de Aragonês.....	35
3.4.2.3	Ensaio 3 - Classificação das idades nos vinhos de Touriga Nacional.....	36
3.4.2.4	Ensaio 4 - Classificação das idades em todos os vinhos.....	36
3.4.2.5	Ensaio 5 - Classificação das idades nos vinhos de Aragonês.....	37
3.4.2.6	Ensaio 6 - Classificação das idades nos vinhos de Touriga Nacional.....	38
4	<i>Conclusões e Perspetivas de trabalho futuro</i>	40
	<i>Referências bibliográficas</i>	42

Índice de figuras

Ilustração 1 – Estrutura química de antocianinas	6
Ilustração 2 - Componentes do espectrómetro de absorção atómica	7
Ilustração 3 – Esquema do funcionamento dum Espectrómetro FTIR	9
Ilustração 4 - Esquema do método de separação de antocianinas por HPLC.....	11
Ilustração 5 - Neurónio artificial.....	19
Ilustração 6 - Rede neuronal de duas camadas	20
Ilustração 7 - Gráfico comparativo das antocianinas nas 2 castas em estudo.....	24
Ilustração 8 - Gráfico comparativo de antocianinas de 3 vinhos da casta de Aragonês.....	25
Ilustração 9 - Gráfico comparativo de antocianinas de 3 vinhos da casta de Touriga Nacional.	26
Ilustração 10 - Valores de antocianinas em 3 vinhos da casta Aragonês (em %)	26
Ilustração 11 - Valores de antocianinas em 3 vinhos da casta de Touriga Nacional (em %)	27
Ilustração 12 - Valores de antocianinas de Aragonês e Touriga Nacional (em %)	28
Ilustração 13 - Esquema simbólico da rede neuronal artificial usada	30
Ilustração 14 - Matriz de confusão na análise da globalidade de antocianinas.....	32
Ilustração 15 - Matriz de confusão na análise das antocianinas discriminantes	33
Ilustração 16 - Matriz de confusão na discriminação por idade de todos os vinhos	34
Ilustração 17 - Matriz de confusão da classificação dos vinhos de Aragonês por idade.	35
Ilustração 18 - Matriz de confusão da classificação dos vinhos Touriga Nacional por idade.	36
Ilustração 19 - Matriz de confusão da separação dos vinhos por idade.....	37
Ilustração 20 - Matriz de confusão da separação dos vinhos de Aragonês por idade.....	38
Ilustração 21 - Matriz de confusão da separação dos vinhos de Touriga Nacional por idade....	38

Índice de tabelas

Tabela 1 - Métodos analíticos de referência.....	10
Tabela 2 – Valores da área de parâmetros químicos obtidos pela técnica cromatográfica HPLC em vinhos da casta de Aragonês e de Touriga Nacional.....	24
Tabela 3 - Parâmetros da rede no Matlab	30

Índice de abreviaturas

DOC – Denominação de Origem Controlada

FANN – Fast Artificial Neural Network

FTIR - Fourier Transform Infrared Spectroscopy

HPLC - High-Performance Liquid Chromatography

IVDP – Instituto dos Vinhos do Douro e Porto

LMS - Least Mean Square

RMSE – Root Mean Square Deviation

RNA – Rede Neuronal Artificial

Resumo

No geral, o interesse da química é o estudo das propriedades, constituição e transformações de substâncias e materiais, sendo os alimentos um foco de grande interesse, estando o vinho dentro desse grande grupo.

O objetivo geral deste trabalho consiste em propor métodos de classificação de vinhos monocasta da região Demarcada do Douro com base em parâmetros químicos. Um primeiro objetivo prende-se com a discriminação da casta predominante num conjunto de amostras constituído por dois tipos de vinho monocasta (Touriga Nacional e Aragonês). Os dados das amostras resultaram de um anterior estudo que envolveu a análise por High-Performance Liquid Chromatography de vinhos das duas castas referidas e que resultou na recolha dos valores de antocianinas em cinquenta e três vinhos diferentes (vinte e um de Aragonês e trinta e dois de Touriga Nacional). O segundo objetivo é discriminar os vinhos por idade, primeiramente no geral (levando em conta as cinquenta e três amostras) e de seguida mais particular, fazendo a classificação por idade apenas nas amostras de cada casta (para as vinte e uma de Aragonês e para as trinta e duas de Touriga Nacional, em separado). Os classificadores usados baseiam-se em redes neuronais artificiais que se confirmaram como uma ferramenta muito útil e fiável na discriminação da casta predominante e da idade dos vinhos.

Os resultados gerais foram bastante aceitáveis, a rondar os 95% de acerto quanto à casta predominante. Independentemente da casta, e na tentativa de discriminação dos vinhos de todas as idades obtiveram-se resultados a rondar os 70%. Na separação dos vinhos entre os mais novos e os mais antigos (um, dois, três e mais que três anos, atingiram-se valores médios de 90%, visto que os vinhos com mais idade começam a ter características, à base de antocianinas, um pouco mais variáveis comparadas com os vinhos mais recentes.

A pouca abundância de amostras foi o que mais penalizou o estudo, visto que uma rede neuronal artificial funciona melhor quanto mais amostras existirem principalmente para realizar o treino. Quanto maior for o número de amostras usadas na fase de treino melhor será o processo de aprendizagem e melhores resultados se obtêm na classificação de novas amostras.

Abstract

The interest of chemistry is the study of the properties, constitution and transformation of substances and materials, been the food is an item of big interest and wine is in this large group.

The general objective of this work is to propose methods of classification of single variety wines from Douro region, based in chemical parameters. The first objective relates to caste prevailing discrimination in a group of samples constituted by two types of single variety wines (Touriga Nacional and Aragonês). Data samples resulted from a previous study which involved High-Performance Liquid Chromatography analyze of two varieties wines that resulted on a collection of values of anthocyanins in fifty-three different wines (twenty-one of Aragonês and thirty-two of Touriga Nacional). The second objective is to discriminate wines by age, first in general (taking into account 53 samples) and then more particularly, making the classification by age only in the samples of each variety (for 21 of Aragonês and 32 of Touriga Nacional separately). The classifiers used are based on artificial neural networks which were confirmed as a useful and reliable in predicting the predominant variety and age of the wine.

The overall results were quite acceptable, around 95% hit as the predominant variety.

Independently of caste, and in an attempt to discriminate wines of all ages yielded results of around 70%. In the separation among younger and older (one, two, three and more than three years), there were reached mean values of 90%, whereas older wines start having characteristics, based of anthocyanins, slightly more variable compared with older wines.

The low abundance of samples was what most penalized the study, whereas an artificial neural network works better as more samples exist, principally for realize the train. As larger is the number of samples used in train phase, will be better the process of learning and will be better results in classification of new samples.

1 Introdução

Os vinhos do Douro pelo facto de serem produzidos na primeira região do mundo que recebeu a denominação de origem, criada na segunda metade do século XVII, assumem nos dias de hoje um renome nacional e internacional, dado o seu carácter único associado à região demarcada - Douro - que comporta uma identidade cultural, patrimonial, gastronómica, paisagística e histórica, classificada como património Mundial pela UNESCO (Oliveira, 2010).

Segundo o IVDP - Instituto dos Vinhos do Douro e Porto - a região vitícola do Douro é a região que mais receitas arrecada fora do país, devido em grande parte ao Vinho do Porto que, constitui, sem dúvida, o mais importante produto da história económica portuguesa do últimos três séculos, sendo representativo de mais de 60% das exportações de vinhos portugueses, cerca de 20% do total das exportações dos produtos agro-alimentares e um pouco menos de 2% do total das exportações do país. O IVDP informa também que, no caso do vinho DOC Douro (vinho de mesa) também houve um aumento de vendas no primeiro semestre do ano, 39 milhões de euros que equivalem a um crescimento de 4,2%. Angola foi o mercado onde as vendas de DOC Douro mais cresceram enquanto a Alemanha comprou mais 25%, atingindo 1,9 milhões. O mercado português retraiu-se ligeiramente mas, perante os valores globais, verifica-se um aumento do preço médio em torno dos 3,9%. Estes números são indicadores da preponderante importância que a produção vitícola traz para a economia da região e do país (Ferreira, 2010). Não se encontrando dados concretos sobre os vinhos do Douro (vinhos de mesa) em anos anteriores e tendo uma carga menos importante no que diz respeito aos valores envolvidos também se sabe que, segundo a Real Companhia Velha, uma seleção de Vinhos de Mesa no espaço de apenas três anos conquistou a exigente crítica especializada, nacional e estrangeira. Com vinhos premiados nos mais prestigiados concursos internacionais, os Vinhos do Douro têm justificado o desenvolvimento de técnicas inovadoras, desde o cultivo da vinha à seleção de castas.

Daqui se pode tirar a conclusão que é muito importante a realização constante de estudos que levem à criação de técnicas inovadoras que conduzam a uma melhoria das condições de produção dos vinhos, garantido assim um produto autêntico e de elevada qualidade. Só deste modo será possível competir, com outros países produtores, como Espanha, França, Itália que têm mostrado um maior interesse pela inovação e possivelmente têm mais recursos que Portugal e em particular, que a região do Douro.

Por outro lado, nos últimos anos assiste-se a um crescimento exponencial dos vinhos monovarietais ou monocasta e por isso torna-se imperioso dispor de técnicas laboratoriais e ferramentas analíticas para poder certificar a sua autenticidade e em consequência proteger os consumidores de fraudes, especialmente quando o rótulo menciona a casta utilizada para a produção de determinado vinho. Várias técnicas vêm sendo propostas para determinar a variedade, origem e tecnologia empregues na produção de vinhos monovarietais. Parece cada vez mais haver um consenso na comunidade científica no uso de ferramentas analíticas baseadas no estudo da cor e de compostos fenólicos como meio de autenticação. Entre estas, a análise de antocianinas e sua classificação são propostas para validar a identidade das uvas usadas durante o processo de vinificação (Neves, 2008).

1.1 Objetivos

Uma das ameaças principais no que concerne à produção de vinho, com a cada vez maior concorrência do mercado interno e externo são as falsificações e imitações a que ele pode ser sujeito. Deste modo é essencial que existam ferramentas que possam impedir que isso mesmo aconteça (Marques, 2010).

Neste enquadramento, os principais objetivos do trabalho são o desenvolvimento de métodos para a:

- Classificação de vinhos monovarietais da região Demarcada do Douro, de acordo com a sua casta.
- Classificação de vinhos monovarietais da região Demarcada do Douro, de acordo com a sua idade.

Os métodos propostos baseiam-se na utilização de redes neuronais artificiais que se têm revelado muito eficazes no tratamento de problemas de reconhecimento de padrões, análise e processamento de sinais, controle de processos, robótica, classificação de dados, etc. Nos métodos desenvolvidos, a rede neuronal faz a classificação do vinho com base num conjunto de parâmetros químicos obtidos através da técnica High-Performance Liquid Chromatography usada na química analítica. A identidade de cada vinho, fornecida pelo seu perfil químico, permite discriminar a sua casta predominante e conhecer a sua idade.

1.2 Organização do documento

Esta dissertação além deste primeiro capítulo é composta por três capítulos. O segundo capítulo faz uma introdução geral aos conteúdos que vão ser abordados durante o documento como por exemplo, a importância do vinho (económica e a nível de saúde), os principais elementos químicos que fazem parte da sua composição e a apresentação dos vários métodos e equipamentos usados para a análise química. Para finalizar, neste capítulo é apresentada uma revisão bibliográfica onde são descritos outros estudos que fazem a análise de vinhos com base na sua composição química. Nalguns destes trabalhos também são usadas redes neuronais artificiais.

No capítulo terceiro apresentam-se dois métodos propostos para a classificação dos vinhos com recurso a redes neuronais artificiais. Neste capítulo são apresentados todos os resultados dos vários ensaios realizados e que cuja finalidade é discriminar o tipo de casta predominante e idade de cada vinho.

No capítulo quatro apresentam-se algumas das conclusões do estudo e são lançadas algumas ideias para futuramente complementar o que é apresentado neste documento.

2 Revisão bibliográfica

2.1 Benefícios do vinho para a saúde

Segundo German & Walzem, 2000, a atividade antioxidante num copo de vinho tinto de 150 ml é equivalente, por exemplo, ao registado em doze copos de vinho branco, duas chávenas de chá, cinco maçãs, sete copos de sumo de laranja ou vinte copos de sumo de maçã o que mostra as vantagens do mesmo.

Afonso, 2009, diz que bebendo moderadamente pode-se ambicionar uma esperança de vida maior que a de um abstémio. Mas um abstémio viverá provavelmente mais que um indivíduo alcoólico.

Segundo múltiplas conclusões, de outros tantos estudos médicos, o consumo diário e moderado de vinho (tinto em particular) trata da saúde sob múltiplos aspetos que já estão perfeitamente identificados, tem efeitos anti-oxidantes, efeito anti-inflamatório, efeito anti-trombose e efeito protetor do endotélio. O vinho tinto ingerido regularmente às refeições, pode ser um dos fatores responsáveis pelo efeito protetor contra doenças cardiovasculares (Afonso, 2009).

