

# Classificação de Leveduras Utilizando Transformada de Hough e Aprendizagem Supervisionada

Diogo Soares da Silva<sup>1</sup>, Lia Nara Balta Quinta<sup>1</sup>, Arnaldo Ibanhe Mongelo<sup>1</sup>,  
Marney Pascoli Cereda<sup>1</sup> e Hemerson Pistori<sup>1</sup>

1 - Universidade Católica Dom Bosco - UCDB  
INOVISAO e Departamento de Biotecnologia - Campo Grande, Brasil<sup>1</sup>  
Web page: [www.gpec.ucdb.br/inovisao](http://www.gpec.ucdb.br/inovisao)

**Resumo**—This work presents a proposal for the location, classification and enumeration of yeasts for the automation of the viability test of the alcoholic fermentation. Thus, we applied computer vision techniques for the location of yeast and machine learning for classification. For the location of the yeasts, Circular Hough transform was used to classify four supervised learning algorithms have been applied, among them the support vector machine, neural networks, k-nearest neighbor and decision trees. The results were compared with the count of an expert to find out if the counts are statistically similar. The results obtained with the method of Spearman to  $\rho$  was 0.9667, with a p value of 0.00016.

**Keywords**-Machine Learning, Hough Transform, Yeasts.

## I. INTRODUÇÃO

A escassez de combustíveis fósseis e as mudanças climáticas causadas pela queima destes combustíveis estão na pauta de discussões em todo mundo. Fontes de energia renovável ganham cada vez mais importância no cenário global, conferindo à agroenergia uma importante fonte de energia para o futuro [1]. O Brasil, em comparação com outros países, apresenta uma das matrizes energéticas mais limpas, tendo na cana-de-açúcar uma cultura de destaque, que representa 14% da energia primária gerada no país [2]. Os fatores que refletem tal importância podem ser relacionados com o clima, solos favoráveis, relevo, apoio governamental e grandes investimentos, que levou, aos grandes produtores, a uma reestruturação produtiva, que está sendo feita por meio do incremento tecnológico, na forma de melhoramento genético e de novos sistemas de colheita, aliadas com outras práticas, proporcionando o aumento de produtividade [3].

Este incremento tecnológico não pode ficar somente na etapa de produção, já que outras etapas como a fermentação alcoólica, que é uma importante etapa na produção do álcool, também merecem uma atenção especial. A fermentação ocorre devido à ação das leveduras, que através de suas necessidades nutricionais, usam o açúcar do mosto (mistura entre melão de açúcar e água) para sua multiplicação e crescimento celular, transformando assim o açúcar em álcool [4]. Durante outras etapas da produção, o mosto pode entrar em contato com outras bactérias, estas consomem o açúcar, porém não trazem nenhum benefício para a fermentação, tornando o controle

da quantidade de bactérias fundamental para a garantia de produtividade.

Este controle, entre bactérias e leveduras, é feito por um especialista, que em determinadas etapas da fermentação faz a coleta de uma amostra do mosto, onde é adicionado um corante para diferenciar as leveduras viáveis das inviáveis, onde as viáveis fazem o processo de fermentação, enquanto as inviáveis não contribuem para a fermentação. Como normalmente as bactérias têm formato diferente das leveduras, estas podem ser diferenciadas com facilidade. Após a preparação da amostra, esta é transferida para uma câmara de Neubauer, que consiste numa lâmina de microscopia dividida em quadrantes de medidas conhecidas, utilizada para facilitar a contagem de bactérias. Assim, é analisada pelo especialista, através de microscópio. Nesta análise é feita a contagem manual das leveduras viáveis, inviáveis e bactérias [5]. Como a contagem pode ser uma tarefa exaustiva, seu resultado pode ser influenciado pelo cansaço, prejudicando o controle das bactérias.

Buscando a melhoria e a automação deste processo, o projeto BioViC foi criado, propondo a utilização de técnicas de visão computacional para o tratamento de imagens microscópicas, com o intuito de contar automaticamente os microorganismos [6].

Este trabalho propõe um módulo baseado em Visão computacional para a localização das leveduras seguido de um classificador baseado em aprendizagem supervisionada para a classificação das leveduras em viáveis e outros (células inviáveis, fundo e bactérias), já que para o controle de viabilidade da fermentação somente as leveduras viáveis são significantes. Como as leveduras possuem um formato circular [7], para a localização das leveduras a transformada Circular de Hough (TCH) foi utilizada. Após a localização, com o centro e a área da levedura, um módulo de extração de atributos é utilizado, para a posterior classificação.

