

Detecção de Anomalias em Padrões Acústicos, de Temperatura e Umidade Sazonais para Abelhas Melíferas (*Apis mellifera* L.)

Ícaro de Lima Rodrigues¹, Davyd B. de Melo¹, Breno M. Freitas², Danielo G. Gomes¹

¹GREat, Departamento de Engenharia de Teleinformática (DETI), Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza - CE, CEP 60.455-970, Brasil.

²Setor de Abelhas, Departamento de Zootecnia, Centro de Ciências Agrárias, Universidade Federal do Ceará, Fortaleza - CE, CEP 60.356-000, Brasil.

icarodelima@alu.ufc.br, {davydmelo, freitas, danielo}@ufc.br

Abstract. *Colony Collapse Disorder (CCD) is a phenomenon related to the sudden honey bees vanishing in managed colonies. Noted in the USA since 2006, some possible causes of CCD are climatic variations, incorrect use of chemical pesticides, diseases, pests, among others. In this sense, creative computational solutions can contribute to a better understanding of the bees health and welfare. Here, we apply machine learning models to detect anomalies in acoustic patterns of Africanized honey bees (*Apis mellifera* L.) and in seasonal temperature and humidity patterns of European beehives. Three predictive models were implemented: Gaussian Mixture Model (GMM), Extreme Learning Machine (ELM) and Support Vector Machine for a class (OC-SVM). We used datasets with temperature and humidity from two European hives located in Bournemouth (England) and Würtzburg (Germany), as well audio from one colony in Fortaleza-CE (Brazil). For temperature and humidity, the ELM reached an average accuracy of 92.6%. Regarding audio, we highlight the GMM (average accuracy of 84.9%) for queenless state detection using acoustic patterns data.*

Resumo. *O Distúrbio do Colapso das Colônias (Colony Collapse Disorder, CCD) é um fenômeno associado ao desaparecimento repentino de abelhas melíferas em colônias manejadas. Registrado nos EUA desde 2006, o CCD tem como possíveis causas desde variações climáticas, uso incorreto de defensivos químicos até pragas e doenças. Neste sentido, soluções computacionais criativas podem contribuir para um melhor entendimento da sanidade e do bem-estar das abelhas. Neste artigo, aplicamos modelos de aprendizagem de máquina para detectar anomalias em padrões acústicos de abelhas melíferas africanizadas (*Apis mellifera* L) e em padrões de temperatura e de umidade sazonais internas em colmeias com abelhas de raça europeia. Três modelos preditivos foram implementados: Modelo de Mistura de Gaussianas (GMM), Máquina de Aprendizado Extremo (ELM) e Máquina de Vetor de Suporte para uma classe (OC-SVM). Utilizamos datasets com dados de temperatura e umidade internas de duas colmeias de abelhas melíferas localizadas nas cidades de Bournemouth (Inglaterra) e Würtzburg (Alemanha) e um dataset com áudio de uma colônia melífera em Fortaleza-CE (Brasil). Para temperatura e umidade, os melhores resultados ocorreram para anomalias sazonais, nas quais o ELM alcançou uma acurácia média de 92,6%. Para os áudios, destacamos o GMM (acurácia média de 84,9%) na detecção de ausência da rainha.*

1. Introdução

Estima-se que cerca de 35% das culturas agrícolas e quase 90% das plantas silvestres com flores dependam de agentes polinizadores [Klein et al. 2007, Ollerton et al. 2011]. Além disso, a polinização para produção comercial de alimentos em escala mundial é de centenas de bilhões de dólares [Lautenbach et al. 2012]. Dentre as abelhas, que são os polinizadores mais importantes [Bezerra et al. 2019], destacam-se as da espécie *Apis mellifera* [Braga et al. 2020][Silva et al. 2020].

Entretanto, o chamado Distúrbio do Colapso das Colônias (*Colony Collapse Disorder*, CCD), um fenômeno associado ao desaparecimento repentino de abelhas operárias melíferas em colônias manejadas, registrado nos EUA desde 2006, tem preocupado cientistas, biólogos, apicultores. As evidências apontam muitas causas para o CCD as quais vão desde mudanças climáticas, perda e fragmentação de habitat, até pragas, patógenos e doenças [Brown et al. 2016], [Sánchez-Bayo and Wyckhuys 2019].