2.2 A química do vinho

No geral, o interesse da química é o estudo das propriedades, constituição e transformações de substâncias e materiais (Mortimer, Machado, & Romanelli, 2000), sendo os alimentos um foco de grande interesse, estando o vinho dentro desse grande grupo. Ao nível químico, o vinho tem na sua constituição dois compostos principais: não flavonóides e flavonóides.

2.2.1 Não Flavonóides

Os compostos não flavonóides compreendem os ácidos fenólicos, benzóicos e cinâmicos, e outros derivados fenólicos como os estilbenos. Estes compostos têm um papel importante nas oxidações que conduzem ao acastanhamento dos mostos e dos vinhos (Cabrita, Ricardo-da-Silva, & Laureano, 2000). Embora não exerçam uma influência direta no gosto dos

vinhos, estão implicados no aparecimento de fenóis voláteis com consequentes alterações aromáticas.

2.2.2 Flavonóides

Os compostos flavonóides compreendem as flavanas, os flavonóis e as antocianinas, revestem-se de grande importância em enologia uma vez que estão relacionados, direta ou indiretamente com a qualidade dos vinhos (Cabrita, Ricardo-da-Silva, & Laureano, 2008). São eles os responsáveis pela cor, corpo e adstringência dos vinhos e são os grandes responsáveis pelas diferenças entre uvas ou vinhos tintos e brancos, pela presença ou ausência de antocianinas.

As características ambientais sob as quais decorre o desenvolvimento dos bagos têm grande influência na quantidade dos compostos responsáveis pela cor, mas a natureza e as percentagens relativas destas substâncias obedecem a uma genética que as torna mais ou menos constantes (Calò, Tomasi, Cravero, & Di Stefano, 1994), tornando possível classificar as castas predominantes num vinho.

Dentro dos flavonóides vai dar-se especial realce às antocianinas que são um dos compostos principais dos flavonóides (são as antocianinas a base de todos os dados que são usados neste estudo como se pode constatar mais à frente) juntamente com os flavonóis, flavanóis e taninos, aos quais não se vai dar tanta importância.

As antocianinas do vinho são os pigmentos responsáveis pela cor nos vinhos tintos. São baseadas numa estrutura catiónica, diferindo pelo grau e posição de hidroxilação. Essa estrutura é apresentada na Ilustração 1.

Existem cinco tipos de antocianinas, a saber: cianidina (R1 = OH, R2 = H), peonidina (R1 = OMe, R2 = H), malvidina (R1 = OMe, R2 = OMe), petunidina (R1 = OMe, R2 = OH) e delphinidina (R1 = OH, R2 = OH) (David Calvo, Rubén Sáenz-López, Purificación Fernández-Zurbano, & Tena, 2004). As suas quantidades relativas variam com a casta, mas a malvidina é sempre maioritária (Cabrita et al., 2000).

Nos vinhos tintos as antocianinas estão na forma de mono e diglicosídeos. Esta variação dá-se pela fixação da molécula de açúcar (glucose).

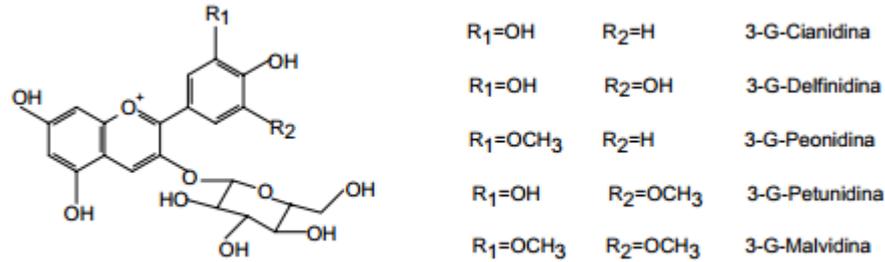


Ilustração 1 – Estrutura química de antocianinas

2.2.3 Compostos fenólicos

Durante a fase de maturação dos vinhos tintos, desde o fim da fermentação até ao engarrafamento, a presença de oxigénio é responsável por transformações químicas dos pigmentos responsáveis pela cor, essenciais ao envelhecimento (Cabrita et al., 2008).

Os compostos fenólicos, as antocianinas e os taninos em particular, são os principais constituintes dos vinhos implicados em fenómenos de oxidação, que se traduzem por alterações de cor (acastanhamento) e por uma evolução do gosto. Ao longo do envelhecimento de um vinho tinto assiste-se a uma diminuição de antocianinas monoméricas que depende tanto das condições de armazenamento como das características iniciais do vinho, podendo constituir uma fonte de informação sobre a idade do vinho (Volp, Renhe, Barra, & Stringueta, 2008).

2.3 Técnicas de análise química ao vinho

Existem diversas técnicas para efetuar a análise química de substâncias, e no caso concreto do vinho pode-se falar em: Espectroscopia de absorção atómica, Cromatografia líquida de alta eficiência, Espectrometria de Infravermelhos, entre outras (Rizzon, 2010).

2.3.1 Espectroscopia de absorção atómica

Em suma, a Espectroscopia de Absorção Atómica de Chama é uma técnica de execução rápida, fácil e com sensibilidade alta para vários elementos da tabela periódica (Paes et al.,

2012). É um método quantitativo de análise elementar, que pode ser usado para quantificar aproximadamente 70 elementos químicos.

Um espectrómetro de absorção atômica pode ser dividido em três partes principais, fonte de radiação, sistema atomizador e sistema de detecção conforme se pode ver na Ilustração 2.

Fonte de radiação: As fontes de radiação mais utilizadas em espectrometria de absorção atômica são as lâmpadas de cátodo oco (HCL). Tais lâmpadas possuem o metal que será analisado (analito) na constituição do cátodo, o que garante linhas de emissão estreitas e a radiação no comprimento de onda que o analito irá absorver. O ar, inicialmente existente nessas lâmpadas é evacuado e depois são preenchidas com gases inertes como argônio (Ar) ou néon (Ne) dependendo do elemento (Castro, 2006).

Ao aplicar uma corrente elétrica na lâmpada, os elétrons gerados irão colidir com os átomos do gás de preenchimento e estes serão acelerados contra o cátodo. Essa mesma colisão irá provocar a saída de alguns átomos do elemento que por sua vez irão colidir com outras espécies e, desta forma, ganharão energia. Esta é a energia emitida pela lâmpada e que as espécies na amostra irão absorver.

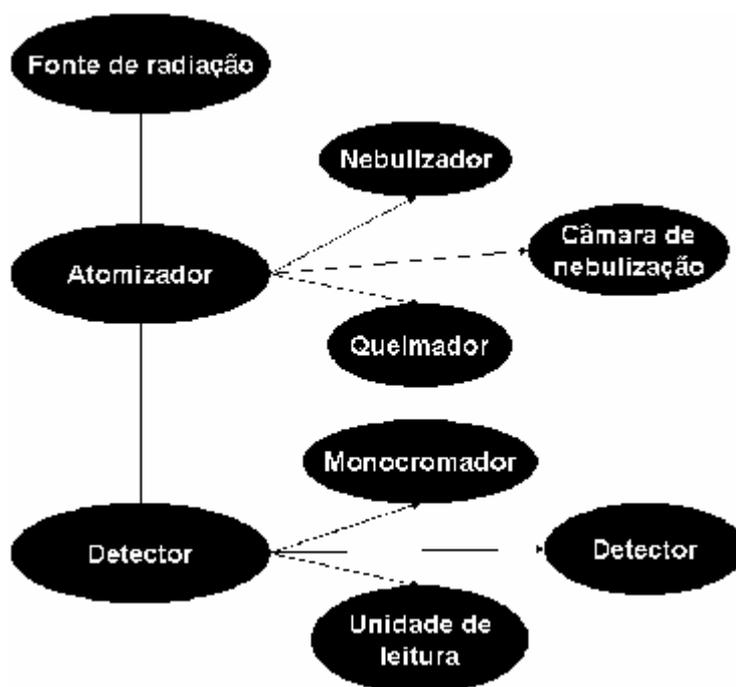


Ilustração 2 - Componentes do espectrómetro de absorção atômica

Pode-se considerar o atomizador como constituído de três partes principais, o nebulizador, a câmara de nebulização e o queimador. O nebulizador, é responsável pela conversão da amostra em aerossol (gotículas extremamente pequenas). Quanto menor forem essas gotículas maior será a eficiência do nebulizador e conseqüentemente um melhor resultado será conseguido. A câmara de nebulização é o local por onde essas gotículas devem passar antes de chegar ao queimador. No interior dessa câmara existem pequenos obstáculos chamados de flow spoiler. Os quais possuem o objetivo de drenar as gotículas maiores.

O queimador é alimentado pelos gases onde ocorrerão alguns processos, como a evaporação do solvente, a vaporização e dissociação da molécula nos seus átomos constituintes. Quanto menor a gotícula maior a velocidade com que estes processos ocorrem e assim é possível garantir a presença do átomo no caminho da radiação.

Na parte do sistema de detecção podemos incluir o monocromador, responsável por isolar o comprimento de onda de interesse, o detetor que transformará o sinal recebido em sinal elétrico e a unidade de leitura que registrará os dados obtidos (Castro, 2006).

2.3.2 Análise por espectrometria de infravermelho

O uso de FTIR (Fourier Transform Infrared Spectroscopy) em Enologia é uma metodologia emergente e muito promissora - deve ser baseado em calibrações analíticas específicas estabelecidas através da quantificação dos parâmetros a analisar, por métodos físico-químicos usuais.

A calibração global desenvolvida para a análise de vinhos permite bons resultados para todos os parâmetros o que torna a metodologia FTIR de bastante confiança.

A espectrometria FTIR permite a determinação automatizada dos parâmetros de maior importância na caracterização de mostos, de vinhos e de vinagres. Sendo uma metodologia que:

- origina rapidamente uma informação sumária sobre o produto (30 seg./amostra),
- não é poluente já que não utiliza reagentes,
- não carece de pessoal especializado para ser operada em rotina,
- é versátil, podendo ser direcionada para outras determinações complementares, a sua utilização em laboratórios de controlo perspectiva-se da maior importância.

O interferômetro utiliza a sobreposição de ondas para criar um padrão de interferência. Esta é uma forma de monocromatizar a luz proveniente de uma fonte policromática. Em espectrômetros modernos FTIR isto é realizado sob a forma de um interferômetro de Michelson, observado na Ilustração 3, que utiliza um divisor de feixe para modular a radiação de infravermelhos de tal forma que a informação de frequência pode ser obtida mais tarde.

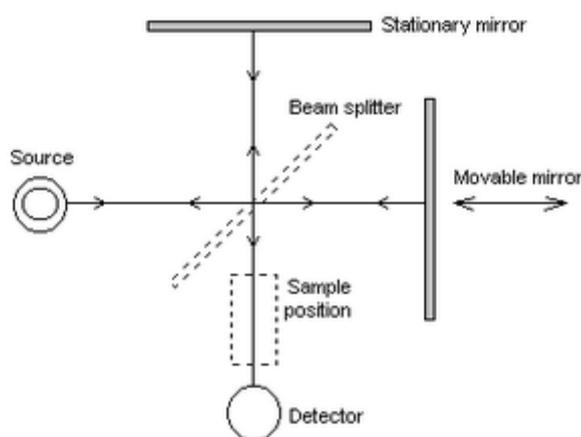


Ilustração 3 – Esquema do funcionamento dum Espectrômetro FTIR

Os espelhos exteriores são totalmente refletores, enquanto que o divisor de feixe é um espelho semi-prateado que tanto transmite como reflete. O percurso da luz a partir da fonte diverge no espelho. Oscilando o espelho móvel, a interferência ocorre quando os feixes se recombinam. A posição do espelho é precisamente controlada pela sensibilidade eletrónica e uma fonte externa monocromática.

Quando as ondas moduladas atingem a amostra, a absorção ocorre, e a onda resultante é projetada como franjas circulares de interferência do eixo em direção ao detector. O padrão de onda modulado detetado é conhecido como um interferograma, e cada ponto de dados contém informações sobre todas as frequências infravermelhas emitidas pela fonte de radiação. Desta forma, a intensidade do interferograma é mutuamente dependente do comprimento de onda e do comprimento do caminho ótico. Depois de realizar a transformada de Fourier, o interferograma pode ser representado como uma soma de todos os comprimentos de onda diferentes para obter o espectro de feixe único, que é dependente da frequência.

Antes de proceder à análise química, é necessário efetuar calibrações analíticas específicas, estabelecidas através da quantificação dos parâmetros a analisar. A seleção de

comprimentos de onda apropriados para quantificação de cada parâmetro é determinante no desenvolvimento de novas calibrações, pelo que se deve dispor de uma grande diversidade de vinhos, com características tão próximas quanto possível dos que irão ser objeto de análise. Para efetuar uma calibração, o equipamento deve ler o espectro de infravermelho de um conjunto de amostras de vinho e os parâmetros a calibrar devem ser quantificados nesses vinhos, de modo a serem usados como valores de referência (Marques, 2010).

Análise por espectrometria de infravermelho está particularmente destinada às situações em que o prazo de resposta é crítico e a utilização de métodos oficiais não se torna imperativa.

