

IDENTIFICAÇÃO AUTOMÁTICA DE DOENÇAS UTILIZANDO IMAGENS DA GLIA ENTÉRICA

Gustavo Zanoni Felipe (PIBIC/CNPq), Yandre Maldonado e Gomes da Costa
(Orientador), e-mail: gzf1996@gmail.com, yandre@din.uem.br.

Universidade Estadual de Maringá / Centro de Tecnologia/Maringá, PR.

Ciências Exatas e da Terra / Ciência da Computação

Palavras-chave: Células da Glia Entérica, Doença, Reconhecimento de Padrões

Resumo:

Quando um indivíduo deve realizar verificações de rotina para conferir a existência de alguma doença, o mesmo submete-se à diferentes exames que variam de acordo com a doença em questão. Este processo pode ser custoso e, na maioria das vezes, demorado. Porém, e se existisse um método de averiguar a existência de diferentes doenças de forma única? Pesquisas recentes envolvendo o Sistema Nervoso Entérico (SNE) mostram que diferentes doenças afetam a glia entérica de forma observável. Desta forma, este projeto teve como objetivo desenvolver um sistema capaz de identificar automaticamente se uma amostra de imagem, cujo conteúdo apresenta células da Glia Entérica, possui características de um indivíduo saudável ou doente. As doenças que foram analisadas são: Câncer (Tumor de Walker-256), o diabetes mellitus e a artrite reumatoide. A base de dados de imagens da glia entérica utilizada neste trabalho foi criada por pesquisadores do Laboratório de Plasticidade Neural Entérica da Universidade Estadual de Maringá. Neste trabalho, os classificadores foram desenvolvidos utilizando técnicas modernas, baseadas em *Deep Learning*, para executar o *representation learning*, de uma forma também conhecida como *non-handcrafted*. Além disso, este trabalho avaliou a performance de características extraídas no modo *handcrafted*, considerando a extração de características por meio de descritores do conteúdo textural de imagens digitais, como: *Local Binary Pattern* (LBP), *Local Phase Quantization* (LPQ) e *Robust Local Binary Pattern* (RLBP). Em adição, foi investigado se a complementaridade entre classificadores criados nos modos *handcrafted* e *non-handcrafted* melhoraria o desempenho do sistema.

Introdução

O intestino delgado é totalmente dependente do Sistema Nervoso Entérico (SNE), sistema que atua no controle do aparelho digestório, para que sejam coordenados os seus diferentes padrões de movimento: propulsão rápida de conteúdos (peristaltismo), movimentos de mistura (segmentação), propulsão lenta e retropulsão (expulsão de substâncias nocivas associadas com vômitos). Os componentes principais do SNE são duas redes ou plexos neuronais: o plexo mioentérico (ou de Auerbach) e o plexo submucoso (ou de Meissner). Neurônios encontrados nos plexos submucosos regulam a secreção e absorção enquanto aqueles do plexo

mioentérico normalmente são responsáveis por controlar a atividade peristáltica. Outras células que também compõem o SNE, como as células da Glia (CGE), desempenham um papel muito importante na homeostasia das funções do Trato Gastrointestinal (TGI), superando a população neurônios entéricos quantitativamente em até quatro vezes (FURNESS, 2006).

Diferentes doenças podem afetar as CGE tais como neoplasias, diabetes mellitus e artrite reumatoide. Neste sentido métodos que avaliem se estas células são afetadas pelas diferentes doenças são de grande importância. Nesse contexto, este trabalho desenvolveu um sistema de identificação automático que teve como objetivo averiguar a presença de doenças em amostras de imagens de CGE do SNE. Dessa forma foi abordada uma metodologia que testa diferentes cenários utilizando *handcrafted* e *non-handcrafted features* como entrada em métodos de classificação.