Para monitorar o estado de saúde das abelhas melíferas, os apicultores periodicamente realizam inspeções para avaliar indicadores-chave (e.g. estágio da cria, presença de pragas, dentre outros) eventualmente realizando intervenções. Estas inspeções normalmente estressam as abelhas e não apresentam ao apicultor análises em tempo real, reduzindo a efetividade das intervenções. De forma a reduzir as inspeções *in loco* e obter informações em tempo real, observa-se um crescente uso da chamada Apicultura de Precisão (AP), a qual consiste em estratégias de gerenciamento de apiários com base em monitoramento das abelhas para reduzir o consumo de recursos e maximizar a produtividade das abelhas [Zacepins et al. 2015].

Dentre os problemas tratados sob o ponto de vista da Apicultura de Precisão, dois deles são de nosso particular interesse neste artigo: (i) a perda da qualidade da termorregulação (i.e., da capacidade das abelhas em manter o microclima da colmeia) e (ii) a ausência da abelha rainha. A ausência da rainha quebra a organização harmônica da colônia por falta dos seus feromônios, leva a ausência de crias e desestrutura o trabalho normal das operárias. Uma possível solução para este problema é o sensoriamento e processamento do áudio (zumbido) da colônia de forma proativa. Observa-se, por exemplo, que o áudio na faixa de 225-285 Hz aumenta em intensidade cerca de duas ou três semanas antes da colônia enxamear [Howard et al. 2013].

Ocorre que a captura de dados indicadores de eventos-problema da apicultura é dificultada pelo altamente provável estado de normalidade das colônias. Adiciona-se a este fato a impossibilidade de estimular a ocorrência artificial de tais eventos de anormalidade tanto por motivos éticos quanto práticos. Desta forma, busca-se técnicas que façam uso intensivo dos dados provenientes da classe normal da colônia a fim de treinar modelos computacionais que reconheçam seu padrão de normalidade de modo que estejam prontos para sinalizar em suas saídas a presença de entradas distintas daquelas aprendidas. Esta classe de problemas é tratada por uma subárea de aprendizagem de máquina chamada de detecção de anomalias (*anomaly detection*) [Pimentel et al. 2014] [Davidson et al. 2020].

O objetivo deste artigo é detectar estados indesejados de termorregulação e de ausência da rainha em colônias de abelhas melíferas. Para tal, propomos o uso de modelos de detecção de anomalias em dados de (i) temperatura e umidade para abordar

a questão da termorregulação e (ii) amostras de áudio capturadas na presença e na ausência da abelha rainha.

2. Material e Métodos

Utilizamos dois conjuntos de dados (*datasets*) em colmeias com abelhas de raça europeia nas cidades de Bournemouth (Inglaterra) e Würzburg (Alemanha) e um *dataset* de áudio de uma colônia do tipo africanizada em uma colmeia de observação localizada na cidade de Fortaleza-CE.

2.1. Rotulagem dos *datasets* de temperatura e umidade (abelhas europeias)

Para as colmeias dos apiários localizados em Bournemouth e Würzburg, os dados foram obtidos através do plataforma HOBOS¹. Para estes apiários, foram analisados os dados de temperatura e umidade interna das duas colmeias sensoriadas. Os dados obtidos registravam amostras a cada 5 minutos durante um ano (2018 para Bournemouth e 2019 para Würzburg). Rotulamos os dados em classes normais e anômalas de acordo com o seguinte procedimento:

$$y = \begin{cases} 1, & \mu - 2\sigma < x < \mu + 2\sigma \\ -1, & \text{caso contrário} \end{cases}$$

em que, y é a classe correspondente a cada amostra x , podendo retornar *positiva ou normal* (1) ou *negativa ou anomalia* (-1). E μ e σ são, respectivamente, a média e o desvio padrão da distribuição analisada.

