

ARTIGO ORIGINAL

Melampus: um modelo *deep learning* para triagem psicológica infantil

Wesley Felipe da Silva*¹ and Mateus Raeder†¹¹Universidade do Vale do Rio dos Sinos (UNISINOS)

*wesleys.felipe@gmail.com; †mraeder@unisinos.br

Submetido: 06/08/2018. Revisado: 07/11/2018. Aceito: 14/11/2018.

Abstract

While mental health related issues usually began in childhood or adolescence, only a small portion of this population receives proper diagnosis and treatment. In part, this scenario is caused due the lack of specialized tools for mental disorders screening, mainly those which reduce cost and time needed. Recently, authors have been analyzing how machine learning could help to build new psychological assessment tools. However, only few researches proposed building specialized tools for groups composed mostly of children. This work aims to propose a model which combines clinical tests and deep learning for supporting child psychological screening. Results suggested that deep learning tools may fit the intended scenario, once the classification models tested performed well even with a small sample size.

Key words: Convolutional Neural Networks; Psychological Screening; Human Figure Drawing – HFD

Resumo

Embora problemas relacionados à saúde mental usualmente tenham início durante a infância ou adolescência, apenas uma pequena parcela desta população recebe diagnóstico e tratamento adequado. Uma das causas para a baixa taxa de identificação de distúrbios mentais é a falta de instrumentos especializados nesta tarefa, especialmente ferramentas que reduzam o custo e o tempo necessário para a execução de processos de triagem psicológica. Na literatura recente, muitos autores vêm analisando como o aprendizado de máquina pode contribuir para a construção de instrumentos de avaliação psicológica, contudo poucas pesquisas se propõem a construir ferramentas válidas para grupos compostos majoritariamente por crianças. O presente trabalho propõe um modelo de ferramenta para apoio à triagem psicológica infantil baseada em testes clínicos e *deep learning*. Tal modelo foi avaliado através de uma implementação que combina o uso de Redes Neurais Convolucionais e um sistema de escalas clínicas para avaliação do Desenho da Figura Humana. Os resultados apresentados pelos modelos de classificação treinados demonstraram bons índices de acerto considerando-se a pequena amostra disponível, o que sugere que ferramentas de *deep learning* podem ser adequadas para o cenário proposto.

Palavras-Chave: Convolutional Neural Networks; Triagem Psicológica; Deep Learning; Desenho da Figura Humana – DFH

1 Introdução

Evidências mostram que a maior parte dos problemas relacionados à saúde mental inicia-se na infância ou adolescência ([Mental Health Foundation; 2016](#); [World](#)

[Health Organization; 2017](#)). Afirma-se também que sintomas destas distúrbios costumam afetar o desempenho dos indivíduos durante a vida adulta ([Falissard; 2016](#)). Apesar da estimativa da OMS de que entre 10% e 20% das crianças e adolescentes

possuem algum tipo de desordem mental (World Health Organization; 2017), apenas uma pequena parcela desta população é diagnosticada e recebe tratamento adequado em idade precoce (Anxiety and Depression Association of America; 2017; Murphy and Fonagy; 2013; Rice et al.; 2017; World Health Organization; 2013).

No contexto brasileiro, apesar dos avanços na área de testagem psicológica, percebe-se que há carência de instrumentos que atendam às necessidades dos profissionais de saúde responsáveis pelos processos avaliativos (Popi et al.; 2016). Segundo os mesmos autores, o aumento da demanda por atendimentos multiprofissionais e a necessidade da redução do tempo e custo dos atendimentos suportam a utilização de instrumentos de triagem psicológica, sendo este um meio de selecionar possíveis candidatos a avaliações completas, evitando assim o desgaste desnecessário dos envolvidos. Esforços empreendidos por de Oliveira (2013) e Viapiana et al. (2016) vem de encontro a necessidade de fornecer um maior portfólio de opções para o exercício da triagem clínica. Estes autores buscam suportar, adaptar e/ou coletar evidências de validade acerca do teste do Desenho da Figura Humana (DFH) voltado especificamente para a população infantil.

Na literatura recente, percebe-se crescente interesse no emprego de aprendizado de máquina, ou *machine learning*, para a construção de instrumentos de avaliação clínica (Seixas et al.; 2014; Jiménez-Serrano et al.; 2015; Carpenter et al.; 2016; Kim et al.; 2017; Zhu et al.; 2017; Hilbert et al.; 2017). Luxton (2014) acredita que aplicações que empregam aprendizado de máquina para a realização de psicodiagnóstico podem ser mais eficientes e sofisticadas em relação aos métodos comuns. Segundo o autor, isto se deve a capacidade de processamento de dados complexos, customização e redução da incerteza provida por este tipo de técnica.

Dentre as disciplinas de aprendizado de máquina mais promissoras, está o *deep learning*, que, segundo LeCun et al. (2015), está fazendo grandes avanços na resolução de problemas que resistiram por anos aos melhores esforços empreendidos pela comunidade científica. Apesar disto, até o presente momento poucos trabalhos têm utilizado esta técnica aplicada a psicologia clínica.

O presente trabalho apresenta um modelo que, no contexto da triagem psicológica infantil, se propõe a realizar a avaliação automática de testes de expressão gráfica através do uso de *deep learning*. Para isto desenvolveu-se uma pesquisa experimental, na qual construiu-se um protótipo para a avaliação e validação dos componentes críticos do modelo.

