

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA POLITÉCNICA/ ENGENHARIA ELÉTRICA
TRABALHO FINAL DE CURSO II

Gabriel do Prado Silva
Matheus de Toledo Silva

DETECÇÃO DE EMOÇÕES DE CRIANÇAS ATRAVÉS DE FEIÇÕES UTILIZANDO
INTELIGÊNCIA ARTIFICIAL

Trabalho Final de Curso como parte dos requisitos para
obtenção do título de bacharel em Engenharia Elétrica
apresentado à Pontifícia Universidade Católica de Goiás.

BANCA EXAMINADORA:

Prof. Me. Gustavo Siqueira Vinhal – Orientador. PUC – Goiás.

Prof. Me. Fernando Gonçalves Abadia. PUC – Goiás.

Prof. Me. Rafael Leal Martins. PUC – Goiás.

Goiânia, 07 de dezembro de 2021.

Detecção de emoções de crianças através de feições utilizando inteligência artificial

Silva, Gabriel do P., Silva, Matheus de T., Vinhal, Gustavo S., Abadia, Fernando G., Martins, Rafael Leal.

Resumo — O reconhecimento facial de imagens por meio de inteligência artificial está se tornando cada vez mais usado. Tendo em vista a alta aplicabilidade, o volume de pesquisas envolvendo essa área teve um crescimento notável, tornando cada vez melhor a qualidade e a precisão desses programas. Algumas crianças possuem dificuldade para se expressarem. Dito isso, o objetivo deste trabalho é implementar um programa de reconhecimento facial capaz de detectar expressões de crianças.

Palavras Chave – Reconhecimento Facial, Processamento de Imagens, Análise de Sentimentos.

Abstract — Facial image recognition through artificial intelligence is becoming increasingly used. In view of the high applicability, the volume of researches involving this area had a remarkable growth, improving the quality and precision of these programs. Some children have difficulty expressing themselves. That said, the aim of this work is to implement a facial recognition program capable of detecting children's expressions.

Keywords – Facial Recognition, Image Processing, Feeling Analysis.

INTRODUÇÃO

Devido a sua ampla aplicabilidade, a inteligência artificial para reconhecimento de expressões faciais tem crescido nos últimos anos. No entanto trata-se de um processo complexo, com um elevado nível de precisão, onde devem ser levados em consideração os mínimos detalhes para a obtenção de uma resposta genuína e confiável.

Naturalmente é notório que as mães têm instintivamente uma aguçada capacidade de entender quais as vontades e necessidades de seu filho apenas olhando para ele. No entanto, esta habilidade na maior parte dos casos, pertence unicamente a mãe, o que gera uma dificuldade quando a mesma não está presente devido a atividades externas, por exemplo o trabalho.

Nos últimos anos a organização social dentro das casas tem sido alterada devido a população feminina estar cada vez mais ativa economicamente, ou seja, as mulheres cada vez mais têm buscado oportunidades de emprego, seja para complementação de renda ou crescimento profissional.

Tendo isto em vista, surge a necessidade e oportunidade de desenvolvimento de um programa que seja capaz de fazer tais análises para facilitar o dia a dia das famílias, diminuindo também a preocupação das mães ao deixarem as suas crianças aos cuidados de terceiros enquanto estão trabalhando.

Neste trabalho, será abordado inteligência artificial para reconhecimento de expressões faciais em crianças, com intuito de corroborar com os responsáveis a conseguirem compreender com maior eficiência e agilidade as

necessidades da criança para que seja tomada as devidas medidas necessárias

I. EXPRESSÕES FACIAIS

As expressões faciais são movimentos feitos com os músculos do rosto, muitas das vezes essas expressões feitas revelam os sentimentos diante de determinadas situações em que o indivíduo se encontra, sendo assim, se torna essencial saber interpretá-las para que se tenha uma melhor comunicação.