A partir de então, criamos duas abordagens para cada conjunto de dados das duas colmeias (*Bournemouth* e *Würzburg*). O primeiro leva em consideração anomalias na perspectiva do ano inteiro, isto é, levando em consideração a média (μ_{anual}) e desvio padrão ($\pm 2\sigma_{\text{anual}}$) das distribuições de temperatura e umidade do respectivo ano. O segundo conjunto foi gerado analogamente, porém, leva em consideração anomalias na perspectiva de cada estação do ano isolada. Ou seja, se uma amostra é distante da média dos dados do período do inverno (μ_{inverno}) por dois desvios padrões de inverno ($\pm 2\sigma_{\text{inverno}}$), ou mais, então ela é uma anomalia na perspectiva daquela estação.

2.2. Padrões acústicos de abelhas africanizadas

Para o apiário de Fortaleza-CE, foram coletadas amostras de áudios gravados dentro de uma colmeia de observação (Figura 1) durante seis dias, dos quais nos cinco primeiros dias havia a presença da rainha e no sexto dia a rainha foi retirada da colmeia. Assim, coletamos amostras de áudio com e sem rainha, constituindo um delineamento entre classes normais (rainha presente) e anômalas (rainha ausente).

Dos dados obtidos, foram extraídas cinco características (*features*) espectrais para análise. O centroide espectral (Figura 2.a) corresponde ao centro de gravidade do espectro; o espalhamento espectral (Figura 2.b) é definida como o segundo momento central do espectro; a entropia espectral (Figura 2.c) diz respeito à entropia do

¹<http://hobos.de>

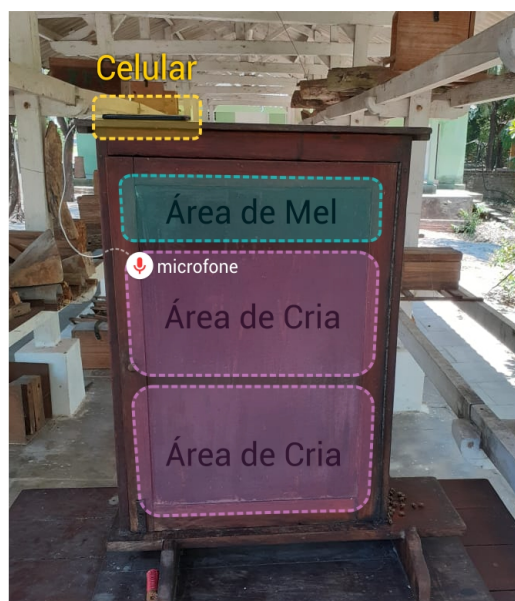


Figura 1. Colmeia de observação destacando o sensoriamento de áudio - apiário em Fortaleza-CE.

conjunto de energias espectrais normalizadas; o fluxo espectral (Figura 2.d) é o quadrado da diferença entre as magnitudes normalizadas entre dois frames sucessivos do espectro; o *Rolloff* espectral (Figura 2.e) corresponde à frequência na qual 90% da distribuição de magnitude do espectro está concentrada.