Materiais e métodos

Base de Dados

Neste trabalho foi utilizada uma base de dados originada a partir de uma parceria com o Laboratório de Plasticidade Neural Entérica (PNE) da Universidade Estadual de Maringá (UEM). Essa conta com amostras de imagens retiradas do SNE de ratos, em que se visualiza as CGE. A base é composta por 1348 amostras de imagens, divididas em três grupos separados de acordo com a doença existente no animal e duas classes por grupo: a das amostras doentes e a das amostras controle/saudável (C). São as doenças trabalhadas: Artrite Reumatoide (AIA), Câncer (TW) e Diabetes Mellitus (D). A divisão das amostras dentre os três grupos, ocorre da seguinte forma: 210 amostras doentes e 208 amostras C para o grupo da doença AIA; 192 amostras doentes e 224 amostras C para o grupo da doença TW; 290 amostras doentes e 224 amostras C para o grupo da doença D.

Metodologia

Este trabalho utilizou de duas abordagens principais. A primeira, baseada em Costa (2013), extrai *handcrafted features* das amostras de CGE por meio de descritores de textura. Dentre os descritores utilizados estão o *Local Binary Pattern* (LBP), *Robust Local Binary Pattern* (RLBP) e o *Local Phase Quantization* (LPQ), de Ojansivu (2008). Tais *features* são utilizadas como entrada nos algoritmos de classificação *Support Vector Machines* (SVM), *Random Forests* (RF), *k-Nearest Neighbors* (k-NN) e *Naive-Bayes* (NB).

A outra abordagem se fundamenta na extração de *non-handcrafted features*. Para que isso fosse concretizado, técnicas de *Deep Learning* foram empregadas para a realização de uma extração inteligente de *features* e classificação. *Convolutional Neural Networks* (CNN) (LECUN, 1989), foram utilizadas para que esta tarefa fosse realizada. Ao total, três arquiteturas CNN foram empregadas, sendo elas: LeNet5, AlexNet e MaxNet. Além disso, foi aplicado o conceito de *Transfer Learning*. Logo, *features* foram extraídas utilizando-se de CNNs pré-treinadas com o conjunto de dados *ImageNet* e estas foram classificadas por um algoritmo de classificação. Para isso, as arquiteturas VGG16, Inception V3 e InceptionResNetV2 foram abordadas. Para reduzir o número de características extraídas por tais arquiteturas, o teste do Qui-quadrado foi empregado como método de *feature selection*.

Além disso, regras de combinação de classificadores foram utilizadas como uma forma de investigar uma possível melhoria nos valores encontrados. Logo, as estimativas de probabilidade geradas ao final de cada classificação foram combinadas por meio das regras da soma, máximo e produto.

Resultados e Discussão

Em um primeiro momento, foi realizado um estudo sobre as amostras da base de dados utilizada. A partir disso, concluiu-se que não seria adequado efetuar a classificação de diferentes doenças como classes de um mesmo problema (problema de multiclasse). Logo, o problema seria composto de três classificações binárias, buscando identificar se uma determinada amostra de células da glia entérica apresentava ou não a doença em questão. Ressalta-se, que foram as doenças estudadas: Artrite Reumatoide (AIA), câncer (TW) e Diabetes Mellitus (D). A primeira abordagem experimentada, extraia *handcrafted features* e os classificava. Para isso, a extração de características foi realizada pelos descritores de textura LBP, RLBP e LPQ, juntamente dos algoritmos de classificação SVM, k-NN, NB e RF. Adicionalmente, diferentes pré-processamentos foram testados, como alterar a coloração presente nas amostras e destacar bordas. Foi possível observar uma tendência a melhores resultados quando utilizados o descritor de textura LPQ e o algoritmo de classificação SVM. A eficiência deste descritor, se dá ao fato de que o mesmo foi desenvolvido para ser eficaz à imagens que apresentam *blur*, características presente nas amostras deste conjunto de dados. Os melhores resultados encontrados para cada doença, podem ser observados na Tabela 1

Tabela 1 – Melhores resultados dos experimentos que classificavam *handcrafted features*.