Na tabela 1 vêem-se variados parâmetros (Carvalheira, 2009) que podem ser analisados em vinhos através do método de espectrometria FTIR seguidos dos nomes das análises correspondentes.

Tabela 1 - Métodos analíticos de referência

Parâmetros	Métodos analíticos de referência
Massa Volúmica	Areometria
Título Alcoométrico Volúmico	Destilação + Areometria
Extracto Seco Total	Densimetria
Açúcares Totais	Iodometria
Acidez Total	Titrimetria
Acidez Volátil	Destilação + Titrimetria
Dióxido de Enxofre Total	Iodometria
pH	Potenciometria
Ácido Cítrico	Colorimetria
Índice de Polifenóis Totais	Espectrofotometria UV/Vis

A não utilização de reagentes, a simplicidade de operação a elevada cadência e os baixos custos de análise e de manutenção tornam a espectrometria FTIR competitiva relativamente a outros equipamentos baseados em tecnologias analíticas sequenciais disponíveis há longos anos no mercado.

2.3.3 Análise cromatográfica

A quantificação de compostos através do processo de cromatografia, consiste no conjunto de métodos físicos de separação, nos quais os componentes de uma mistura são distribuídos entre duas fases, uma fixa (fase estacionária), e outra móvel (fase móvel), como mostra a Ilustração 4.

Durante este processo de separação, os componentes da mistura são arrastados pela fase móvel ao longo da fase estacionária, sendo que a velocidade de migração destes depende da respetiva afinidade relativa para as duas fases.

Na cromatografia por HPLC, as partículas de enchimento da coluna são extremamente pequenas (diâmetros de 3 a 10 μm), sendo necessário recorrer a pressões muito elevadas para intensificar a deslocação da fase móvel ao longo da coluna. De notar que o poder de separação deste tipo de colunas é muito elevado. Depois de passadas as duas fases (estacionária e móvel) assiste-se à deteção dos dados e apresentação dos mesmos. A cromatografia em coluna moderna – HPLC foi utilizada como método de separação das antocianinas nos mais variados estudos feitos na área (Dopico-Garcia, Valentão, Guerra, Andrade, & Seabra, 2006).

As antocianinas localizam-se na película dos bagos, com exceção de algumas variedades tintureiras, que possuem a polpa corada (Rogerson et al., 2001). Em virtude da sua grande concentração nos vinhos, as antocianinas serão identificadas por injeção direta no aparelho de HPLC-PDA. Assim, as amostras são pipetadas e colocadas no aparelho onde se realiza a identificação (Mermet, Otto, & Widmer, 1997) visto no esquema da Ilustração 4.

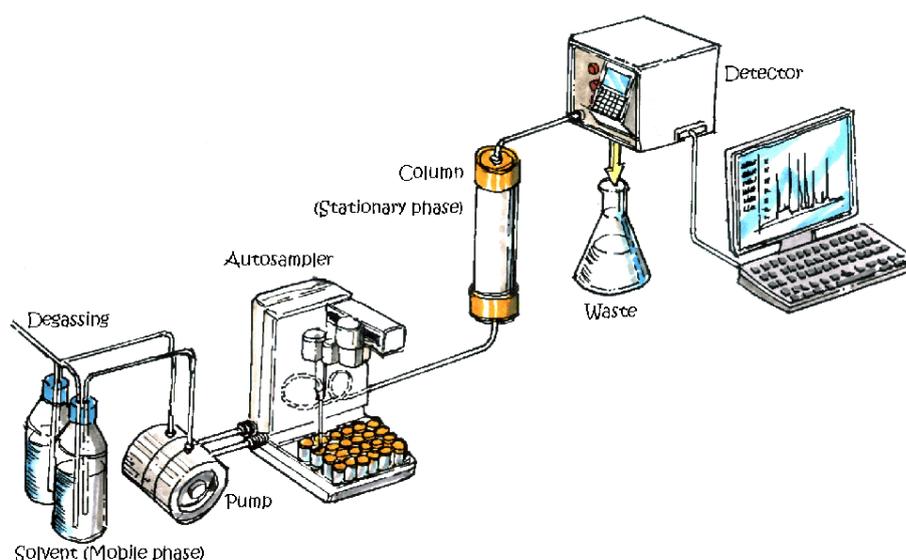


Ilustração 4 - Esquema do método de separação de antocianinas por HPLC

2.4 Vinhos Monocasta e Multicasta

Existem muitas variedades de castas de vinhos produzidas na Região do Douro, que servem para a conceção dos melhores vinhos do Mundo, nesse sentido têm-se vinhos monocasta e vinhos multicasta que como o próprio nome indica, os vinhos monocasta são constituídos por uma única casta e os vinhos multicasta são compostos por diferentes castas. Cada um deste tipo de vinhos tem as suas características próprias que de alguma forma estão relacionadas com a sua composição química.

A visão tradicional, portuguesa e europeia, insiste nas virtudes do lote (multicasta), nos benefícios evidentes da mistura de castas, na divisão de tarefas e riscos. O aproveitamento do que cada casta oferece de melhor, combinar as virtudes e minorar as imperfeições são as vantagens evidentes do lote. Se numa combinação de castas, todas diferentes, todas particulares, todas únicas, pode-se captar e combinar a estrutura ácida de uma casta com a estrutura de taninos de outra e com a matéria corante de aquela outra, o conjunto será sempre melhor do que a individualidade de cada casta (Falcão, 2007). O princípio é elementar e de fácil reconhecimento e sempre, ou quase sempre, foi essa a lógica nacional e europeia. Por isso, sempre se plantaram as vinhas misturadas, com castas misturadas, e, em muitas ocasiões, com castas brancas e tintas associadas no mesmo talhão. Preenchiam-se assim os defeitos individuais de cada casta, numa associação proveitosa para o conjunto.

Mas, por paradoxal que tal se possa afigurar, a argumentação é controversa e incerta nos paradigmas. E é questionável porque a prática se encarrega de, por vezes, desmentir a sustentabilidade lógica do raciocínio. Tanto em Portugal como no mundo! Atente-se nos exemplos dos Alvarinhos da sub-região de Monção ou dos Moscatel de Setúbal. Repare-se no exemplo dos Chardonnay e Pinot Noir da Borgonha, nos Riesling da Alemanha, Áustria e Alsácia, nos Gruner Veltliner da Áustria, nos Sauvignon Blanc e nos Chenin Blanc do Loire. Tantos exemplos de vinhos de uma só casta, conotados com a excelência, com o melhor que se faz no mundo.

2.4.1 Castas do Douro

Atualmente, nas novas plantações durienses tem-se optado por um número mais reduzido de castas, eleitas pelas suas características particulares. As principais castas da região

são: a Tinta Amarela, Tinta Barroca, Tinta Roriz, Touriga Francesa, Touriga Nacional e Tinto Cão. (Guerra & Abade, 2008) . Em seguida faz-se uma breve descrição de cada uma das castas referidas.

Touriga Nacional: chamada por alguns autores de “o Cabernet português”, pela alta qualidade, adaptabilidade a vários solos e personalidade marcante. Existe hoje em Portugal um consenso de que esta é a casta mais importante do país. No Douro a Touriga Nacional ainda ocupa uma área mínima (2,5%), mas proporcionalmente é a casta que mais cresce em área plantada na região. No passado esta casta era indesejada pois produzia muito pouco e era muito suscetível a doenças. Os vinhos de Touriga Nacional normalmente têm grande capacidade de envelhecimento, mantendo por muito tempo as suas características e por isso esta casta reina nos Portos Vintage. Além disso é uma das poucas castas do Douro que segue muito bem em “carreira solo”, em vinhos de mesa sem a adição de outras castas, o que chamamos de monocastas.

Touriga Franca: É a casta mais plantada do Douro, no passado chamada de Touriga Francesa (apesar de não ter nenhuma relação com a França). Tem muitas qualidades: adapta-se bem a qualquer tipo de solo, resiste bem ao calor, por isso é muito utilizada no Douro Superior, e resiste bem aos anos secos. No paladar geralmente são estruturados, com muitos e bons taninos, com boa capacidade de envelhecimento. Como é a casta mais plantada na região e muito usada nos Portos Vintage, geralmente nos anos em que a Touriga Franca vai bem há bons Vintages de todos os produtores.

Tinta Roriz: também chamada de Aragonês, entre vários outros nomes. É uma das castas mais conhecidas e cultivadas de Portugal, graças a sua adaptabilidade aos solos e climas e ao seu bom rendimento. No Douro é a segunda casta mais plantada, depois da Touriga Franca, e lá diz-se que esta casta é muito “aneira” (depende do ano), e é “boa quando lhe apetece”. A Tinta Roriz gera vinhos com menor intensidade de cor, mas que nos melhores anos têm aromas intensos, que podem chegar a ótima estrutura de taninos, elevado teor alcoólico, baixa acidez e pela sua estrutura de taninos, ótima capacidade de envelhecimento.

Tinta Barroca: É uma casta antiga, remonta ao século XVII e hoje é uma das mais plantadas do Douro. Gera vinhos de média intensidade de cor, taninos macios, alcoólicos e encorpados. É muito usada em lotes de Portos Vintage, pela sua graduação alcoólica, concentração de cor e boa capacidade de envelhecimento em garrafas. É mais comum encontrá-la no Baixo Corgo e Cima Corgo. No Douro Superior desidrata-se facilmente, atingindo a sobrematuração, podendo dar vinhos mais grosseiros.

Tinto Cão: Esta casta gera vinhos mais leves, de pouca cor e pouco expressivos na juventude mas que com a idade mostram grandes qualidades, resistentes à oxidação. É quase sempre misturada a outras castas.

Tinta Amarela: muito presente no Baixo Corgo. É uma casta de difícil cultivo, pois é suscetível a doenças, mas que pode gerar vinhos de ótima cor, com ótima fragrância, aromas intensos, com boa capacidade de envelhecimento. É também muito plantada no Alentejo.

2.5 Trabalhos relacionados

Fruto de uma pesquisa de artigos relacionados com o que se trata neste estudo, foram encontrados vários trabalhos em que se empregam métodos estatísticos e baseados em redes neuronais para “classificar” vinhos com base em parâmetros químicos.

Projeto VINHO DB

A identificação da origem geográfica de vinhos foi o objetivo do projeto VINHO DB, financiado pela União Europeia, que procurou estabelecer uma base de dados com os parâmetros analíticos de vinhos em diferentes países e cujo foco principal é a discriminação de amostras de vinho em relação à sua origem geográfica usando apenas alguns parâmetros químicos. A base de dados que contém cerca de 600 vinhos brancos e tintos comerciais de quatro países foi criada ao longo de um período de 3 anos, no período 2001-2004. O estudo refere que foram considerados sessenta e três parâmetros químicos para cada um desses exemplos.

A análise estatística dos dados envolveu métodos de análise de dados univariada descritiva e exploratória, bem como métodos multivariados. Os autores deste estudo, consideraram três métodos de discriminação e classificação de dados multivariados: a árvore de classificação e regressão (CART), a análise discriminante regularizada (RDA), que inclui a análise discriminante linear (LDA), a análise discriminante quadrática (ADQ), e a análise discriminante quadrática menos parcial (PLS-DA) com a eliminação de variáveis não informativas (PLS-UVE). No início a atenção dos investigadores centrou-se na discriminação dos vinhos dos quatro países, a fim de confirmar a expectativa a partir da avaliação univariada, e a partir das análises de anos individuais, tanto os vinhos autênticos como os vinhos comerciais sul-africanos foram discriminados muito facilmente dos de países do Leste Europeu. A aplicação de CART e RDA nos dados de vinho dos três Países do Leste Europeu confirmaram que a discriminação de vinhos entre a Hungria, República Checa, e a Roménia foi muito mais difícil por causa da sua localização geográfica. Finalmente os métodos de classificação PLS foram usados para construir um modelo discriminante de "um contra todos" (Römisch et al., 2009). Dando um passo adiante, Charlton et al., 2010, investigaram a possibilidade de discriminar os vinhos da base de dados de acordo com as estações de colheita e variedades de uva utilizando árvores de classificação e regressão, para identificar alguns parâmetros químicos, incluindo alguns isótopos, amins biogénicas e elementos de terras raras que discriminam os vintages e algumas variedades de uvas para os vinhos produzidos nos mesmos países. Estes parâmetros foram utilizados na avaliação da autenticidade dos vinhos.

Nariz eletrónico para análise química de vinhos italianos

A utilização de um nariz eletrónico é apresentada por Penza & Cassano, 2003, e consiste no seguinte: nove amostras de vinhos italianos (três brancos, três tintos e três rosés) de diferentes denominações de origem foram analisados para tentarem classificá-los por meio da caracterização quimiométrica dos dados obtidos a partir de uma matriz de multi-sensores de películas finas. Todos os vinhos também foram analisados para medir a sua condutividade iónica, pH e teor alcoólico. Um nariz electrónico compreendendo quatro semicondutores de óxido de metal de sensores de película fina foi utilizado para gerar uma impressão digital química típica (padrão) dos compostos voláteis presentes nos vinhos. Análise de componentes principais e redes neuronais artificiais foram aplicadas aos padrões gerados para alcançar tarefas de classificação diferentes. O desempenho da classificação de nove algoritmos diferentes de pré-processamento foi estudado com base em três parâmetros de diferentes sensores e três técnicas de normalização diferentes. A geração de padrões no vinho com

matrizes de sensores de sinal e de tratamento quimiométrico são rápidos e simples, fornecendo uma taxa de reconhecimento e uma taxa de predição bastante elevada como 100% e 78%, respetivamente. Estes resultados podem ser considerados satisfatórios e aceitáveis, com as variáveis úteis selecionadas para diferenciar estes vinhos pela sua classe.