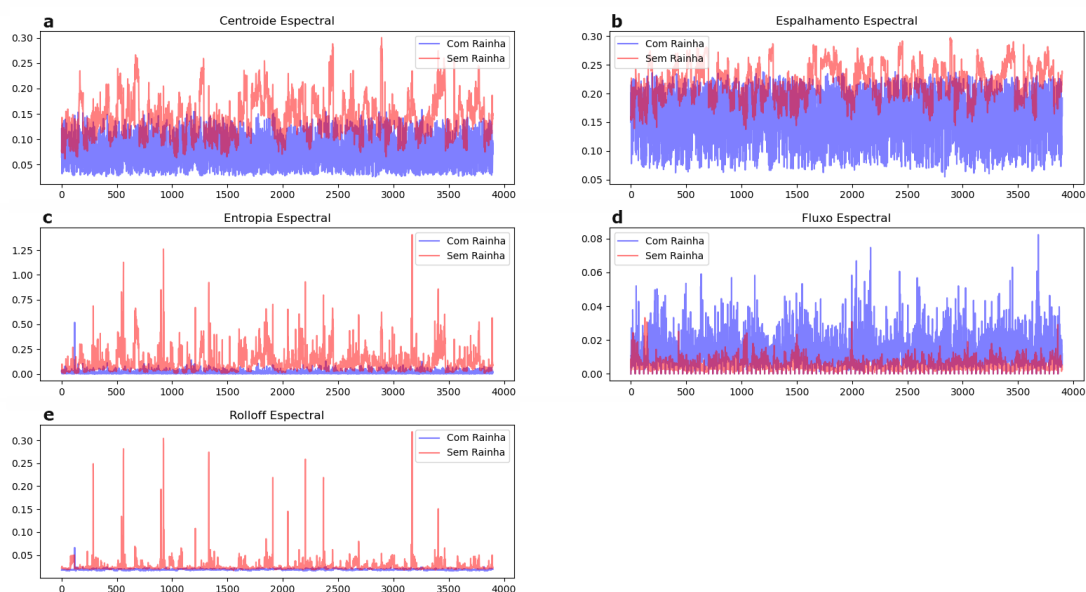


Figura 2. Características espectrais (*features*) de áudio de abelhas melíferas africanizadas observadas em Fortaleza-CE (a) centroide espectral (b) segundo momento central do espectro (c) entropia espectral (d) fluxo espectral e (e) *Rolloff* espectral.

2.3. Modelos Preditivos

Foram escolhidos três modelos de aprendizagem de máquina preditivos baseados em suas abordagens distintas e aplicabilidade para o problema de detecção de novidade: (i) OC-SVM, um método baseado em domínio, (ii) GMM, um método de detecção de novidade da classe probabilística e (iii) ELM, o qual é uma rede neural não-recorrente de única camada oculta.

Os modelos foram avaliados de acordo com a métrica de acurácia média, a qual foi calculada ao longo de 100 iterações e, para cada uma delas, a parte dos dados normais do conjunto de teste foi reamostrada randomicamente. Consequentemente, uma nova acurácia foi calculada para cada iteração. Desta maneira, temos uma noção mais real do desempenho do modelo para dados de teste que variam um pouco ao invés de analisar apenas o melhor caso.

Muito embora as abordagens tradicionais seguirão de maneira direta e usual, no caso da abordagem sazonal, os modelos serão levemente diferentes. Nesta abordagem, um modelo (*ELM*) se subdivide em réplicas do mesmo para cada uma das quatro estações do ano (e.g. $ELM_{ver\tilde{a}o}$, ELM_{outono}). Isto é, cada modelo preditivo, na verdade, são quatro modelos que são alternados conforme a o período do ano da amostra analisada. Sendo assim, cada submodelo treina e prediz com os dados da estação do ano em que ele é especializado.

2.3.1. OC-SVM

As máquinas de vetores de suporte para uma classe são métodos de detecção de novidade baseadas em domínio. Esse tipo de modelo descreve as bordas, ou domínio, da classe alvo ao invés da densidade [Pimentel et al. 2014]. O objetivo é usar uma hiperesfera para descrever o domínio dos dados de maneira que seu volume seja o menor possível mas que inclua a maior parte dos dados de treino [Chen et al. 2001].

Temos dois parâmetros utilizados para o OC-SVM, um deles é o *gamma*, já bem conhecido do SVM convencional. No caso de detecção de novidade, temos um outro parâmetro chamado *nu*. O *nu* é um valor de ajuste para que o hiperplano possa rejeitar alguns dados normais disponibilizados durante o treino a fim de se tornar mais preciso para o teste. Os valores para cada parâmetro foram definidos através de testes. Para o *gamma* foram testados valores entre 10^{-6} e 10^6 . Enquanto para o *nu* foram experimentados valores entre 0.05 e 0.5, com passo de 0.05 unidades. O valor escolhido foi 0,1 tanto para o *nu*, quanto para o *gamma*. Foram avaliados ainda o tipo de *kernel* entre linear, polinomial e função de base radial (rbf). Dentre eles, o *kernel* rbf apresentou maior acurácia.

2.3.2. GMM

Os modelos de misturas gaussianas são métodos probabilísticos em detecção de novidade. Estes métodos buscam estimar a função densidade de probabilidade geral ou *probability density function* (PDF) dos dados através de somas ponderadas de componentes gaussianas [Reynolds 2009].

