

RELATÓRIO DE PESQUISA:

Avaliação de metodologias para reconhecimento facial

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA

RELATÓRIO DE PESQUISA:
Avaliação de metodologias para reconhecimento facial

AUTOR: JOÃO GUILHERME DO NASCIMENTO TELES

ORIENTADORA: ELIANA PANTALEÃO

Patos de Minas
2019

Resumo

Diversas áreas de pesquisa voltadas para a análise de imagens foram criadas a partir do desenvolvimento do processo de digitalização. A abundância de dados a serem explorados, derivados desse processo, possibilitaram análises das mais diversas formas. Entre essas análises, o reconhecimento facial é uma das áreas de pesquisa com diversas aplicações voltadas para segurança. Em virtude disso, sistemas que realizam esse tipo de análise possuem diversas etapas, visto que as aplicações necessitam de um alto grau de confiança. Essas etapas se dividem em: detecção da face, adequação da imagem ao algoritmo, descrição e classificação. Este trabalho analisa metodologias de classificação para o reconhecimento facial, buscando utilizar os dados da face da forma mais útil possível. Foram avaliadas duas metodologias de classificação de dados, as redes neurais artificiais e a máquina de vetores de suporte, utilizando como descritor o nível de cinza dos pixels da imagem. A partir disso, foram realizados testes utilizando 3 bases de dados com 30, 60 e 90 imagens. Além disso, foram variados os seguintes parâmetros dos classificadores: arquitetura de rede para a rede neural artificial e o tipo de *kernel* e a constante C para a máquina de vetores de suporte. Após a realização dos testes, a maior taxa de acerto para a RNA foi de 96,67% e para a SVM foi 93,33%.

Palavras chave: Reconhecimento facial. Aprendizagem de máquina. Redes neurais artificiais. Máquina de vetores de suporte.

Abstract

Several areas of research focused on image analysis were created from the development of the digitization process. The abundance of data to be explored, derived from this process, made possible analyses in many different ways. Among these analyzes, facial recognition is one of the areas of research with several security applications. As a result, systems that perform this type of analysis have several steps, as applications require a high degree of reliability. These steps are divided into: face detection, image suitability to algorithm, description and classification. This paper analyzes classification methodologies for facial recognition, seeking to use face data as usefully as possible. Two data classification methodologies were evaluated, the artificial neural networks and the support vector machine, using as a descriptor the gray level of the image pixels. From this, tests were performed using 3 databases with 30, 60 and 90 images. In addition, the following parameters of the classifiers were varied: network architecture for the artificial neural network and kernel type and the constant C for the support vector machine. After the tests, the highest hit rate for RNA was 96.67% and for SVM was 93.33%.

Key words: Face recognition. Machine Learning. Artificial neural networks. Support vector machine.

Sumário

1. Introdução	5
2. Fundamentação teórica	5
3. Metodologia	7
4. Resultados.....	9
4.1. Rede neural artificial.....	9
4.2. Máquina de vetores de suporte	11
4.3. Comparação entre os algoritmos	12
5. Conclusão	13
6. Referências	14

1. Introdução

A criação de imagens digitais possibilitou inúmeros estudos de análise de imagens. Entre esses campos de pesquisa, um que atualmente tem bastante relevância é o estudo dos métodos de reconhecimento facial, que une outras duas áreas de pesquisa em uma só: o estudo de imagens digitais, e o reconhecimento de padrões.

Essa área de estudo se mostrou bastante promissora e com diversas aplicações, entre elas: o desenvolvimento de sistemas de segurança [1], sistemas de reconhecimento para a diminuição da criminalidade [2] e até sistemas de organização de imagens [3]. Essa diversidade de aplicações atraiu um grande investimento para essa área de estudo, possibilitando o desenvolvimento de diversos sistemas robustos e eficazes no reconhecimento facial.