Uma criança pode ter dificuldade em se comunicar, pois o desenvolvimento da criança para que a mesma possa se auto avaliar é um processo que leva um tempo muito individual, onde algumas amadurecem de forma rápida e outras necessitam de mais tempo. A timidez ou insegurança, podem levar a criança a ficar menos comunicativa, utilizando pouco da comunicação verbal e muito da não verbal e expressão corporal.

Tendo isso em vista, é notória a dificuldade que o responsável por uma criança pode enfrentar para identificar quais as necessidades e emoções da criança. Isso torna o processo de tomada de decisão mais complicado, gerando uma demora no tempo de resposta para solucionar o problema de maneira eficaz.

Na Figura 1 é observar uma expressão facial de uma criança.



Fig. 1. Expressão de felicidade em uma criança.
Fonte: [2].

Como pode se observar na Figura 1, tem-se uma expressão facilmente reconhecida, felicidade, é possível notar linhas de expressões ocasionadas devido ao movimento do sorriso, e a partir destas linhas de expressões será montada a parte experimental do trabalho.

II. PROCESSAMENTO DE IMAGENS

Com o advento dos computadores a área de processamento de imagens apresentou um enorme crescimento em suas

funcionalidades, permitindo que várias áreas conseguissem utilizá-la conforme suas necessidades.

O Processamento Digital de Imagens (PDI) obteve sua importância devido à alta versatilidade que possui, sendo muito utilizada em aperfeiçoamento de imagens coletadas por sondas e missões espaciais para melhorar e possibilitar a interpretação humana. Ele é muito utilizado nos aparelhos de Raio-X e ultrassonografia, gerando uma melhora na precisão dos diagnósticos na medicina. Imagens capturadas por câmeras de segurança podem ser tratadas com o intuito de melhorar a nitidez, facilitando o reconhecimento da área filmada.

Abaixo será explicado brevemente o processamento digital de imagens e suas respectivas etapas, em seguida falaremos sobre estrutura de um sistema de visão computacional e seus respectivos passos.

A. Sistema de Processamento de Imagens

A Figura 2 abaixo apresenta o sistema de processamento de imagens e suas respectivas etapas:

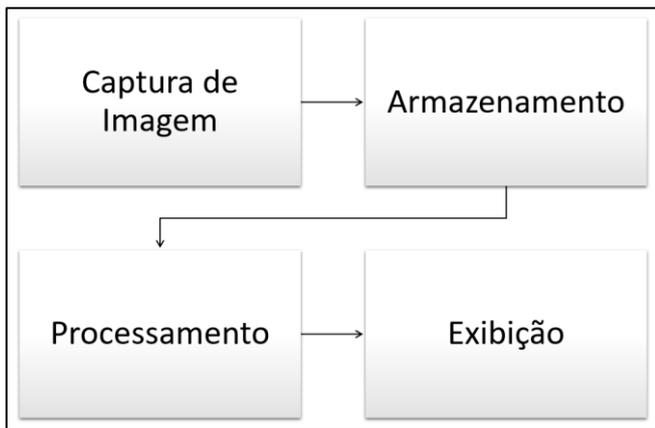


Fig. 2. Diagrama do processamento de imagens.
Fonte: [1].

De acordo com a Figura 2, o sistema de processamento de imagens é dividido em quatro etapas:

i) Captura de Imagem

Trata-se de um dispositivo físico sensível a uma faixa de energia no espectro eletromagnético (como raio X, ultravioleta, espectro visível ou raios infravermelhos), que produz na saída um sinal elétrico proporcional ao nível de energia detectado [1].

