



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE ZOOTECNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA

JULLYANE IVO GARCIA DA SILVA

ANÁLISE DE IMAGENS E BIOMETRIA NA IDENTIFICAÇÃO DE MATERNIDADE
DE OVOS DE CODORNA

FORTALEZA

2020

JULLYANE IVO GARCIA DA SILVA

ANÁLISE DE IMAGENS E BIOMETRIA NA IDENTIFICAÇÃO DE MATERNIDADE DE
OVOS DE CODORNA

Dissertação apresentada ao Programa de Pós-Graduação em Zootecnia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Zootecnia. Área de concentração: Produção e Melhoramento Animal.

Orientador: Prof. Dr. Luciano Pinheiro da Silva.

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S58a Silva, Jullyane Ivo Garcia da.
Análise de imagens e biometria na identificação de maternidade de ovos de codorna / Jullyane Ivo Garcia da Silva. – 2021.
32 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de Pós-Graduação em Zootecnia, Fortaleza, 2021.
Orientação: Prof. Dr. Luciano Pinheiro da Silva.
1. Análise de imagem. 2. Antiagrupamento. 3. Aprendizado de máquina. 4. Coturnix. 5. Fenômica. I. Título.

CDD 636.08

JULLYANE IVO GARCIA DA SILVA

ANÁLISE DE IMAGENS E BIOMETRIA NA IDENTIFICAÇÃO DE MATERNIDADE DE
OVOS DE CODORNA

Dissertação apresentada ao Programa de Pós-Graduação em Zootecnia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Zootecnia. Área de concentração: Produção e Melhoramento Animal.

Aprovada em: 21 / 12 / 2020.

BANCA EXAMINADORA

Prof. Dr. Luciano Pinheiro da Silva (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Fabyano Fonseca e Silva
Universidade Federal de Viçosa (UFV)

Prof. Dr. Jeferson Corrêa Ribeiro
Instituto Federal Goiano (IF Goiano)

AGRADECIMENTOS

Primeiramente agradeço a toda minha família, que sem ela nenhum dos meus sonhos teriam se tornado realidade.

A todos meus amigos da vida, e em especial Gisa Herbster e Claudia Vanessa, pela companhia, amor e afeto durante esses anos, compartilhando os momentos alegres e aflitos, além de me darem forças para continuar quando as coisas não iam bem, obrigado por terem cuidado e zelado por essa amizade, amo muito vocês.

A minha grande amiga e também companheira, Joyce Felix, por sempre estar disposta a oferecer um ombro amigo e a ouvir sem hesitar as aflições dessa estudante ansiosa, serei eternamente grata por seu cuidado, afeto e por todo carinho dado a mim.

Ao professor Luciano Pinheiro da Silva, agradeço por todos os anos de orientação e por ser um orientador que sempre esteve disposto a ajudar e a ensinar.

À Universidade Federal do Ceará pelo apoio a execução deste trabalho e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), não somente pelo apoio financeiro, mas principalmente por crer na ciência como futuro promissor do nosso país e de seus pesquisadores.

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

“Apressa-te a viver bem e pensa que cada dia é,
por si só, uma vida.”

Sêneca.

RESUMO

Codornas são animais modelo para diversas áreas das ciências da vida, bem como uma importante espécie para produção de carne e ovos ao redor do mundo. A produção de ovos, seja como alimento ou como meio reprodutivo, frequentemente se dá em gaiolas coletivas, dificultando a identificação individual para controle da produção e em programas de melhoramento genético. O objetivo do presente trabalho foi testar algoritmos de aprendizado estatístico e esquemas de alojamento de codornas que otimizem a identificação da produção, baseada em características externas de seus ovos. Foram utilizados dados de 90 aves, com no mínimo dez ovos cada, sendo testados quatro algoritmos de aprendizado estatístico usando validação cruzada, além de verificar a influência do número de codornas por gaiola e métodos para designar aves a cada gaiola. Os modelos de melhor desempenho consistem no uso de dez variáveis do ovo: peso, largura, altura, proporção de área da casca com padrões, intensidade de vermelho, de azul e de verde, matiz, saturação e luminosidade da cor de fundo dos ovos. A acurácia da classificação é aumentada em gaiolas com menor número de codornas (máxima com três aves) e com direcionamento para aumento da variância dentro de gaiola. O método apresentado mostra viabilidade para uso prático e tem possibilidade de melhoria pelo uso futuro de novas variáveis e métodos mais avançados.

Palavras-chave: Análise de imagem. Antiagrupamento. Aprendizado de máquina. Coturnix. Fenômica.

ABSTRACT

Quail are animal models for many fields of life sciences, as well as an important species for meat and egg production worldwide. Egg production, both as food or as for breeding purposes is often based on multiple-hen cages, hindering individual identification for control of production and in-breeding programs. The aim of the present study was to test algorithms of statistical learning and housing schemes for quail that optimize individual laying control based on quail egg external features. 90 birds were used, with a minimum of ten eggs each, four statistical learning algorithms with cross-validation were tested, as well as verifying the influence of number of quail per cage and methods to assign the birds to each cage. Model with better performance consist in the use of ten variables per egg: weight, height, width, eggshell ratio of patterned area, hue, saturation, lightness, intensity of red, green, and blue of egg background color. The classification accuracy increases when cages have less quail (maximum of three birds) and aimed to increase inside-cage variance. The present method shows feasibility for real-world data and with possibility of improvements with new features and more advanced methods.

Keywords: Anticlustering. Coturnix. Image analysis. Machine learning. Phenomics.

LISTA DE FIGURAS

- Figura 1 – Exemplo de segmentação de imagem em preto e branco para quantificação de área coberta por máculas em casca de ovo de codorna..... 16
- Figura 2 – Exemplo de separação de canais RGB de imagem de um ovo. A primeira imagem é a original, seguida das bandas de vermelho, verde e azul, respectivamente..... 17
- Figura 3 – Comparação de algoritmos de aprendizado estatístico para classificação em bancos de dados sintéticos..... 18
- Figura 4 – Ilustrações do uso dos métodos de agrupamento (parte inferior) e anti-agrupamento (parte superior) para formação de subconjuntos de observações com o objetivo de reunir pela variância (parte esquerda) ou pela diversidade (parte direita) 22
- Figura 5 – Ganho em precisão pela inclusão de recursos para classificação de ovos: classificados por importância obtidos por meio do algoritmo RFE..... 27

LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas das características do ovo.....	26
Tabela 2 – Correlação entre características do ovo.....	26
Tabela 3 – Métricas de validação cruzada de algoritmos de aprendizado de máquina utilizados pelo número de codornas por gaiola na classificação de ovos com base em suas características.....	27
Tabela 4 – Validação de conjuntos externos utilizando diferentes sistemas de alocação de gaiolas (Aleatório ou Antiagrupamento) e considerando a classificação por dia de postura.....	28

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Fatores que afetam a pigmentação da casca	13
<i>1.1.1</i>	<i>Fatores Genéticos</i>	13
<i>1.1.2</i>	<i>Fatores Fisiológicos</i>	14
1.2	Análise de imagem	15
<i>1.2.1</i>	<i>Segmentação de imagem em preto e branco</i>	15
<i>1.2.2</i>	<i>Sistemas RGB e HSL</i>	16
1.3	Aprendizado estatístico seus algoritmos	17
<i>1.3.1</i>	<i>Análise Discriminante</i>	18
<i>1.3.2</i>	<i>Random Forest</i>	18
<i>1.3.3</i>	<i>k-vizinho mais próximo</i>	19
<i>1.3.4</i>	<i>Máquinas de Vetores de Suporte</i>	20
1.4	Métricas de avaliação de modelos de aprendizado estatístico	20
<i>1.4.1</i>	<i>Acurácia</i>	20
<i>1.4.2</i>	<i>Coeficiente kappa</i>	21
1.4	Antiagrupamento	21
2	OBJETIVOS	23
2.1	Objetivo geral	23
2.2	Objetivos específicos	23
3	MATERIAL E METÓDOS	24
3.1	Coleta de dados e definição de variáveis	24
3.2	Métodos de classificação	24
4	RESULTADOS	25
5	DISCUSSÃO	28
6	CONCLUSÃO	30
	REFERÊNCIAS	31

1 INTRODUÇÃO

A codorna é uma pequena ave galiforme domesticada por volta do século XII, principalmente como ave canora, recentemente usada na produção industrial de alimentos, fornecendo carne e ovos, e para estudos científicos em genética, fisiologia, comportamento, embriologia entre outros, usada como modelo animal para aves e até mesmo para humanos (MINVIELLE, 2009). As vantagens do uso científico e comercial das codornas são devido ao seu pequeno tamanho, baixo consumo de alimentos, reprodução precoce e um número relativamente alto de animais em pequenas instalações (HUSS; POYNTER; LANSFORD, 2008).