O crescimento da pesquisa sobre reconhecimento facial, em face às suas diversas aplicações, proporcionou a criação de várias vertentes dentro deste campo de estudo. Essas subáreas envolvem a utilização de diferentes descritores [4] e de diferentes classificadores [5,6], bem como as várias combinações possíveis entre descritor e classificador. Existem, ainda, abordagens que utilizam mais de um descritor para melhorar a taxa de acertos [7].

Este trabalho buscou avaliar alguns métodos existentes de reconhecimento facial em imagens digitais e teve como foco a prototipagem de um sistema de segurança que possa ser utilizado instituições que necessitem de serviços com essa camada de segurança.

2. Fundamentação teórica

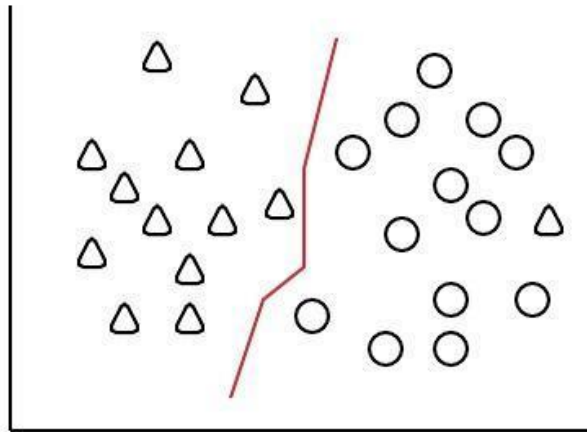
Para o desenvolvimento de um sistema de reconhecimento facial, dois conceitos são importantes: descrição e classificação. A descrição é a maneira com a qual você quer representar a face, ou seja, como deseja descrevê-la. Alguns exemplos de descritores são: níveis de cinza da imagem RGB, níveis de cinza da imagem monocromática, distância euclidiana de pontos pré-determinados da face, histograma dos níveis de cinza e localização de determinados pontos da face. Ou seja, o processo de descrição na área de reconhecimento facial nada mais é que encontrar um método para representar a face numericamente. Por outro lado, o processo de classificação se trata das diversas formas de utilizar os dados que representam a face para treinar um algoritmo que seja capaz de, posteriormente, classificar outras faces.

O descritor utilizado nesse trabalho são os níveis de cinza da imagem monocromática. Uma imagem colorida de $i \times j$ pixels é convertida para tons de cinza e representada por um vetor cujo tamanho é o produto de i por j . Esse vetor, por sua vez, representa os dados de entrada do classificador.

Os classificadores testados nesse trabalho foram a máquina de vetores de suporte (SVM) e as redes neurais artificiais (RNAs). Além de diferentes classificadores, conjuntos de dados de treinamento de diferentes tamanhos também foram utilizados para realizar o treinamento do classificador. Nos próximos parágrafos é explicada a fundamentação básica para a compreensão dos dois classificadores utilizados [8].

A SVM é uma técnica de aprendizagem de máquina que tem como embasamento um modelo geométrico dos dados. Basicamente, essa técnica estabelece uma função capaz de separar o objeto de estudo em classes diferentes, visando a utilização dessa função para prever a classe de outro objeto. Esse tipo de classificação é ilustrada na Figura 1.

Figura 1: Representação gráfica da SVM.

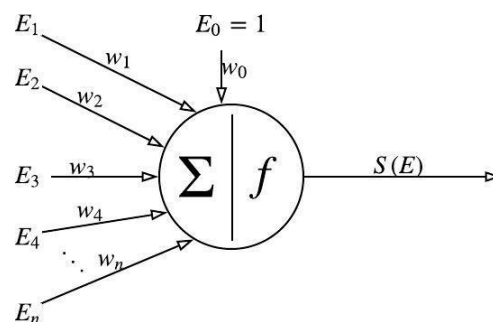


Fonte: O autor.

A Figura 2 mostra um problema simples de classificação onde duas classes, “triângulo” e “círculo”, são projetadas em um gráfico de dois eixos. A curva em vermelho representa a função de separação, também chamada de fronteira de decisão, explicada no parágrafo anterior. Essa fronteira, estabelecida pelos dados de treinamento, é posteriormente utilizada para classificar outros objetos.