ii) Armazenamento

O armazenamento de imagens digitais progrediu muito na última década. Atualmente é possível armazenar uma grande quantidade de bytes, possibilitando utilizar imagens com altas resoluções. Existem três tipos de armazenamento: de curta duração, consistindo na utilização parcial da memória de acesso aleatório (RAM) ou *frame buffers*; utilizando discos magnéticos encontrados em discos rígidos (HD's), podendo suportar até 96TB (também pode-se utilizar discos magneto-ópticos, estes já atingiram marcas de Petabytes (PB)); e, por fim, tem-se os discos ópticos WORM (*Write-Once, Read Many Media*), que gravam uma grande quantidade de bytes, que não podem ser reescritos ou apagados.

iii) Processamento

O processamento de imagem é uma forma de processar dados (imagens) utilizando algoritmos e técnicas matemáticas. É possível implementar *softwares* para realização deste processamento. Caso este processo esteja lento será necessário a utilização de *hardwares* especializados.

iv) Exibição

O monitor de vídeo é um elemento fundamental de um sistema de processamento de imagens [1]. De acordo com a utilização empregada, será utilizada uma resolução e gama de cores distinta para melhor atender as necessidades de cada processo.

B. Estrutura de um Sistema de Visão Computacional

Um Sistema de visão computacional é utilizado para realizar a interpretação e o processamento de imagens do mundo real a fim de solucionar e automatizar processos e problemas. Partindo sempre de uma situação prática, o processo é dividido em sete etapas, conforme mostrado na Figura 3.

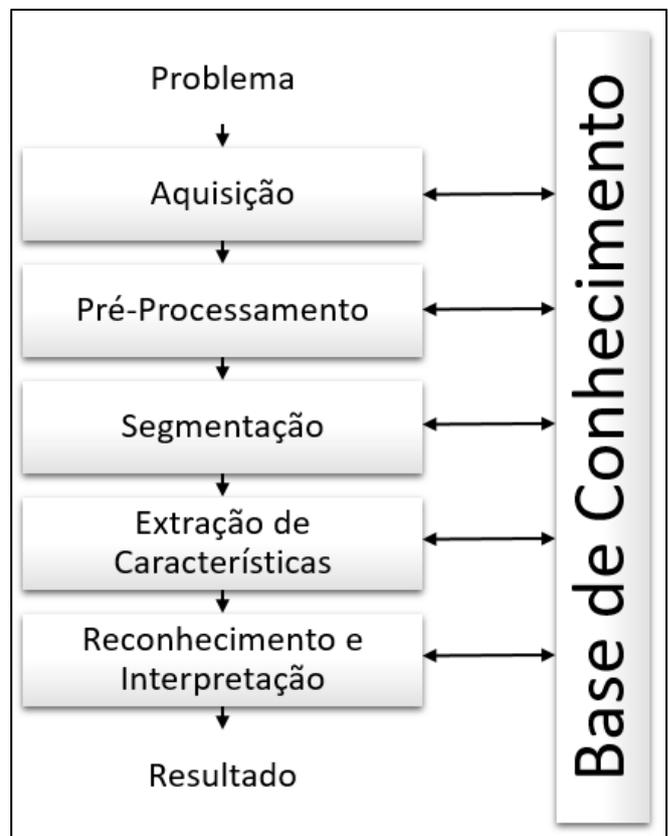


Fig. 3. Diagrama Estrutura de um Sistema de Visão Computacional.
Fonte: [1].

De acordo com a Figura 3, o sistema de visão computacional possui as seguintes etapas:

i) Problema e Resultado

O problema será a problemática a ser resolvida. O resultado esperado é a automatização do processo de forma a retornar apenas a informação necessária.

ii) Aquisição da Imagem

A aquisição de imagem é o procedimento de coletar as imagens referentes a problemática proposta. Ruídos podem

ocorrer devido à grande variabilidade de aparelhos telefônicos, câmeras fotográficas e as condições do ambiente em que a foto será capturada (Baixa luminosidade).

iii) Pré-Processamento

A imagem resultante da etapa anterior pode conter imperfeições tais como: ruído e brilho inadequado. As operações efetuadas nesta etapa são ditas de baixo nível porque trabalham diretamente com os valores de intensidade dos pixels [1].