Ovos de aves pertencentes à família dos Fasianídeos que incluem as codornas, costumam apresentar manchas de formas assimétrica, também conhecidas como máculas (KILNER, 2006), sendo principalmente compostas por protoporfirina e biliverdina. As protoporfirinas dão origem a cores que são tons de vermelho, amarelo e marrom, enquanto os compostos de biliverdina fornecem tons de azul e verde (KENNEDY; VEVERS, 1976).

A protoporfirina e a biliverdina são produtos do catabolismo do grupamento heme, que atua em várias vias metabólicas devido a sua capacidade de se ligar e liberar oxigênio. Portanto, essas moléculas podem ser derivadas diretamente dos glóbulos vermelhos ou podem ser sintetizadas *de novo* no útero das fêmeas, e apenas adicionadas à casca pouco antes da postura (SPARKS, 2011).

Sabe-se que a criação de codornas, tanto para produção de ovos como para reprodução, é feita de forma coletiva, onde vários animais ficam alojados em uma mesma gaiola. Esse tipo locação costuma vir acompanhado de interações sociais negativas, devido principalmente ao estresse causado por lesões, bicagem, além da competição por alimento, fazendo com que a produtividade em grupo possa estar comprometida, nós mostrando que a correlação existente entre produção de aves em grupos e individual é negativa (MUIR; BIJMA; SCHINCKEL, 2013; MUIR; CRAIG, 1998).

Quando criadas em gaiolas de criação coletiva, os ovos são coletados sem a identificação da codorna que realizou a postura. Isso implica que a codorninha nascida de um ovo específico terá genealogia desconhecida e que não haverá informações sobre o número total de ovos individuais para o manejo da produção.

A variação da cor entre as ninhadas de diferentes fêmeas é maior do que a variação dentro de uma ninhada, mostrando um possível método de identificação da maternidade pela

cor do ovo, podendo ser realizado até mesmo por profissional treinado (SEZER; TEKELIOGLU, 2009).

O reconhecimento de padrões é uma ciência que tem como objetivo a classificação de objetos em várias categorias ou classes já definidas, sendo parte integrante da maioria dos sistemas de inteligência de máquina que são desenvolvidos para tomada de decisões. Dependendo da área esses objetos podem ser imagens, formas, dígitos ou qualquer tipo de medição que precise ser classificada (THEODORIDIS e KOUTROUMBAS, 2008).

A correlação entre as variáveis do ovo registradas ao longo do tempo pode indicar previsibilidade de um registro futuro, e essa correlação é chamada de repetibilidade. Altos coeficientes de repetibilidade foram relatadas para peso do ovo ($> 0,76$), comprimento do ovo ($> 0,94$), índice de forma do ovo (proporção largura / altura $> 0,52$) e para cor do ovo (entre 0,58 e 0,85) (AKPA *et al.*, 2008; AMAO; OJEDAPO; OGUNDIPE, 2013; GOODMAN, 1965; SATO; IDA; INO, 1989; SEZER, 2007; SEZER; TEKELIOGLU, 2009).

1.1 Fatores que afetam a pigmentação da casca

1.1.1 Fatores Genéticos

As assinaturas para reconhecimento são comuns em diversos táxons, podendo ser químicos como feromônios, auditivos como vocalizações ou visuais como coloração, são usadas principalmente para reconhecimento individual dos organismos com base em características distintivas. Nas aves os padrões de coloração podem ser assinaturas visuais de identidade da prole, permitindo que os pais possam reconhecer seus ovos na ninhada ou no substrato do ninho (JOHNSTONE, 1997; KILNER, 2006; STODDARD; KILNER; TOWN, 2014; TIBBETTS; DALE, 2007).

O reconhecimento é necessário para uma gama de comportamentos sociais. Esse reconhecimento envolve uma ampla dimensão, como o reconhecimento próprio, do parceiro, gênero, vizinho, oponente, amigo, espécie, predador e presa. Já o reconhecimento individual (RI) se refere a um subconjunto de reconhecimento que acontece quando um organismo identifica outro a partir de suas características individuais distintas (DALE; LANK; REEVE, 2001). Durante o RI, o sinalizador é reconhecido por características de reconhecimento que são

únicas e o receptor aprende essas características e usa, para identificar o sinalizador em outras possíveis interações (TIBBETTS; DALE, 2007).

No caso das codornas, a coloração da casca pode ser usada pelas aves como estratégia de camuflá-los e identificá-los no substrato. A manutenção da cor da casca do ovo, pode ser um comportamento adaptativo que evoluiu para facilitar a camuflagem do ovo e diminuir o risco de predação em espécies como codornas japonesas que põem ovos maculados e nidificam no solo (DUVAL *et al.*, 2013).

Muitas vezes não se sabe qual característica fenotípica específica dos ovos, as aves usam como sistema de reconhecimento, pois existe variação da coloração em relação a forma, cor média de ovo inteiro e padrões de maculação (LYON, 2003). Alguns trabalhos nos mostram que a variação da forma, tamanho e das máculas entre codornas é alta, mais a variação dentro de ovos da mesma ninhada é baixa, auxiliando no reconhecimento (SEZER; TEKELIOGLU, 2009).

Medidas quantitativas de pigmentação da casca são utilizadas na indústria de produção de ovos com o objetivo de torná-los mais atrativos para o consumidor, como mensurações feitas com refletômetro estão associadas a luminosidade e matiz. Características relacionadas a estes padrões apresentam herdabilidade entre 0,46 e 0,50 sendo usadas principalmente para melhoramento da coloração de ovos vermelhos (FÖRSTER *et al.*, 1996; ZHANG *et al.*, 2005).

Herdabilidades elevadas garantem que fenotipicamente as colorações sejam bem similares ao longo do tempo, devido a menor influência de efeitos ambientais nestas características. A grande variedade de padrões de cores está relacionada a adaptação das codornas aos substratos, sendo que essa intensidade de cor e tamanho da cobertura de mácula levam ao comportamento de busca de um micro-habitat mais adequado, garantindo a conservação da diversidade de cores e padrões nas codornas (LOVELL *et al.*, 2013).

1.1.2 Fatores Fisiológicos

A coloração da casca do ovo em codornas está fortemente relacionada com a condição fisiológica da fêmea, em particular, com sua capacidade antioxidante (MORENO; OSORNO, 2003). Essa afirmação é baseada no estudo dos principais pigmentos que compõem a coloração da casca, a biliverdina, um pigmento azul, com potente ação antioxidante (KAUR *et al.*, 2003), e a protoporfirina, principal pigmento responsável pela cor das máculas da casca

(GOSLER; HIGHAM; JAMES REYNOLDS, 2005), conhecido por possuir propriedades oxidantes (AFONSO; VANORE; BATLLE, 1999).

Para explicar esse contexto, um estudo com codornas japonesas, mostrou que fêmeas em dietas de alta qualidade, ou seja, *ad libitum*, elevaram a deposição de biliverdina na casca, propondo que somente fêmeas com condições físicas superiores podem enfrentar o balanceamento entre pigmentação das cascas dos ovos e controle do estresse oxidativo. Já fêmeas que enfrentam desafios de condições ambientais flutuantes, como disponibilidade de alimentos, baixa condição corporal e estresse elevado depositam menor quantidade de biliverdina nas cascas dos seus ovos, para facilitar a redução da oxidação (DUVAL *et al.*, 2013; MORENO; OSORNO, 2003).

A deposição de quantidades crescentes de protoporfirina na casca do ovo em fêmeas em baixa condição corporal, ocorre como uma forma de eliminar mais eficientemente ao serem depositadas na casca, removendo esse oxidante do corpo (DUVAL *et al.*, 2013; MORENO; OSORNO, 2003). Ressaltando que o acúmulo em excesso de protoporfirina no fígado induz o estresse oxidativo, levando a um rápido aumento na atividade das enzimas antioxidantes e seu acúmulo pode causar danos no órgão (AFONSO; VANORE; BATLLE, 1999). Portanto como vimos, a coloração da casca dos ovos de codorna, podem refletir tanto a qualidade do ovo, como a condição do corpo da fêmea (DUVAL *et al.*, 2013).