O principal hiperparâmetro do GMM é o número de componentes, ou número de gaussianas que o modelo usa para o aprendizado. Neste trabalho, o número de componentes é escolhido entre uma faixa de 1-10, sendo escolhido aquele que retornar o menor valor de Akaike Information Criterion (AIC). Até seria possível percorrer números de componentes mais elevados (> 10), contudo, isto torna o modelo mais complexo sem necessariamente agregar um aprendizado mais eficiente. O valor limiar de normalidade foi configurado de maneira que rejeitasse 10% das pontuações de normalidade mais inferiores vistas durante o treino.

2.3.3. ELM

A máquina de aprendizado extremo é uma rede neural com arquitetura não-recorrente com única camada oculta ou *Single-hidden layer Feedforward Neural Network* (SLFN). A arquitetura do ELM permite um aprendizado muito rápido onde os pesos entre a camada de entrada e a camada oculta são gerados aleatoriamente. O aprendizado é determinado de maneira analítica através do cálculo da matriz pseudo-inversa de Moore-Penrose [Huang et al. 2006].

Para o ELM implementado neste trabalho, foram utilizados 30 neurônios na camada oculta e apenas um nó na camada de saída. Já na camada de entrada, são escolhidos o número de nós de acordo com o número de características utilizadas para o modelo. A escolha destes parâmetros da arquitetura do ELM se deram através de testes e observações durante o experimento que retornaram a melhor acurácia. A função de ativação escolhida foi a tangente hiperbólica e os dados também foram normalizados para este caso. Também utilizamos o limiar de normalidade para rejeitar 10% dos pontuações de normalidade mais inferiores disponíveis no treino.

3. Resultados

3.1. Bournemouth - 2018

Para os dados de temperatura e umidade da colônia de Bournemouth, durante o ano de 2018, obtivemos acurácias médias entre 89,5 e 90,4% entre os modelos na abordagem anual. Analogamente, obtivemos, para a abordagem sazonal de anomalias no bem-estar das abelhas, acurácias médias entre 90,7 e 92,6%. Dentre as duas abor-

Tabela 1. Acurácia média de cada modelo para Bournemouth.

Bournemouth			
Anual	OC-SVM	GMM	ELM
	89,5%	90,4%	89,9%
Sazonal	OC-SVM	GMM	ELM
	90,7%	91,6%	92,6%

dagens, a sazonal foi superior na identificação de anomalias (Tabela 1). O GMM é o melhor entre os três modelos para a abordagem anual (Figura 3) enquanto o ELM é o melhor na sazonal (Figura 4).

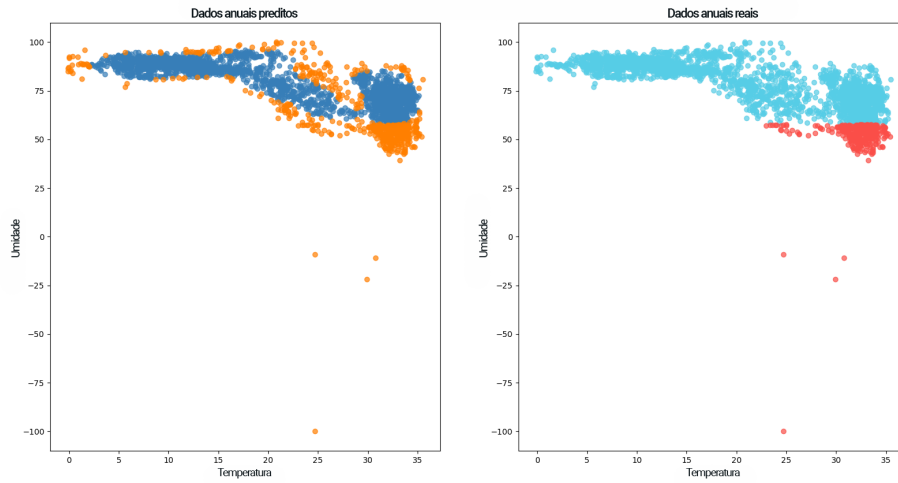


Figura 3. Colônia de Bournemouth: dados preditos pelo GMM (à esquerda) vs. dados reais (à direita) - dados normais em azul, dados anômalos em vermelho/laranja.