A RNA “[...]” é um modelo de grafo em que nós representam neurônios artificiais e as arestas orientadas denotam as conexões entre as entradas e as saídas dos neurônios.” [9]. Esse tipo de sistema de aprendizagem possui uma grande gama de aplicações envolvendo o reconhecimento de padrões e, por isso, foi utilizada nesse trabalho.

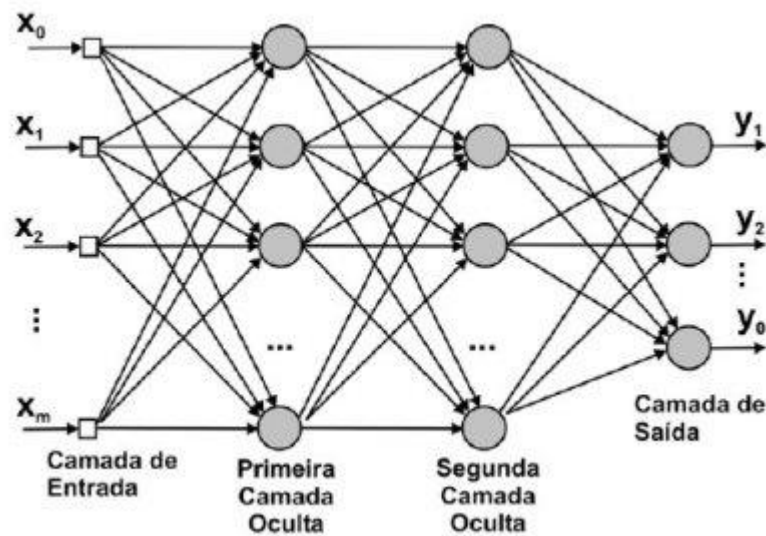
Figura 2: Diagrama de um neurônio de uma rede neural artificial.



Fonte: adaptado de [9].

No diagrama da Figura 3, pode-se dizer que: $E = [E_1 \ E_2 \ E_3 \ E_4 \ ... \ E_n]$ é o vetor de características, $w = [w_1 \ w_2 \ w_3 \ w_4 \ ... \ w_n]$ são os pesos utilizados para o treinamento da rede neural artificial e $S(E)$ é a saída do sistema [10]. Essa figura mostra um sistema simples, utilizando apenas um neurônio. No entanto, sistemas mais complexos, como o da Figura 3, podem ser desenvolvidos, variando não só o número de neurônios, mas também o número de camadas e também utilizando realimentações.

Figura 3: Diagrama de um sistema de rede neural artificial.



Fonte: [11].

No sistema desenvolvido nesse trabalho, o vetor de características é representado pelo vetor de tamanho $i \times j$, onde cada elemento do vetor representa o nível de cinza de um pixel da imagem. A resposta em função da quantidade de imagens utilizadas para o treinamento, da função de ativação (f) e do esquema de neurônios foi abordado na implementação do código.

3. Metodologia

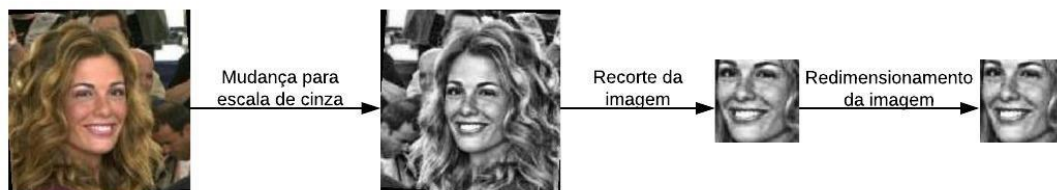
Um sistema de reconhecimento facial é composto por diversas etapas, cada uma com sua devida importância. Essas etapas podem ser divididas, de modo geral, em: detecção da face, adequação da imagem ao algoritmo, descrição e classificação.