Fatores como a uniformidade de manejo, controle de temperatura, circulação de gases, diminuição de estresse e fornecimento de alimento de qualidade podem favorecer a constância dos fenótipos ao longo da vida do animal, contribuindo para maior acurácia de métodos de identificação através do ovo.

1.2 Análise de imagem

Sistemas de cores são utilizados frequentemente em sistemas de visão computacional para gerar parâmetros quantitativos úteis para uso no aprendizado estatístico. No presente estudo focamos na segmentação através de imagens transformadas em preto e branco, o sistema RGB e sistema HSL.

1.2.1 Segmentação de imagem em preto e branco

Uma das possíveis técnicas para separação de formas, ou segmentação, de imagens é a uniformização de cores similares. A uniformização de cores pode ser feita através

de métodos multivariados, tais como o processo dos k-vizinhos mais próximos (CELEBI, 2009) ou através de transformação da imagem em preto e branco através de um limiar.

Para segmentação em preto e branco é necessária a partir de uma imagem original, a conversão para uma escala de cinza e posterior estabelecimento de um limiar em que se separa branco do preto como cores únicas da imagem. Geralmente se utiliza o ponto médio da escala de cor, que vai do zero ao 255 (Figura 1). A partir deste processo é possível quantificar a proporção da área que é coberta por cores próximas ao preto ou ao branco.

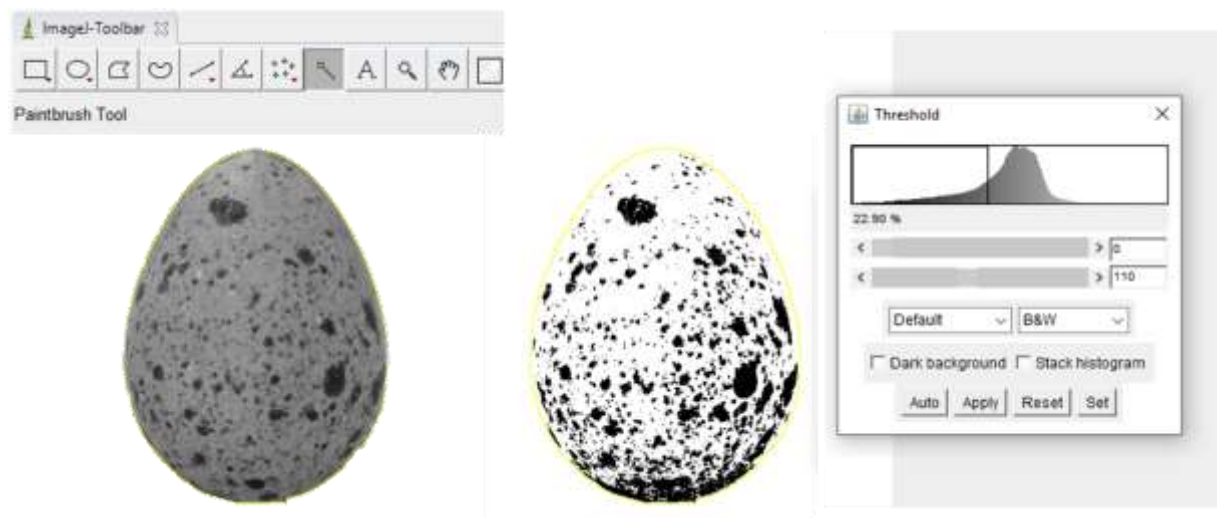


Figura 1: Exemplo de segmentação de imagem em preto e branco para quantificação de área coberta por máculas em casca de ovo de codorna.

1.2.2 Sistemas RGB e HSL

RGB é a sigla do sistema de cor que representa o espaço de cores aditivas, fundamentado nas três cores primárias, do inglês Red, Green e Blue. Introduzido por Maxweel (1860), no seu trabalho sobre a teoria da cor, foi responsável por introduzir os princípios para o modelo de cores RGB, amplamente utilizado no cenário atual como em imagens, telas e câmeras (LOESDAU *et al.*, 2014). O sistema é baseado na percepção de cores pelo olho humano e pode ser baseado na escala 0-255 ou 0-1 (Figura 2).

O sistema HSL, por sua vez é baseado na parametrização de cores através da tonalidade cromática ou matiz (H) que define o comprimento de onda dominante de uma cor, saturação (S), que mede a pureza ou grau de mistura de luz branca e da intensidade ou luminância (L) que está relacionada a quão brilhante é a cor.

1.3 Aprendizado estatístico seus algoritmos

Métodos de aprendizado estatístico para classificação buscam, a partir de um conjunto de características, classificar corretamente determinado sujeito. Essa forma de classificação é chamada de aprendizado supervisionado, onde ocorre previamente um treinamento, sendo fornecidos dados que servem como exemplo da relação entre as características, além de fornecer os rótulos em que os dados podem ser classificados. A validação é realizada para testar a capacidade do modelo de classificar corretamente os objetos, e consiste em usar os dados que não foram utilizados no treinamento para validar.

Tipicamente esse tipo de relação entre dados de treino e teste pode ser estabelecida por uma partição inicial do banco de dados em que os dados usados no teste nunca aparecem no conjunto de dados de validação ou no caso da validação cruzada em que diversas subdivisões

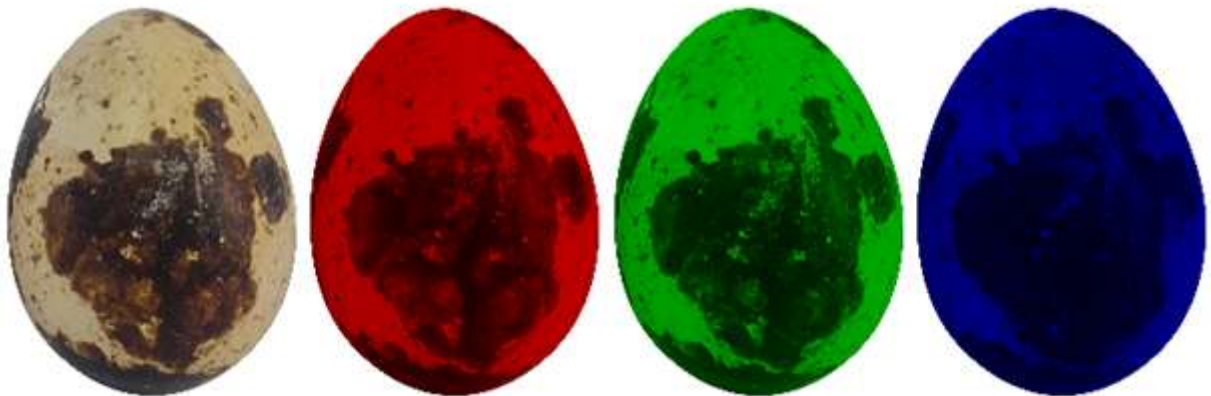


Figura 2: Exemplo de separação de canais RGB de imagem de um ovo. A primeira imagem é a original, seguida das bandas de vermelho, verde e azul, respectivamente.

do conjunto de dados são alternadas entre grupos de treinamento e validação, obtendo no final estatísticas das métricas obtidas em cada “k” grupo que foi usado na validação.

Algoritmos diferem no modo de divisão dos rótulos em um banco de dados, adotando critérios únicos para definir a classificação final. Na Figura 3 e nos subtópicos a seguir são apresentadas especificidades dos algoritmos.

1.3.1 Análise Discriminante

É uma técnica da estatística multivariada bastante aplicada para distinguir e classificar objetos (VARELLA, 2008) tendo como alvo encontrar uma função discriminante (ou variável estatística discriminante) que irá segregar da melhor maneira os objetos nos grupos já definidos *a priori* (HAIR JR *et al.*, 2009).

Existem duas fases que compõem o processo da análise discriminante. A primeira etapa é o processo de discriminação, nela são determinadas as características de maior peso, responsáveis por alocar os objetos nos grupos já existentes de forma que maximize a diferença entre os grupos (HAIR JR *et al.*, 2009; VARELLA, 2008). E a segunda fase, de classificação, consiste em um conjunto de regras que serão empregadas para atribuir novos objetos aos grupos previamente definidos (JOHNSON; WICHERN, 2002).

