

Detecção de pessoas em imagem de praia.

Leonardo A. Monte, João Paulo S. M Lima
Departamento de computação - DC
Universidade Federal Rural de Pernambuco - UFRPE
{leopk98@gmail.com, joao.mlima@gmail.com}

Resumo: Monitoramento de ambientes aquáticos, por conta de riscos como afogamento e ataque de tubarão é uma tarefa que requer grande atenção e constante trabalho por equipe de resgates. Um software de sistema inteligente pode atuar diretamente no reconhecimento de banhistas em zonas de riscos e alertar uma equipe de resgate, de forma que, acidentes sejam evitados e haja uma redução no quantitativo de mão de obra necessário para a tarefa. Neste artigo é apresentado os resultados da aplicação dos algoritmos de classificação (MÁQUINA DE VETORES DE SUPORTE, REGRESSÃO LOGÍSTICA e MLP) em conjunto com os descritores de características (HOG, HU MOMENTS e LOCAL BINARY PATTERNS) em comparativo com os resultados da aplicação de uma CNN no reconhecimento de pessoas em imagens tiradas na praia de Boa viagem, Pernambuco. O conjunto de imagem é composto por fotos de pessoas e não pessoas em diferentes tamanhos. Os resultados sugerem que a utilização da CNN em comparação com os outros modelos é vantajoso e na abordagem clássica o melhor modelo é o da MLP.

Abstract: Aquatic environments monitoring because the risks of drownings and shark attacks, is a task that requires big attention and constant job from rescues teams. A intelligent software system can act directly in the recognize of bathers in danger zones and generate an alert to the rescue team, thus that, accidents can be avoided and can occur a reduction in the necessary amount of manpower to the job. In this work is presented the application results of the classification algorithms (SUPPORT VECTOR MACHINE, LOGISTIC REGRESSION and MLP) together with the characteristics descriptors (HOG, HU MOMENTS and LOCAL BINARY PATTERNS) in comparative with an application of a CNN in the recognize of persons in images taken in Boa viagem beach, Pernambuco. The set of images is composed by person and non person photos in different sizes. The result suggest that the utilization of an CNN against the other models is worthy and in the classical approach the MLP has better results.

I. INTRODUÇÃO

Com a construção do porto de Suape no estado de Pernambuco, Brasil, e conseqüentemente a destruição dos mangues (local onde os tubarões da

espécie cabeça-chata tinham os seus filhotes) os tubarões que ali residiam tiveram que buscar novos lugares para alimentação e reprodução. Outros fatores como abundância de alimentos e grandes navios cargueiros que servem como uma espécie de “guia” levaram a um alto nível de incidentes de ataques de tubarão nas praias de Pernambuco. Um dos locais mais afetado por tais problemas é a cidade do Recife, localizada em Pernambuco, estado do nordeste brasileiro, com uma população média de 4 milhões de habitantes, que tem como um forte ponto econômico o turismo. Recife representa cerca de 5,15% do número total de turistas no Brasil, possui uma costa de cerca de 20 km de águas de temperatura quente, que atrai pessoas de todo o mundo a procura dessas praias. De 1992 a 2006, foram registrados cerca de 47 ataques de tubarão, incluindo 17 fatalidades, que foram altamente noticiadas por canais de mídia brasileiros e internacionais, prejudicando economicamente o mercado de turismo pernambucano[1]. Correntes marítimas são fatores de riscos para banhistas desavisados que não conhecem as mediações da praia e tendem a ser arrastados para zonas de risco de morte por afogamento. Existem nas imediações das praias salva vidas que fazem a vigia e fiscalização da área, de modo que seja evitado o maior número possível de acidentes. Porém, tal tarefa não é de fácil realização devido ao grande número de banhistas e a grande extensão das áreas restritas de banho. Avisos em tais zonas são constantes e numerosos, porém, nem sempre são relevados pela população gerando assim acidentes e uma constante necessidade de monitoramento. Um algoritmo eficiente de classificação de banhistas e não banhistas, é um primeiro passo para um sistema de monitoramento de câmeras em tempo real e eficiente utilizando inteligência artificial, garantindo uma melhor fiscalização e

