

# Q3.2018 - Metodologia de pesquisa - Revisão sistemática

João Carlos Pandolfi Santana

November 2018

## 1 Informações gerais

### 1.1 Título

Uma revisão sistemática de *transference learning* e *visão computacional* voltada para classificação e segmentação de espécies de mosquitos.

### 1.2 Descrição

Revisão sistemática com o foco em *transference learning* e *visão computacional* baseada em *redes neurais*, com o intuito de classificar espécies de mosquitos. Também pretendo abranger materiais relacionados a estrutura corporal, cladística e filogenética dos insetos em questão.

### 1.3 Objetivos

Revisar a literatura a fim de encontrar material de estudo e apoio na área de *transference learning*, além de embasar e direcionar minha pesquisa levando em consideração trabalhos relacionados. Também tenho como objetivo identificar possíveis caminhos a serem seguidos durante a condução do meu projeto de mestrado, procurando não esbarrar ou gastar energia desnecessária em frentes de pesquisa já consolidadamente exploradas, mas sim usá-las como ferramental para a minha contribuição.

## 2 Proposta

### 2.1 Problema de pesquisa

O problema abordado é a identificação por imagem de espécies de mosquito, de forma a agilizar e aumentar a eficiência de agentes de saúde e pesquisadores de campo em relação a epidemiologia e controle dos vetores de doença.

## 2.2 Objetivos

O objetivo principal é fazer uma revisão da literatura, onde possa reunir os trabalhos relacionados e partir de um problema em aberto ou otimizar alguma etapa ineficiente. Outro objetivo é mapear as técnicas utilizadas e propor uma nova e mais eficiente para classificação de mosquitos. Tenho como objetivo final, desenvolver uma ferramenta mobile, no qual facilite a identificação do espécime capturado. Por fim, também pretendo revisar conceitos de aprendizado de máquina, que possam ser aplicados na pesquisa, como por exemplo, *transference learning* (transferência de aprendizado) onde possa utilizar sistemas especialistas em outras áreas, mas que possuam um certo grau de proximidade computacional do problema, de modo que sejam aplicados para otimização e melhoria no aprendizado do modelo.

## 2.3 Hipótese

A hipótese inicial é que não haja muita pesquisa relacionada ao tema proposto, devido ao fato dos estudos serem voltados, na sua maior parte, a doenças e migração dos vetores não na identificação computacional dos espécimes, também levo em conta que pelos insetos serem muito pequenos e dependerem de captura para análise, os pesquisadores que fazem a coleta já são capacitados para identificar o espécime, não abrindo tanta abertura para o problema de classificação destes. No entanto, visto desenvolver uma ferramenta para facilitar e reduzir o viés desta classificação, facilitando assim a análise em regiões que não possuam muitos especialistas e também permitir um pessoal menos treinado a fazer a primeira etapa da pesquisa, a identificação do espécime.

## 2.4 Métodos

Como a hipótese se baseia em não ter muita pesquisa relacionada ao tema proposto, procurarei conteúdo relacionado a classificação de imagem utilizando ferramental computacional que pretendo usar na pesquisa, assim como procurar material relacionado a estrutura corporal e classificação morfológica dos insetos a serem estudados.

## 2.5 Resultados esperados

Espero encontrar uma grande quantidade de conteúdo sobre classificação de imagens e conteúdo relacionado a morfologia de insetos. Assim como modelos e métodos previamente validados de *transference learning* e visão computacional para basear minha pesquisa. Pretendo encontrar também, caso haja, pesquisas relacionadas ao tema proposto, de forma que possa seguir por caminhos consolidados e explorar conceitos ainda não estudados, assim, possa contribuir não só com o produto final produzido, mas também com técnicas computacionais novas ou aplicadas em áreas inexploradas.

## 3 Identificação de estudos

### 3.1 Palavras chave

*Transference learning, machine learning, computational vision, neural network, convolutional network, image classification cross validation ,classification, image segmentation, mosquito, mosquitos, insect, visão computacional, aprendizado de máquina, redes neurais, classificação de imagens, segmentação de imagem, insetos*

### 3.2 Strings de busca

Strings de busca utilizadas durante a revisão sistemática

- Parsifall
  1. "transference learning" AND ("neural network" OR "neural networks") AND (mosquito OR mosquitos OR insect OR insects)
  2. ("transference learning" OR "machine learning" ) AND ("mosquito" OR "mosquito")
  3. ("classification" ) AND ("mosquito" OR "mosquito")
  4. ("neural network" ) AND ("classification") AND ("mosquito" OR "mosquitos")
  5. "transference" AND "learning" AND "machine"
- Google Scholar
  1. transference learning machine
  2. transductive learning
  3. transductive learning on insects classification
  4. image mosquito classification

### 3.3 Critério de seleção das fontes de busca

O primeiro critério para seleção da fonte, foi a base ser internacional, assim eu tenho acesso a uma quantidade de artigos maior e consequentemente das maiores universidades e centros de pesquisa.