Este estudo está alinhado ao quarto objetivo do plano de ação para a saúde mental criado pela OMS para os anos 2013-2020, objetivo este que incentiva a busca pelo fortalecimento dos sistemas de informação, evidências e pesquisa para a saúde mental (World Health Organization; 2013).

O trabalho está estruturado da seguinte forma: A seção 2 trata da fundamentação teórica e apresentação de conceitos relevantes para o estudo. Na seção 3 são apresentados alguns trabalhos relacionados. O modelo proposto é detalhado na seção 4. A seção 5 destina-se a metodologia adotada. A sexta seção é reservada para a apresentação dos resultados.

A última seção destina-se às considerações finais, contribuições e trabalhos futuros.

2 Fundamentação Teórica

Esta seção apresenta os conceitos-chave para a concepção do modelo Melampus. A primeira subção apresenta o Desenho da Figura Humana como instrumento de psicodiagnóstico através de expressão gráfica. Já a segunda subseção apresenta uma visão geral sobre Redes Neurais Convolucionais como uma ferramenta de *deep learning*.

2.1 Desenho da Figura Humana

Segundo Gauy and Guimarães (2006), um dos motivos pelos quais muitas crianças não recebem tratamento é a avaliação inadequada ou a não avaliação dos comportamentos problemáticos. Segundo o mesmo autor, avaliar transtornos na infância é uma tarefa complexa, devido principalmente às dificuldades de expressão verbal e reconhecimento das próprias emoções por parte da criança.

Avaliações psicológicas são geralmente conduzidas no formato de entrevistas ou através de inventários (Muller and Erford; 2012). Profissionais e pesquisadores da área da saúde mental que trabalham com crianças acreditam que a arte como técnica de expressão gráfica também pode ser uma ferramenta útil para acessar aspectos psicológicos e emocionais dos pacientes, especialmente quando a comunicação falha (Hammond and Gantt; 1998; Gabriels et al.; 2000; Rollins; 2005).

Dentre os testes de expressão gráfica mais conhecidos e utilizados na psicologia clínica brasileira está o Desenho da Figura Humana (DFH) (Saur et al.; 2010). Abrangente, simples e barato (Hutz and Bandeira; 1995); o DFH é aplicado geralmente em crianças (Borsa and Bauermann; 2013).

Entendendo que a avaliação psicológica deve contar com diferentes técnicas e instrumentos além dos já padronizados, Borsa and Bauermann (2013) consideram que o DFH pode contribuir significativamente para o psicodiagnóstico, permitindo à criança expressar de forma livre e não controlada as emoções que permeiam seus comportamentos. Na Figura 1 é apresentado um exemplo de DFH produzido por uma criança.

Segundo Hutz and Bandeira (1995) e Viapiana et al. (2016), estudos com o DFH para diagnóstico de problemas emocionais empregam diferentes formas de avaliação dos desenhos, distinguindo-se em três estratégias de análise: a) aspectos globais; b) aspectos estruturais, como tamanho da figura e localização na página; e c) itens específicos, tais como presença ou ausência de elementos como cabeça, braços, etc. Atualmente, apenas dois sistemas de avaliação do DFH estão aprovados pelo SATEPSI¹, ambos voltados a avaliação da maturação cognitiva das crianças. Porém, diversos sistemas têm sido propostos por pesquisadores da área (de Oliveira; 2013; Arteche; 2006; Viapiana et al.; 2016).

¹Sistema de avaliação de testes psicológicos desenvolvido pelo Conselho Federal de Psicologia (CFP).

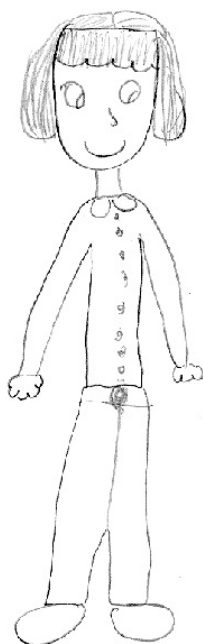


Figura 1: DFH produzido por uma criança

Em seu trabalho, [de Oliveira \(2013\)](#) propõe construir escalas clínicas do DFH para crianças de 6 a 12 anos de idade, sendo o objetivo destas identificar a necessidade (ou não) de uma avaliação psicológica mais completa. O autor parte da premissa de que, de acordo com estudos brasileiros, o DFH parece servir mais como um indicador da existência problemas psicológicos do que um método de diagnóstico para um tipo de problema específico. Os resultados desta pesquisa indicam que as escalas apresentam adequada capacidade de discriminar grupos e validade para indicação de uma avaliação psicológica. O método de avaliação sugerido pelo autor consiste basicamente na análise de itens específicos do DFH, tais como presença de linhas tremidas ou pesadas no desenho. A presença ou ausência de cada um destes itens é atribuída um peso, que quando somado com os pesos referentes aos demais itens determina o escore final. Este escore pode ser interpretado como um dos seguintes resultados: (i) Indicadores insuficientes para uma avaliação psicológica; (ii) Indicação de uma entrevista/observação; (iii) Indicação de uma avaliação psicológica; (iv) Necessidade de uma avaliação psicológica; e (v) Fortes indicadores para uma avaliação psicológica. Os itens analisados variam conforme o grupo amostral ao qual o indivíduo testado pertence, sendo que ([de Oliveira; 2013](#)) determina escalas para quatro grupos que variam em gênero e idade (crianças mais jovens [6 a 8 anos] e crianças mais velhas [9 a 12 anos]).