uma diminuição nos riscos de acidentes e ataques de tubarões na costa pernambucana. Tal problema encontra barreiras na detecção devido ao clima e ambiente da praia, como por exemplo: Variação na luminosidade dependendo do horário do dia, posição do sol referente ao movimento de rotação da terra, banhistas parcialmente submersos na água e a posição da câmera. Pensando nisso, este artigo apresenta a utilização de um conjunto de algoritmos de aprendizado de máquina (MÁQUINA DE VETORES DE SUPORTE, REGRESSÃO LOGÍSTICA, MLP e CNN) em imagens tiradas na praia de Boa viagem, Recife, Pernambuco, de banhistas e não banhistas.

Foram utilizados três descritores de características na construção deste artigo, o descritor Hu moments, o HOG e o LBP (Local Binary Patterns). Todos os três foram utilizados separados e de forma conjunta. A técnica de PCA (Análise de componentes principais) foi utilizado para uma redução na quantidade de dados e melhoria no desempenho e velocidade de execução dos algoritmos

II. BACKGROUND

Esta seção discute os métodos de extração de características usados e os algoritmos de classificação.

A. Descriptores de características

- 1) HOG(Histograma de Gradientes Orientados): A ideia principal por trás do descritor HOG, é que uma imagem pode ser representada por seus gradientes, mais especificamente contando a ocorrência de padrões de gradientes locais.
- 2) Hu moments: O conjunto Hu de momentos invariantes é um dos mais antigos e bem estabelecidos descritores de imagens. Seus valores permanecem imutáveis mesmo sob mudança de tamanho, rotação ou translação. Os sete momentos são definidos da seguinte forma:

$$\begin{aligned}
\phi_1 &= m_{20} + m_{02} \\
\phi_2 &= (m_{20} - m_{02})^2 + 4m_{11}^2 \\
\phi_3 &= (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2 \\
\phi_4 &= (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2 \\
\phi_5 &= (m_{30} - 3m_{12})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{12} + m_{03})^2) \\
&\quad + (3m_{21} - m_{03})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} - m_{03})^2) \\
\phi_6 &= (m_{20} - m_{02})((m_{30} + m_{12})^2 - (m_{21} + m_{03})^2) + 4m_{11}(m_{30} + m_{12})(m_{21} + m_{03}) \\
\phi_7 &= (3m_{21} - m_{03})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{12} + m_{03})^2) \\
&\quad - (m_{30} - 3m_{12})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2)
\end{aligned}$$

Fig1. Definição dos sete momentos Hu

- 3) LBP (Local Binary Patterns): Local Binary Patterns ou LBP é um descritor de texturas que se tornou popular a partir do trabalho de Ojala et al. no seu artigo do ano de 2002, entretanto o conceito de LBP foi introduzido em meados de 1993. LBP computa uma representação local de textura, essa representação é construída comparado cada pixel com seus pixels em zonas arredores.

5	8	1
5	4	1
3	7	2

→

0	0	1
0		1
1	0	1

Fig2. Exemplo ilustrativo LBP

Na figura acima fica visível o funcionamento do descritor em uma matriz 3x3. É escolhido o pixel central para comparação com o seu arredor e se o pixel comparado ao central for menor ou igual, a nova matriz de descrição na posição comparada receberá 1, caso não receberá zero. Após esse processo a nova matriz é utilizada para o cálculo da representação decimal de forma em que cada pixel escolhido possua um novo valor que é usado para a construção do histograma que é usado como a descrição LBP da imagem referente.

B. Classificação

Este artigo 4 algoritmos de classificação supervisionados os quais são:

- 1) Máquina de vetores de suporte: É um algoritmo utilizado para regressão e classificação, onde constrói um hiperplano ótimo para a separação de diferentes classes de dados com a maior margem possível. São usados no projeto o SVM com kernel Linear e com Kernel RBF.