Diferenciação de dois vinhos Andalusos de acordo com o seu conteúdo metálico usando métodos de reconhecimento de padrões

Álvarez, Moreno, Jos, Cameán, & González, 2007, usam o teor de metal (Ca, Mg, Sr, Ba, K, Na, P, Fe, Al, Mn, Cu e Zn) de vários vinhos “finos” (pertencentes a duas Denominações de Origem Andalusas (DO) determinado pelo ICP-OES) como descritores químicos discriminatórios visto que desempenham um papel fundamental na discriminação devido à sua correlação com a natureza do solo, origem geográfica e variedade da uva. Os dois vinhos de Denominação de Origem Andalusas estudados foram o Jerez-Xérès-Sherry e Manzanilla-Sanlúcar de Barrameda (classe D) e do Condado de Huelva (classe C). A Análise Linear Discriminante (ALD) e procedimentos baseados em Redes Neurais Artificiais (RNA) levaram a uma separação perfeita de classes, especialmente quando se aplicaram ‘Perceptrons’ multi-camada à RNA treinada por back-propagation.

Redes neuronais artificiais: uma ferramenta promissora para avaliar a autenticidade dos vinhos

(Álvarez et al., 2007), usam RNA na classificação dos diferentes processos de fabricação de vinho da variedade Vinhão (*Vitis vinifera*) para as culturas entre os anos de 2000 e 2004. Depois de ser treinado empregando os dados correspondentes aos anos de 2000 a 2004, as RNA demonstraram um índice de erro quadrático médio (RMSE) entre os dados reais e os calculados sempre inferiores a 0,14. Além disso, a sua operação foi verificada usando os dados anteriormente reservados de dez vinhos famosos. Como resultado, um índice de RMSE entre os dados observados e calculados sempre inferiores a 0,17 foi obtido em todos eles, confirmando a capacidade da RNA como um modelo de previsão de processos de vinho para esta variedade muito eficaz.

Caraterização Analítica do aroma de 5 vinhos

A solução mais eficiente para a avaliação estimada da idade do vinho destilado com ênfase sobre as propriedades do aroma e do sabor, enquanto um modelo matemático apropriado não for encontrado, segundo Escudero, Campo, Fariña, Cacho, & Ferreira, 2007, é o de desenvolvimento de sistemas periciais adequados e fiáveis, com base na lógica difusa e redes neuronais. Um classificador difuso e uma rede neuronal foram propostos para a classificação de vinhos destilados para cada um dos dois elementos distintos dos produtos, o aroma e o sabor. O classificador difuso é baseado num algoritmo, enquanto o sistema neuronal é uma rede sigmoidal 'feedforward' multicamadas, que é treinada usando o método de 'backpropagation'. Os resultados mostraram que ambos os sistemas de classificação difusa e neuronal realizaram bem a avaliação do aroma e do sabor dos produtos com resultados entre os 95% e 99.9%.

Classificação de vinhos produzidos em regiões específicas por Espectroscopia de UV visíveis

De acordo com Acevedo, Jiménez, Maldonado, Domínguez, & Narváez, 2007, discriminar os vinhos de acordo com a sua denominação de origem, utilizando técnicas de custo-benefício é algo que atrai a atenção de diferentes sectores industriais. Em busca da simplicidade, foram usadas técnicas espectrofotométricas de UV visíveis e diferentes técnicas estatísticas multivariadas com resultados admissíveis para caracterizar vinhos produzidos em regiões específicas. No entanto, a maior parte dos métodos relatados de classificação não exploram todas as relações estatísticas no conjunto de dados investigados e são inerentemente afetados pela presença de 'outliers'. As vantagens de tal ferramenta na discriminação são demonstradas quando as taxas de classificação são comparadas para um grande número de vinhos tintos e brancos espanhóis e as taxas de classificação superiores a 96% são alcançados. A metodologia proposta também permite a seleção dos comprimentos de onda mais relevantes para a discriminação da amostra.

3 Métodos desenvolvidos para a classificação de vinhos

3.1 Redes neuronais

3.1.1 História

As redes neurais foram desenvolvidas, originalmente, na década de 40, pelo neurofisiologista Warren McCulloch, do MIT, e pelo matemático Walter Pitts, da Universidade de Illinois, os quais, fizeram uma analogia entre células nervosas vivas e o processo eletrônico num trabalho publicado sobre "neurónios formais". O trabalho consistia num modelo de resistências variáveis e amplificadores que representavam ligações sinápticas de um neurónio biológico (Tafner, 1998).

Desde então, a partir da década 80, diversos modelos de redes neuronais artificiais têm surgido com o propósito de os aperfeiçoar e os aplicar em problemas de diversas áreas. Algumas destas propostas tendem a aperfeiçoar mecanismos internos da rede neuronal para aplicação na indústria e negócios, outras procuram aproximá-las ainda mais dos modelos biológicos originais (Tafner, 1998).

3.1.2 Definição

As Redes Neurais Artificiais são modelos matemáticos que se assemelham às estruturas neuronais biológicas e que tem capacidade computacional adquirida por meio de aprendizagem e generalização (R. César Román, O. Gonzalo Hernández, & Urtubia, 2011).

Basicamente, uma rede neuronal consiste numa estrutura de ligações, na qual o processamento se encontra distribuído por um grande número de pequenas unidades densamente interligadas. Por analogia com os sistemas biológicos estas unidades processadoras são chamadas neurónios.

Resultante do desejo dos investigadores de desenvolver sistemas com verdadeira inteligência artificial e beneficiando dos progressos da neurobiologia, os quais permitiram compreender melhor muitos dos processos do cérebro, bem como criar modelos da sua principal unidade constituinte, o neurónio; as redes neuronais artificiais constituem uma das abordagens mais inovadoras a esse problema.

Um neurónio biológico, consiste numa única célula capaz de realizar um forma simples de processamento. Cada neurónio é estimulado por uma ou mais ligações vindas de outros neurónios, chamadas sinapses, dependendo o sinal produzido, quer da força das ligações bem como da sua natureza (inibitória, excitatória, etc). Este sinal é propagado ao longo do axónio indo, por sua vez, estimular outros neurónios. O funcionamento dos neurónios artificiais baseia-se, na generalidade dos casos, neste modelo simplificado dos neurónios biológicos.

Hoje em dia, a variedade de redes neuronais artificiais a serem experimentadas é enorme, com a modelização de cada neurónio a variar desde a simples soma pesada das entradas a conjuntos complexos de equações diferenciais (Daniel Osorio, J. Ricardo Pérez-Correa, Agosin, & Cabrera, 2008); a informação a circular na rede apenas num sentido ou em vários; os neurónios atualizados simultaneamente ou com intervalos temporais entre eles, etc, resultando daqui um número quase ilimitado de combinações possíveis.

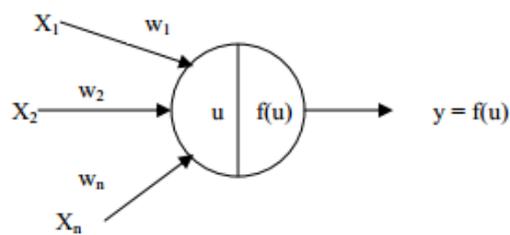


Ilustração 5 - Neurónio artificial

Vem da ilustração 5 que as entradas do neurónio correspondem ao vetor de entrada $X = [x_1, x_2, \dots, x_n]$ de dimensão n . Para cada entrada x_i , existe um peso correspondente w_i na entrada dos neurónios. Da soma das entradas x_i ponderadas pelo seu peso correspondente w_i chama-se saída linear u , onde $u = \sum w_i x_i$. A saída y do neurónio, de seu nome saída de ativação, é obtida pela aplicação de uma função f à saída linear u , indicada por $y = f(u)$. A função f é chamada de função de ativação e pode assumir várias formas, geralmente não-lineares. Das funções de ativação, as mais utilizadas são: linear, rampa, degrau e sigmóide.

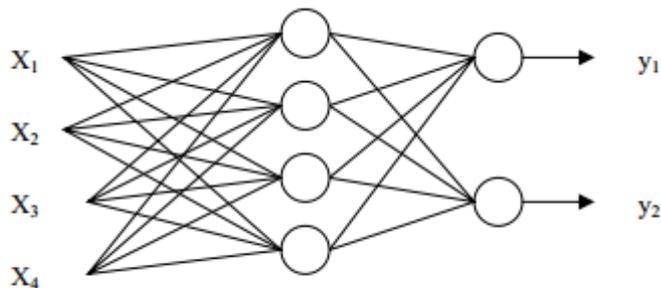


Ilustração 6 - Rede neuronal de duas camadas

Uma Rede Neural é, portanto, formada por elementos processadores simples. Cada elemento processador executa uma função simples (Jordan, 1996), mas a RNA como um todo tem capacidade computacional para resolução de problemas complexos. A estrutura apresentada na Ilustração 6 possui quatro entradas (x_1, x_2, x_3, x_4), duas saídas (y_1, y_2) e quatro neurónios na camada intermediária, e é capaz de resolver problemas de classificação ou predição.

Eis então algumas das áreas em que a performance das redes neuronais artificiais mais se faz notar:

Classificação: Uma das possibilidades mais explorada das redes neuronais consiste na sua utilização para classificar padrões de entradas em diferentes categorias (R. César Román et al., 2011).

Redução de ruído: Uma rede neuronal pode ser treinada de forma a reconhecer um determinado número de padrões de entradas, reproduzindo-os nas suas saídas. Desta forma pode ser utilizada para reproduzir o padrão original quando, este lhe é apresentado com ruído (B. Debska & Guzowska-Swider, 2011).

Predição: Um problema muito comum, em áreas como a economia ou a meteorologia, consiste em prever o valor futuro de uma variável a partir de valores históricos dessa variável ou de outras variáveis com ela relacionada. Também neste domínio sistemas baseados em redes neuronais se têm muitas vezes mostrado mais eficazes que técnicas mais tradicionais (Astray, Castillo, Ferreiro-Lage, Gálvez, & Mejuto, 2010).

É ainda de realçar que, em todas as tarefas mencionadas, as redes neuronais apresentam uma grande robustez e flexibilidade, mostrando-se capazes de lidar com alguma

incerteza (fuzzy) ou com grande margem de erro; encontrando padrões subtis, obscuros para muitas outras técnicas e resistindo a alguma não-linearidade nos dados.

Muitas outras aplicações estão constantemente a ser encontradas para as redes neuronais ao mesmo tempo que os seus fundamentos teóricos vão-se tornando cada vez mais robustos, devido ao trabalho incessante de um número cada vez maior de investigadores, a quem esta área tem despertado interesse (Raptis, Siettos, Kiranoudis, & Bafas, 1999).

Resumidamente, hoje em dia, muitos tipos de redes neuronais artificiais apresentam tanto uma sólida base teórica como encontram os mais variados campos de aplicação. Um exemplo disso está no estudo do estado da arte. A importância destes resultados, combinado com a robustez deste tipo de rede no que diz respeito a erros nas entradas, transforma-a numa ferramenta extremamente poderosa, como mais à frente no caso deste estudo se confirmará.

3.1.3 Fases

As redes neuronais para o seu bom funcionamento passam por duas fases essenciais: Treino e Teste, podendo em alguns casos incluir uma terceira fase (Validação), dependendo do software utilizado ou da rede que se pretenda implementar.

3.1.3.1 Fase de treino ou aprendizagem

Na fase de treino a RNA realiza o seu processo de aprendizagem através de exemplos que lhe são apresentados. Para um determinado conjunto de dados, o algoritmo de treino deve ser responsável pela adaptação dos parâmetros da rede, de maneira que, em um número finito de iterações do algoritmo, haja convergência para uma solução (Horner & Kúhn, 1997). O critério de convergência varia de acordo com o algoritmo e com o paradigma de teste, mas pode envolver, por exemplo, a minimização de uma função-objetivo, a variação do erro de saída ou mesmo a variação das magnitudes dos vetores de peso da rede. Basicamente, considera-se que o processo de aprendizagem tem como característica a ocorrência de estímulo da rede pelo meio externo através da apresentação do conjunto de dados (Dash, Panda, Remesan, & Sahoo, 2010). Como consequência deste estímulo, o algoritmo de teste provoca mudança nos parâmetros da rede e, finalmente, a mudança nos parâmetros

proporciona a mudança no comportamento da RNA. Espera-se que a mudança gradual no comportamento da rede resulte em melhoria gradativa do seu desempenho. A modificação do processamento ou da estrutura de conhecimento de uma rede neuronal envolve a alteração do seu padrão de interligação. Em princípio, isto pode ser feito de três maneiras: desenvolvimento de novas ligações; perda de ligações existentes na rede ou modificação dos pesos das conexões já existentes.