Para obter sucesso nos passos seguintes é crucial que todas as imperfeições sejam tratadas a fim de passar uma informação real adiante.

iv) Segmentação

A segmentação consiste em obter informações de forma direcionada, para isto o programa deverá identificar e separar as informações necessárias daquilo que será descartado.

v) Extração de Características

Utiliza-se as informações que foram obtidas na etapa anterior para em seguida realizar a extração das características utilizando descritores. Os descritores devem ser representados por estruturas de dados adequadas com o algoritmo que será utilizado.

É importante observar que nesta etapa a entrada ainda é uma imagem, mas a saída é um conjunto de dados correspondentes àquela imagem [1].

vi) Reconhecimento e Interpretação

No reconhecimento é utilizado uma forma de codificação para cada característica coletada anteriormente nos descritores. A interpretação consiste em a partir da codificação, utilizar uma base de conhecimento para comparar as informações coletadas com as demais informações previamente definidas, podendo assim retornar com a informação mais parecida com a de entrada.

IV. METODOLOGIA UTILIZADA

Neste trabalho a técnica utilizada será a Máquina de Vetores Suporte (SVM's). As Máquinas de Vetores de Suporte constituem uma técnica de aprendizado que vem recebendo crescente atenção da comunidade de Aprendizado de Máquina (AM) [3]. Os resultados da aplicação dessa técnica são comparáveis e muitas vezes superiores aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (RNAs) [4, 5].

Essa teoria estabelece uma série de princípios que devem ser seguidos na obtenção de classificadores com boa generalização, definida como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu [6].

As técnicas de AM empregam um princípio de inferência denominado indução, no qual obtém-se conclusões genéricas a partir de um conjunto particular de exemplos. O aprendizado indutivo pode ser dividido em dois tipos principais: supervisionado e não-supervisionado [6].

No aprendizado supervisionado tem-se a figura de um professor externo, o qual apresenta o conhecimento do ambiente por conjuntos de exemplos na forma: entrada saída desejada [5]. O algoritmo de AM extrai a representação do conhecimento a partir desses exemplos. O objetivo é que a representação gerada seja capaz de produzir saídas corretas para novas entradas não apresentadas previamente [6].

No aprendizado não-supervisionado não há a presença de um professor, ou seja, não existem exemplos rotulados. O algoritmo de AM aprende a representar (ou agrupar) as entradas submetidas segundo uma medida de qualidade. Essas técnicas são utilizadas principalmente quando o objetivo for encontrar padrões ou tendências que auxiliem no entendimento dos dados [7].

O desempenho desejado de um classificador f é que o mesmo obtenha o menor erro durante o treinamento, sendo o erro mensurado pelo número de previsões incorretas de (f) . Sendo assim definimos como risco empírico $R_{emp}(f)$, como sendo a medida de perda entre a resposta desejada e a resposta real. A Equação 1 mostra a definição do risco empírico.

$$R_{emp}(f) = \frac{1}{n} \sum_{i=1}^n c(f(x_i), y_i) \quad (1)$$

Onde $c(\cdot)$ é a função de custo relacionada a previsão $f x_i$ com a saída desejada y_i [6], onde um tipo de função de custo é a "perda 0/1" definida pela Equação 2. O processo de busca por uma função f' que represente um menor valor de R_{emp} é denominado de minimização do risco empírico.

$$c(f(x_i), y_i) = \begin{cases} 1, & \text{se } y_i f(x_i) < 0 \\ 0, & \text{caso contrário} \end{cases} \quad (2)$$

Sobre a hipótese de que os padrões de treinamento (x_i, y_i) são gerados por uma distribuição de probabilidade $P(x, y)$ em $R^N \times \{-1, +1\}$ sendo P desconhecida. A probabilidade de classificação incorreta do classificador f é denominada de Risco Funcional, que quantifica a capacidade de generalização, conforme é mostrado pela Equação 3 [8].

$$R(f) = \int c(f(x_i), y_i) dP(x_i, y_i) \quad (3)$$

Durante processo de treinamento, $R_{emp}(f)$, pode ser facilmente obtido, ao contrário de $R(f)$, pois em geral a distribuição de probabilidades P é desconhecida [6].