Na etapa de localização, uma série de algoritmos de pré-processamento são utilizados para otimizar os resultados da TCH. Como algumas das técnicas utilizadas possuíam grande variação de parâmetros, alguns experimentos foram realizados para definir o melhor pré-processamento. Para a classificação, foram utilizados quatro tipos de algoritmos de aprendizagem

supervisionada. Estes foram treinados com um conjunto de teste, onde os mesmos atributos foram utilizado para os experimentos, sendo utilizado o teste de Friedman [23] [24] para verificar se existe diferença significativa entre os classificadores. O conjunto de testes e atributos utilizados no treinamento e classificação serão descritos na seção 4. Os algoritmos utilizados para a aprendizagem foram: K vizinhos mais próximos [8], árvores de decisão [9], máquinas de vetores de suporte [10] e redes neurais artificiais [11]. Uma descrição detalhada dos experimentos realizados será apresentada posteriormente.

O restante deste trabalho está organizado da seguinte maneira. Seção 2 com alguns trabalhos correlatos. A apresentação do método proposto pode ser visto na Seção 3. Todo o delineamento experimental utilizado neste trabalho pode ser conferido na Seção 4, com os resultados obtidos no experimento na Seção 5. Na Seção 6, os resultados são analisados e discutidos, deixando a conclusão do trabalho no final desta seção.

## II. TRABALHOS CORRELATOS

Tradicionalmente as análises de imagens microscópicas são feitas de forma manual, com inspeção visual, sendo uma atividade cansativa é propensa a erros. A análise digital de amostras de microscopia vem ganhando cada vez mais espaço, já que a quantidade de dados gerados para a análise estão cada vez maiores [12]. Vários exemplos de aplicações utilizando visão computacional e aprendizagem de máquina para a localização, análise ou classificação de amostras, podem ser citados.

Garg [13] e sua equipe utilizaram a aprendizagem de máquina na previsão da localização subcelular [14]. Eles utilizaram um conjunto de 135,527 proteínas para selecionar seu conjunto de treinamento, onde destas, 26.156 seqüências de proteínas com 11 diferentes localizações foram utilizadas para o treinamento do algoritmo e mais 1.660 seqüências foram utilizadas para o teste. Os algoritmos de aprendizagem utilizados pela equipe foram o k-nearest neighbor (k-NN) e as Redes Neurais Probabilísticas. Segundo os autores, os baixos valores obtidos na localização de alguns pontos decorrem da falta de exemplos no conjunto de treinamento, já que para outros regiões a localização ocorre de forma satisfatória.

Outro trabalho interessante foi realizado por Jones et. al. [15], que desenvolveram um software de código aberto para a exploração e análise de grandes quantidades de dados. Este software possui uma grande quantidade de ferramenta para a análise de imagens de células, podendo gerar dados estatísticos de diferentes características da mesma.

Já Pan et. al. [16] dedicam sua pesquisa na detecção de células troncos em imagens, onde a célula pode assumir diversas variedades de formas. Os autores propõem uma técnica de aprendizagem estatística para o aumento da adaptabilidade do algoritmo para reconhecer diferentes formas de células. O algoritmo consiste em encontrar pontos que são candidatos a célula através de um pré-processamento, e então extrair suas características para o cálculo de um modelo discriminante que faz o treinamento de um classificador, e determina se

este ponto faz parte ou não de uma célula. Outro módulo extrai características da imagem para o cálculo de um novo modelo discriminativo que prevê se dois pontos distintos pertencem à mesma célula. Por fim, é utilizado um algoritmo de agrupamento, que agrupa os pontos pertencentes à mesma célula. Os autores conseguiram uma medida de F-Score em torno de 90% no conjunto de testes utilizado.