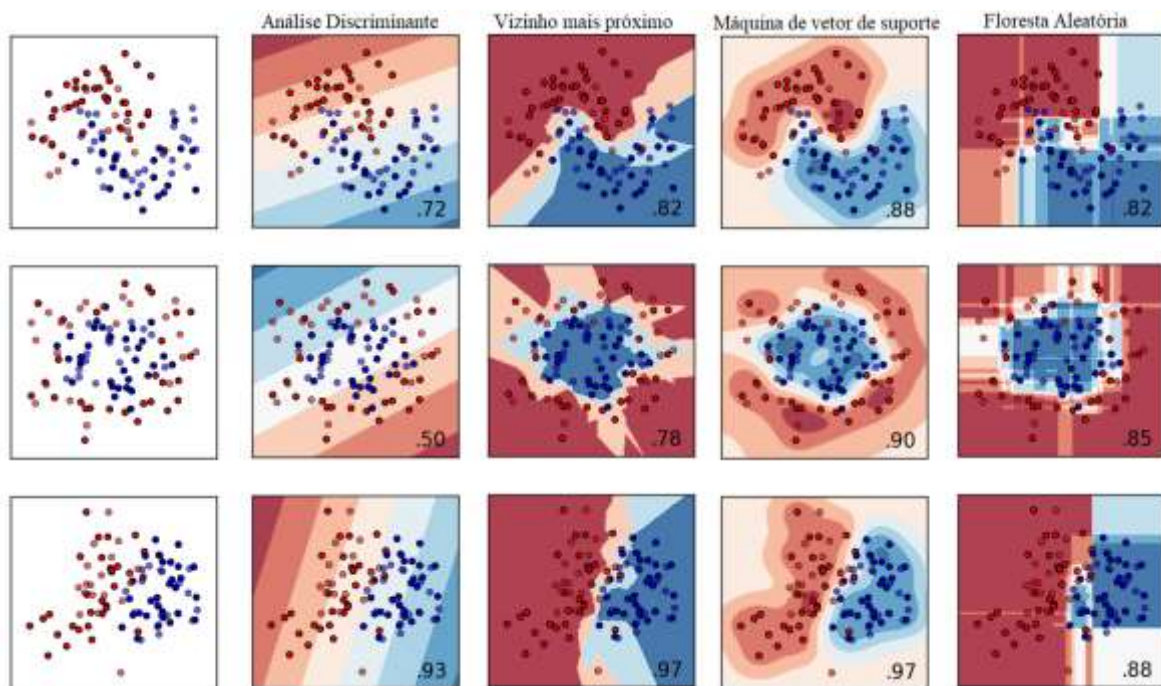


Figura 3. Comparação de algoritmos de aprendizado estatístico para classificação em bancos de dados sintéticos. Adaptado de <http://scikit-learn.org>.

1.3.2 Random Forest

Antes de começar a descrever como funciona a Random Forest ou Florestas Aleatórias é necessário elucidar o funcionamento das árvores de decisão. Essa é uma técnica

baseada em identificar os principais fatores que diferenciam as classes baseado em sucessivas perguntas que auxiliam na predição do novo objeto a classe em que ele pertence (QUINLAN, 1986).

O método funciona em uma estrutura de árvore, por isso leva esse nome, começando pela “raiz”, onde se encontra o objeto que se deseja classificar, passando pelos “nós”, que são as sucessivas perguntas que compõem o processo de divisão da árvore, terminando nas “folhas”, que são os pontos finais de cada ramificação, ou seja, a classe final em que aquele objeto será atribuído (SAFAVIAN; LANDGREBE, 1991; ZHANG; MA, 2012).

Desta forma, a Random Forest é um algoritmo de classificação que consiste no uso de um conjunto de árvores de decisão (BREIMAN, 2001). Seu funcionamento se baseia na atribuição da classe final, através do voto majoritário composto por todas as árvores da floresta, processo denominado de *bagging* (BREIMAN, 1996).

São chamadas de aleatórias, pois cada árvore da floresta é treinada usando um subconjunto aleatório de dados de treinamento, o que faz com que as árvores não sejam relacionadas entre si (ZHANG; MA, 2012) tornando a previsão mais precisa e estável, pois mesmo que uma árvore atribua a classe erroneamente, a classificação final somente é atribuída a classe com número majoritário de votos.

1.3.3 k-vizinho mais próximo

O classificador k-vizinhos mais próximos, é um dos métodos mais simples e amplamente utilizados para solucionar problemas de classificação de padrões (SILVA, 2008) é um método que classifica um objeto desconhecido através de alguma métrica de distância, geralmente a distância Euclidiana (COVER; HART, 1967).

O procedimento é feito identificando os elementos do banco de dados de treinamento que estão mais próximos do objeto que se deseja classificar, ou seja, seus vizinhos mais próximos, que já possuem classe definida. Após calculada as distâncias, a classe atribuída será aquela em que o vizinho k apresenta menor distância em relação ao elemento desconhecido (COVER; HART, 1967).

1.3.4 Máquinas de Vetores de Suporte

Máquinas de vetores de suporte são uma metodologia usada para classificação de grupos (CORTES; VAPNIK, 1995). Conceitualmente a máquina funciona da seguinte forma: mapeiam de forma não linear os vetores de entrada em um espaço de características de alta dimensão escolhido *a priori*, denominados hiperplanos. Esses dados no espaço de entradas são separados em duas ou mais classes, podendo ser feito através do uso de funções lineares, polinomiais e de base radial.

Um hiperplano ideal é definido como a função de decisão linear com margem máxima de separação entre os vetores das duas classes. Para construir esses hiperplanos ideais (supondo dados de treinamento linearmente separáveis, sem superposição), é necessário levar em consideração apenas uma pequena quantidade dos dados de treinamento, os chamados vetores de suporte, que determinam essa margem, de forma a maximizar a distância de separação entre os grupos.

1.4 Métricas de avaliação de modelos de aprendizado estatístico

1.4.1 Acurácia

Primeiramente definida por Gauss, a acurácia é o grau em que a medida representa o fenômeno de interesse, ou seja, indica a concordância entre o resultado da mensuração e valor de referência ou valor verdadeiro, aquele que é usado como parâmetro de comparação da grandeza ou medida (CALDEIRA *et al.*, 2007; MONICO *et al.*, 2009).

A acurácia em modelos de aprendizagem estatística pode ser definida como a porcentagem de classificações corretas sobre todas as classificações feitas.

$$\text{Acurácia} = \frac{\text{Verdadeiro Negativo} + \text{Verdadeiro Positivo}}{\text{Verdadeiro Positivo} + \text{Falso Positivo} + \text{Verdadeiro Negativo} + \text{Falso Negativo}}$$

1.4.2 Coeficiente kappa

Primeiramente introduzido por Jacob Cohen (COHEN, 1960), a estatística kappa ou coeficiente kappa, é bastante usada para avaliar a precisão de um classificador, pois corrige o grau de concordância entre previsões do classificador e a realidade, ao considerar a proporção de previsões que podem ocorrer ao acaso (BEN-DAVID, 2006; LUCAS, 2007; WITTEN *et al.*, 2016).

Resumindo, a estatística kappa de Cohen varia de -1 (discordância total), 0 (classificação aleatória), a 1 (concordância total) (BEN-DAVID, 2006).

$$k = \frac{p_o - p_e}{1 - p_e}$$

Onde p_o é a concordância relativa observada entre avaliadores e p_e é a probabilidade hipotética da concordância casual.

1.4 Antiagrupamento

O antiagrupamento é um método usado quando é necessário que haja heterogeneidade dentro do grupo estudado, (PAPENBERG; KLAU, 2020; SPÄTH, 1986; VALEV, 1983, 1998).

O método de antiagrupamento atua no sentido oposto ao dos métodos de agrupamento (*clustering*) sendo que para obter máxima homogeneidade dentro do grupo pode ser usado o k-vizinho mais próximo com o objetivo de maximizar a variância, com o objetivo é minimizar a soma de quadrados das distâncias entre o centro do grupo e seus elementos. Por outro lado, é possível adotar o *clustering*, com objetivo de maximizar a diversidade, como a soma das distâncias par a par entre elementos pertencentes ao mesmo grupo (Figura 4).