Em problemas de aproximação, classificação e predição, o conjunto de treino é composto por pares de entrada e saída desejada (x, y_d) , sendo caracterizada previamente a saída desejada y_d para um determinado valor de entrada x . O ajuste de pesos deve modificar a saída y de forma que a diferença entre y e y_d , ou seja, o erro, diminua a cada iteração.

3.1.3.2 Fase de Validação

Na fase de validação, a RNA calcula a saída com base nas entradas e nos pesos, validando a fase de treino que aconteceu antes. Nesta fase a rede já não será alterada, independentemente dos resultados serem positivos ou negativos.

3.1.3.3 Fase de teste

A fase de teste consiste na introdução na rede neuronal, antecipadamente treinada, de dados desconhecidos ou não. Usando dados novos poder-se-á ter uma noção mais completa do seu bom funcionamento. Nesta fase a rede jamais será alterada independentemente dos resultados obtidos sendo apenas verificados se as amostras utilizadas são bem ou mal classificadas.

3.1.4 Classificação

As redes neuronais podem ser classificadas por diferentes meios: método de treino, características e estrutura. Vai dar-se mais realce à classificação pelo método de treino que é considerada a mais importante (Karrer, Cameira, Vasques, & Benzecry, 2003).

No treino supervisionado são sucessivamente apresentadas à rede, conjuntos de padrões de entrada e correspondentes padrões de saída. Durante o processo, a rede realiza um ajustamento dos pesos das ligações entre os elementos de processamento, segundo uma determinada algoritmo de aprendizagem, até que o erro entre os padrões de saída gerados pela rede alcance um valor mínimo desejado. Por exemplo, *perceptron*, *adaline* e *madaline*,

backpropagation, são algumas de entre as dezenas de leis de treino supervisionado (Portugal & Fernandes, 1995).

No treino não-supervisionado a rede “analisa” os conjuntos de dados apresentados, determina algumas propriedades dos conjuntos de dados e “aprende” a refletir estas propriedades na sua saída. A rede utiliza padrões, regularidades e correlações para agrupar os conjuntos de dados em classes. As propriedades que a rede vai “aprender” sobre os dados podem variar em função do tipo de arquitetura utilizada e da lei de aprendizagem.

Mapa Auto-Organizável de Kohonen, Redes de Hopfield e Memória Associativa Bidirecional, são alguns métodos de treino não-supervisionado. As redes neuronais também podem ser classificadas quanto às suas características: contínua, discreta, determinística e estocástica, ou quanto à sua estrutura: redes de múltiplas camadas - multilayer feedforward network, em que o fluxo de dados segue uma única direção, e redes recursivas - recurrent network.

3.2 Caracterização do conjunto de amostras usadas no estudo

Neste estudo foi utilizada uma amostra composta por cinquenta e três vinhos de duas das castas mais importantes da região do Douro: vinte e uma amostras de vinho da casta de Aragonês e trinta e duas amostras de vinho da casta de Touriga Nacional. A idade destes vinhos varia entre um a nove anos. Os dados são considerados poucos mas eram os únicos disponíveis para estudo pelo Departamento de Química da Universidade de Trás-os-Montes e Alto Douro. Através da técnica de cromatografia HPLC obtiveram-se os valores de cianidina, peonidina, delphinidina, petunidina e malvidina relativos às antocianinas monoglucosiladas, acetilglucosiladas e cumarilglucosiladas, expressas em área ocupada por cada antocianinas.

Nos vinhos tintos as antocianinas estão na forma de mono e diglicósidos. O monoglicósido de malvidina é o constituinte principal da matéria corante das variedades de uvas tintas *Vitis vinifera*, por outro lado o diglicósido de malvidina é encontrado somente nas variedades americanas e em alguns híbridos e por esse motivo não se vai dar qualquer realce neste estudo a este último. A concentração dos monoglucósidos de malvidina nos vinhos tintos das variedades *Vitis vinifera* varia muito com a idade dos vinhos e das cepas das quais os vinhos provêm.

Juntamente com os monoglucósidos também se obtiveram os valores das antocianinas esterificadas com o ácido acético (acetilglucósido) e ácido p-cumárico (cumarilglucósido).

O objetivo deste trabalho é desenvolver um método capaz de discriminar a casta predominante da amostra e a sua idade com base nestes parâmetros químicos obtidos por cromatografia HPLC.

De modo a entenderem-se os resultados obtidos mostra-se na tabela seguinte os valores obtidos pela técnica de cromatografia HPLC em duas amostras de castas diferentes.

Tabela 2 – Valores da área de parâmetros químicos obtidos pela técnica cromatográfica HPLC em vinhos da casta de Aragonês e de Touriga Nacional

Amostra	Série	Delfinidina	Cianidina	Petunidina	Peonidina	Malvidina
Aragonês	Glucósidos	1.246	61.577	115.39	55.281	796.837
	Acetilglucósido	15.806	1.463	1.707	4.652	103.664
	Cumarilglucósido	4.038	0	1.148	13.473	115.607
Touriga Nacional	Glucósidos	10.08	25.304	48.609	70.63	882.15
	Acetilglucósido	19.717	10.487	6.533	6.62	289.225
	Cumarilglucósido	0.196	6.296	0	10.395	149.129

Para se perceber mais claramente que existem diferenças entre os valores de ambas as castas é apresentado um gráfico de seguida (Ilustração 8) que ajuda a compreender melhor algumas discrepâncias de valores em algumas antocianinas.

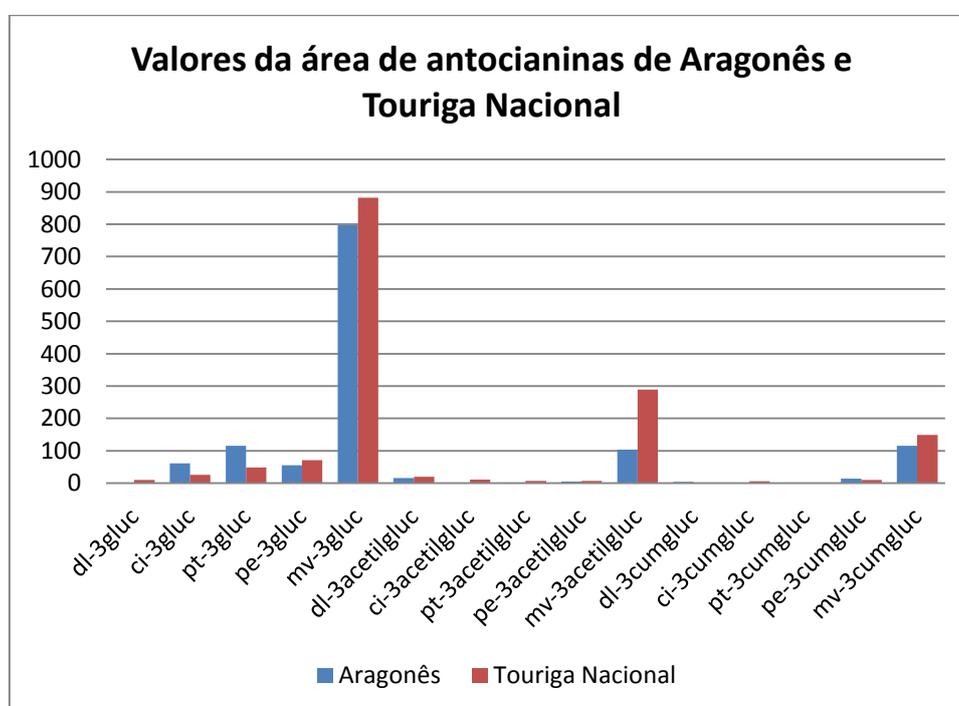


Ilustração 7 - Gráfico comparativo das antocianinas nas 2 castas em estudo

Pela Ilustração 8 pode-se ver que os valores de antocianinas de dois vinhos de castas diferentes, um vinho de Aragonês e outro de Touriga Nacional têm valores em alguns casos bastante diferentes. De seguida, são apresentados gráficos semelhantes mas mostrando a diferença de valores de antocianinas entre três vinhos da mesma casta, começando pelo vinho da casta Aragonês, seguindo-se o vinho da casta de Touriga Nacional.

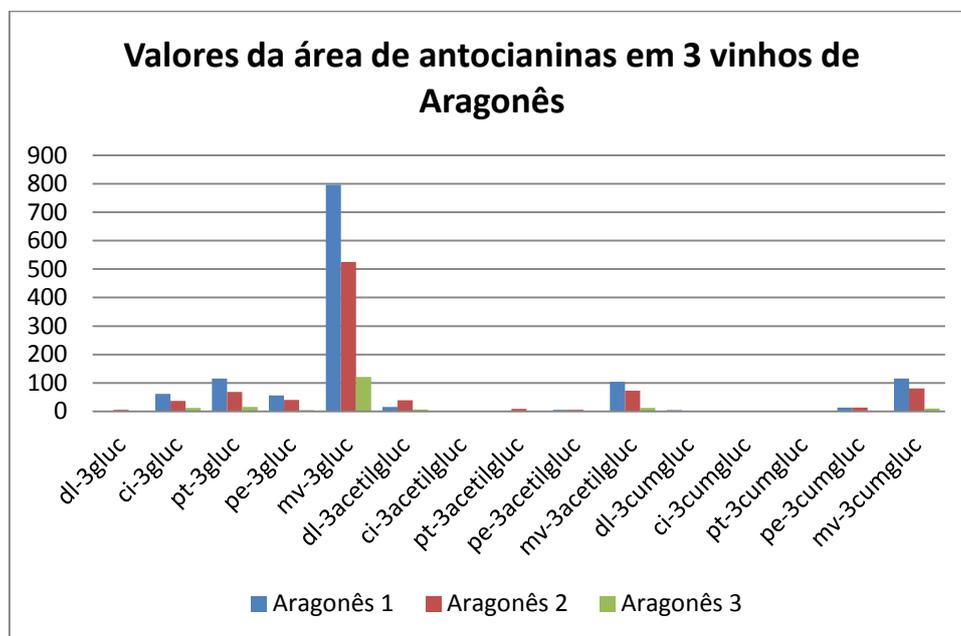


Ilustração 8 - Gráfico comparativo de antocianinas de 3 vinhos da casta de Aragonês

Através da Ilustração 8, pode ver-se que os valores de antocianinas em 3 vinhos da casta Aragonês, todos eles com 1 ano de idade, também diferem muito entre si. Cerca de metade dos valores são impercetíveis por se aproximarem muito do valor nulo, mas os que têm valores mais altos, como a antocianina malvidina-3glucósida, observa-se que o valor do primeiro vinho ronda os 800 de área, do segundo ronda os 500 de área e do terceiro muito perto dos 100 de área. Pode-se reparar que não é tão linear como se poderia pensar, fazer a distinção entre as duas castas, visto que na Ilustração 7, no gráfico dos 2 vinhos de casta diferente temos valores mais aproximados entre eles.

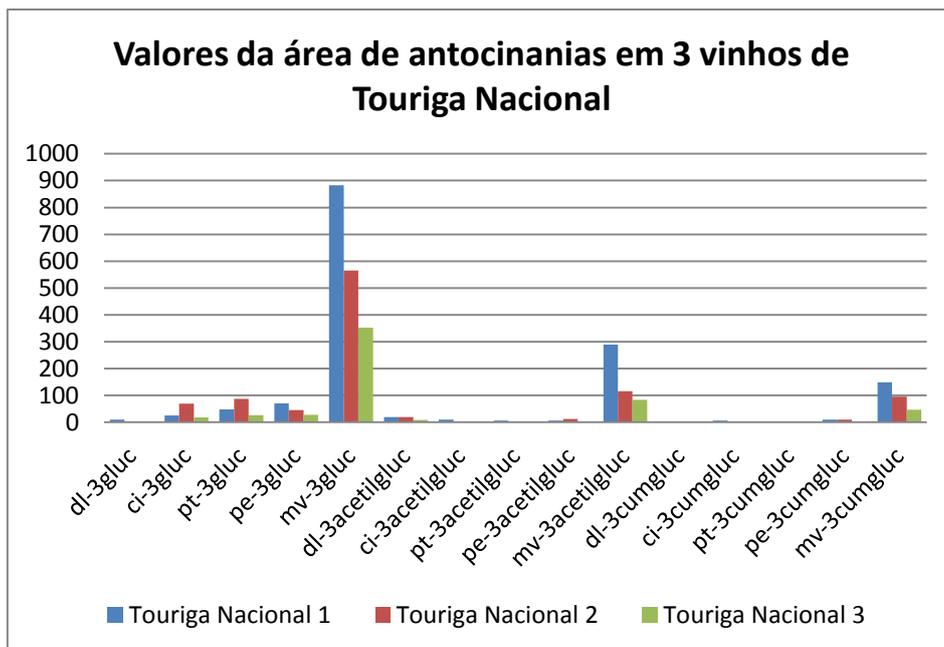


Ilustração 9 - Gráfico comparativo de antocianinas de 3 vinhos da casta de Touriga Nacional

Tal como aconteceu nos vinhos da casta de Aragonês, também acontece nos vinhos da casta de Touriga Nacional, vêm-se valores muito desiguais. Como exemplo, observa-se na Ilustração 9 valores da antocianina de malvidina-glucósida do primeiro vinho com valores a rondar os 900 de área, do segundo vinho a rondar os 600 de área e do terceiro perto dos 350 de área. Uma possibilidade de analisar estes dados é mostrar os valores percentuais de antocianinas nos mesmos vinhos que foram explicados acima. Tal informação é mostrada de seguida.