A partir disto, dado um conjunto de dados de treinamento (x_i, y_i) com $x_i \in R^N$ e $y_i \in \{-1, +1\}$, $i = \{1, 2, \dots, n\}$, sendo x_i o vetor de entrada e y_i o rótulo da classe.

O objetivo então é estimar uma função $f: R^N \rightarrow \{-1, +1\}$. Caso nenhuma restrição seja imposta na classe de funções em que se escolhe a estimativa f , pode ocorrer que a função obtenha um bom desempenho no conjunto de treinamento, porém não tendo o mesmo desempenho em padrões desconhecidos, sendo este fenômeno denominado de "overfitting". Em outras palavras, a minimização apenas do risco empírico $R_{emp}(f)$ não garante uma boa capacidade de generalização, sendo desejado um classificador f^* tal que $R(f^*) = \min_{f \in F} R(f)$, onde F é o conjunto de funções f possíveis [9].

A Figura 4 mostra um exemplo onde uma classe de funções pode ser utilizada para separar padrões linearmente separáveis. É necessário determinar uma função que minimize o R_{emp} na figura como a reta mais escura [9].

Um hiperplano é um subespaço afim, que divide um espaço em duas partes, correspondendo a dados de duas classes distintas [10].

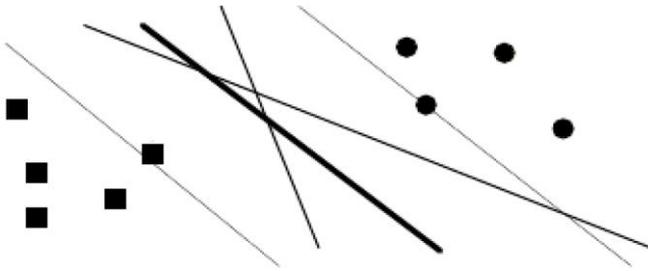


Fig. 4. Classe de hiperplanos com um hiperplano ótimo.
Fonte: [9].

A TAE provê formas de limitar a classe de funções (hiperplanos), com o intuito de prevenir modelos ruins, ou seja, que levem ao "overfitting", implementando uma função com a capacidade adequada para o conjunto de dados de treinamento [11]. Estas limitações são impostas ao risco funcional da função.

V. RESULTADOS E DISCUSSÕES

A técnica SMV utilizada demonstrou um bom resultado quando se tem um banco de dados mais extensivo, que contemple a maior quantidade possível de possibilidades dentro do problema proposto.



Fig. 5. Diagrama do processo do código utilizado.
Fonte: Autoria Própria.

i) Início

Antes da utilização do banco de dados, foi necessário realizar alguns procedimentos: Primeiramente, foi reduzida a escala da imagem para 160x160 pixels, em seguida especificou-se o diretório de cada emoção, separando assim, cada emoção em sua respectiva nomenclatura, por fim foi utilizada a intensidade de 32 bits.

ii) Leitura da Imagem

Esta etapa consiste no processo no qual o programa realiza a análise do conteúdo da imagem, para a partir disto melhor interpretá-la.

iii) Conjunto de Treinamento

Para realização do treinamento, são escolhidas aleatoriamente uma quantidade n de fotos do banco de dados, apenas selecionando uma quantia igual entre as emoções a serem utilizadas, a partir destas fotos o programa aprende a distinguir os padrões entre elas, quanto mais vasto o treinamento realizado, menor será a taxa de erro obtida.