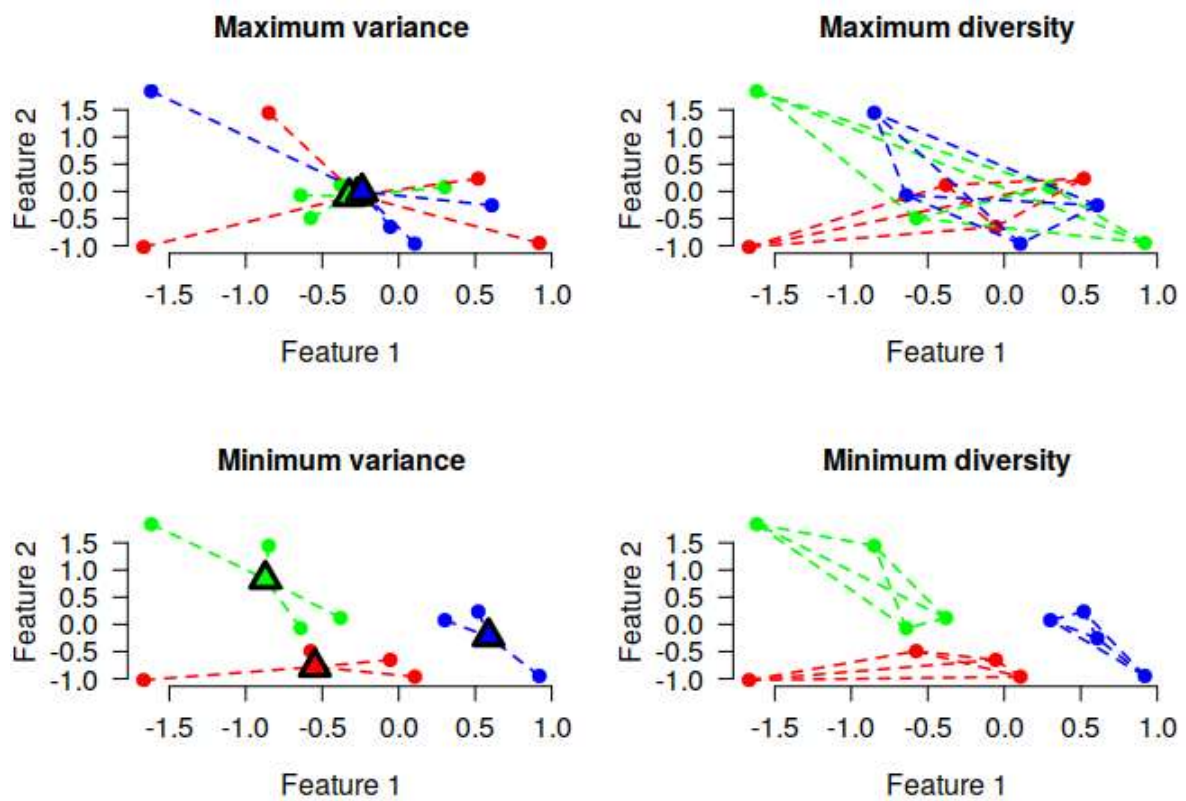


Figura 4: Ilustrações do uso dos métodos de agrupamento (parte inferior) e antiagrupamento (parte superior) para formação de subconjuntos de observações com o objetivo de reunir pela variância (parte esquerda) ou pela diversidade (parte direita).

2 OBJETIVOS

2.1 Objetivo geral

- Verificar a eficiência do método de identificação de fêmeas, baseado em dados métricos e na imagem do ovo.

2.2 Objetivos específicos

- Testar algoritmos para identificação de maternidade através do padrão de maculação da casca do ovo.
- Avaliar a eficiência da identificação pelo método dos dados métricos e imagem do ovo, comparando com animais individualmente identificados.

3 MATERIAL E METÓDOS

3.1 Coleta de dados e definição de variáveis

Os ovos são provenientes de uma linhagem sintética do programa de melhoramento de codornas experimental desenvolvido na Universidade Federal do Ceará, em Fortaleza, Ceará, Brasil. As codornas em fase reprodutiva são alojadas em gaiolas individuais para o controle do pedigree, garantindo assim a coleta de informações sobre a maternidade dos ovos.

Ao longo de 60 dias, ovos de 90 fêmeas foram coletados diariamente e marcados com uma etiqueta removível para identificação. Além disso, foram feitos registros de altura e comprimento, com auxílio de um paquímetro digital (0,01mm), peso, com balança de precisão semi-analítica (0,01g) e fotografias (12MP). As fotografias foram tomadas de uma face aleatória do ovo e, em seguida, giradas 180° para registrar a face oposta, de modo que há dois registros de imagem de cada ovo.

O processamento de imagens foi realizado no ImageJ (ABRÀMOFF; MAGALHÃES; RAM, 2004) via Bio7 (AUSTENFELD; BEYSCHLAG, 2012), transformando em tons de cinza e depois convertido em binário de preto e branco com um limiar de intensidade de 110. Em cada face, a proporção da área com máculas foi definida pela proporção de área escura para a área total e a medida utilizada para a análise dos dados foi a média desses valores.

A quantificação da coloração RGB (vermelho, verde e azul) da cor média de ovo inteiro, foi obtida através da segmentação de imagem usando pacote gridExtra em R (AUGUIE; ANTONOV, 2017). Os valores de HSL (matiz, saturação e luminosidade) foram obtidos pela conversão dos valores de RGB por meio do pacote Colorscience em R (GAMA; DAVIS, 2019).

3.2 Métodos de classificação

Dez variáveis foram utilizadas para compor um algoritmo de classificação, a saber: peso, altura e largura do ovo, proporção da área com máculas na casca do ovo, intensidade de vermelho, verde e azul da cor média de ovo inteiro e matiz, saturação e luminosidade da cor média de ovo inteiro.

Correlações de Pearson par a par entre todas as variáveis foram calculadas e a seleção de variáveis foi feita através da eliminação recursiva de variáveis (RFE), sendo consideradas as variáveis que aumentaram a acurácia da classificação utilizando validação cruzada de 10 vezes para classificação de 1475 ovos de 90 aves. Todas as análises foram feitas usando o programa R e a escolha do conjunto de variáveis através da função RFE do pacote caret (KUHN, 2008). Repetibilidades de todas as variáveis foram preditas usando Ave como grupo em modelo aleatório usando o pacote rptR (STOFFEL; NAKAGAWA; SCHIELZETH, 2017)

A escolha do melhor algoritmo foi feita entre quatro algoritmos de aprendizagem estatística: k-vizinhos mais próximos, análise linear discriminante, random forest e máquinas de vetores de suporte. Todos os algoritmos foram analisados considerando o desempenho de classificação simulando diferentes tamanhos de gaiola, ou seja, nove gaiolas com dez codornas cada ou 18 gaiolas com cinco codornas cada ou 30 gaiolas com três codornas cada. As métricas para tomada de decisão foram a acurácia e o kappa médios para todas as gaiolas, vindos de uma validação cruzada de 10 vezes, separada para os três tamanhos de gaiola. Métricas de desempenho de classificação foram obtidas via pacote caret (KUHN, 2008).

Por último, o desempenho da classificação foi avaliado em um cenário de alocação aleatória das aves comparada com o método de antiagrupamento com objetivo de maximização de variância dentro de gaiola (SPÄTH, 1986; VALEV, 1983, 1998) usando o pacote anticlust em R (PAPENBERG; KLAU, 2020). Depois foram separados do banco de dados de treino/validação dois ovos por cada ave para que fossem utilizados como conjunto de validação externa, limitando a classificação a apenas um ovo por ave por dia, usando probabilidades ordenadas baseadas em uma função de limiar do pacote utiml (RIVOLLI; DE CARVALHO, 2019).

4 RESULTADOS

As estatísticas descritivas para as características do ovo são apresentadas na Tabela 1 e correlação entre as características e suas repetibilidades na Tabela 2. A altura e largura do ovo apresentaram menor coeficiente de variação entre todas as características e correlação moderada com o peso do ovo.

A Figura 5 apresenta as principais variáveis por ordem de importância sem consideração de gaiolas ou dia de postura, da mais influente para menos influente e o incremento de acurácia na classificação com a adição de cada de cada variável, desta forma, o uso de todas as variáveis presentes no estudo, leva a incremento da acurácia, sendo essa a abordagem escolhida.