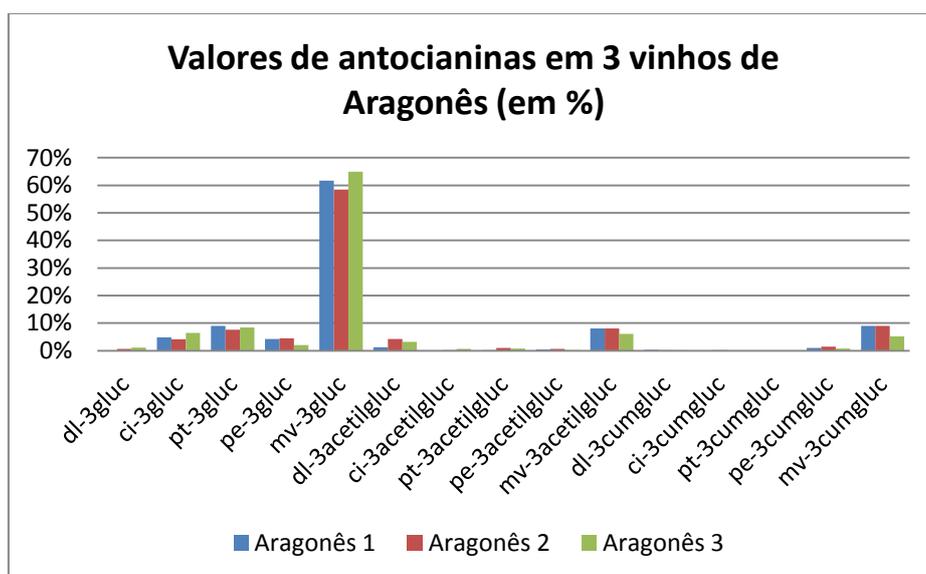


Ilustração 10 - Valores de antocianinas em 3 vinhos da casta Aragonês (em %)

Percentualmente já se consegue reconhecer (na ilustração 10) um padrão, mesmo a olho nu, com os valores em percentagem de antocianinas a aproximarem-se bastante dum valor fixo. Por exemplo, nos valores da antocianina petunidina-glucósida obtém-se um valor quase invariável para os 3 vinhos, a rondar os 8% para todos eles.

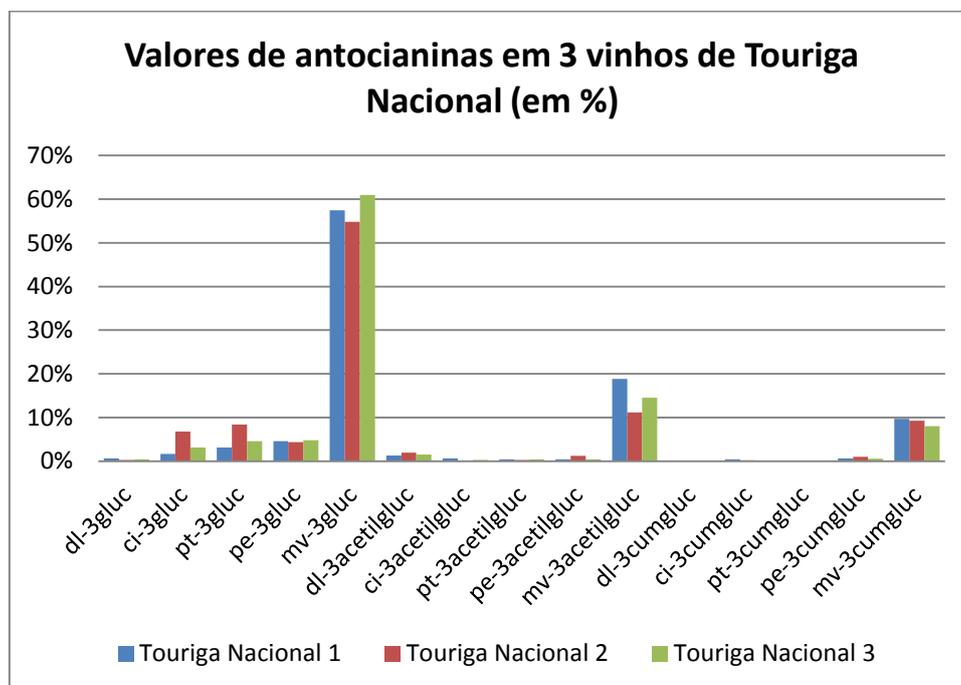


Ilustração 11 - Valores de antocianinas em 3 vinhos da casta de Touriga Nacional (em %)

Do mesmo modo que sucedeu com os vinhos de casta de Aragonês, pode-se provar através da Ilustração 11, que o mesmo sucede quando se trata de vinhos da casta de Touriga Nacional. Por exemplo, a antocianina de peonidina-glucósida apresenta para os três vinhos um valor percentual muito aproximado entre eles e que não chega a 1%.

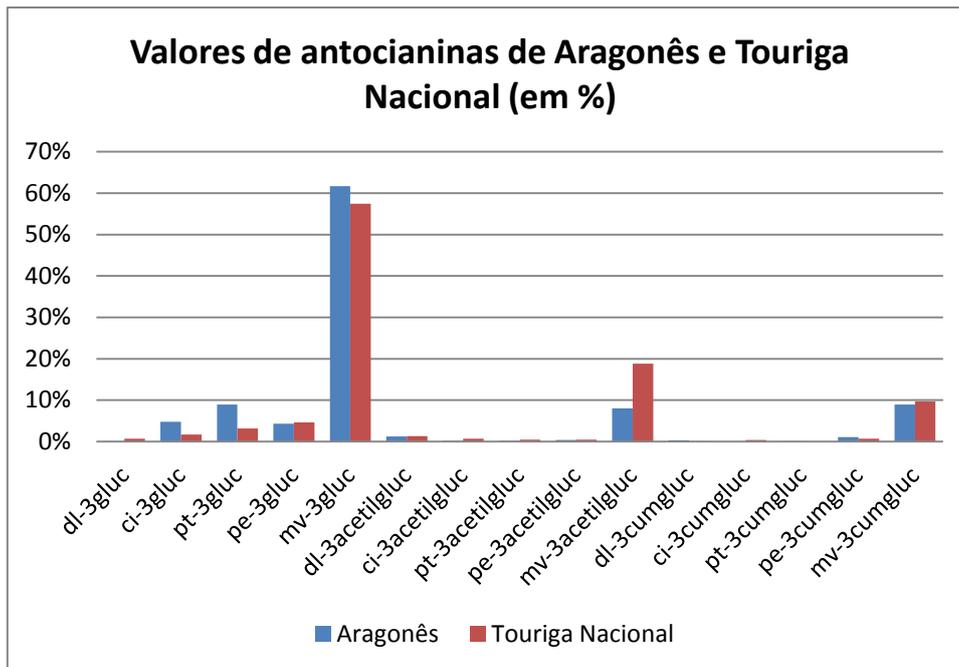


Ilustração 12 - Valores de antocianinas de Aragonês e Touriga Nacional (em %)

Para finalizar a comparação entre antocianinas de ambos os vinhos, de castas diferentes tem-se ilustração 12 que mostra a diferença em percentagem de antocianinas existentes em dois vinhos da mesma idade, mas de castas diferentes. Pode ver-se também aqui valores semelhantes entre eles, dificilmente seria possível distinguir por observação direta dos valores, dois vinhos de castas diferentes e dizer qual deles seria o Aragonês ou Touriga Nacional.

Do exposto torna-se evidente a necessidade de usar um método que possa auxiliar na classificação das amostras. De notar que os valores em análise serão os valores das áreas de antocianinas e não em percentagem. Por experiência feita, a taxa de sucesso na discriminação das castas era semelhante, usando a área ocupada e a percentagem da área de antocianinas. Optou-se ter como entradas os valores das áreas ocupadas por cada antocianina nas diferentes amostras porque é o resultado original, oferecido pelo aparelho usado na análise.

3.3 Implementação

Para a realização deste estudo usou-se o Matlab que é um ambiente de programação para desenvolvimento de algoritmos, análise de dados, visualização e cálculo numérico. Pode resolver problemas de computação numérica mais rápido do que com linguagens de programação tradicionais, como C, C++ e Fortran. O Matlab tem uma ampla gama de aplicações, incluindo processamento de sinal e imagem, comunicação, teste e medição, modelagem e análise financeira, e biologia computacional. O Matlab é a linguagem de computação técnica usada por mais de um milhão de engenheiros e cientistas a nível profissional e acadêmico (MathWorks, 2011).

Algumas características do Matlab, tendo em conta as redes neuronais e os dados que são caso de estudo:

- Existem três fases de ensaio (Treino, Validação e Teste).
- Existe a possibilidade de não haver mão humana na escolha os dados das diferentes fases.
- Em cada treino efetuado existe aleatoriedade das amostras para todas as fases, dando resultados sempre diferentes e não muito variáveis.
- Há acesso ao código, mesmo sendo criado pelo próprio software, existe a possibilidade de ser alterado de acordo com as expectativas de cada caso.
- Existe uma grande variedade de informação que se pode consultar, tanto gráfica como matricial.
- Os resultados têm uma representação amigável, até pela grande escolha que se tem, de acordo com as preferências pessoais e que mais se adequam aos dados.
- Existe muita e variada informação de apoio.

3.3.1 Configuração e parametrização da rede neuronal artificial

Neste ponto apresentam-se as configurações e a parametrizações da rede neuronal artificial usada para a classificação das amostras de vinho, de acordo com a sua casta e idade. Para cada objetivo estabelecido foi treinada uma RNA fornecendo os parâmetros químicos que lhe permite obter melhores resultados.

O primeiro objetivo é discriminar qual a casta predominante para cada amostra Como imagem ilustrativa da rede utilizada apresenta-se a Ilustração 13. Têm-se várias entradas x que

correspondem aos valores de antocianinas, as ligações a uma camada intermédia de neurónios ocultos e as saídas y onde se obterá o resultado do processamento das entradas e onde se saberá qual a casta predominante em determinado vinho, cujos valores de antocianinas foram colocados nas entradas x .

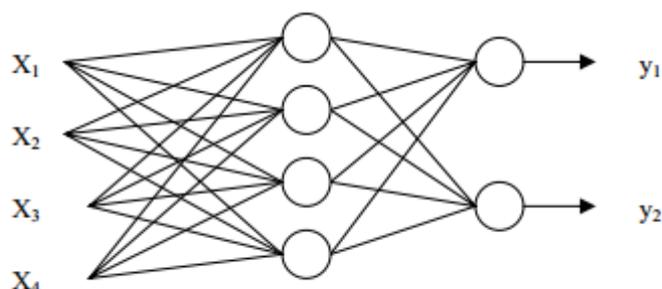


Ilustração 13 - Esquema simbólico da rede neuronal artificial usada

Antes disso são apresentadas as configurações que se verificaram mais positivas para o que é desejado, depois de uma análise anterior de muitas outras combinações possíveis.

Na tabela 4 mostram-se as configurações gerais da rede neuronal para todos os exemplos apresentados de seguida.

Tabela 3 - Parâmetros da rede no Matlab

Tipo de rede		Reconhecimento de padrões
Método de treino		Backpropagation
Função de performance		MSE
Divisão das amostras	Treino	85%
	Validação	5%
	Teste	10%

Tipo de rede: O reconhecimento de padrões é o mais ajustado para o problema em estudo porque, como o próprio nome indica, e por definição é o ato de partir de dados misturados para fazer ou tomar uma decisão baseada na categoria do padrão (Pinheiro, 2008).

Método de treino: Supervisionado com Backpropagation, consiste em ajustar os pesos de acordo com as camadas seguintes. Deste modo os erros obtidos na camada de saída serão usados para ajustar os pesos entre as ligações da camada oculta e a camada de saída. Por sua vez, o erro calculado da camada escondida será usado para ajustar os pesos da camada de entrada. Este processo será repetido até que a camada oculta esteja ajustada. Desta forma, os erros são retropropagados camada a camada com as devidas correções. O processo é repetido

por um número determinado de vezes para cada padrão de dados durante o treino até que o erro total da saída tenha convergido para um valor mínimo (Florianópolis, 2002). Após experimentar diversos métodos de treino aconselhados, este método proporcionou melhores resultados para o conjunto de dados em estudo.

Função de performance: MSE (Mean Squared Error) é uma função de performance da rede. Mede a performance de acordo com o erro quadrático médio, ficando o treino completo ao ser atingido esse mesmo nível de performance.

Divisão das amostras por grupos: No Matlab existem três fases para o funcionamento das RNA, às quais vão ser dadas percentagens diferentes conforme a importância da fase. Numa rede neuronal a fase mais importante é o Treino e é onde vão ser colocadas o maior número de amostras (85%). A seguir tem-se a fase de Teste que é onde vamos poder verificar o bom funcionamento, ou não, da rede e onde vai ser dada uma importância de 10%. Por ser menos importante e por isso para último e com menos percentagem de amostras para a sua fase, está a Validação com 5% das amostras disponíveis.