O processo de detecção tem como objetivo “encontrar” uma face na imagem, utilizando quaisquer métodos para esse fim. O estudo desse processo não foi o alvo deste trabalho, por isso, um método mais simples foi utilizado para realizar a etapa de detecção. A biblioteca Dlib [12] possui as ferramentas necessárias para a realização da detecção da face, e foi a forma mais adequada encontrada para a realização dessa tarefa.

Uma das etapas de adequação de imagem compreende no alinhamento da face para que o reconhecimento seja devidamente efetuado. O estudo e desenvolvimento dessa etapa também não é o objetivo desse trabalho, por isso, foi necessário escolher um banco de imagens que já possuísse imagens alinhadas. Com isso, após pesquisas, o banco escolhido foi o LFW [13,14], que possui imagens devidamente alinhadas.

No entanto, mesmo com imagens já alinhadas foi necessária uma etapa no projeto que adequasse as imagens para a classificação. Primeiramente foi necessário transformar a imagem, inicialmente em escala RGB, para escala de cinza. Posteriormente, houve a necessidade de selecionar uma área de interesse nas fotos que tivesse um número pixels fixo de 10000. Essa área de interesse foi determinada a partir da biblioteca Dlib, que a partir de um arquivo de extensão *.dat* pré-determinado, realiza a detecção a partir 68 pontos da face. Utilizando esses 68 pontos foi possível recortar a imagem e redimensioná-la de forma que fosse obtido um número de pixels fixo em todas as imagens, para por fim, serem utilizados na etapa de classificação. A Figura 4 mostra toda a etapa de adequação da imagem.

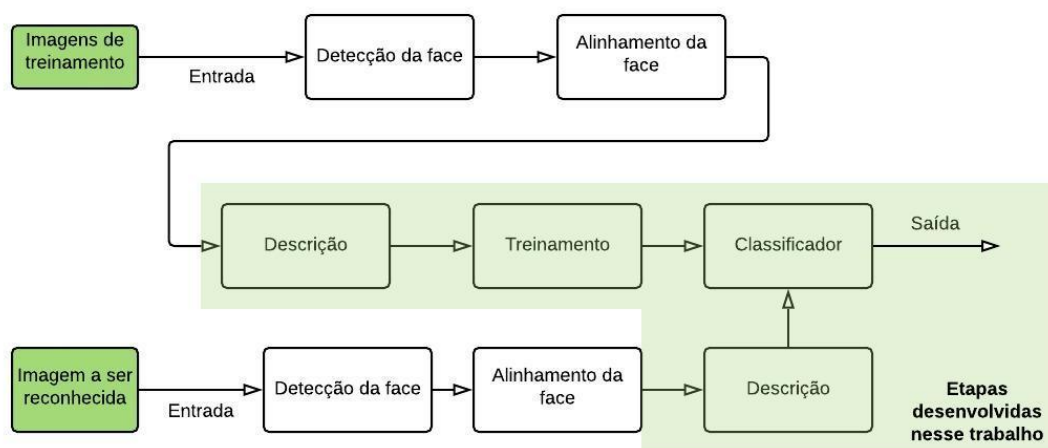
Figura 4: Etapa de adequação da imagem.



Fonte: O autor.

O trabalho desenvolvido teve como principal objetivo o desenvolvimento das duas últimas etapas do processo de reconhecimento: a descrição e a classificação. O descritor utilizado foi o nível de cinza dos pixels das imagens. E, os métodos de classificação foram a SVM e as RNAs, como já dito no segundo tópico deste mesmo trabalho. A Figura 5 mostra as etapas do sistema de reconhecimento facial implementado, bem como evidencia as etapas desenvolvidas.

Figura 5: Etapas de um sistema para reconhecimento facial.



Fonte: O autor.

Todo o código foi implementado na linguagem de programação Python, necessitando das seguintes bibliotecas: *Open CV* e *scipy* para manipulação de imagens, *Sklearn* e *Open CV* para implementação dos métodos de classificação, *Dlib* para a etapa de detecção facial, *Numpy* para manipulações numéricas e *Glob* para manipulação de diretórios e arquivos. Além disso, foi implementado um bloco de automatização para que esse código fosse mais intuitivo para leigos.