2.2 Deep Learning e Redes Neurais Convolucionais

Por muito tempo, desenvolver sistemas de aprendizado de máquina significou construir também um extrator de aspectos (*features*) capaz de transformar dados brutos; tal como pixels de uma imagem, em representações internas adequadas

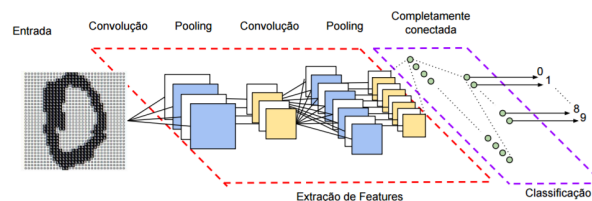


Figura 2: Exemplo de uma rede neural convolucional e suas diferentes camadas

para servirem de entrada para um método de aprendizado ([LeCun et al.; 2015](#)). Este processo requer considerável conhecimento do domínio no qual o agente deverá atuar ([LeCun et al.; 2015](#)).

O aprendizado de representação, ou *representation learning*, é um conjunto de métodos que permitem a detecção automática de *features* necessárias para a execução de procedimentos de classificação ou detecção ([LeCun et al.; 2015](#)). Para [Goodfellow et al. \(2016\)](#), estas representações aprendidas geralmente apresentam performance superior à obtida através de métodos convencionais de seleção de *features*.

O aprendizado de representação, ou *representation learning*, é um conjunto de métodos que permitem a detecção automática de *features* necessárias para a execução de procedimentos de classificação ou detecção ([LeCun et al.; 2015](#)). Para [Goodfellow et al. \(2016\)](#), estas representações aprendidas geralmente apresentam performance superior à obtida através de métodos convencionais de seleção de *features*.

Um dos modelos que têm apresentado grande sucesso na detecção, segmentação e reconhecimento de objeto e regiões em imagens são as *Convolutional Neural Networks* (CNNs ou ConvNets) ([LeCun et al.; 2015](#)). Segundo [LeCun et al. \(2015\)](#), as ConvNets são projetadas para processar dados na forma de múltiplos vetores e seguem três princípios arquiteturais para garantir invariância espacial: campos receptivos locais (*local receptive fields*), compartilhamento de pesos (*shared weights*) e subamostragem espacial (*spatial subsampling*) ([Noda et al.; 2014](#)). Na Figura 2 é apresentada a arquitetura geral de uma ConvNet para a classificação de imagens.

Em uma ConvNet, o objetivo da convolução é extrair *features* da camada anterior ([LeCun et al.; 2015](#)). No caso de imagens, esta operação preserva a relação espacial entre os pixels através do aprendizado de *features* utilizando pequenos quadros dos dados de entrada ([LeCun et al.; 1998](#)). A operação de convolução é representada na Figura 3. Diferentes *kernels* produzem diferentes saídas sobre o mesmo dado de entrada ([LeCun et al.; 2010](#)). Durante a fase de treinamento, as ConvNets são capazes de aprender os valores dos *kernels*, embora seja necessário especificar uma série de parâmetros, tais quais o número e tamanho dos *kernels* ([Vargas et al.; 2016](#)).

Segundo [Vargas et al. \(2016\)](#), é comum após a camada de convolução aplicar uma função de ativação, sendo que normalmente esta função apresenta algum grau de não-linearidade. Segundo os autores, esta não-linearidade permite que as categorias de saída sejam linearmente separáveis.

As camadas de *pooling* juntam semanticamente

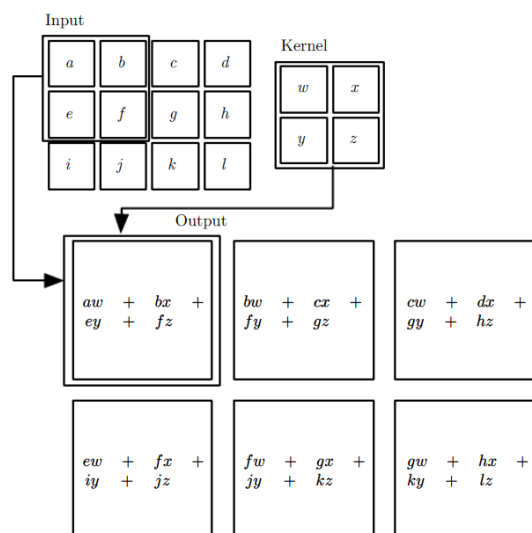


Figura 3: Processo de convolução

features semelhantes em uma única feature (LeCun et al.; 2015). Ainda segundo os mesmos autores, uma unidade de pooling tipicamente computa os máximos locais das entradas recebidas. Para LeCun et al. (1998), esta etapa é a responsável pela redução da precisão espacial, permitindo que o resultado da classificação seja menos sensível a diferentes posicionamentos e distorções em imagens.

Através de uma camada totalmente conectada, as features identificadas alimentam um classificador capaz de relacionar estas informações com as classes de resposta (Vargas et al.; 2016). Em seu trabalho, Cireşan et al. (2012) utilizam a função de ativação *softmax* nesta última camada, pois a saída pode ser interpretada como a probabilidade da imagem de entrada pertencer a uma das classes desejadas.