Os resultados obtidos são apresentados através de uma matriz de confusão que por definição: de uma hipótese h , oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas *versus* as classificações erradas para cada classe, sobre um conjunto de exemplos, T :

- O número de acertos, para cada classe, localiza-se na diagonal principal da matriz, em valor absoluto e percentual;
- Os demais elementos representam erros na classificação; a matriz de confusão de um classificador ideal possui todos esses elementos iguais a zero uma vez que ele não comete erros.

3.4 Resultados

3.4.1 Objetivo 1 – Classificação dos vinhos por castas

Como objetivo de discriminar os vinhos por casta fizeram-se vários ensaios de modo a definir qual a melhor parametrização para a RNA realizar a classificação.

3.4.1.1 Ensaio 1 – Classificação das castas nos 53 vinhos com todos os resultados de antocianinas

Neste ensaio foi usada uma RNA para classificação das cinquenta e três amostras de vinho, de acordo com a sua casta, usando todos os parâmetros químicos obtidos na análise HPLC. A rede é composta por quinze entradas correspondentes aos valores de antocianinas das amostras e duas saídas, que entre elas existe a camada oculta composta por sete neurónios e a camada de saída composta por dois neurónios, do tipo Ilustração 13.

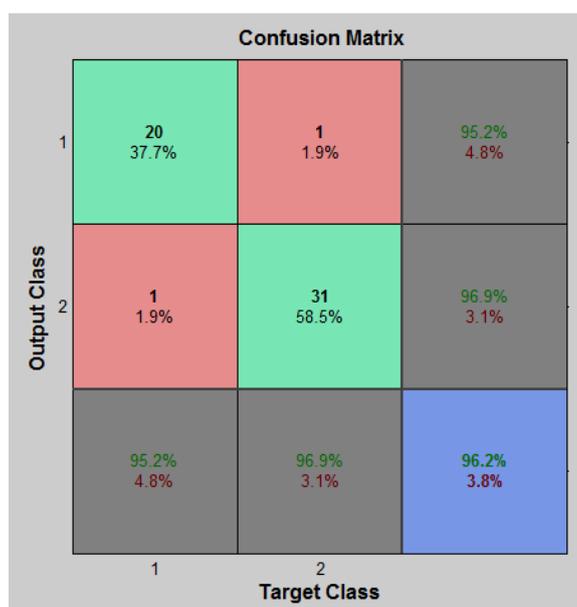


Ilustração 14 - Matriz de confusão na análise da globalidade de antocianinas

Na Ilustração 14 apresentam-se os resultados da matriz de confusão para este ensaio.

De entre os vinte e um vinhos da casta de Aragonês e os trinta e dois da casta de Touriga Nacional com todos os valores de antocianinas obteve-se uma taxa de discriminação a rondar os 95%. O ensaio repetiu-se várias vezes, obtendo-se em média uma taxa de acerto superior a 90%. No pior dos ensaios observou-se que quatro vinhos foram incorretamente classificados quanto à sua casta.

Na matriz de confusão apresentada na Ilustração 14 pode-se confirmar que no total das castas houve um erro de classificação em cada uma delas. Pela primeira coluna percebe-se que da casta de Aragonês, a rede neuronal artificial acertou em vinte amostras e errou um, num total de vinte e uma amostras de vinhos dessa casta. A segunda coluna da matriz de confusão mostra exatamente o mesmo em relação à casta de Touriga Nacional, houve acerto em trinta e uma amostras e erro numa amostra num total de trinta e duas. No quadrado azul da matriz (na última posição da última linha) é apresentado o resultado global da análise aos

cinquenta e três vinhos em que existiu erro em duas amostras, tendo por isso uma percentagem de acerto de 96.2%.

3.4.1.2 Ensaio 2 - Classificação das castas em todos os 53 vinhos com resultados de antocianinas selecionadas

Neste ensaio foi usada uma RNA para classificação das cinquenta e três amostras de vinho, de acordo com a sua casta, usando os valores que (Neves, 2008) considera mais discriminantes na totalidade das antocianinas. A rede é composta por quatro entradas correspondentes aos valores das antocianinas delfidina-glucósido, cianidina-glucósido, malvidina-glucósido, petonidina-cumarilglucósido. A camada oculta é composta por sete neurónios enquanto na camada de saída existem dois neurónios.

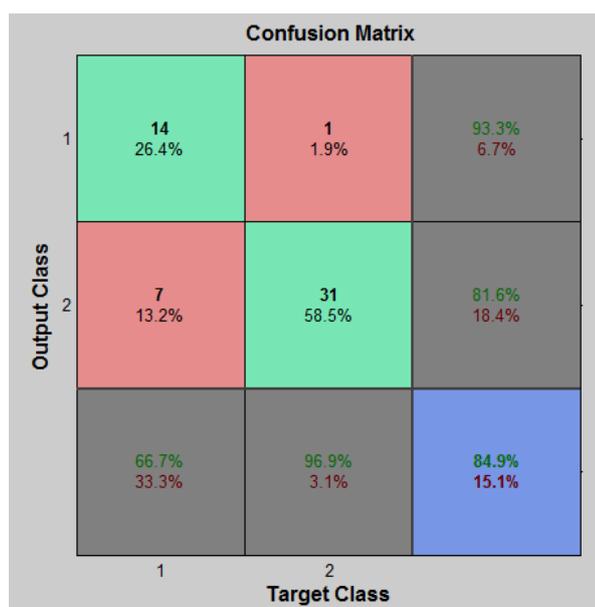


Ilustração 15 - Matriz de confusão na análise das antocianinas discriminantes

Como se pode ver pela matriz da Ilustração 15, as antocianinas que (Neves, 2008) considerou serem as mais discriminantes não se obtêm valores percentuais tão elevados como quando analisado a globalidade dos dados, caso do exemplo anterior (Ilustração 14) em que as percentagens rondavam os 95%. A explicação desta matriz é semelhante à anterior, das vinte e uma amostras de vinhos de casta Aragonês, a rede neuronal artificial deu catorze positivos e sete negativos. Quanto aos vinhos da casta de Touriga Nacional, das trinta e duas amostras

apenas se obteve uma classificação errada. Na globalidade dos resultados obteve-se uma percentagem de acertos de 84.9%.

3.4.2 Objetivo 2 – Classificação dos vinhos por idade

Como objetivo de discriminar os vinhos por idade fizeram-se vários ensaios de modo a definir qual a melhor parametrização para a RNA realizar a classificação.

3.4.2.1 Ensaio 1-Classificação dos vinhos das duas castas por idade

Numa primeira experiência fez-se a divisão dos vinhos nas seguintes sete classes: um ano, dois anos, três anos, quatro anos, cinco anos, seis anos e por fim a classe com amostras de oito e nove anos de idade, respetivamente. Os resultados são mostrados na ilustração 16 com quinze entradas correspondentes a todos os valores de antocianinas obtidos por HPLC. Na camada de saída foram colocados dois neurónios correspondentes aos vinhos das duas castas (Aragonês e Touriga Nacional). Este ensaio funcionou com quinze neurónios na camada oculta.

Output Class	1	2	3	4	5	6	7	
1	13 24.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	1 1.9%	9 17.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	90.0% 10.0%
3	1 1.9%	2 3.8%	6 11.3%	1 1.9%	1 1.9%	0 0.0%	0 0.0%	54.5% 45.5%
4	0 0.0%	0 0.0%	4 7.5%	6 11.3%	0 0.0%	1 1.9%	1 1.9%	50.0% 50.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 1.9%	4 7.5%	1 1.9%	66.7% 33.3%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 1.9%	100% 0.0%
	86.7% 13.3%	81.8% 18.2%	60.0% 40.0%	85.7% 14.3%	0.0% 100%	80.0% 20.0%	33.3% 66.7%	73.6% 26.4%
	1	2	3	4	5	6	7	
	Target Class							

Ilustração 16 - Matriz de confusão na discriminação por idade de todos os vinhos

Nos anos em que o número de amostras é mais significativo têm-se naturalmente, melhores e mais fiáveis resultados que é o que acontece com os vinhos de idade mais baixa

onde se concentram o maior número amostras. Como por exemplo temos a primeira coluna em que treze das quinze amostras de vinhos (Aragonês e Touriga Nacional em conjunto) de primeiro ano são classificadas corretamente. Em contrapartida temos a quinta coluna em que só existem dois vinhos com a idade de cinco anos não sendo nenhuma das amostras classificada com a idade correta. Como resultado final tem-se uma taxa de acerto de 73.6% na globalidade das amostras.

3.4.2.2 Ensaio 2 - Classificação das idades nos vinhos de Aragonês

De seguida, classificaram-se as idades dos vinhos para cada casta em particular, começando pela casta de Aragonês, cujos resultados estão apresentados na Ilustração 17. A rede neuronal foi construída com quinze neurónios na entrada e dois na camada de saída. O número de neurónios na camada oculta continua a ser sete.

	1	2	3	4	5	6	7	
1	5 23.8%	1 4.8%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	83.3% 16.7%
2	0 0.0%	3 14.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	4 19.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	3 14.3%	1 4.8%	0 0.0%	1 4.8%	60.0% 40.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 9.5%	1 4.8%	66.7% 33.3%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	100% 0.0%	75.0% 25.0%	100% 0.0%	100% 0.0%	0.0% 100%	100% 0.0%	0.0% 100%	81.0% 19.0%
	1	2	3	4	5	6	7	

Ilustração 17 - Matriz de confusão da classificação dos vinhos de Aragonês por idade.

Enquanto o ensaio 1 englobava todos os vinhos na sua classificação por idade, este ensaio classifica somente as idades dos vinhos de casta Aragonês nas sete classes referidas anteriormente. (um, dois, três, quatro, cinco e seis anos de idade nas primeiras seis colunas e sete e oito anos, juntos na última coluna). Têm-se resultados mais corretos em idades mais baixas por existir um maior número de amostras sendo classificadas corretamente todos os

vinhos de um, três, quatro e seis anos. (A taxa de sucesso do ensaio foi de 81% de na discriminação dos vinhos de Aragonês pela idade.

3.4.2.3 Ensaio 3 - Classificação das idades nos vinhos de Touriga Nacional

Do mesmo modo que se procedeu no ensaio 2 para os vinhos da casta de Aragonês, procede-se de seguida à classificação da idade nos vinhos de Touriga Nacional, com os resultados mostrados na Ilustração 18. O número de neurónios na camada de entrada foi de quinze e dois neurónios na saída. O número de neurónios na camada oculta é sete.

Output Class	1	2	3	4	5	6	7	
1	9 28.1%	1 3.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	90.0% 10.0%
2	0 0.0%	5 15.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	1 3.1%	1 3.1%	6 18.8%	4 12.5%	1 3.1%	0 0.0%	0 0.0%	46.2% 53.8%
4	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 9.4%	1 3.1%	75.0% 25.0%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	90.0% 10.0%	71.4% 28.6%	100% 0.0%	0.0% 100%	0.0% 100%	100% 0.0%	0.0% 100%	71.9% 28.1%
	1	2	3	4	5	6	7	
	Target Class							

Ilustração 18 - Matriz de confusão da classificação dos vinhos Touriga Nacional por idade.

Este ensaio classifica as idades dos vinhos de casta Touriga Nacional nas sete classes referidas anteriormente. Os resultados podem ser vistos na Ilustração 18 e onde se repara que nos vinhos do primeiro ano apenas existiu um erro de classificação, nas amostras do segundo ano ocorreram dois erros, obtendo-se os melhores resultados nos vinhos de três e seis anos em que a percentagem de sucesso na sua classificação, atingiu os 100%. Como resultado final tem-se uma taxa de acerto perto dos 72% na discriminação dos vinhos de Touriga Nacional pela sua idade.

3.4.2.4 Ensaio 4 - Classificação das idades em todos os vinhos

	1	2	3	4	
1	14 26.4%	1 1.9%	0 0.0%	0 0.0%	93.3% 6.7%
2	1 1.9%	9 17.0%	0 0.0%	0 0.0%	90.0% 10.0%
3	0 0.0%	1 1.9%	9 17.0%	1 1.9%	81.8% 18.2%
4	0 0.0%	0 0.0%	1 1.9%	16 30.2%	94.1% 5.9%
	93.3% 6.7%	81.8% 18.2%	90.0% 10.0%	94.1% 5.9%	90.6% 9.4%
	1	2	3	4	

Ilustração 19 - Matriz de confusão da separação dos vinhos por idade.

Para melhorar os resultados em relação à idade fez-se um ensaio com apenas quatro classes, resultantes da junção das amostras de idades mais velhas. São definidas as seguintes classes: um ano, dois anos, três anos e a classe de idade superior a três anos. Os resultados podem ser vistos na Ilustração 19 onde se observa que nos vinhos de um, dois e quatro e mais anos de idade uma amostra mal classificada em cada (primeira, segunda e quarta colunas da matriz). Nos vinhos com três anos de idade obtêm-se duas amostras mal classificadas (terceira coluna da matriz). Neste caso obtiveram-se 90.6% de acertos usando quinze neurónios na camada oculta. Resolveu-se optar também por esta abordagem porque vinhos com idade superior a três anos (por observação de ensaios anteriores) começam a ter características, à base de antocianinas, um pouco mais variáveis comparadas com os vinhos mais recentes.