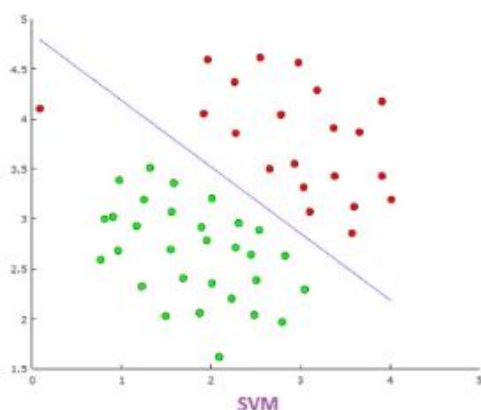


Fig2. Exemplo SVM.

- 2) Regressão logística: É uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica.

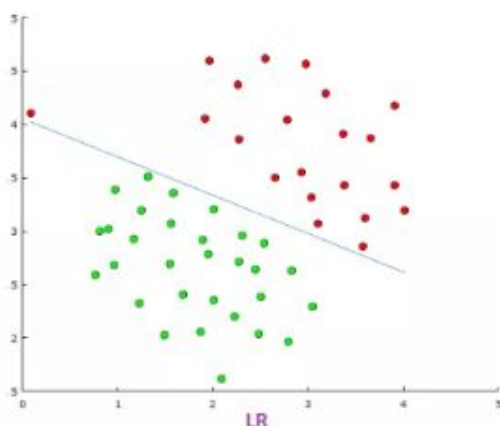


Fig3. Exemplo Regressão logística

- 3) Multilayer perceptron: É uma classe de redes neurais artificiais feedforward, que consiste em ao menos três camadas de nós, uma de entrada, uma escondida e uma de saída.

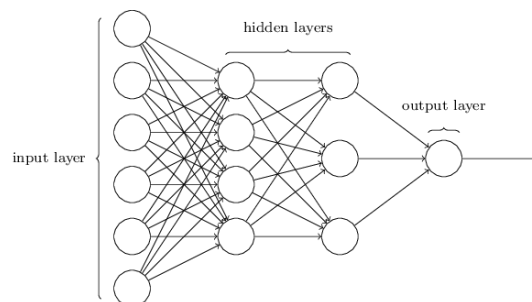


Fig4. Exemplo de uma MLP

- 4) CNN (Convolutional Neural Network): Redes neurais convolutivas é um tipo de rede neural das redes neurais profundas, que possui uma camada de convolução, a qual realiza a partir dos pesos das redes, filtragens em kernels de tamanho escolhidos pelo autor da rede de forma a extrair características da imagem. A cada nova convolução pode ser encontrada camadas de pooling, que diminui a dimensionalidade da imagem de forma em que a próxima convolução consiga descobrir informações em escalas diferentes.

Outras técnicas são utilizadas em conjuntos nessa rede, de forma em que seja obtido melhores resultados para o problema. Ao final das convoluções, é utilizado uma camada totalmente conectada que irá receber a matriz resultante em formato de vetor unidimensional para a classificação do problema.

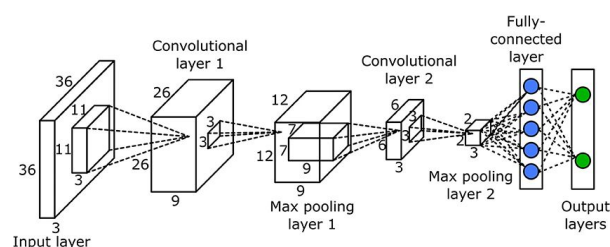


Fig5. Exemplo de uma CNN

III. METODÓLOGIA

Nesta seção será abordada a metodologia utilizada para a realização dos experimentos.

As imagens consiste numa base de dados formada por 1446 imagens de objetos que podem ser facilmente confundíveis com banhistas, como guarda sol, cadeiras de praia ou coqueiros e fotos recortadas de pessoas nesse mesmo ambiente. As imagens são pré processadas de modo em que se tenha apenas duas dimensões e estejam em formato de binário de representação. Essas imagens são redimensionadas para a dimensão 64x128 para permitir a utilização do descritor HOG.

Após o processo de pré-processamento dessas imagens, são utilizados de forma separada e conjunto os 3 descritores escolhidos para a tarefa. Tendo os vetores características representando cada imagem, é aplicado a técnica de dimensionalização PCA para a diminuição do tamanho dos conjuntos de dados.