Diferentes arquiteturas repetem e combinam camadas de *pooling* e convolução. Por exemplo, enquanto a LeNet-5 possui duas camadas de convolução seguidas de *pooling*, seguidas novamente de uma camada de *pooling* (LeCun et al. (1998)), a GoogleNet possui cinco camadas de convolução seguidas de *pooling* (Szegedy et al.; 2015).

3 Trabalhos Relacionados

Diversos trabalhos tem explorado o uso de ferramentas computacionais no âmbito da avaliação psicológica. Zhu et al. (2017) propõe o uso de ConvNets na construção de um sistema para diagnóstico de depressão baseado na análise de expressões faciais em vídeos. Os autores empregam duas ConvNets; uma para classificação da aparência e outra para as dinâmicas faciais. O grau de depressão é determinado por uma camada de integração que combina o resultado dos dois modelos. Em comparação a outros sistemas de diagnóstico que apoiam-se apenas na análise visual dos vídeos e que foram treinados nas mesmas bases de dados, os autores concluem que a abordagem proposta apresentou desempenho superior. Sugere-se que analisar também o áudio presente nos vídeos pode incrementar o desempenho do obtido.

Partindo da premissa de que o uso do espaço em desenhos feitos à mão pode revelar informações úteis sobre o indivíduo que o desenhou, in Kim, Han and Oh (2012) propõe um sistema que, baseado em modelos de regressão, seja capaz de analisar esta característica em artefatos de arte bidimensional. A vantagem fornecida pelo sistema é a redução da subjetividade e do erro humano na análise quantitativa desta variável. A avaliação do sistema construído se deu através de sua aplicação na avaliação do nível de demência em uma população formada por indivíduos idosos. Os resultados demonstraram que o uso de espaço nos desenhos está significativamente relacionado ao nível de demência dos indivíduos da amostra.

Aplicando técnicas de *machine learning* sobre dados previamente coletados, Carpenter et al. (2016) desenvolveram um sistema capaz de identificar crianças sob risco de desenvolver transtornos de ansiedade. A análise dos resultados demonstra que o sistema classificou corretamente os indivíduos em 96% dos casos. O instrumento de avaliação psicológica utilizado foi o PAPA, que consiste basicamente de uma ferramenta para a análise de informações obtidas através de entrevistas com os pais de crianças com idades entre 2 e 5 anos. Apesar de os autores terem utilizado ADTrees como método de classificação dos dados, estes sugerem que outros algoritmos podem ser utilizados e que estes podem inclusive produzir resultados superiores.

Em seu trabalho, Seixas et al. (2014) propõe um modelo baseado em Redes Baysianas para o diagnóstico de Alzheimer, demência e comprometimento cognitivo leve. O conjunto de dados utilizado para o treinamento do modelo é composto basicamente de dados demográficos e resultados de testes neuropsicológicos dos indivíduos que compõe a amostra. Os autores se apoiam na premissa de que Redes Baysianas são adequadas para a representação de incerteza e causalidade, ambos elementos presentes no diagnóstico clínico. Por fim, os resultados demonstram que o modelo atingiu desempenho superior quando comparado a outros classificadores conhecidos.

O trabalho de Jiménez-Serrano et al. (2015) possui dois objetivos: (i) desenvolver um modelo para a detecção de depressão pós-parto durante a primeira semana após o nascimento da criança e (ii) baseado no modelo criado, desenvolver uma aplicação mobile para mães que recém deram a luz e clínicos que desejam monitorá-las. A ferramenta utilizada para o diagnóstico da desordem foi o *Edinburgh Postnatal Depression Scale* (EPDS), um questionário composto por perguntas de fácil entendimento. Com base nos resultados do EPDS e dados demográficos e socioeconômicos de cada indivíduo, foram experimentados quatro classificadores: *Naive Bayes*, *Logistic Regression*, SVM, e RNA. Para os autores, o desempenho do classificador é determinado pela relação entre sensibilidade e especificidade, e com base nisto, *Naive Bayes* foi o método que apresentou melhor desempenho.

Hilbert et al. (2017) utilizam técnicas de *machine learning* para classificar indivíduos em três diferentes grupos: portadores de depressão, portadores de ansiedade generalizada e indivíduos não portadores de desordens mentais. Os dados coletados da amostra foram respostas de questionários clínicos, níveis de

Tabela 1: Pesquisas anteriores em avaliação psicológica com apoio de instrumentos de computação

	Técnica Computacional	Método de diagnóstico	Desordem testada	Amostra		
				Faixa etária	País	Tamanho
in Kim, Han and Oh (2012)	Processamento gráfico / Modelagem estatística	Expressão Gráfica: PPAT/FEATS	Demência	Idosos	Coreia do Sul	145
Seixas et al. (2014)	Aprendizado de máquina	Dados demográficos / Testes neuropsicológicos	Demência / Alzheimer/ Comprometimento cognitivo leve	Idosos	EUA/Brasil	3.145
Jiménez-Serrano et al. (2015)	Aprendizado de máquina	EPDS	Depressão pós-parto	Adultos	Espanha	1.397
Carpenter et al. (2016)	Aprendizado de máquina	PAPA	Ansiedade	Crianças	EUA	1.224
Zhu et al. (2017)	Deep Learning	Expressões e dinâmicas faciais / Vídeo	Depressão	Adultos	Alemanha	292
Hilbert et al. (2017)	Aprendizado de máquina	Questionários de cortisol / Níveis de Ressonância Magnética	Depressão / Ansiedade Generalizada	Adultos	Alemanha	57
Presente Trabalho	Deep Learning	DFH	Triagem psicológica	Crianças	Brasil	804

Fonte: Elaborado pelos autores.

cortisol e volumes de massa cinzenta e branca. Estas informações foram analisadas tanto de forma isolada quanto combinada. O algoritmo de classificação utilizado foi o SVM, com o qual foi possível obter 90,10% de precisão ao determinar de se um indivíduo é saudável ou não. Determinar se o indivíduo possui depressão ou ansiedade generalizada se mostrou uma tarefa mais complexa, sendo que a precisão máxima alcançada foi de 67,46%.