4. Resultados

Para efeito de comparação, foram utilizados três conjuntos de dados de diferentes tamanhos, com trinta, sessenta e noventa fotos cada. O processo de treinamento foi realizado utilizando dois terços dos dados da base de dados, e a predição com os dados restantes. A partir disso foi feito um código que automatizasse todo o processo de treinamento e predição.

4.1. Rede neural artificial

Utilizando a RNA, alguns parâmetros foram importantes para a realização do processo: a função de ativação (f), citada no terceiro tópico desse trabalho, cujos testes mostraram que a melhor escolha seria a função sigmoide, e a arquitetura da rede neural, que mostra a quantidade de camadas e de neurônios em cada camada. A Tabela 1 mostra a quantidade de dados da base utilizada e as diferentes arquiteturas da RNA em relação com a taxa de acerto do algoritmo.

Tabela 1: Taxa de acerto em função da arquitetura da rede neural artificial e da quantidade de fotos da base de dados utilizadas.

Arquiteturas	Quantidade de fotos da base de dados		
	30 fotos	60 fotos	90 fotos
	Taxa de acerto (%)		
10000;4000;100;10;1	70,00	75,00	86,67
10000;3000;100;10;1	70,00	65,00	86,67
10000;2000;100;10;1	70,00	70,00	73,33
10000;500;100;10;1	70,00	75,00	86,67
10000;250;100;10;1	70,00	75,00	93,33
10000;100;100;10;1	70,00	85,00	96,67

Fonte: O autor.

É importante deixar claro o que representa os números da coluna da arquitetura. Eles representam os neurônios e a quantidade de números representa a quantidade de camadas da RNA. Como visto, foram utilizadas 5 camadas, em que a primeira representa os pixels das imagens (10000 neurônios) e camada de saída (1 neurônio) que mostra o resultado do reconhecimento. Ou seja, foram utilizadas 3 camadas intermediárias nesse processo, sendo os números de neurônios escolhidos arbitrariamente. É importante ressaltar que a última camada da RNA possui apenas um neurônio porque o objetivo é indicar se a pessoa a ser reconhecida é ou não quem queremos reconhecer, ou seja, é um problema com duas classes.

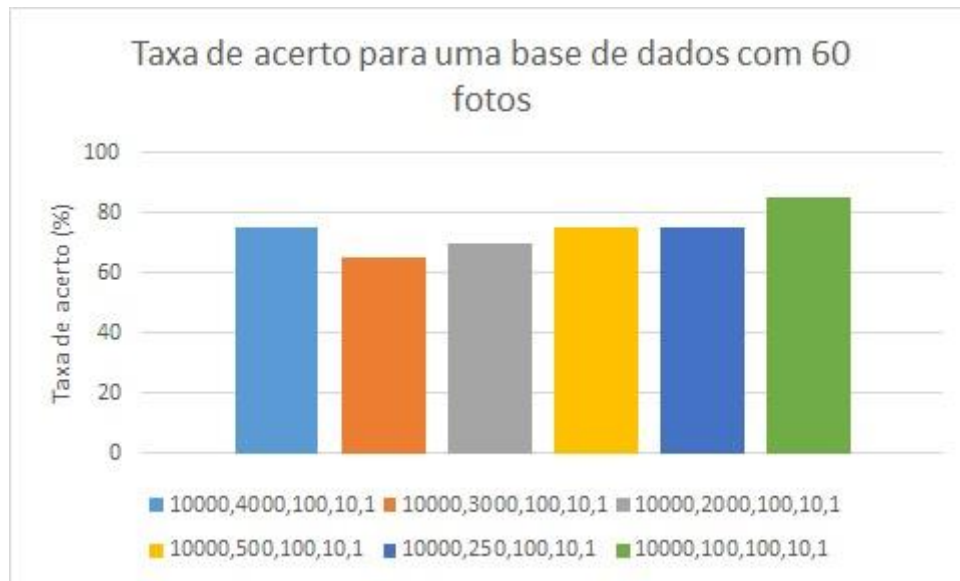
As Figuras 6, 7 e 8 mostram graficamente as taxas de acerto do algoritmo desenvolvido. Com isso, fica claro que o melhor caso de combinação entre arquitetura e quantidade de fotos da base de dados é entre a arquitetura 10000;100;100;10;1 e a base de 90 fotos.