Foi utilizado a técnica de 10-Fold cross-validation para a análise dos resultados. Cada 10 iterações do cross-validation foi realizada 5 vezes, tendo como ponto de partida uma nova distribuição dos dados (Foi utilizado uma técnica de shuffle após o término dos experimento dos 10-Folds) gerando assim um conjunto de 50 resultados de acurácia de teste de cada algoritmo.

A partir desses resultados foi realizado o teste de friedman para comparação estatística dos resultados dos algoritmos e foi gerado um boxplot da distribuição para melhor observação dos resultados.

Diferentemente dos algoritmos clássicos, a CNN não precisa da etapa de pré-processamento e extração de características, o próprio algoritmo aprende quais características devem ser utilizadas e otimiza o processo durante a execução. A dimensão das imagens utilizadas para a CNN foi 32x32x3 e o método de avaliação do algoritmo também foi o 10-Fold cross-validation.

IV. RESULTADOS

Nesta seção serão abordados os resultados obtidos pelos algoritmos de classificação.

É levado em consideração a média da acurácia, mediana da acurácia, desvio padrão da acurácia, o tempo de execução e o teste estatístico.

A. Hu moments:

Algoritmo	Média	Mediana	Desvio Padrão	Tempo médio
SVMLINEAR	56,91	56,42	0,0440	0,03
SVMRBF	71,92	71,78	0,0333	0,05
LOGREG	71,44	71,42	0,0377	0,006
MLP	76,67	76,78	0,0492	1,373

Fig6.. Resultados da classificação usando Hu moments

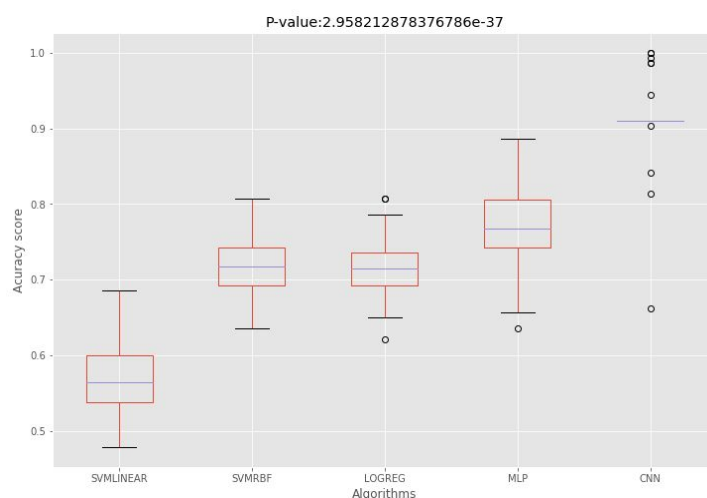


Fig7.. Boxplot dos resultados + teste estatístico + resultados cnn

B. HOG:

Algoritmo	Média	Mediana	Desvio Padrão	Tempo médio
SVMLINEAR	81	81,42	0,0292	0,07
SVMRBF	84,48	84,28	0,0247	0,134
LOGREG	81,05	80,71	0,0309	0,0254
MLP	88,77	89,28	0,0412	1,1376