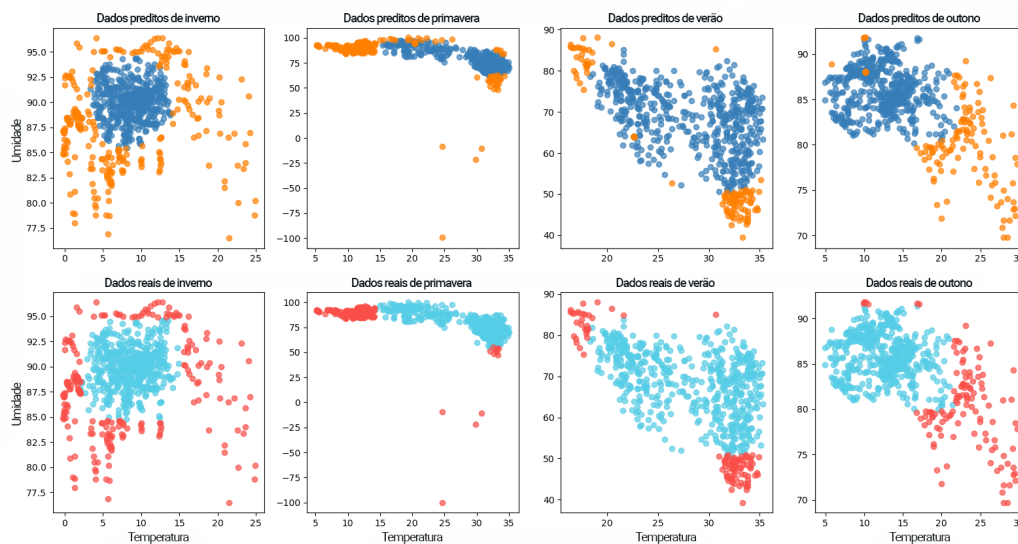


Figura 4. Colônia de Bournemouth: dados preditos pelo ELM (acima) vs dados reais (abaixo) - dados normais em azul, dados anômalos em vermelho/laranja.

3.2. Würzburg - 2019

Dando sequência aos dados de temperatura e umidade da colônia de Würzburg, no ano de 2019. Para este caso as acurácias médias mostravam valores entre 86,3 e 89,4% entre os modelos na abordagem anual. Registrando valores um pouco inferiores a Bournemouth. Em contrapartida, para a abordagem sazonal de anomalias, as acurácias médias ficaram entre 89,6 e 92,6%. Novamente, entre as duas abordagens, a sazonal foi superior na identificação de anomalias (Tabela 2). Assim como para os dados de Bournemouth, o GMM foi o melhor na abordagem anual (Figura 5) enquanto o ELM foi superior na sazonal (Figura 6).

Tabela 2. Acurácia média de cada modelo para Würtzburg.

Württemberg			
Anual	OC-SVM 86,3%	GMM 89,4%	ELM 85,7%
Sazonal	OC-SVM 89,6%	GMM 91,0%	ELM 92,6%