Dell'Anna e sua equipe [17] utilizaram a aprendizagem de máquina para a classificação de grãos de pólen após a aplicação da Transformada Infravermelha de Fourier (FT-IR) nas amostras de ar. Foram utilizadas amostras de 11 espécies de pólen coletadas de diferentes espécies de plantas em diferentes regiões. Todos os espectros foram pré-processados, a equipe utilizou análises estatísticas para encontrar espectros com boa discriminância e foi aplicado um classificador baseado em k-NN para identificar espécies de pólen não pertencentes à amostra utilizada. Por fim, um classificador também baseado em k-NN é utilizado para a classificação das amostras. O grupo conseguiu uma precisão global para a classificação de 84%.

Já para a análise de amostras de leveduras, Kim et. al. [18] utilizaram técnicas de visão computacional para o rastreamento de leveduras com o intuito de monitorar o comportamento e mudanças fenotípicas (características visíveis) através do tempo decorrido e das novas gerações de leveduras. Kvarnström [19] utilizou sua pesquisa para encontrar os contornos da célula de levedura em imagens de fluorescência, seu objetivo era encontrar a região onde as leveduras se encontram para facilitar a análise das células com fluorescência. Já Holloway [20] e colaboradores utilizaram classificadores baseados em aprendizagem de máquinas para a análise bioquímica das leveduras.

Embora todos os trabalhos aqui citados façam análises nas células de leveduras, nenhum faz a contagem ou teste de viabilidade das leveduras. Para comprovar a eficiência do método apresentado, os resultados obtidos pelo algoritmo proposto serão comparados com a contagem realizada por um especialista da área.

## III. MÉTODO PROPOSTO

Como as amostras utilizadas para a classificação são imagens do mosto a ser analisado, e as leveduras possuem um padrão de forma [7], um módulo baseado em Visão Computacional foi desenvolvido com o objetivo de detectar e localizar as leveduras nas imagens. Com as informações referentes à posição e área da região de interesse (ROI), um segundo módulo entra em ação para a extração dos atributos das ROIs encontradas. Finalmente, com os atributos extraídos, o módulo de aprendizagem é utilizado para classificar as leveduras como viável, inviável e fundo. Todo o processo realizado pelo algoritmo será mais bem descrito nos tópicos a seguir.

### A. Pré-processamento

Ainda que a Transformada de Hough seja uma técnica muito conhecida de casamento de modelos, ela necessita de algumas técnicas de pré-processamento para seu perfeito

funcionamento [21] [22]. Assim, antes da utilização da TCH, algumas técnicas de pré-processamento são utilizadas para a otimização da localização. Estas técnicas possuem dois princípios básicos, realçar as áreas de interesse, e segmentar estas regiões.

Algo que pode ocorrer na segmentação da imagem é a presença de objetos desnecessários para a detecção das ROIs, como bactérias, bolhas de ar que ocorrem no momento da preparação da amostra, sujeiras na amostra ou diferenças muito grande na luminosidade da imagem. Este tipo de interferência pode ser chamada de ruído, e torna a segmentação um processo ainda mais complexo. Um método de amenizar este problema é a utilização de algum filtro de suavização na imagem original que diminui sua variação de brilho e reduz as falhas. É importante salientar que o filtro de suavização que pode tornar a imagem um tanto borrada, fazendo com que alguns dos detalhes de forma sejam perdidos.

Para o pré-processamento da imagem, uma sequência de técnicas foram utilizadas. Para o realce das bordas das leveduras, um aumento da saturação da imagem foi realizado. Após o aumento na saturação, um filtro de suavização Gaussiana é utilizada para reduzir os ruídos. Posteriormente, a imagem é convertida para tons de cinza e por fim o filtro de Canny é utilizado para encontrar as bordas das leveduras. A Figura 1 representa a execução do pré-processamento, onde em (a) é apresentada a imagem original, (b) ocorre aumento na saturação da imagem, (c) a imagem é suavizada pelo filtro Gaussiano, (d) a imagem é convertida em tons de cinza e em (e) as bordas são localizadas pelo filtro de Canny.

### B. Detecção

Como já foi citado, as leveduras possuem um formato padrão normalmente circular. Assim para a localização das mesmas, a transformada circular de Hough (TCH) foi utilizada. A TCH implementada utiliza uma *Hash Table*, para armazenar os valores dos cálculos dos *senos* e *cosenos*, reduzindo o tempo de processamento, já que o cálculo é feito para cada iteração. Ela também realiza a limpeza dos vizinhos dos pontos máximos do espaço de parâmetros de Hough, com o intuito de amenizar as redundâncias dos círculos marcados.