Na Tabela 1 é apresentada uma comparação entre as principais características dos trabalhos anteriormente citados e a abordagem adotada no presente trabalho. Ante o exposto, percebe-se que a maior parte dos trabalhos analisados fazem uso de técnicas de *machine learning* para a elaboração de modelos preditivos.

Em uma análise mais generalista da produção científica com foco na avaliação psicológica computadorizada, percebe-se que a grande maioria dos modelos preditivos construídos utilizam dados provenientes de exames bioquímicos, EEGs (eletroencefalogramas) e ressonâncias magnéticas (Hosseini et al.; 2013; Foland-Ross et al.; 2015; Patel et al.; 2015; Jiang et al.; 2016; Shim et al.; 2016; Li et al.; 2016; Hilbert et al.; 2017; Zheng et al.; 2017). Apenas um dos trabalhos encontrados pelos autores empregou testes baseados em produção de desenhos como ferramenta de psicodiagnóstico (in Kim, Kang, Chung and Joo Hong; 2012). Destaca-se também que a população infanto-juvenil foi o foco de apenas dois trabalhos encontrados (Carpenter et al.; 2016; Foland-Ross et al.; 2015).

Ante o exposto, os autores entendem que a abordagem proposta no presente trabalho ainda não foi explorada na literatura, não havendo ainda um modelo que, no contexto do psicodiagnóstico, buscase combinar técnicas de *deep learning* com qualquer modalidade de teste baseado em expressão gráfica.

4 Modelo Proposto

O Melampus² é uma ferramenta modelada para atuar como apoio ao processo de triagem psicológica de crianças. O objetivo desta ferramenta é identificar em uma população de interesse quais são os indivíduos que demonstram maior propensão a apresentar

problemas psíquicos, permitindo aos profissionais de saúde mental conduzir intervenções menos extensivas em grandes grupos de pessoas.

O Melampus automatiza a análise de testes de expressão gráfica que utilizam a análise estática do desenho produzido e possuem um sistema de escalas bem definidas para a avaliação dos itens que as compõe, tal qual o método proposto por de Oliveira (2013) para o DFH. Para tal análise, o modelo sugere o treinamento de classificadores de imagem especializados na avaliação de tais itens.

Por oferecer avaliação automatizada e padronizada dos desenhos, o Melampus permite que a triagem seja passível de aplicação por pessoas com pouco ou nenhum treinamento em psicodiagnóstico. Esta característica aumenta significativamente o número de indivíduos capazes de aplicar a ferramenta em grandes grupos, uma vez que se torna possível que profissionais que tradicionalmente possuem contato com crianças, tais como educadores e assistentes sociais, se tornem agentes de triagem.

Por fim, destaca-se que o Melampus não objetiva identificar a natureza específica dos problemas apresentados pelos indivíduos testados, muito menos substituir o profissional de saúde mental no processo de diagnóstico clínico, uma vez que somente este pode oferecer um parecer oficial sobre cada caso. Além disto, por ser um modelo genérico, admite-se que possam ser construídos diferentes implementações deste, com diferentes testes de expressão gráfica ou mesmo ferramentas de *deep learning*, desde que estes atendam as necessidades do modelo.

4.1 Decisões de Projeto

O processo de triagem proposto pelo Melampus envolve a ação de quatro agentes:

- Melampus: sistema responsável pela avaliação automática do desenho. Recebe como dados de entrada o desenho produzido, idade e gênero do avaliado. A saída produzida será o resultado da avaliação e instrução associada a este resultado;
- Aplicador do Teste: responsável por conduzir o teste e fornecer os materiais e instruções para a sua execução. Alimenta o Melampus com as entradas necessárias para então obter um resultado. O aplicador não precisa ser necessariamente um especialista em psicodiagnóstico;
- Avaliado: no contexto proposto, o avaliado é a

²Adivinho e curandeiro da mitologia grega

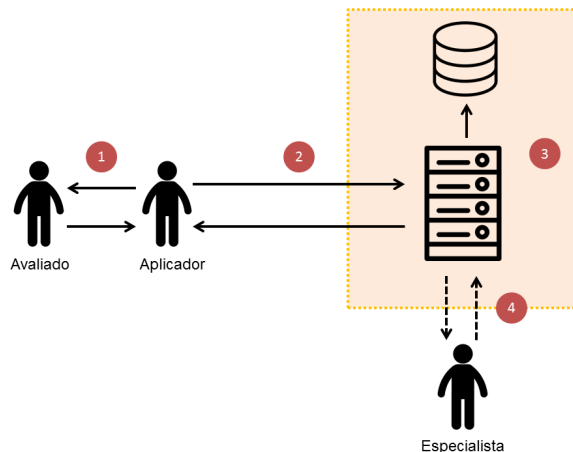


Figura 4: Etapas do processo de triagem no modelo proposto

criança submetida ao processo de triagem;

- **Especialista:** profissional de saúde mental com expertise na avaliação do teste. O especialista é acionado sempre que o sistema não pode determinar o resultado final com grau satisfatório de certeza.