Figura 6: Taxa de acerto em função da arquitetura da rede.



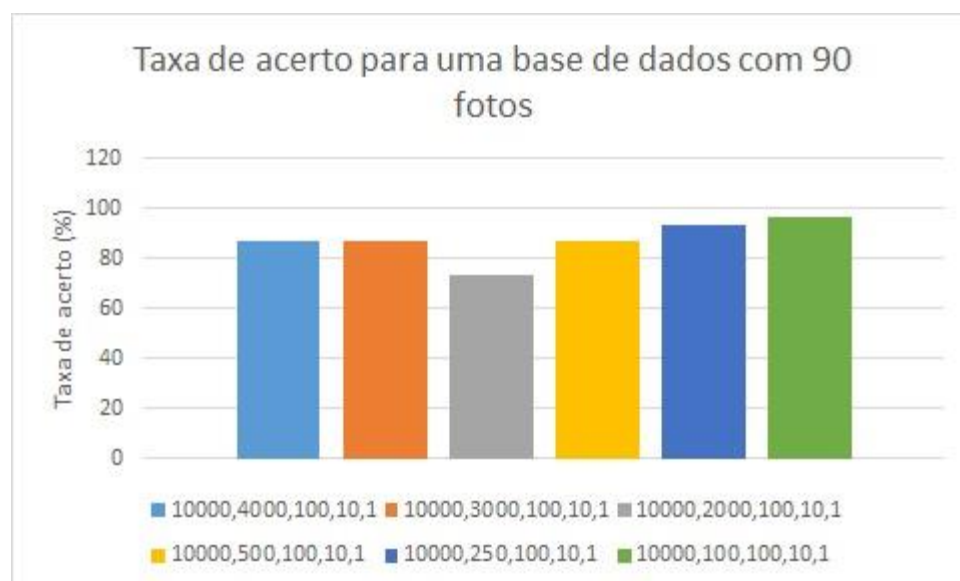
Fonte: O autor.

Figura 7: Taxa de acerto em função da arquitetura da rede.



Fonte: O autor.

Figura 8: Taxa de acerto em função da arquitetura da rede.



Fonte: O autor.

4.2. Máquina de vetores de suporte

A máquina de vetores de suporte também conta com diversos parâmetros configuráveis. Neste trabalho, dois desses parâmetros foram variados para que verificássemos qual sua influência nos resultados.

O primeiro parâmetro a ser configurado é a constante C que determina, basicamente, o quanto é desejado classificar incorretamente um dado de treinamento. Valores grandes de C determinam hiperplanos de margem menor se o processo de treinamento for bem sucedido. Por outro lado, valores baixos de C fazem com que o hiperplano tenha uma margem maior, possibilitando a classificação errônea de alguns dados. A vantagem de se obter essa classificação errônea é que por meio dela é possível tornar o sistema mais flexível, fazendo com que o mesmo não seja tão sensível a qualquer variação. Isso atenua um possível condicionamento do sistema a um modelo fixo de dados [8]. Além disso, é importante ressaltar que esse parâmetro só influencia na função de *kernel* polinomial.

O segundo e último parâmetro é a função *kernel* a ser utilizada, que define simplificada e a forma com que os dados são mapeados [15, 16]. As funções utilizadas neste trabalho foram: RBF, Linear e Polinomial.

A Tabela 2 mostra as diferentes combinações entre os dois parâmetros e as diferentes bases de dados. Com isso, é possível afirmar que os melhores casos situam-se nos testes realizados com a base de 90 fotos.

Tabela 2: Taxa de acerto em função do *kernel* e da constante C e da quantidade de fotos da base de dados utilizadas.