Após o pré-processamento, a TCH é utilizada, mapeando os pixels da imagem gerada pelo filtro de Canny no espaço de Hough, onde a cada novo pixel gera um círculo no espaço de Hough. Este círculo possui o tamanho do raio igual ao raio do padrão a ser encontrado. Assim a matriz de Hough vai sendo incrementada, já que cada ponto da borda dos círculos gerados descreve o índice da matriz que deve ser incrementada ou votada. Após a realização dos votos na matriz de Hough, os pontos máximos são encontrados através de um limiar pré-estabelecido. Com o término da execução da transformada todas as posições dos pontos máximos de Hough são armazenadas junto com o valor de cada raio, para o cálculo da área da levedura. Com estes valores, o módulo de extração é utilizado para a extração dos atributos das leveduras. A execução da transformada de Hough pode ser visualizada na Figura 1, onde em (f) é marcado os círculos encontrados.

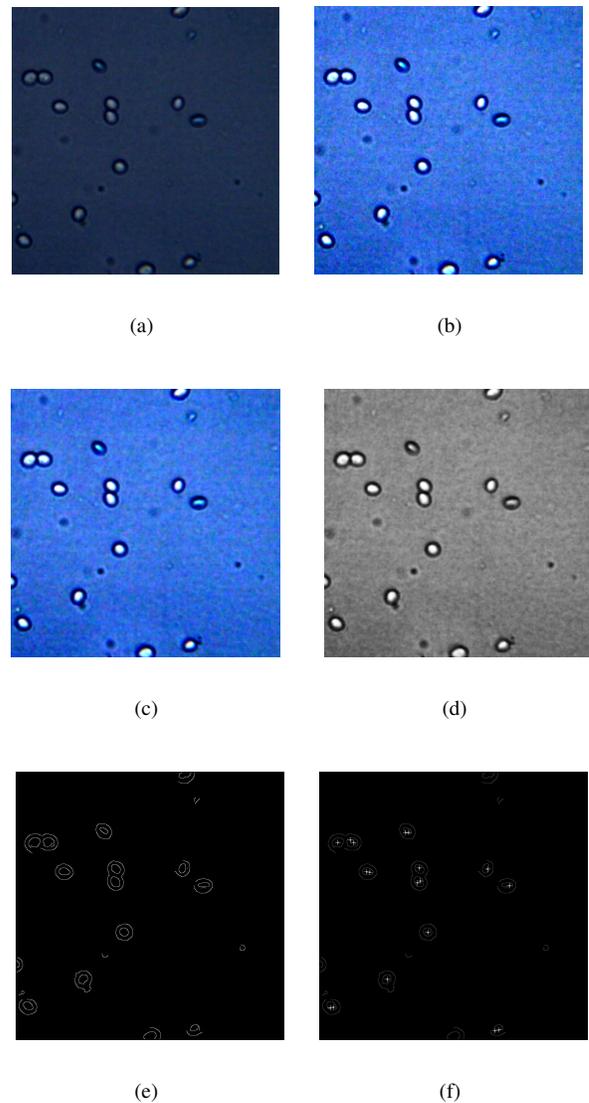


Figura 1: (a) Imagem original, (b) Aumento na Saturação, (c) Suavização, (d) Conversão para tons de cinza, (e) Aplicação do Filtro de Canny, (f) Transformada de Hough

### C. Extração de Atributos e Classificação

Com as ROIs encontradas, informações de diferentes atributos podem ser retirados da imagem. No trabalho proposto, os atributos extraídos foram atributos de forma, cor e textura. Para a extração de atributos de forma foram aplicados os algoritmos de K-curvatura e *Shape Descriptors*, já para os atributos de cor foram utilizados os modelos de cor HSB, e para os atributos de textura foi utilizado mapas de Interação.

Após a extração de atributos o módulo de classificação é utilizado para o treinamento e classificação das leveduras. O classificador é baseado em aprendizagem supervisionada e foi testado com quatro técnicas de aprendizagem, sendo elas. Redes Neurais Artificiais [11], Árvores de Decisão [9], Máquinas de Vetores de Suporte [10] e K-Vizinhos

mais próximos [8]. Somente o classificador com melhor discriminância foi utilizado no módulo final.