O segundo critério foram bases de dados relacionadas a publicações na área de ciência da computação, ou seja, não foram incluídas bases de dados que não tem a ver com o tema de pesquisa, esse critério foi adicionado para evitar ruído na pesquisa.

O terceiro e último critério, foi o uso do *google scholar* para refinamento final da pesquisa, onde o objetivo foi pesquisar com um espectro mais amplo, vai de encontro com o segundo critério onde o objetivo é exatamente não ficar preso no espectro definido. O objetivo é somente verificar a relevância e qualidade da

*string* de busca, para garantir a assertividade, ou seja, se ao buscar numa base mais abrangente o resultado for próximo do encontrado na base mais específica, a *string* está boa.

### 3.4 Estratégia de busca

A primeira etapa, foi a definição das strings de busca. O critério de qualidade de cada *string*, foi definida pela quantidade de resultados em relação a correspondência do material apresentado com o tema da pesquisa. A sequência de *strings* utilizadas se encontram na seção 3.2.

Após identificar as *strings* mais promissoras, utilizei a ferramenta de busca avançada *parsif.al*, onde adicionei os resultados em uma planilha no formato *.csv*. Filtrei os resultados obtidos pelo título com os critérios de *inclusão* 4.1 e *exclusão* 4.2.

Por final, utilizei a ferramenta *Google Scholar* para pesquisar artigos relacionados e verificar a qualidade das *strings* de busca, ao ponto que dependendo da correlação dos resultados obtidos com o tema de pesquisa, pude verificar se as buscas anteriores foram assertivas.

## 4 Seleção e avaliação de estudos

### 4.1 Critérios de inclusão

1. Conter as palavras chave
2. Ser relacionado com:
  - Machine Learning
  - Transference Learning
  - Transductive Learning
  - Image Processing
  - Image Classification
  - Mosquito Classification
  - Insect Classification

### 4.2 Critérios de exclusão

1. Fugir dos temas
2. Não abranger sistemas de informação
3. Não utilizar machine learning
4. Não ter acesso ao pdf

### 4.3 Estratégia para seleção

Após o filtro pelo título dos trabalhos, fui para a leitura do *abstract*, onde utilizei novamente os critérios de *inclusão* e *exclusão* para selecionar os artigos que fizessem sentido. Por fim, após as três etapas de filtragem citadas, parti para leitura dos arquivos selecionados. Pesquisei no google scholar para não ficar na bolha das bases selecionadas e usar o algoritmo de sugestão do google.

### 4.4 Avaliação da qualidade

- Quantidade de citações
- Correlação com o conteúdo pesquisado
- Profundidade da pesquisa
- Ferramentas utilizadas parecidas com as que pretendo usar
- Bons resultados
- Relevância da publicação
- Tipo de publicação
- Citação por artigo relevante, ou seja, um artigo relevante cita

## 5 Discussão

Ao contrário do que previ na hipótese inicial encontrei bastante material relacionado a pesquisa, em contrapartida a maioria dos resultados das buscas foram sobre as doenças vetorizadas por mosquitos e estudos sobre migração e comportamento dos insetos. No quesito de aprendizado computacional, obtive bastante conteúdo de qualidade, o que irá ajudar bastante no desenvolvimento do projeto.

Tive bastante dificuldade ao selecionar conteúdos relacionados a computação, em vista que, muito material tem sido produzido na última década devido ao aumento do poder de processamento dos computadores e popularização das técnicas de aprendizado computacional. Para chegar aos artigos mais relevantes, dei ênfase aos mais citados e me vali de consultoria a professores da área para indicação e direcionamento da pesquisa.

Quanto aos trabalhos que têm como tema os insetos e sua morfologia, apesar de encontrar bastante artigo sobre o assunto, a maioria destes apontavam para livros, exceto os que tratavam de algum assunto específico estudado, como o [1] que trata da geometria morfométrica de mosquitos (no qual, acredito eu, ser uma chave para identificação de *features* da pesquisa da qual conduzirei). Os livros foram um pouco mais difíceis de se obter, em vista que são comprados, porém consegui acesso aos conteúdos e consegui adicionar a minha base de pesquisa.

## 6 Conclusão

A pesquisa foi mais árdua do que imaginei que seria, contudo, obtive resultados que se mostram promissores, devido ao fato de alguns deles serem diretamente relacionados com a minha proposta de pesquisa e apresentarem resultados positivos. Assim como consegui materiais de suporte para desenvolvimento direcionado das etapas que precisarei transpor para concluir a pesquisa.

Em relação a classificação de mosquitos, encontrei estudos que desenvolveram o que propus estudar, porém não englobam todo o escopo que havia definido, os que tornam uma ótima base de pesquisa e pontapé inicial, assim como prova de conceito que de viabilidade.

### 6.1 Agradecimentos

## References

- [1] Camila Lorenz, Fabio Almeida, Fernanda Almeida-Lopes, Caroline Louise, Stella N. Pereira, Vivian Petersen, Paloma O. Vidal, Flávia Virginio, and Lincoln Suesdek. Geometric morphometrics in mosquitoes: What has been measured? *Infection, Genetics and Evolution*, 54:205–215, 2017.