Doença	Configuração	Precision	Recall	F-Measure	Acurácia
AIA	LPQ ₁₁ + SVM	0,8331	0,8325	0,8325	0,8325
TW	LPQ ₁₃ + SVM	0,9712	0,9712	0,9712	0,9712
D	LPQ ₁₃ + SVM	0,8968	0,8969	0,8967	0,8968

Observação: para as doenças AIA e D, um pré-processamento fora realizado nas amostras, onde as mesmas foram convertidas para tons de cinza (AIA) e pseudo-coloridas para o *colormap* HSV (D).

Já para a segunda abordagem experimentada, *non-handcrafted features* foram extraídos por métodos de *Deep Learning*. Experimentos utilizando de CNNs apresentaram taxas de classificação regulares, porém, inferiores às até então encontradas. Foi pelo uso de Transfer Learning que melhores taxas de classificação foram geradas. Estas, extraíam características utilizando de CNNs pré-treinadas e as classificava por meio do algoritmo SVM. Os melhores resultados para estes experimentos, podem ser vistos na Tabela 2. Assim, nota-se uma pequena melhora, quanto à valores anteriores, por parte das doenças AIA e D.

Por último, classificadores de ambas as abordagens foram combinados. Diferentes disposições de combinações foram montadas e combinadas pelas regras da soma, máximo e produto. Foi por meio desta abordagem que os melhores resultados foram encontrados neste trabalho. As acurácias finais, foram de 88,99%, 98,56% e 94,36% para as doenças AIA, TW e D. Estas utilizaram respectivamente de três, cinco e três

classificadores. Em todos os casos, foram apresentados classificadores que utilizavam de *handcrafted* e *non-handcrafted features*.

Tabela 2 – Melhores resultados dos experimentos que classificavam *non-handcrafted features*.

Doença	CNN	Precision	Recall	F-Measure	Acurácia
AIA	InceptionV3	0,8474	0,8469	0,8468	0,8469
TW	VGG16	0,9327	0,9327	0,9327	0,9327
D	VGG16	0,9051	0,9047	0,9043	0,9046

Conclusões

Este trabalho objetivou desenvolver um método de diagnóstico que utilizasse de imagens da CGE para averiguar a presença de diferentes doenças em diversas amostras. Para que isso fosse alcançado, foram extraídas características em dois modos (*handcrafted* e *non-handcrafted*) e estas seriam utilizadas como entrada em modelos de classificação. Adicionalmente, uma metodologia híbrida foi abordada, onde classificadores formados a partir de tais modos seriam combinados com o intuito de averiguar uma possível complementaridade entre os mesmos. Ao fim deste trabalho, foi possível observar uma efetividade da abordagem que combinava classificadores gerados a partir de *handcrafted* e *non-handcrafted features*. Por meio desta, os melhores valores de acurácia foram alcançados. Sendo estes de 88,99%, 98,56% e 94,36% respectivamente para os grupos das doenças AIA, TW e D. Mostrando assim, uma complementaridade entre tais tipos de *features* e demonstrando taxas de classificação satisfatórias para o problema abordado. Em trabalhos futuros, pretende-se expandir o conjunto de dados para outros tipos de células, como Neurônio Entérico.

Agradecimentos

Ao professor Yandre Maldonado e Gomes da Costa pela orientação e pela oportunidade proporcionada. Ao CNPq pelo apoio financeiro.

Referências

COSTA, Y. M. G. **Reconhecimento de gêneros musicais utilizando espectrogramas com combinação de classificadores**. 2013. 125 f. Tese (Doutorado) – Programa de Pós-graduação em Informática, Universidade Federal do Paraná, Curitiba, 2013.

FURNESS, J. B. The enteric nervous system. **Blackwell Publishing**, p. 290, 2006.

LECUN, Y. Generalization and network design strategies. Em: PFEIFER, R. et al (Ed.). **Connectionism in perspective**. [S. l.]: Elsevier, 1989.

OJANSIVU, V.; HEIKKILÄ, J. Blur insensitive texture classification using local phase quantization. **Image and Signal Processing**. p. 236-243, 2008.