iv) Conjunto de Teste

O conjunto de teste consiste em uma parcela de dados conhecida que será utilizada para realizar o teste de precisão nas respostas a serem obtidas através do algoritmo. Como o valor dos dados de teste é conhecido, em casos de erro o programa é capaz de validar se sua resposta foi correta ou errada.

v) Aplicação da Técnica SVM

A técnica SVM consiste em uma técnica de aprendizado, portanto, a partir do treinamento realizado a técnica é capaz de prever corretamente a classe de novos dados de um mesmo domínio em que o aprendizado ocorreu.

vi) Resultado

O resultado obtido representa a taxa de acerto que o algoritmo alcançou, sendo assim, o resultado esperado é que o classificador obtenha o menor erro possível durante a sua fase de treinamento.

As especificações do computador utilizado para realização dos testes são: Processador: Intel Core I7-4510U, Memória RAM: 16GB e CPU: 2Ghz.

O banco de dados utilizado neste trabalho foi concedido através da faculdade Rutgers University, Newark, NJ, USA. O banco de dados contemplava 948 fotos de crianças expressando diferentes emoções, dentre elas temos: Raiva, Nojo, Medo, Felicidade, Neutro, Tristeza e Surpresa.

Em um primeiro teste foram utilizadas 129 fotos, sendo 25 para teste e 104 para treinamento, contemplando as emoções: Felicidade, Raiva e Neutro. A Tabela 1 apresenta os resultados obtidos.

	Taxa (%)
Precisão de treinamento	100
Precisão de teste	64,00

Tab. 1 - Teste com o primeiro conjunto de dados.
Fonte: Autoria Própria.

De acordo com a Tabela 1, neste teste foi possível observar que uma baixa quantia de fotos para treinamento do programa fez com que o resultado fosse inferior ao esperado, 64,0% de precisão.

A Figura 6 é uma demonstração gráfica da precisão do teste realizado.



Fig. 6. Demonstrativo do Resultado – Teste 01.

As estrelas são os dados originais e os quadrados representam os resultados com a máquina de vetor suporte. O acerto é representado por uma estrela seguida de um quadrado, formando pares, os demais padrões representam um erro.

A partir deste resultado, foi decidido que a utilização de mais fotos acarretaria uma maior precisão. Então foi definido que as emoções para o segundo teste seriam: Felicidade, Raiva e Neutro. Ao todo foram utilizadas 544 fotos distribuídas dentro das três emoções. Destas imagens foram utilizadas 108 para teste e 436 para treinamento.

A Tabela 2 apresenta os resultados obtidos.

	Taxa (%)
Precisão de treinamento	100
Precisão de teste	71,54

Tab. 2 - Teste com o segundo conjunto de dados.
Fonte: Autoria Própria.

De acordo com a Tabela 2, ao aumentar a quantidade de imagens, houve uma melhora, tendo a taxa de acerto de 71,54%. Portanto, foi comprovado que utilizando uma maior quantidade de dados na base a precisão consequentemente irá aumentar.

A Figura 7 apresenta a demonstração gráfica da precisão do teste realizado.

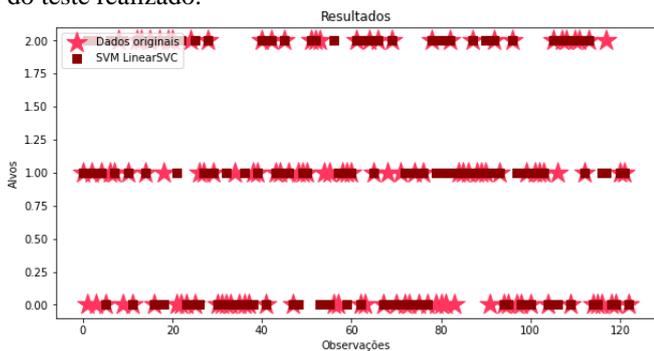


Fig. 7. Demonstrativo do resultado – Teste 02.