A acurácia e o coeficiente kappa aumentaram conforme o número de aves por gaiola diminuiu. Para grupos de 10 codornas a acurácia variou de 0,55 a 0,66, em grupos de cinco codornas de 0,69 a 0,76, e para 3 grupos de codornas de 0,80 a 0,83 (Tabela 3).

Tabela 1. Estatísticas descritivas das características do ovo.

Características	Mínimo	Máximo	Média	Mediana	Desvio Padrão
Peso do ovo (g)	4,8600	20,370	12,702	12,695	1,4756
Altura do ovo (mm)	24,080	43,150	33,734	33,560	1,9911
Largura do ovo (mm)	10,890	37,590	25,635	25,500	1,9222
Proporção de área padronizada (%)	37,702	92,844	62,230	61,415	9,5258
R/Vermelho	0,3472	0,7762	0,6122	0,6108	0,0676
G/Verde	0,2958	0,7524	0,5610	0,5625	0,0703
B/Azul	0,1868	0,6884	0,4292	0,4310	0,0762
H/Matiz	0,0725	0,2477	0,1230	0,1197	0,0157
S/Saturação	0,0320	0,3948	0,2081	0,2120	0,0615
L/Luminância	0,2924	0,7323	0,5207	0,5214	0,0671

Tabela 2. Correlação de Pearson (abaixo da diagonal), p-valores do teste de correlação (acima da diagonal) e repetibilidades (diagonal, em negrito) para características de ovos de codorna.

	PO	AO	LO	PAP	R	G	B	H	S	L
PO	0,371	0,000	0,000	0,005	0,000	0,005	0,976	0,894	0,000	0,060
AO	0,717	0,303	0,000	0,000	0,182	0,627	0,186	0,043	0,001	0,920
LO	0,443	0,483	0,153	0,006	0,858	0,548	0,354	0,001	0,253	0,655
PAP	-0,073	-0,091	0,071	0,197	0,000	0,000	0,000	0,000	0,000	0,000
R	0,098	0,032	-0,005	-0,442	0,302	0,000	0,000	0,035	0,000	0,000
G	0,073	0,009	0,016	-0,373	0,939	0,429	0,000	0,000	0,376	0,000
B	-0,002	-0,039	0,024	-0,224	0,742	0,910	0,512	0,000	0,000	0,000
H	0,003	-0,054	0,085	0,148	0,052	0,341	0,494	0,473	0,000	0,000
S	0,115	0,087	-0,030	-0,278	0,295	-0,015	-0,381	-0,624	0,317	0,003
L	0,048	-0,007	0,012	-0,348	0,924	0,990	0,942	0,311	-0,070	0,426

PO: peso do ovo, AO: altura do ovo, LO: largura do ovo, PAP: proporção de área padronizada, R: vermelho, G: verde, B: azul, H: matiz, S: saturação e L: luminosidade. N: não significativo considerando um nível de significância de 0,05 para hipótese nula.

A previsão diária do conjunto de validação externa atingiu os valores mais altos de acurácia (Tabela 4) em comparação com o método anterior, onde todos os ovos postos foram classificados em conjunto.

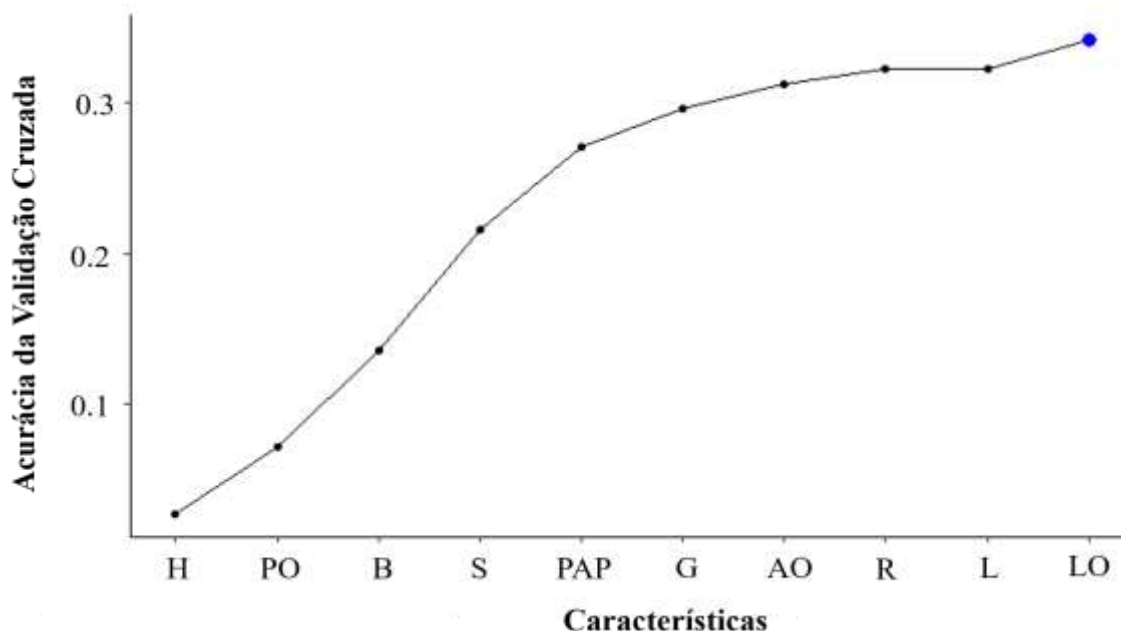


Figura 5. Ganho em precisão pela inclusão de recursos para classificação de ovos classificados por importância obtidos por meio do algoritmo RFE. PO: peso do ovo, AO: altura do ovo, LO: largura do ovo, PAP: proporção da área padronizada, R: vermelho, G: verde, B: azul, H: matiz, S: saturação e L: luminosidade.

Tabela 3. Intervalos de confiança de 95% para acurácia e kappa em validação cruzada de algoritmos de aprendizado de máquina utilizados pelo número de codornas por gaiola na classificação de ovos com base em suas características.

Aves/gaiola	Algoritmo	Acurácia	Intervalo de Confiança da Acurácia	Kappa
10	KNN	0,6029	[0,5578; 0,6480]	0,5525
	LDA	0,5468	[0,5017; 0,5919]	0,5468
	RF	0,6574	[0,6287; 0,6860]	0,6574
	SVM	0,6286	[0,5805; 0,6767]	0,5248
5	KNN	0,7219	[0,6905; 0,7533]	0,6412
	LDA	0,6907	[0,6577; 0,7236]	0,6006
	RF	0,7559	[0,7323; 0,7795]	0,6850
	SVM	0,7485	[0,7073; 0,7896]	0,6728
3	KNN	0,7991	[0,7707; 0,8273]	0,6841
	LDA	0,7963	[0,7608; 0,8318]	0,6828
	RF	0,8343	[0,8089; 0,8598]	0,7399
	SVM	0,8129	[0,7823; 0,8436]	0,7022

LDA: Análise Linear Discriminante, KNN: k - Vizinhos mais Próximos, SVM: Máquinas de vetores de suporte, RF: Random Forest.

Tabela 4. Validação de conjuntos externos utilizando diferentes sistemas de alocação de gaiolas (Aleatório ou Antiagrupamento) e considerando a classificação por dia de postura.

Aves/gaiola	Alocação	Acurácia	Intervalo de Confiança da Acurácia	Kappa
10	Aleatório	0,7641	[0,6982, 0,8218]	0,7379
	Antiagrupamento	0,7551	[0,6887, 0,8136]	0,7279
5	Aleatório	0,8135	[0,7512, 0,8658]	0,7669
	Antiagrupamento	0,8281	[0,7672, 0,8786]	0,7852
3	Aleatório	0,8757	[0,8193, 0,9195]	0,8136
	Antiagrupamento	0,9135	[0,8634, 0,9498]	0,8703

5 DISCUSSÃO

Podemos dividir os recursos usados no presente trabalho em métricas (peso, altura e comprimento) e recursos relacionados à imagem (HSL, RGB e proporção da área padronizada). Esses dois grupos mostram correlação interna moderada a alta e correlação baixa com características do outro grupo. Esperava-se que os recursos HSL e RGB tivessem algum grau de correlação, visto que esses dois sistemas são conversíveis um no outro, mas alguns aspectos são distintos entre os sistemas, então o uso de ambos melhorou a acurácia.