#### IV. DELINEAMENTO EXPERIMENTAL

Neste trabalho, foram realizados dois experimentos, um para encontrar o classificador que melhor discrimina as classes e outro para a comparação com a contagem humana. Para o teste de desempenho dos algoritmos de aprendizagem supervisionada de máquina, vários experimentos foram realizados, a fim de encontrar qual dos algoritmos alcançavam a melhor discriminância das classes analisadas. Para todas as técnicas utilizadas o mesmo conjunto de treinamento foi utilizado. Os resultados foram tabulados e o teste de Friedman foi utilizado para verificar se existe diferença significativa entre os classificadores. Já no experimento para a comparação humana, um teste de correlação foi utilizado para saber se as contagens são semelhantes estatisticamente.

##### A. Data Sets

Para o treinamento dos classificadores foram utilizadas 178 imagens que possuíam 1436 amostras de leveduras viáveis e 732 de outros, formando 2 classes para a classificação. Estas leveduras foram retiradas da mesma amostra que o especialista preparou para a contagem, porém as imagens foram retiradas de uma região fora da área de contagem. As leveduras foram recortadas automaticamente das imagens com a utilização da transformada de Hough. A Figura 2 (a) mostra alguns exemplos dos recortes das leveduras viáveis. Já a Figura 2 (b) ilustra os recortes gerados pelos outros elementos na imagem. Estas amostras foram divididas em três conjuntos, sendo o Brix 3 com menor quantidade de açúcar, Brix 6 e Brix 12 com a maior quantidade de açúcar.

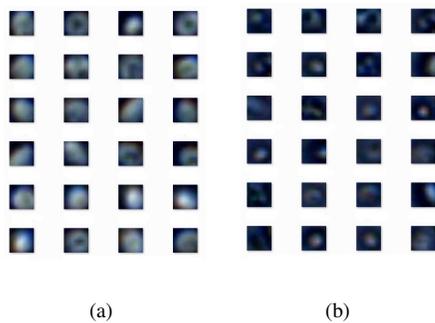


Figura 2: (a) Amostras Viáveis, (b) Amostras Outros

No experimento para a comparação do método proposto com a contagem humana, um conjunto de 225 imagens de uma amostra foi utilizado. Este conjunto de imagens foi retirado de uma amostra do mosto da cana de açúcar em diferentes Brix, definidos pelo especialista da área. As amostras utilizadas para a contagem de leveduras pelo especialista foram as de Brix 3, 6 e 12, onde cada uma das amostras foi feita três repetições, o que originou 9 conjuntos de 25 imagens para a contagem. Onde o especialista e o software realizaram a contagem de

cada repetição, onde a contagem da amostra do Brix é a média da contagem das três repetições.

##### B. Método de Amostragem

Em cada teste feito com os classificadores, o método de amostragem de validação cruzada foi utilizado para a organização dos conjuntos para o treinamento e testes, com número de dobras igual a 10. A validação cruzada consiste em dividir a amostra em  $X$  conjuntos, sendo  $X$  o número de dobra. Cada conjunto com o mesmo tamanho. Após a divisão dos conjuntos,  $X-1$  conjuntos são utilizados para o treinamento do classificador, enquanto o conjunto remanescente é utilizado para o teste, este processo é repetido  $X$  vezes, considerando um conjunto diferente para o teste em cada iteração. O erro da validação cruzada é a média dos erros em cada iteração.

##### C. Métricas Utilizadas

A métrica para avaliar o desempenho dos classificadores utilizados, foi *F-Score*, que é a média harmônica dos valores de *Recall* e *Precision* [25]. Seu cálculo pode ser conferido na equação 1.

$$F = 2 * \frac{precision * recall}{precision + recall} \quad (1)$$

Após a avaliação dos acertos e erros dos classificadores e cálculos de seus *F-Score*, o teste de Friedman é utilizado para verificar se existe diferença entre os resultados obtidos pelos classificadores. Caso a hipótese nula, onde todos os classificadores são estatisticamente iguais, seja rejeitada, deve ser aplicado o teste de Post-hoc para descobrir qual o classificador que melhor discrimina as classes.