3.4.2.5 Ensaio 5 - Classificação das idades nos vinhos de Aragonês

De seguida são apresentados os resultados mais importantes no que diz respeito à separação dos vinhos por idade, utilizando-se em separado as amostras dos vinhos da casta Aragonês (Ilustração 20) e dos vinhos da casta Touriga Nacional (Ilustração 21).

Output Class	1	2	3	4	
1	5 23.8%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	4 19.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	4 19.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	8 38.1%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
	1	2	3	4	
	Target Class				

Ilustração 20 - Matriz de confusão da separação dos vinhos de Aragonês por idade.

Este ensaio classifica as idades dos vinhos de casta Aragonês nas quatro classes referidas anteriormente. Os resultados podem ser vistos na Ilustração 20 onde se repara que, tanto as amostras dos vinhos com um, dois, três e superiores a três anos de idade foram classificadas corretamente. Como resultado final tem-se uma taxa de acerto ideal que equivale a 100% na discriminação dos vinhos de Aragonês pela sua idade. Neste caso foram utilizados sete neurónios na camada oculta.

3.4.2.6 Ensaio 6 - Classificação das idades nos vinhos de Touriga Nacional

Output Class	1	2	3	4	
1	9 28.1%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	1 3.1%	7 21.9%	0 0.0%	0 0.0%	87.5% 12.5%
3	0 0.0%	0 0.0%	6 18.8%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	9 28.1%	100% 0.0%
	90.0% 10.0%	100% 0.0%	100% 0.0%	100% 0.0%	96.9% 3.1%
	1	2	3	4	
	Target Class				

Ilustração 21 - Matriz de confusão da separação dos vinhos de Touriga Nacional por idade.

Este ensaio é idêntico ao anterior com a diferença de ser realizado para os vinhos da casta Touriga Nacional. Pela Ilustração 21 observa-se facilmente que à exceção de uma amostra de primeiro ano que é classificada incorretamente (primeira coluna da matriz), todas as outras amostras são bem classificadas pela rede neuronal artificial, tanto as de segundo, de terceiro e superiores a três anos de idade (segunda, terceira e quarta colunas da matriz). O acerto observado resultou numa percentagem total perto dos 97%. Como na maioria dos ensaios, também neste foram utilizados sete neurónios na camada oculta.

4 Conclusões e Perspetivas de trabalho futuro

Neste trabalho foram propostos métodos baseados em rede neuronais que discriminam quanto à casta predominante e quanto à idade amostras de vinho da região do Douro. Em termos de resultados gerais e considerando todos os ensaios realizados, verificaram-se resultados médios na ordem dos 85%. Tendo chegado a 100% ou muito próximo em alguns dos ensaios e havendo outros com valores a rondar os 70%.

Dentro do objetivo 1 que é classificar os vinhos por castas (Aragonês ou Touriga Nacional), no ensaio em que entram todos os valores das quinze antocianinas dos dois vinhos em estudo (vinte e uma amostras dos vinhos de Aragonês e trinta e duas amostras dos vinhos de Touriga Nacional) obteve-se um resultado positivo de 96%. O ensaio em que se levam a estudo somente os valores de antocianinas que são consideradas mais discriminantes por Carlos Neves (a delfidina-glucósida, cianidina-glucósida, malvidina-glucósida, petonidina-cumarilglucósida) obtiveram-se resultados menos bons, com valores a rondar os 85%.

Dentro do objetivo 2 que é a classificação dos vinhos por idade têm-se resultados mais baixos quando se tentam classificar os vinhos por todas as idades possíveis e resultados superiores quando se faz a classificação da idade dos vinhos em duas classes, uma com vinhos até três anos e outra de vinhos com três ou mais anos. Passando de percentagens de 70% no primeiro caso, para 95% no segundo.

Retirou-se como conclusão menos positiva, a baixa quantidade de amostras disponíveis para fazer o treino das várias redes, não permitiu uma correta generalização dos padrões apreendidos, prejudicando os resultados nalguns ensaios. Este problema foi significativamente mais notório aquando da tentativa de classificação de vinhos com idades mais avançadas, em alguns dos casos existia apenas uma amostra pertencente a determinado ano. Haveriam todas as condições de obter resultados bastante superiores, caso houvesse um número significativamente superior de amostras.

Este estudo pretende precaver possíveis enganar. Uma das principais ameaças na produção de vinho, com a cada vez maior concorrência dos mercados, são as falsificações e imitações a que ele pode ser sujeito. A ferramenta apresentada é uma possível solução para os casos identificados (falsificação e imitação).

Como possível trabalho futuro, fica a ideia de construir uma rede neuronal que treine com um maior número de castas, ou as principais da região do Douro, e com um número de amostras de cada classe significativamente superior.

Com isso podia pensar-se chegar a um modelo de classificação para qualquer vinho monocasta, obtendo informação suplementar da idade de vinificação.

Referências bibliográficas

- Acevedo, F. J., Jiménez, J., Maldonado, S., Domínguez, E., & Narváez, A. (2007). Classification of Wines Produced in Specific Regions by UV-Visible Spectroscopy Combined with Support Vector Machines [Electronic Version], 55.
- Afonso, J. (2009). Saúde e Vinho. *Revista de Vinhos*.
- Álvarez, M., Moreno, I. M., Jos, Á., Cameán, A. M., & González, A. G. (2007). Differentiation of 'two Andalusian DO 'fino' wines according to their metal content from ICP-OES by using supervised pattern recognition methods. *Microchemical Journal*, 87, 72-76.
- Astray, G., Castillo, J. X., Ferreira-Lage, J. A., Gálvez, J. F., & Mejuto, J. C. (2010). Artificial neural networks: a promising tool to evaluate the authenticity of wine [Electronic Version]. Retrieved 8-2-2012, from <http://dx.doi.org/10.1080/19476330903335277>
- B. Debska, & Guzowska-Swider, B. (2011). Application of artificial neural network in food classification. *ScienceDirect*.
- Cabrita, M. J., Ricardo-da-Silva, J., & Laureano, O. (2000). *Os Compostos Polifenólicos das Uvas e dos Vinhos*. Universidade Técnica de Lisboa.
- Cabrita, M. J., Ricardo-da-Silva, J., & Laureano, O. (2008). *Os Compostos Polifenólicos das Uvas e dos Vinhos*. Instituto Superior de Agronomia, Universidade Técnica de Lisboa.
- Calò, A., Tomasi, D., Cravero, M. C., & Di Stefano, R. (1994). Contributo alla caratterizzazione e classificazione varietale (*Vitis* sp), attraverso la determinazione degli antociani e degli acidi idrossicinnamoi tartarici della buccia di varietà a bacca rossa [Electronic Version], XXV,
- Carvalho, J. (2009). Análise físico-química sumária de vinhos [Electronic Version], Espectrometria de Absorção Atômica 2006).
- Charlton, A. J., Wrobel, M. S., Stanimirova, I., Daszykowski, M., Grundy, H. H., & Walczak, B. (2010). Multivariate discrimination of wines with respect to their grape varieties and vintages [Electronic Version]. Retrieved 9-2-2012,
- Daniel Osorio, J. Ricardo Pérez-Correa, Agosin, E., & Cabrera, M. (2008). Soft-sensor for on-line estimation of ethanol concentrations in wine stills. *Journal of Food Engineering*, 87.
- Dash, N. B., Panda, S. N., Remesan, R., & Sahoo, N. (2010). Hybrid neural modeling for groundwater level prediction. *Neural Comput & Applic*, 19.
- David Calvo, Rubén Sáenz-López, Purificación Fernández-Zurbano, & Tena, M. T. (2004). Migration order of wine anthocyanins in capillary zone electrophoresis. *ScienceDirect*.

- Dopico-Garcia, M. S., Valentão, P., Guerra, L., Andrade, P. B., & Seabra, R. M. (2006). Experimental design for extraction and quantification of phenolic compounds and organic acids in white "Vinho Verde" grapes. *ScienceDirect*.
- Escudero, A., Campo, E., Fariña, L., Cacho, J., & Ferreira, V. (2007). Analytical Characterization of the Aroma of Five Premium Red Wines. Insights into the Role of Odor Families and the Concept of Fruitiness of Wines. *Journal of Agricultural and Food Chemistry*, 55.
- Falcão, R. (2007). Vinhos monocasta ou vinhos de lote? Retrieved 23-7-2012, from <http://www.essenciadovinho.com/bluewine/php/apreciar.php?familia=3&id=79>
- Ferreira, C. M. S. (2010). *O Alto Douro entre o Livre-Cambismo e o Proteccionismo: A «Questão Duriense» na Economia Nacional*. Faculdade de Letras da Universidade do Porto, Porto.
- Florianópolis. (2002). Introdução as Redes Neurais Artificiais [Electronic Version],
- German, J. B., & Walzem, R. L. (2000). The health benefits of wine. *Annual Review Nutrition*, 20:561-93.
- Guerra, J., & Abade, E. (2008). Castas do Douro - Caracterização enológica de castas tintas [Electronic Version]. Retrieved 8-2-2012, from http://www.drapn.min-agricultura.pt/drapn/conteudos/fil_trab/Castas%20Tintas%20do%20Douro.pdf
- Horner, H., & Kúhn, R. (1997). *Neural Networks*.
- Jordan, M. I. (1996). Neural Networks [Electronic Version],
- Karrer, D., Cameira, R. F., Vasques, A. S., & Benzecry, M. d. A. (2003). *Redes neurais artificiais: conceitos e aplicações*. Paper presented at the Profundão – IX Encontro de Engenharia de Produção da UFRJ.
- Marques, J. P. T. M. (2010). *A Valorimetria dos Stocks de Vinho do Porto: uma abordagem contabilístico-fiscal*. Universidade Aberta.
- MathWorks. (2011). Matlab - The Language of Tecchnical Computing [Electronic Version]. Retrieved 2-8-2012, from mathworks.com/products/matlab
- Mermet, J.-M., Otto, M., & Widmer, H. M. (1997). *Analytical Chemistry: WILEY-VCH*.
- Mortimer, Machado, & Romanelli. (2000). Química Nova [Electronic Version],
- Neves, C. (2008). Universidade de Trás-os-Montes e Alto-Douro, Vila Real.
- Oliveira, J. d. M. L. B. (2010). *Denominações de Origem e Indicações Geográficas - protecção e impacto socio-económico*. Instituto Superior de Economia e Gestão.
- Paes, M. A. T., Rocha, M. d. A., Couto, M. B. d., Silva, T. A. d., Azevedo, T. U. P. d., Junior, A. I. d. S., et al. (2012). Determinação de Metais em Vinhos Brasileiros por Absorção Atômica com Chama.

- Penza, M., & Cassano, G. (2003). Chemometric characterization of Italian wines by thin-film multisensors array and artificial neural networks [Electronic Version]. Retrieved 8-2-2012, from <http://www.journals.elsevier.com/food-chemistry/#description>
- Pinheiro, A. M. G. (2008). Introdução ao Reconhecimento de Padrões [Electronic Version], Portugal, M., & Fernandes, L. (1995). Redes Neurais Artificiais e Previsão de Séries Econômicas: uma introdução.
- R. César Román, O. Gonzalo Hernández, & Urtubia, U. A. (2011). *Prediction of problematic wine fermentations using artificial neural networks*. Paper presented at the Bioprocess Biosyst Eng.
- Raptis, C. G., Siettos, C. I., Kiranoudis, C. T., & Bafas, G. V. (1999). Classification of aged wine distillates using fuzzy and neural network systems [Electronic Version]. *Journal of Food Engineering*, from www.elsevier.com/locate/jfoodeng
- Rizzon, L. A. (2010). Metodologia para análise de vinho [Electronic Version],
- Rogerson, F. S. S., Castro, H., Fortunato, N., Azevedo, Z., Macedo, A., & Freitas, V. A. P. D. (2001). Chemicals with Sweet Aroma Descriptors Found in Portuguese Wines from the Douro Region: 2,6,6-Trimethylcyclohex-2-ene-1,4-dione and Diacetyl. *Journal of Agricultural and Food Chemistry*, 49.
- Römisch, U., Jäger, H., Capron, X., Lanteri, S., Forina, M., & Smeyers-Verbeke, J. (2009). Characterization and determination of the geographical origin of wines. Part III: multivariate discrimination and classification methods. *Eur Food Res Technol*.
- Tafner, M. A. (1998). Redes Neurais Artificiais: Aprendizado e Plasticidade. *Cérebro & Mente*.
- Volp, A. C. P., Renhe, I. R. T., Barra, K., & Stringueta, P. C. (2008). Flavonóides antocianinas: características e propriedades na nutrição e saúde. *Rev Bras Nutr Clin*, 23.
- Zorzan, C. (2006). *Comparação Físico-química e sensorial de vinhos tintos varietais Ancellotta e Tannat*. Centro Federal de Educação Tecnológica de Bento Gonçalves.