A Figura 4 ilustra as quatro etapas que compõem este processo, sendo que a primeira destas acontece sem a intervenção da ferramenta. Nesta etapa, o aplicador deve instruir, fornecer os materiais necessários (folha, lápis e borracha), e acompanhar o avaliado durante a realização do teste. Após a conclusão deste, o aplicador deve coletar o desenho produzido, e demais dados necessários para a análise.

A segunda etapa consiste na digitalização e submissão do desenho produzido para a análise da ferramenta. Além do arquivo de imagem, será necessário informar os demais dados coletados durante a etapa 1. Todo acesso à ferramenta deve ser autenticado, a fim de garantir que apenas usuários com permissão tenham acesso aos resultados emitidos pelo sistema e aos dados dos indivíduos testados.

Na etapa 3, o sistema armazenará os dados recebidos em sua base de conhecimento e em seguida, a imagem é processada por classificadores previamente treinados, que identificam a presença ou ausência dos itens que compõem o sistema de escalas implementado. Estes classificadores retornarão um valor entre 0 e 1, onde 0 indica ausência do item e 1 indica a presença deste. Para cada um dos classificadores haverá um ponto de corte previamente parametrizado que determina se aquele valor representa a presença ou ausência do item. Além do ponto de corte, será parametrizado também uma margem de erro para cada item. A margem determina um intervalo para mais ou para menos onde o resultado emitido pelo classificador será considerado inconclusivo. Por exemplo, para um determinado item A, com ponto de corte igual a 0.75 e margem de erro de 0.05, resultados maiores ou iguais a 0.80 representam a presença do item, valores inferiores ou iguais a 0.70 indicam ausência do item, e resultados entre 0.70 e 0.80 são inconclusivos.

Antes de retornar o resultado para o aplicador, o

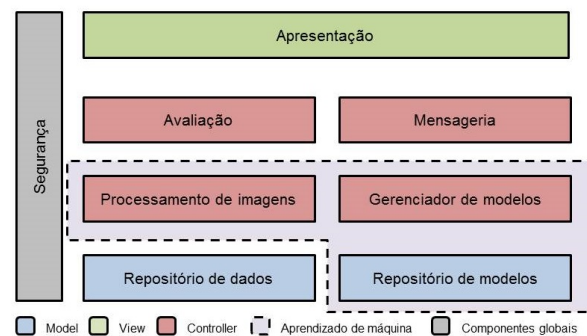


Figura 5: Arquitetura do modelo proposto

Melampus verificará se existe confiança suficiente ou não para tal ação. Para isso, o sistema verificará se existem classificadores que tenham produzido resultados inconclusivos e que por sua vez devam ser então desconsiderados na composição do resultado final. Analisa-se então se estes resultados inconclusivos afetam de alguma forma no resultado final obtido. Em casos negativos, o resultado final da avaliação é apresentado ao avaliador e o processo é encerrado. Em casos negativos, a etapa quatro do processo é iniciada.

Na quarta etapa, o especialista é notificado de que o sistema precisa de apoio humano para avaliar um teste. O especialista então acessa a plataforma e avalia os itens o desenho. Após a submissão da análise, o resultado é enviado para o avaliador e o armazenado na base de dados do Melampus. Observa-se que toda a quarta etapa representa uma comunicação assíncrona entre a ferramenta, o aplicador e o especialista.

Devido ao fato de o Melampus armazenar todos as informações dos testes executados na plataforma, obtêm-se dois benefícios adicionais: (i) facilidade de armazenamento e recuperação dos dados; e (ii) possibilidade de aumento no desempenho dos classificadores, uma vez que estes podem ser treinados novamente conforme a base de dados é enriquecida.

4.2 Arquitetura

Os módulos do Melampus estão organizados em camadas no padrão MVC (*Model-View-Controller*), conforme mostra a Figura 5. Este padrão arquitetural foi escolhido devido a este prover facilidade para a criação de soluções altamente flexíveis e desacoplamento entre as tarefas de interação com o usuário e as tarefas de processamento de dados (Curry and Grace; 2008).

Os módulos básicos que compõem o sistema proposto são: (i) *Apresentação*, permite a interação com os usuários do sistema; (ii) *Segurança*, garante autenticação e autorização; (iii) *Repositório de dados*, módulo responsável pelo armazenamento e recuperação de dados do sistema (imagens, dados dos avaliados, resultados de avaliações, usuários); (iv) *Processamento de imagens*, responsável pela tarefa de identificar a ausência/presença dos itens da escala implementada através de classificadores de imagem; (v) *Mensageria*, responsável por atividades que requerem o envio de mensagens aos usuários,

tais como notificações ou e-mails; (vi) *Avaliação*, contém as regras que determinam os modelos a serem utilizados durante o processamento da imagem para cada grupo específico, realizando também o cálculo e interpretação do escore final; (vii) *Gerenciador de modelos*, responsável pelo gerenciamento dos modelos de classificação aprendidos durante a fase de treinamento e pela execução de novos ciclos de treino; e (viii) *Repositório de modelos*, armazena e recupera os modelos de classificação aprendidos.