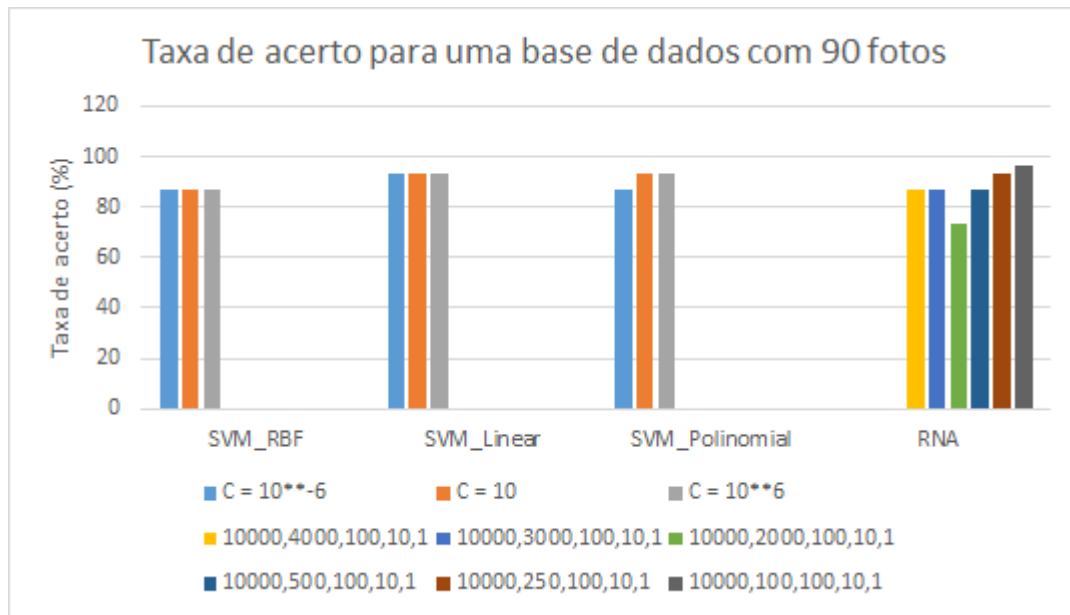
Função <i>kernel</i>	C	Quantidade de fotos da base de dados		
		30 fotos	60 fotos	90 fotos
Taxa de acerto (%)				
RBF	-	70,00	80,00	86,67
	-	80,00	80,00	86,67
	-	80,00	75,00	86,67
Linear	-	80,00	85,00	93,33
	-	80,00	85,00	93,33
	-	80,00	85,00	93,33
Polinomial	10^{-6}	70,00	75,00	86,67
	10	80,00	80,00	93,33
	10^6	80,00	80,00	93,33

Fonte: O autor.

4.3. Comparação entre metodologias de classificação

Para efeito de comparação, foi selecionada a base que obteve maior taxa de acertos para a construção do gráfico da Figura 9, mostrando assim, que o tipo de algoritmo que obteve maior taxa de acerto para a base de dados selecionada foi o que utilizava a RNA com arquitetura 10000;100;100;10;1.

Figura 9: Taxa de acerto em função do tipo de classificador



Fonte: O autor.

5. Conclusão

Nesse trabalho foi desenvolvido um sistema de reconhecimento facial com o objetivo de analisar dois métodos de classificação de dados, bem como o impacto de alguns parâmetros desses métodos, e a influência da quantidade de dados fornecido para o sistema no resultado.

A partir disso, pode-se concluir que para base de dados escolhida o algoritmo que obteve melhor desempenho utilizou como classificador uma RNA com arquitetura 10000;100;100;10;1, alcançando taxa de acerto de 96,67%. No entanto, os testes realizados com o classificador SVM também mostraram resultados expressivos, sendo importantes para o desenvolvimento do trabalho.

É importante frisar que as imagens da base de dados foram escolhidas arbitrariamente, assim como os conjuntos de fotos utilizadas para treinamento e validação do algoritmo.

Os principais requisitos do trabalho foram desenvolvidos, e os resultados foram satisfatórios. No entanto, diversas áreas de estudo ainda são de grande relevância para o desenvolvimento de algoritmos de reconhecimento facial.