O algoritmo RFE se classifica quase alternadamente entre as melhores variáveis a serem usados para classificação variáveis métricas e baseadas em imagens, mostrando uma complementaridade notável entre esses recursos. O uso de todos os recursos registrados aumentou a acurácia sem nenhum problema perceptível de sobreajuste.

No contexto do reconhecimento de ovos pelas aves, as codornas são capazes de reconhecer seus próprios ovos, mas isso depende da semelhança fenotípica das características das outras codornas ao redor, atingindo classificação correta em mais de 80% das vezes (PIKE, 2011). O sucesso do reconhecimento natural do ovo está relacionado à evolução dos padrões dos ovos e sugere-se que venha da experiência, aprendendo com os ovos postos por si próprios, sendo o padrão de maculação (manchas pretas sobre a casca do ovo) e a forma do ovo as principais características relacionadas à rejeição de um estrangeiro ovo (LÓPEZ-DE-HIERRO; MORENO-RUEDA, 2010).

A existência de um mecanismo que evoluiu para lidar com a competição de parasitas ou companheiros de ninho implica um efeito genético subjacente nas características do ovo. Embora neste estudo não tenha sido estimadas as herdabilidades, os coeficientes de

repetibilidade estimados, que indicam o limite superior da herdabilidade da característica, estão entre 0,15 e 0,51, bem abaixo do estimado para peso, altura e largura do ovo, com herdabilidades estimadas variando de 0,63 a 0,83 (SEZER, 2007; SEZER; TEKELIOGLU, 2009).

Estimativas de alta herdabilidade e repetibilidade asseguram uma homogeneidade dos fenótipos das feições analisadas dentro da ninhada e uma heterogeneidade entre as diferentes ninhadas de codornas, tornando essas características altamente discriminatórias. Menores repetibilidades indicam uma maior proporção de efeitos aleatórios residuais sobre o fenótipo.

Os algoritmos de predição usados neste estudo geram acurácias razoáveis e, até onde sabemos, nenhum algoritmo de aprendizado estatístico é descrito na literatura para a classificação de ovos para codornas que desovaram. No entanto, a classificação feita por humanos é comum e os resultados deste estudo podem ser usados para minimizar erros e economizar custos de mão de obra (SEZER; TEKELIOGLU, 2009) e novas aplicações que virão.

Algumas abordagens para identificar animais por seus ovos foram desenvolvidas principalmente no contexto da ecologia de interações parasita-hospedeiro e ninhos comuns usando algoritmos de extração de imagem, por ex. SIFT e cálculo de pontuações de similaridade por técnicas multivariadas (CARIELLO *et al.*, 2004; PIKE, 2011; STODDARD; KILNER; TOWN, 2014).

O número de aves por gaiola afetou a acurácia da classificação. Em grandes grupos de aves, é mais provável encontrar ovos de diferentes camadas que são semelhantes em várias características, portanto, quando os algoritmos são aplicados em tais grupos são mais suscetíveis a erros de classificação.

As diferenças entre algoritmos são maiores em grupos com mais animais. O algoritmo Random Forest apresentou os mais altos níveis de acurácia para todos os cenários de tamanho de grupo, mas o alcance de sua acurácia é bem próximo ao algoritmo SVM em todos os cenários de gaiolas simuladas.

Apesar da acurácia superior para Random Forest, a demanda computacional de memória e tempo de cálculo é maior quando comparada aos métodos mais simples, por exemplo, KNN e LDA, e se for aplicado a um dispositivo independente, o uso de energia deve ser considerado.

Os métodos adotados neste estudo aumentaram consideravelmente a acurácia da identificação de codornas com base em características do ovo que podem ser facilmente automatizadas para implementação em escala. Outros recursos podem ser adicionados ou substituir os atuais para aumentar a acurácia ou tornar a coleta de dados mais fácil.

Aspectos como mudanças na iluminação, presença de ruído na imagem e padrões de máculas que parecem maiores ou menores devido a mudanças no tamanho, forma e peso do ovo podem contribuir para que seja necessário um segundo treinamento após a mudança dessas características na curva de postura das aves (DIONELLO et al., 2011; STODDARD; KILNER; TOWN, 2014).

6 CONCLUSÃO

Para máxima acurácia na identificação das fêmeas, recomenda-se utilização de três de aves por gaiola, considerando classificação baseada em postura diária e alojando-as de modo a maximizar a variância das características dentro da gaiola, e utilizar todas as variáveis analisadas para ajuste do algoritmo Random Forest.

REFERÊNCIAS

- ABRÀMOFF, M. D.; MAGALHÃES, P. J.; RAM, S. J. **Image processing with imageJ** *Biophotonics International*, 2004.
- AFONSO, S.; VANORE, G.; BATLLE, A. **Protoporphyrin IX and oxidative stress**. *Free Radical Research*, v. 31, n. 3, p. 161–170, 7 jan. 1999.
- AKPA., KAYE, J., ADEYINKA, I. A., & KABIR, M. **The relationships between laying age and repeatability of egg quality traits in Japanese quails (*Coturnix coturnix japonica*)**. *International Journal of Poultry Science*, v. 7, n. 6, p. 555–559, 2008.
- AMAO, S.; OJEDAPO, L.; OGUNDIPE, R. **Repeatability Estimates of Egg Quality Traits of Pharaoh Quail (*Coturnix Coturnix Japonica*) Reared In Derived Savanna Zone of Nigeria**. *International Journal of Agriculture and Biosciences*, v. 2, n. 4, p. 156–159, 2013.
- AUGUIE, B.; ANTONOV, A. **Package gridExtra**. Disponível em: <https://cran.r-project.org/web/packages/gridExtra/gridExtra.pdf>. Acesso em: 10 ago. 2020.
- AUSTENFELD, M.; BEYSCHLAG, W. **A Graphical User Interface for R in a Rich Client Platform for Ecological Modeling**. *Journal of Statistical Software*, v. 49, n. 4, p. 1–19, 2012.
- BEN-DAVID, A. **Expert Opinion What’s Wrong with Hit Ratio?** *IEEE Intelligent Systems*, v. 21, n. 6, p. 68–70, 2006.
- BREIMAN, L. **Bagging predictors**. *Machine Learning*, v. 24, p. 123–130, 1996.
- BREIMAN, L. **Random Forests**. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.
- CARIELLO, M. O. R. LIMA, M., G. SCHWABL, H., & H. MACEDO, R. **Egg characteristics are unreliable in determining maternity in communal clutches of guira cuckoos *Guira guira***. *Journal of Avian Biology*, v. 35, n. 2, p. 117–124, 2004.
- CELEBI, M. Emre. **Effective initialization of k-means for color quantization**. In: 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE, 2009. p. 1649-1652.
- COHEN, J. **A Coefficient of Agreement for Nominal Scales**. *Educational and Psychological Measurement*, v. 20, n. 1, p. 37–46, 2 abr. 1960.
- CORTES, C.; VAPNIK, V. **Support-vector networks**. *Machine Learning*, v. 20, n. 3, p. 273–297, set. 1995.
- COVER, T. M.; HART, P. E. **Nearest Neighbor Pattern Classification**. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21–27, jan. 1967.
- DALE, J.; LANK, D. B.; REEVE, H. K. **Signaling Individual Identity versus Quality: A Model and Case Studies with Ruffs, Queleas, and House Finches**. *The American Naturalist*, v. 158, n. 1, p. 75–86, jul. 2001.