##### D. Organização dos Testes

Como foram utilizados quatro algoritmos para a análise de discriminância, sendo que cada um possui uma grande variação de parâmetros, foram realizados 45 testes. Os testes foram realizados da seguinte maneira: 25 testes realizados com as Redes Neurais Artificiais (RNA), 10 com os K Vizinhos mais próximos (KNN), 5 com as máquinas de Vetores de Suporte (SVM) e 5 com as Árvores de Decisão. Para os testes realizados com as RNA, o valor do parâmetro *Momentum* foi de 0.2 até 1.0, variando em 0.2. Para cada valor de *Momentum* o *LearningRate* foi modificado de 0.1 até 0.5, variando em 0.1. Para todos os testes com as RNAs o valor do tempo de treinamento não foi variado, sendo mantido em 500. Já os K Vizinhos mais próximos, o valor de  $K$  variou de 5 até 15, sendo incrementado em 1. Com as Máquinas de Vetores de Suporte a complexidade do valor de *Cost* ficou entre 0.4 e 1.4, variando em 0.2, com kernel polinomial com expoente igual a 1. E por último, nas Árvores de Decisão, o valor de *ConfidenceFactor* foi de 0.05 até 0.55, variando em 0.1. Sendo estes valores os limites superior e inferior dos classificadores, já que acima ou abaixo destes valores o resultado não é satisfatório.

Com o classificador com a melhor discriminância das três classes, o segundo experimento pode ser executado, utilizando o segundo *data set*, para a comparação da contagem do método proposto com a contagem humana.

## V. RESULTADOS

Os resultados dos melhores valores para cada classificador no primeiro experimento podem ser conferidos na Tabela I. Onde J.48 são as Árvores de Decisão, IBk os k-vizinhos mais próximos, SMO as Máquinas de Vetores de Suporte e a RNA as Redes Neurais Artificiais. O valor em parênteses é o desvio padrão para cada técnica.

	J.48	IBk	RNA	SMO
Brix 3	0.984(0.010)	0.987(0.009)	0.986(0.010)	0.987(0.010)
Brix 6	0.966(0.020)	0.971(0.017)	0.967(0.019)	0.969(0.018)
Brix 12	0.972(0.012)	0.973(0.012)	0.969(0.011)	0.969(0.011)

Tabela I: Resultados do teste em diferentes classificadores para Brix 3, 6 e 12

Com a aplicação do teste de Friedman, o valor obtido para p-valor foi de 0.1282, como o nível de confiança adotado foi 5%, a hipótese nula é aceita, já que o valor-p é maior do que 10%. Assim não existe diferença significativa entre os classificadores.

Por obter um valor maior na maioria dos experimentos, normalmente com um menor desvio padrão, o classificado k-NN, com k igual a 13, foi utilizado para a contagem das leveduras. A Tabela II mostra os resultados obtidos pelo método proposto na classificação de leveduras viáveis, em comparação com a contagem humana.

	Hough + ML	Humana
Brix 3 R1	191	190
Brix 3 R2	216	233
Brix 3 R3	257	255
Brix 6 R1	194	197
Brix 6 R2	178	178
Brix 6 R3	73	67
Brix 12 R1	270	259
Brix 12 R2	265	239
Brix 12 R3	239	210

Tabela II: Resultados obtidos na classificação das leveduras viáveis.

Para a análise do coeficiente de correlação entre as contagens, foi utilizado o método de Spearman [26]. Na análise dos resultados obtidos com as leveduras viáveis, o valor de  $\rho$  foi de 0.9667, demonstrando uma grande correlação entre os dados, já que quanto mais perto de 1, maior a correlação dos dados. Com um valor-p de 0.00016, como o grau de confiança assumido foi de 5%, pode-se assumir que a contagem do classificador e a contagem humana é estatisticamente a mesma. Demonstrando a eficiência do método apresentado.

## VI. CONCLUSÃO

Como pode ser visto nos resultados obtidos proporcionaram uma contagem das leveduras viáveis estatisticamente semelhante à contagem humana, o controle de viabilidade das leveduras pode ser encontrada pelo cálculo da quantidade de leveduras viáveis por litro, assim o método proposto pode ser aplicado à automação do controle de viabilidade na fermentação alcoólica.

## AGRADECIMENTOS

Este trabalho recebeu apoio financeiro da Universidade Católica Dom Bosco, UCDB, da Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul, FUNDECT e do Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq.