Fig8.. Resultados da classificação usando HOG

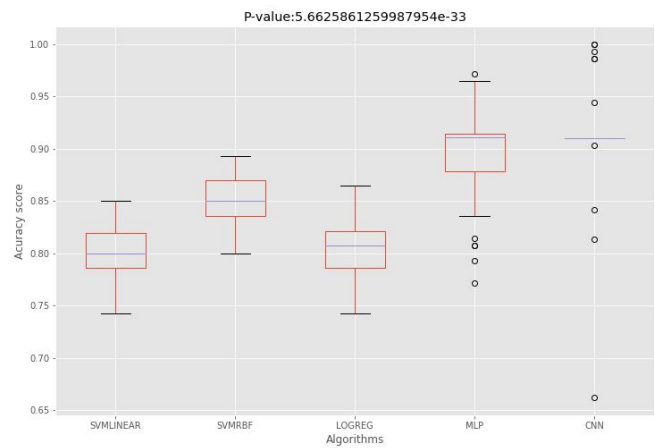


Fig11. Boxplot dos resultados + teste estatístico + resultados cnn

D. Hog e LBP:

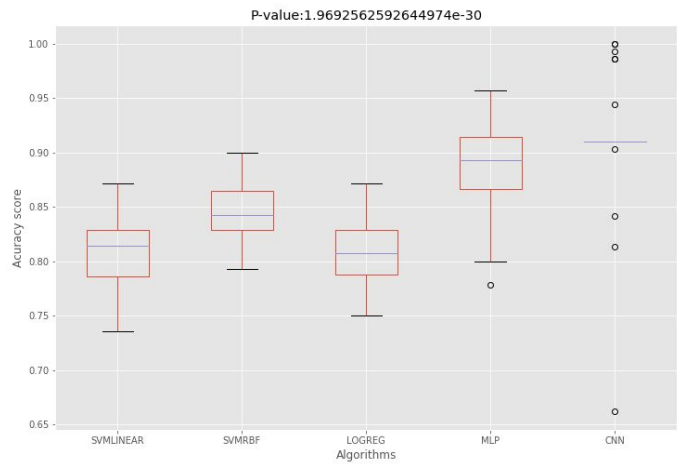


Fig9.. Boxplot dos resultados + teste estatístico + resultados cnn

Algoritmo	Média	Mediana	Desvio Padrão	Tempo médio
SVMLINEAR	79,62	78,57	0,0301	0,07
SVMRBF	82,95	82,85	0,0326	0,135
LOGREG	80,08	79,28	0,0318	0,02
MLP	89,91	91,42	0,0402	1,40

Fig12. Resultados da classificação usando HOG e LBP

C. HOG e Hu moments:

Algoritmo	Média	Mediana	Desvio Padrão	Tempo médio
SVMLINEAR	80,12	80	0,0255	0,07
SVMRBF	85,17	85	0,0229	0,136
LOGREG	80,44	80,71	0,0279	0,024
MLP	89,80	91,07	0,0430	1,37

Fig10. Resultados da classificação usando HOG e Hu.

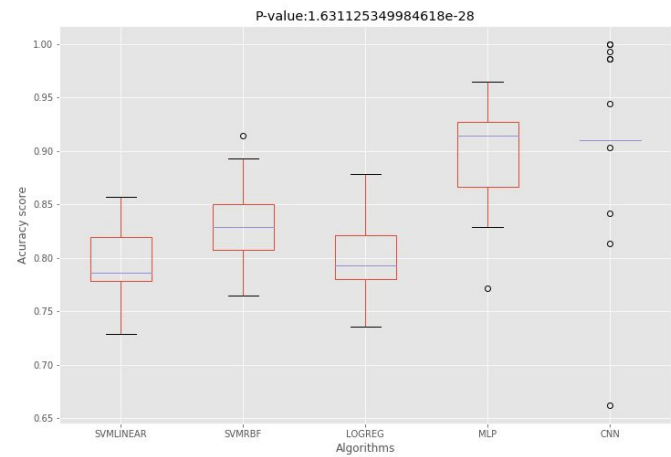


Fig13. Boxplot dos resultados + teste estatístico + resultados cnn

E. LBP:

Algoritmo	Média	Mediana	Desvio Padrão	Tempo médio
SVMLINEAR	79,47	80	0,0339	0,04
SVMRBF	80,68	81,42	0,0320	0,072
LOGREG	78,68	78,57	0,0299	0,02
MLP	84,07	83,92	0,0324	1,45