5 Metodologia

Para Wazlawick (2014), um método é composto por uma sequência de etapas necessárias para a demonstração de que um determinado objetivo proposto foi alcançado. Para o autor, se os passos do método forem executados, os resultados deverão ser convincentes. Wazlawick (2014) complementa que o método deve indicar quais ações serão desenvolvidas, tais como a construção de protótipos, tipos de experimentos realizados e organização dos dados.

O presente trabalho caracteriza-se por uma pesquisa experimental, onde buscou-se implementar os módulos críticos do modelo proposto como meio de avaliar a sua viabilidade. Os autores entendem os módulos de Avaliação e Processamento de imagens são os mais importantes do modelo Melampus. Para que estes itens pudessem ser completamente funcionais, fez-se necessário implementar também os módulos de Apresentação e Repositório de modelos.

Para o protótipo, tomou-se o sistema de escalas para avaliação do DFH propostas por de Oliveira (2013) como instrumento de triagem adotado. A escolha se deve ao fato de que este se mostra aderente ao modelo proposto por apresentar uma maneira estática de avaliação do desenho baseada na presença/ausência de itens gráficos. Para a construção dos classificadores de imagem, os autores optaram pelo uso de Redes Neurais Convolucionais (ConvNets).

Sendo assim, os passos necessários para a condução do experimento foram (i) organização e unificação de bases de dados; (ii) treinamento dos classificadores de imagem necessários para a implementação de um protótipo, (iii) construção de um protótipo que implementa os módulos críticos do modelo Melampus; e (iv) coleta e avaliação de métricas obtidas a partir da execução do protótipo em laboratório.

5.1 Bases de dados

Os dados utilizados durante as fases de construção e avaliação do protótipo provêm de dois bancos de dados distintos, que em conjunto totalizam 804 desenhos de crianças com idades entre 6 e 12 anos de ambos os sexos e residentes no Estado do Rio Grande do Sul. Nestas bases, as crianças foram classificadas em dois grupos: clínico e não clínico. O grupo clínico é composto por indivíduos que, a época da coleta dos dados, estavam em processo de tratamento psicológico, enquanto que o grupo não clínico reúne indivíduos que não recebiam tratamento psicológico, não frequentavam classes especiais, e

que foram indicados pelos professores como não tendo problemas comportamentais significativos.

Uma das bases de dados pertence ao Laboratório de Mensuração da Universidade Federal do Rio Grande do Sul (UFRGS), e é composta por 606 desenhos coletados entre os anos de 1984 e 1990. Metade dos desenhos pertence a crianças que compõe o grupo clínico. O outro banco de dados utilizado é composto por 198 desenhos coletados no ano de 2005 pertencentes ao Grupo de Estudo, Aplicação e Pesquisa em Avaliação Psicológica (GEAPAP) da UFRGS (Arteche; 2006). Ambas bases de dados também foram utilizadas por de Oliveira (2013) para o desenvolvimento de suas escalas de avaliação do DFH. Segundo Bandeira et al. (2012), o desempenho das crianças nos indicadores cognitivos dos desenhos coletados nos dois períodos não apresentam diferenças significativas, sugerindo certa estabilidade na produção do desenho das crianças através do tempo.

Antes da construção do protótipo, se fez necessário a unificação das bases de dados de forma que fosse possível utilizá-las para treinar os classificadores de imagem. Todas as imagem (no formato de arquivo BMP) receberam então um identificador numérico.

5.2 Treinamento dos Classificadores de Imagem

Para a construção do protótipo e avaliação do modelo, realizou-se o treinamento de modelos de classificação capazes de indicar a ausência/presença dos itens que compõe algumas das escalas propostas por de Oliveira (2013). A escala selecionada foi a EC-DFH-F-9-12 (Escala Clínica do Desenho da Figura Humana para Meninas de 9 a 12 anos de idade) devido a esta ser a escala composta por menor número de itens dentre as quatro escalas disponíveis.

Um dos desafios encontrados durante a fase de treinamento dos classificadores foi o grande desbalanceamento da base de dados. Para reduzir o impacto deste fator no desempenho dos modelos treinados fez-se uso da técnica de atribuição de pesos para as classes do problema (*weighting*), sendo esta uma opção nativa da API³ utilizada para os experimentos. Assim como para Wang et al. (2015), devido a relativamente pequena base de dados optou-se pelo uso da técnica de *dropout* durante a fase de treinamento. *Dropout* consiste basicamente de um algoritmo para treinamento de redes neurais que se baseia em estocasticamente "deixar de fora" neurônios durante o treinamento com o objetivo de evitar *overfitting* (Baldi and Sadowski; 2013). Witten and Frank (2005) definem *overfitting* como o fenômeno que acontece quando o modelo atinge bons resultados quando executado para dados que compõe a base de teste, mas não é capaz de repetir estes resultados para dados que nunca antes alimentaram o modelo.