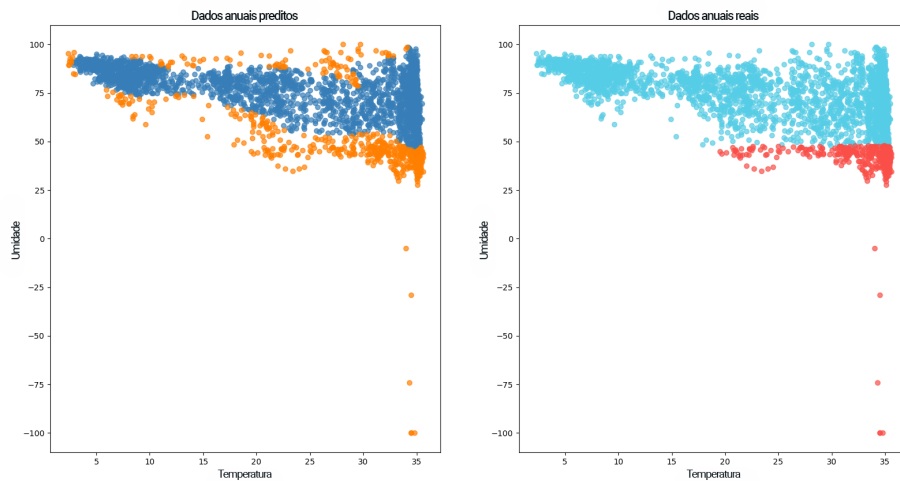


Figura 5. Colônia de Würtzburg: dados preditos pelo GMM (à esquerda) versus dados reais (à direita) - dados normais em azul, dados anômalos em vermelho/laranja.

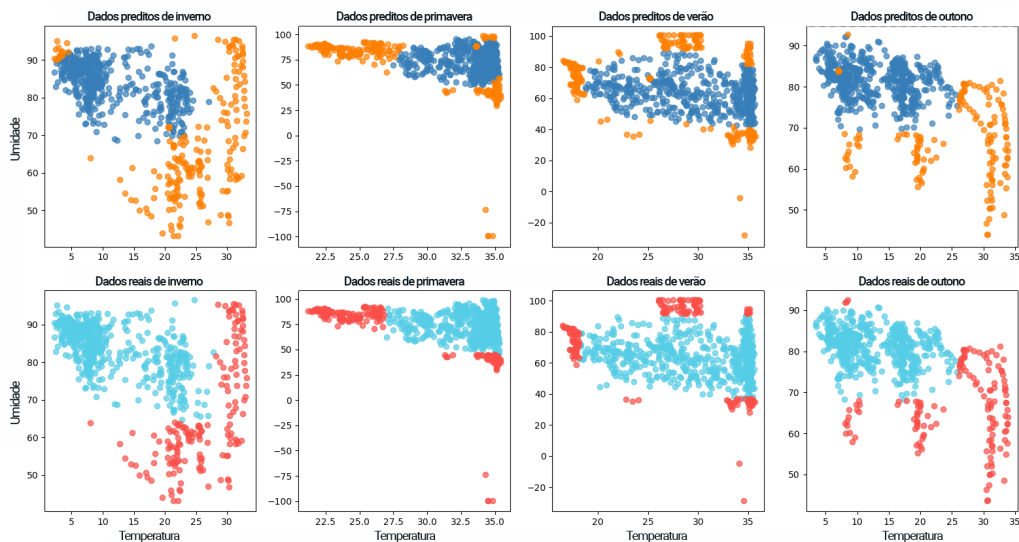


Figura 6. Colônia de Würtzburg: dados preditos pelo ELM (acima) versus dados reais (abaixo) - dados normais em azul, dados anômalos em vermelho/laranja.

3.3. Fortaleza - 2020

Para os dados de áudio produzidos pela colônia de abelhas melíferas africanizadas de Fortaleza, obtivemos resultados que chegaram próximos aos 80 e 85% de acurácia mé-

dia. Com os modelos *OC-SVM*, *GMM*, *ELM* obtendo, respectivamente, 79,5%, 84,8% e 80,2% de acurácia média. Como a anomalia de teste, em questão, foi o padrão acústico emitido pelas abelhas no dia em que a rainha estava ausente da colônia, podemos dizer que a detecção de anomalia prediz com uma taxa de acerto média de 84,9% a ausência ou presença de rainha na colônia.

4. Discussão

Com dados de temperatura e umidade internas e anomalias em perspectiva anual, o *GMM* obteve os melhores resultados (90,4% e 89,5% de acurácia média para Bourne-mouth e Würtzburg, respectivamente). Com os mesmos dados, porém, em perspectiva sazonal, o *ELM* foi o modelo com os melhores resultados (92,6% de acurácia média para ambas as colmeias). Para a identificação de anomalias nos dados de áudio coletados de colônias de abelhas africanizadas (Fortaleza), o *GMM* foi superior aos demais com 84,9% de acurácia média. Sendo assim, é factível e eficiente aplicar modelos de detecção de anomalias no bem-estar de colônias de abelhas melíferas. Como mostrado neste trabalho, trata-se de uma abordagem com aplicações bem generalizadas (e.g. tipos de abelhas, local do apiário, tipo de dados, perspectiva de anomalias).