## REFERÊNCIAS

- [1] E. Campelo and I. L. Michels, "A expansão da agroindústria canavieira em mato grosso do sul," *Revistade Política Agrícola*, no. 3, 2009.
- [2] P. e. A. Ministério da Agricultura, "Balço nacional de cana-de-açúcar e agroenergia." 2007.
- [3] A. M. S. Oliveira, "As novas fronteiras do agronegócio no brasil." Fórum da Cana de Açúcar: impactos sócio-ambientais da cana de açúcar, 2007.
- [4] F. J. Ribeiro, J. J. Lopes, and S. E. Ferrari, "Complementação de nitrogênio de forma contínua no processo de fermentação alcoólica." *Brasil Açucareiro*, vol. 105, no. 1, 1987.
- [5] S. R. C. Antonini, "Método de análise e monitoramento microbiológico em laboratório de destilaria." Usina de Açúcar Santa Terezinha Ltda., 2004.
- [6] L. N. B. Quinta, "Desenvolvimento de um sistema de visão computacional para o controle microbiano em produção de etanol." Master's thesis, Centro de Ciências Exatas e da Terra, Universidade Católica Dom Bosco, 2009.
- [7] E. Baron and N. C. for Clinical Laboratory Standards, *Abbreviated identification of bacteria and yeast: proposed guideline*, ser. NCCLS document. NCCLS, 2000.
- [8] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "Knn model-based approach in classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*. Springer, 2003.
- [9] P. E. Utgoff, "Incremental induction of decision trees," *Mach. Learn.*, vol. 4, 1989.
- [10] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, 1999.
- [11] X. Zhang, M. Kanematsu, H. Fujita, X. Zhou, T. Hara, R. Yokoyama, and H. Hoshi, "Application of an artificial neural network to the computer-aided differentiation of focal liver disease in mr imaging." *Radiological Physics and Technology*, vol. 2, 2009.
- [12] L. Coelho, E. Glory-Afshar, J. Kangas, S. Quinn, A. Shariff, and R. Murphy, "Principles of bioimage informatics: Focus on machine learning of cell patterns," in *Linking Literature, Information, and Knowledge for Biology*, 2010.
- [13] P. Garg, V. Sharma, P. Chaudhari, and N. Roy, "Subcellprot: predicting protein subcellular localization using machine learning approaches." *In Silico Biol*, vol. 9, 2009.
- [14] S. M. Chabregas, "Caracterização da localização subcelular da proteína thi1 de arabisidopsis thaliana." Master's thesis, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo., 2001.
- [15] T. R. Jones, I. H. H. Kang, D. B. Wheeler, R. A. Lindquist, A. Papallo, D. M. Sabatini, P. Golland, and A. E. Carpenter, "CellProfiler Analyst: data exploration and analysis software for complex image-based screens." *BMC bioinformatics*, vol. 9, 2008.
- [16] J. Pan, T. Kanade, and M. Chen, "Learning to detect different types of cells under phase contrast microscopy," *Microscopic Image Analysis with Applications in Biology (MIAAB)*, 2009.
- [17] R. Dell'Anna, P. Lazzeri, M. Frisanco, F. Monti, F. Malvezzi Campeggi, E. Gottardini, and M. Bersani, "Pollen discrimination and classification by fourier transform infrared (ft-ir) microspectroscopy and machine learning." *Anal Bioanal Chem*, vol. 394, 2009.
- [18] K. Kim, A. C. Rowat, and A. E. Carpenter, "Automated tracking of yeast cell lineages," 2010.
- [19] M. Kvarnström, K. Logg, A. Diez, K. Bodvard, and M. Käll, "Image analysis algorithms for cell contour recognition in budding yeast," *Opt. Express*, vol. 16, 2008.
- [20] D. T. Holloway, M. Kon, and C. Delisi.
- [21] R. O. Duda and P. E. Hart, "Use os the hough transformation to detect lines and curves in pictures," *Graphics and Image Processing*, vol. 15, no. 01, 1972.
- [22] P. V. C. Hough, "Method and means for recognizing complex patterns," *USENIX Security Symposium*, 1962.

- [23] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, 2006.
- [24] S. García and F. Herrera, "An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons," *Journal of Machine Learning Research* 9, 2008.
- [25] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Advances in Information Retrieval*, 2005, vol. 3408.
- [26] S. Lyerly, "The average spearman rank correlation coefficient," *Psychometrika*, 1952.