Fig14. Resultados da classificação usando LBP

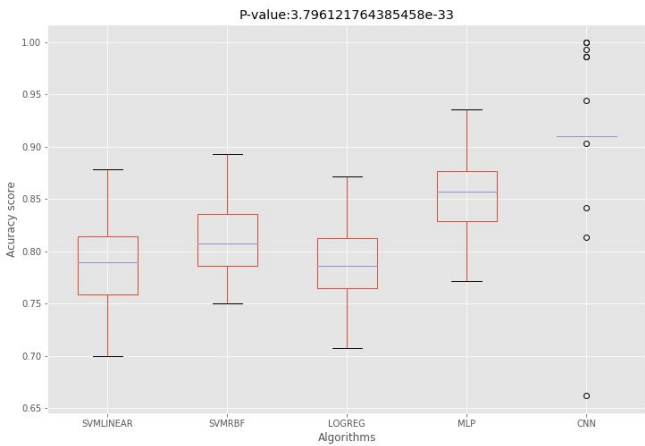


Fig1. Boxplot dos resultados + teste estatístico + resultados cnn

G. LBP, HOG e HU:

Algoritmo	Média	Mediana	Desvio Padrão	Tempo médio
SVMLINEAR	80,05	80	0,3142	0,07
SVMRBF	84,27	85	0,0322	0,138
LOGREG	80,34	80,35	0,3150	0,024
MLP	89,18	89,64	0,0392	1,33

Fig18. Resultados da classificação usando LBP, HOG e HU

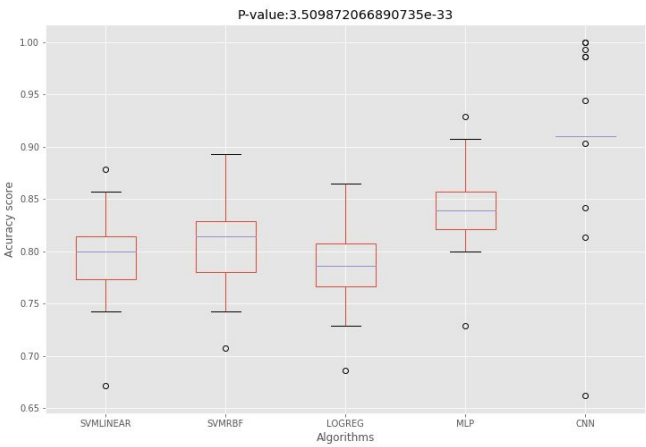


Fig15. Boxplot dos resultados + teste estatístico + resultados cnn

F. LBP e HU:

Algoritmo	Média	Mediana	Desvio Padrão	Tempo médio
SVMLINEAR	80,11	80,71	0,0324	0,05
SVMRBF	80,2	80	0,0337	0,08
LOGREG	79,61	0,8	0,0332	0,025
MLP	83,14	83,57	0,0384	1,42

Fig16. Resultados da classificação usando LBP e HU

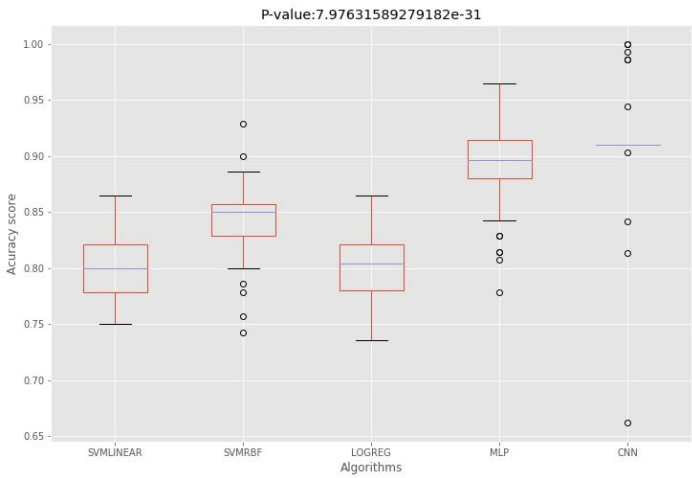


Fig19. Boxplot dos resultados + teste estatístico + resultados cnn

V. CONCLUSÕES

É possível notar a partir da análise das tabelas, que o resultado teve melhores resultados na utilização das redes neurais, MLP e CNN. O melhor resultado foi encontrado utilizando a CNN, chegando a ocorrer predições com 100% de acerto. Na abordagem clássica o melhor resultado foi utilizando em conjunto o descritor HOG com o descritor LBP.

V. Referências.

[1] F. H. V. Hazin, G. H. Burgess, and F. C. Carvalho, "A shark attack outbreak off Recife, Pernambuco, Brazil: 1992-2006," *Bulletin of Marine Science*, vol. 82, no. 2, 2008.