A partir da Figura 7 é possível concluir que a quantidade de acertos em comparação com a figura 6 é consideravelmente maior, tendo em vista que mais pares foram formados.

Um problema recorrente na base de dados são os *outliers*: imagens classificadas originalmente como uma emoção, mas que representa outra. A Figura 8 apresenta um exemplo encontrado.



Fig. 8. Demonstrativo de *outlier*. Fonte: [2].

A imagem apresentada na Figura 8 foi, originalmente, classificada como Felicidade. Como observa-se, há um erro de classificação. Não houve correção para manter mais fielmente possível os dados originais.

III. CONCLUSÃO

Este trabalho propõe o desenvolvimento de programa capaz de detectar expressões faciais de crianças. Primeiramente foram apresentados conceitos de processamento de imagens e o algoritmo SVM, em seguida foram apresentados os testes realizados.

Devido a quantidade de imagens presente no banco de dados, em um primeiro momento, optou-se por utilizar uma quantidade menor de imagens, porém com quatro emoções. Devido a baixa taxa de acerto, foi decidido por reduzir a quantidade de emoções em prol do aumento da quantidade de imagens.

Os resultados obtidos foram satisfatórios, ultrapassando os 70% de precisão, no entanto, para obtenção de uma maior precisão será necessária a utilização de um extenso banco de dados. Quanto maior a quantidade de dados, melhor será o resultado obtido.

Outro problema da base de dados são os *outliers*. Manter eles na base de dados para treinamento e teste podem influenciar o algoritmo negativamente. Porém, ajuda a não causar *overfitting* (sobreajuste, ou seja, quando o modelo se adapta muito bem aos dados de treinamento e não consegue prever novos resultados).

Como trabalhos futuros, sugere-se a implementação de outras técnicas de inteligência artificial para melhorar a taxa de acerto. Além disso, sugere-se a implementação de algoritmos para detecção de emoções através de feições em imagens de recém-nascidos.

REFERÊNCIAS

- [1] Filho, Ogê Marques., Neto, Hugo Vieira, Processamento digital de imagens. Rio de Janeiro – Brasport, 1999.
- [2] Rutgers Univeristy, Newark, NJ, USA – data base.
- [3] T. Mitchell. Machine Learning. MCGraw Hill, 1997.
- [4] A. Braga A.C.P.L.F Carvalho, and T.B Ludermir. Redes Neurais Artificiais: Teoria e Aplicações. Editora LTC, 2000.
- [5] S. Haykin. Neural Networks – A Comprehensive Foundation. Prentice-Hall, New Jersey 2nd edition, 1999.
- [6] LORENA, A. C.; CARVALHO, A. C. P. L. de Introdução às Máquinas de Vetores Suporte. São Carlos – SP. Abril 2003.
- [7] M. C. P Souto A. C. Lorena A. C. B. Delbem and A. C. P. L. F. Carvalho. Técnicas de Aprendizagem de Máquina para problemas de Biologia Molecular, pages 103-152. Minicurso de Inteligência Artificial, Jornada de Atualização Científica em Inteligência Artificial, XXIII Congresso da Sociedade Brasileira de Computação. 2003.
- [8] SMOLA, A. J. et al. Introducion to large margin classifiers. In: [S.I]: Morgan_Kauffman, 1999. Cap. 1, p. 1-28.
- [9] GONÇALVES, A. R. Máquina de Vetores Suporte - Unicamp

- [10] LIMA, A. R. G. Máquinas de Vetores Suporte na Classificação de Impressões Digitais. Dissertação (Mestrado) – Universidade Federal do Ceará, Fortaleza, Ceará, 2002.
- [11] HEARST, M. A. et al. Support vector machines. IEEE Intelligent Systems, IEEE Computer Society, Los Alamitos, CA, USA, v. 13, n. 4, p 18-28, 1998. ISSN 1094-7167.