5. Conclusão

A principal contribuição deste artigo é a detecção de anomalias em padrões acústicos produzidos por abelhas melíferas africanizadas. Embora os dados aqui utilizados tenham sido relativos à ausência de rainha na colônia, os modelos treinados são capazes de identificar quaisquer outros tipos de dados anômalos que se distanciem suficientemente da normalidade (e.g. risco de enxameação, patógenos, abandono). Podemos destacar como segunda contribuição a identificação de anomalias em temperatura e umidade sob uma perspectiva sazonal, sendo uma abordagem mais eficiente e capaz de melhor agregar contextualização ao aprendizado dos algoritmos.

Por fim, acreditamos que este artigo contribui para definir parâmetros de bem-estar e de sanidade para a abelha africanizada em clima tropical, algo ainda inédito. O estado da arte e da prática tem tratado somente da abelha europeia de clima temperado.

Agradecimentos

O presente artigo foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Danielo G. Gomes e Breno M. Freitas agradecem o suporte financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) através dos processos 432585/2016-8, 310317/2019-3, 308948/2016-5 e 308948/2019-8.

Referências

Bezerra, A. D. M., Pacheco Filho, A. J., Bomfim, I. G., Smagghe, G., and Freitas, B. M. (2019). Agricultural area losses and pollinator mismatch due to climate changes endanger passion fruit production in the neotropics. *Agricultural Systems*, 169:49–57.

- Braga, A. R., Gomes, D. G., Freitas, B. M., and Cazier, J. A. (2020). A cluster-classification method for accurate mining of seasonal honey bee patterns. *Ecological Informatics*, 59:101107.
- Brown, M. J., Dicks, L. V., Paxton, R. J., Baldock, K. C., Barron, A. B., Chauzat, M.-P., Freitas, B. M., Goulson, D., Jepsen, S., Kremen, C., et al. (2016). A horizon scan of future threats and opportunities for pollinators and pollination. *PeerJ*, 4:e2249.
- Chen, Y., Zhou, X. S., and Huang, T. S. (2001). One-class svm for learning in image retrieval. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, volume 1, pages 34–37. IEEE.
- Davidson, P., Steininger, M., Lautenschlager, F., Kobs, K., Krause, A., and Hotho, A. (2020). Anomaly detection in beehives using deep recurrent autoencoders. *arXiv preprint arXiv:2003.04576*.
- Howard, D., Duran, O., Hunter, G., and Stebel, K. (2013). Signal processing the acoustics of honeybees (*apis mellifera*) to identify the "queenless" state in hives. *Proceedings of the Institute of Acoustics*, 35:290–297.
- Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3):489–501.
- Klein, A.-M., Vaissiere, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., and Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the royal society B: biological sciences*, 274(1608):303–313.
- Lautenbach, S., Seppelt, R., Liebscher, J., and Dormann, C. F. (2012). Spatial and temporal trends of global pollination benefit. *PLoS one*, 7(4):e35954.
- Ollerton, J., Winfree, R., and Tarrant, S. (2011). How many flowering plants are pollinated by animals? *Oikos*, 120(3):321–326.
- Pimentel, M. A., Clifton, D. A., Clifton, L., and Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99:215–249.
- Reynolds, D. (2009). *Gaussian Mixture Models*, pages 659–663. Springer US, Boston, MA.
- Silva, D., Ícaro Rodrigues, Braga, A., Nobre, J., Freitas, B., and Gomes, D. (2020). An autonomic, adaptive and high-precision statistical model to determine bee colonies well-being scenarios. In *Anais do XI Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais*, pages 31–40, Porto Alegre, RS, Brasil. SBC.
- Sánchez-Bayo, F. and Wyckhuys, K. A. (2019). Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation*, 232:8–27.
- Zacepins, A., Brusbardis, V., Meitalovs, J., and Stalidzans, E. (2015). Challenges in the development of precision beekeeping. *Biosystems Engineering*, 130:60–71.