Um estudo que pode ser desenvolvido é o de comparação de métodos de classificação, possuindo aplicações não só na área de reconhecimento facial, mas também na área de reconhecimento de padrões. O desenvolvimento de pesquisas que comparem diferentes classificadores, variando seus diversos parâmetros e verificando sua relação com o sistema proposto, são de extrema importância.

O estudo de diferentes descritores para o reconhecimento facial também pode ser explorado. O uso de descritores geométricos, ou até mesmo a combinação de descritores,

foi visto ao longo desse trabalho, como algo promissor. Além disso, algumas técnicas relevantes para a análise de imagens digitais como a análise de componentes principais também é uma área de pesquisa que pode oferecer melhorias para os sistemas de reconhecimento facial.

6. Referências

- [1] ACESSO DIGITAL. Acesso Digital, 2019. **Homepage da empresa Acesso Digital**. Disponível em: <<https://home.acessodigital.com.br>>. Acessado em julho de 2019
- [2] CAMARAPATOS. **Câmara Municipal de Patos de Minas: Projeto “Olho Vivo” é inaugurado em Patos de Minas**. Disponível em: <http://www.camarapatos.mg.gov.br/index.php?option=com_content&view=article&id=530:projeto-olho-vivo-e-inaugurado-em-patos-de-minas&catid=59:noticias&Itemid=101>. Noticiado em 09 de junho de 2014. Acessado em julho de 2019
- [3] GOOGLE. Google Fotos: todas as suas fotos organizadas e fáceis de encontrar. **Homepage do aplicativo Google Fotos**. Disponível em: <<https://www.google.com/intl/pt-BR/photos/about/>>. Acessado em julho de 2019
- [4] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037 – 2041, Outubro 2006
- [5] Guodong Guo, Stan Z. Li, e Kapluk Chan. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. Agosto, 2002
- [6] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman. Deep Face Recognition. *BmvcVol. 1. No. 3*. 2015.
- [7] Yaniv Taigman, Ming Yang, Marc’Aurelio Ranzato e Lior Wolf. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Junho, 2014
- [8] Sergios Theodoridis e Konstantinos Koutroumbas: Pattern Recognition. 4. Ed. Orlando: Academic Press, Inc., 2008
- [9] Hélio Pedrini e William Robson Scharzt: Análise de Imagens Digitais: princípios, algoritmos e aplicações. 1. Ed. São Paulo: Editora Thomson Learning, 2008
- [10] Almir O. Artero: Inteligência Artificial: Teórica e Prática. 1. Ed. São Paulo: Editora Livraria da Física, 2009
- [11] Matheus Facure. **Introdução às Redes Neurais Artificiais: Uma apresentação teórica e intuitiva às redes neurais artificiais**. Disponível em: <<https://matheusfacure.github.io>>. Acessado em agosto de 2019

- [12] Davis E. King. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10, pp. 1755-1758, 2009
- [13] Gary B. Huang, Vidit Jain, and Erik Learned-Miller. Unsupervised joint alignment of complex images. *International Conference on Computer Vision (ICCV)*, 2007.
- [14] Gary B. Huang, Marwan Mattar, Honglak Lee, and Erik Learned-Miller. Learning to Align from Scratch. *Advances in Neural Information Processing Systems (NIPS)*, 2012.
- [15] Saptashwa Bhattacharyya. Support Vector Machine: Complete Theory of Support Vectors. **Medium**, 2018. Disponível em:<
<https://towardsdatascience.com/understanding-support-vector-machine-part-1-lagrange-multipliers-5c24a52ffc5e> >. Acessado em agosto de 2019
- [16] Saptashwa Bhattacharyya. Support Vector Machine: Kernel Trick; Mercer's Theorem. **Medium**, 2018. Disponível em:<
<https://towardsdatascience.com/understanding-support-vector-machine-part-2-kernel-trick-mercers-theorem-e1e6848c6c4d>>. Acessado em agosto de 2019