- DIONELLO, N. J. L., ROLL, V. F. B., XAVIER, E. G., VAYEGO, S., & CORRÊA, G. D. S. **S. Produção em número, peso e massa de ovos em codornas de corte até 406 dias de idade.** *Pubvet*, v. 5, n. 27, p. 1–15, 2011.
- DUVAL, C. CASSEY, P., LOVELL, P. G., MIKSIK, I., REYNOLDS, S. J., & SPENCER, K. A. **Eggshell appearance does not signal maternal corticosterone exposure in Japanese quail: An experimental study with brown-spotted eggs.** *PLoS ONE*, v. 8, n. 12, 2013.
- FÖRSTER, A. JAENECKE, D. WITTMANN, M. FLOCK, D. K. KREUZER, M. **Untersuchungen zur Eignung fotometrisch ermittelter Schalenfarbparameter für die Selektion auf marktgerechte braunschalige Eier.** *Archiv für Geflügelkunde*, v. 60, n. 1, p. 1–6, 1996.
- GAMA, J.; DAVIS, G. **Colorscience: Color Science Methods and Data**, 2019. Disponível em: <https://cran.r-project.org/package=colorscience>. Acesso em: 16 nov. 2020.
- GOODMAN, B. L. **Repeatability of Egg Quality Traits in the Coturnix Quail.** *Poultry Science*, v. 44, n. 5, p. 1338–1339, 1965.
- GOSLER, A. G.; HIGHAM, J. P.; JAMES REYNOLDS, S. **Why are birds' eggs speckled?** *Ecology Letters*, v. 8, n. 10, p. 1105–1113, out. 2005.
- HAIR JR, J. F. **Análise multivariada de dados.** Ed. 6. p. 687. Porto Alegre. 2005.
- HUSS, D.; POYNTER, G.; LANSFORD, R. **Japanese quail (*Coturnix japonica*) As a Laboratory Animal Model.** *Lab Animal*, v. 37, n. 11, p. 513–519, 2008.
- JOHNSON, R. A., & WICHERN, D. W. **Applied multivariate statistical analysis.** Upper Saddle River, NJ: Prentice hall, 2002.
- JOHNSTONE, R. A. **Recognition and the evolution of distinctive signatures: when does it pay to reveal identity? Proceedings of the Royal Society of London.** *Biological Sciences*, v. 264, n. 1387, p. 1547–1553, 22 out. 1997.
- KAUR, H., M. N., Green, C. J., NAUGHTON, P., FORESTI, R., & MOTTERLINI, R. **Interaction of bilirubin and biliverdin with reactive nitrogen species.** *FEBS Letters*, v. 543, n. 1–3, p. 113–119, 22 maio 2003.
- KENNEDY, G. Y.; VEVERS, H. G. **A survey of avian eggshell pigments. Comparative Biochemistry and Physiology.** *Comparative Biochemistry*, v. 55, n. 1, p. 117–123, 1 jan. 1976.
- KILNER, R. M. **The evolution of egg colour and patterning in birds.** *Biological Reviews*, v. 81, n. 03, p. 383, 2 ago. 2006.
- KUHN, M. **Building Predictive Models in R Using the caret Package.** *Journal of Statistical Software*, v. 28, n. 5, p. 1–26, 2008.
- LOESDAU, Martin; CHABRIER, Sébastien; GABILLON, Alban. **Hue and saturation in the RGB color space. In: International conference on image and signal processing.** Springer,

Cham. p. 203-212, 2014.

LÓPEZ-DE-HIERRO, M. D. G.; MORENO-RUEDA, G. **Egg-spot pattern rather than egg colour affects conspecific egg rejection in the house sparrow (*Passer domesticus*)**. Behavioral Ecology and Sociobiology, v. 64, n. 3, p. 317–324, 23 jan. 2010.

LOVELL, P. G. RUXTON, G. D., LANGRIDGE, K. V., & SPENCER, K. A. **Egg-Laying Substrate Selection for Optimal Camouflage by Quail**. Current Biology, v. 23, n. 3, p. 260–264, 4 fev. 2013.

LUCAS, L. S. **Joint segmenting consumers using both behavioral and attitudinal data**. Proceedings of the Sawtooth Software Conference, n. October, p. 111–126, 2007.

LYON, B. E. **Egg recognition and counting reduce costs of avian conspecific brood parasitism**. Nature, v. 422, n. 6931, p. 495–499, abr. 2003.

MAXWELL, James Clerk. IV. **On the theory of compound colours, and the relations of the colours of the spectrum**. Philosophical Transactions of the Royal Society of London, n. 150, p. 57-84, 1860.

MINVIELLE, F. **What are quail good for in a chicken-focused world?** World's Poultry Science Journal, v. 65, n. 4, p. 601–608, 2009.

MORENO, J.; OSORNO, J. L. **Avian egg colour and sexual selection: does eggshell pigmentation reflect female condition and genetic quality?** Ecology Letters, v. 6, n. 9, p. 803–806, set. 2003.

MUIR, W. M.; BIJMA, P.; SCHINCKEL, A. **Multilevel selection with kin and non-kin groups, experimental results with japanese quail (*coturnix japonica*)**. Evolution, v. 67, n. 6, p. 1598–1606, 2013.

MUIR, W. M.; CRAIG, J. V. **Improving Animal Well-Being Through Genetic Selection**. Poultry Science, v. 77, n. 12, p. 1781–1788, 1998.

PAPENBERG, M.; KLAU, G. W. **Using Anticlustering to Partition Data Sets Into Equivalent Parts**. Psychological Methods, p. 1–44, 2020.

PIKE, T. W. **Egg recognition in Japanese quail**. Avian Biology Research, v. 4, n. 4, p. 231–236, dez. 2011.

QUINLAN, J. R. **Induction of Decision Trees**. Machine Learning, v. 1, n. 1, p. 81–106, 1986.

RIVOLLI, A.; DE CARVALHO, A. C. P. L. F. **The utiml package: Multi-label classification in R**. R Journal, v. 10, n. 2, p. 24–37, 2019.

SAFAVIAN, S. R.; LANDGREBE, D. **A Survey of Decision Tree Classifier Methodology**. IEEE Transactions on Systems, Man and Cybernetics, v. 21, n. 3, p. 660–674, 1991.

SATO, K.; IDA, N.; INO, T. **Genetic parameters of egg characteristics in Japanese quail**. Jikken dobutsu. Experimental animals, v. 38, n. 1, p. 55–59, 1989.

SEZER, M. **Heritability of Exterior Egg Quality Traits in Japanese Quail**. Journal of Applied Biological Sciences, v. 1, n. 2, p. 37–40, 2007.

SEZER, M.; TEKELIOGLU, O. **Quantification of Japanese Quail eggshell colour by image analysis**. Biological Research, 2009.

SILVA, M. M. **Uma Abordagem Evolucionária Para o Aprendizado Semi-Supervisionado em Máquinas de Vetores de Suporte**. Universidade Federal de Minas Gerais, 2008.

SPARKS, N. H. C. **Eggshell Pigments—from Formation to Deposition**. Avian Biology Research, v. 4, n. 4, p. 162–167, 27 dez. 2011.

SPÄTH, H. **Anticlustering: maximizing the variance criterion**. Control and Cybernetics, v. 15, n. 2, p. 213–218, 1986.

STODDARD, M. C.; KILNER, R. M.; TOWN, C. **Pattern recognition algorithm reveals how birds evolve individual egg pattern signatures**. Nature Communications, v. 5, p. 1–18, 2014.

STOFFEL, Martin A.; NAKAGAWA, Shinichi; SCHIELZETH, Holger. **An introduction to repeatability estimation with rptR**. 2017.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**, 4th Ed. Elsevier, p. 984, 2008.

TIBBETTS, E. A.; DALE, J. **Individual recognition: it is good to be different**. Trends in Ecology & Evolution, v. 22, n. 10, p. 529–537, 1 out. 2007.

VALEV, Ventseslav. **Set partition principles revisited**. In: **Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)**. Springer, Berlin, Heidelberg, 1998. p. 875-881.

VALEV, V. **Set partition principles**. **Transactions of the ninth Prague conference on information theory, statistical decision functions, and random processes**. Anais Prague: Springer Netherlands, 1983.

VARELLA, C. A. A. **Análise multivariada aplicada as ciências agrárias**. Disponível em: [http://www.ufrj.br/institutos/it/deng/varella/Downloads/multivariada aplicada as ciências agrarias/](http://www.ufrj.br/institutos/it/deng/varella/Downloads/multivariada%20aplicada%20as%20ci%C3%AAncias%20agrarias/). Acesso em: 14 dez. 2020.

WITTEN, Ian H.; FRANK, Eibe. **Data mining: practical machine learning tools and techniques with Java implementations**. Acm Sigmod Record, v. 31, n. 1, p. 76-77, 2002.

ZHANG, Cha; MA, Yunqian (Ed.). **Ensemble machine learning: methods and applications**. Springer Science & Business Media, 2012.

ZHANG, L.C. NING, Z. H., XU, G. Y., HOU, Z. C., & YANG, A. N. **Heritabilities and genetic and phenotypic correlations of egg quality traits in brown-egg dwarf layers**. Poultry Science, v. 84, n. 8, p. 1209–1213, ago. 2005.