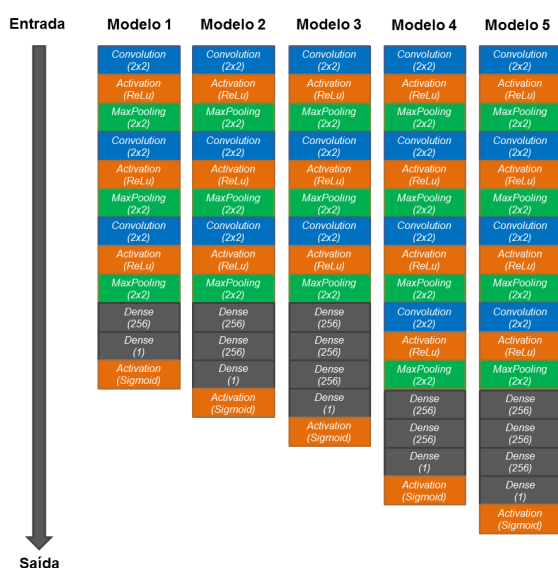
Para cada item que compõe a escala, foram treinados cinco modelos diferentes de ConvNets, afim de selecionar o modelo que apresentasse melhor desempenho. As arquiteturas adotadas são representadas na Figura 6. Estes modelos

³Application Programming Interface

Tabela 2: Distribuição da amostra em relação a escala EC-DFH-F-9-12

Item	Total		Conjunto de Treinamento (70%)		Conjunto de Avaliação (20 %)		Conjunto de Testes (10%)	
	Presença	Ausência	Presença	Ausência	Presença	Ausência	Presença	Ausência
Linha Tremida	154	650	108	455	31	130	15	65
Linha Pesada	58	746	41	522	12	149	5	75
Dificuldade de Integração	41	763	29	534	8	153	4	76
Figura Baixa	143	661	100	463	29	132	14	66
Transparência	14	790	10	553	3	158	1	79
Adaptação do Cabelo	21	783	15	548	4	157	2	78
Penteado	303	501	212	351	61	100	30	50
Ênfase da Face	43	761	30	533	9	152	4	76
Boca Cortada	54	750	38	525	11	150	5	75
Dedos Juntos	253	551	177	386	51	110	25	55
Problemas de Representação de Roupa	280	524	196	367	56	105	28	52

Fonte: Elaborado pelos autores.

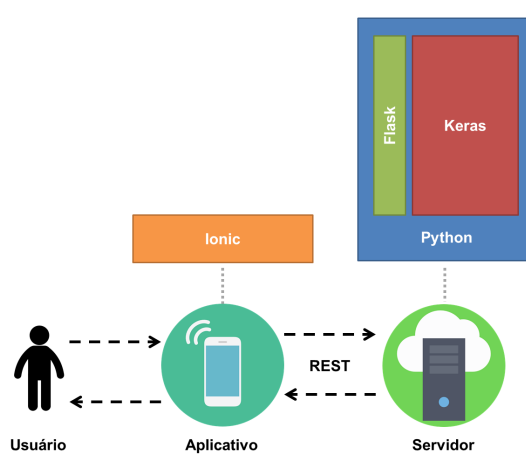
**Figura 6:** Arquitetura geral das ConvNets avaliadas

foram implementados através da API Keras⁴ para Python⁵, que por sua vez atua como uma camada de abstração para a biblioteca de *deep learning* Tensorflow⁶.

A fase de treinamento foi composta por 50 épocas. Durante esta fase também fez-se uso da técnica de *data augmentation*. Esta técnica consiste basicamente de aumentar o tamanho da amostra através da criação de réplicas ligeiramente modificadas dos exemplos que compõe a base de treinamento. Segundo Krizhevsky et al. (2012), esta é a forma mais fácil e comum de evitar *overfitting*.

5.3 Protótipo

Para fins de avaliação da viabilidade técnica do Melampus, construiu-se um protótipo composto pelos principais módulos do modelo: Apresentação, Avaliação, Processamento de Imagens e Repositório de Modelos. A Figura 7 apresenta uma visão em

**Figura 7:** Arquitetura geral do protótipo

alto nível da arquitetura do protótipo desenvolvido, sendo este composto por um aplicativo *mobile* e um *webservice*. O aplicativo e o servidor se comunicam através do protocolo REST (*Representational State Transfer*). O aplicativo permite que os desenhos produzidos pelos indivíduos avaliados sejam digitalizados através da câmera do dispositivo no qual este executa.

Para a construção do aplicativo fez-se uso do *framework* de desenvolvimento *mobile* híbrido Ionic⁷. Já o *webservice* foi escrito na linguagem Python, fazendo uso das bibliotecas Keras e Flask⁸, responsáveis pelas tarefas de *machine learning* e exposição dos serviços REST, respectivamente. Todos os *frameworks* previamente citados são *open source*.

5.4 Métricas Avaliadas

Métricas de qualidade para classificadores são construídas a partir de matrizes de confusão (Sokolova et al.; 2006). Estas matrizes são estruturas compostas por quatro categorias de resultados: Verdadeiros Positivos (VP); Falsos Positivos (FP); Verdadeiros Negativos (VN) e Falsos Negativos (FN) (Davis and Goadrich; 2006). De acordo com Sokolova et al. (2006), uma das métricas mais comumente

⁴<https://keras.io/>

⁵<https://www.python.org/>

⁶<https://www.tensorflow.org/>

⁷<https://ionicframework.com/>

⁸<http://flask.pocoo.org/>

utilizadas é a acurácia (ACU), sendo esta definida pela equação 1.

$$ACU = \frac{VP + VN}{VP + FP + VN + FN} \quad (1)$$

Para Provost et al. (1998), examinar o desempenho de classificadores com base apenas na acurácia pode ser enganoso. Os autores recomendam o uso de curvas ROC (*Receiver Operator Characteristic*) para este tipo de avaliação em classificadores binários. Segundo Castanho et al. (2004), o ideal é que a área sob a curva ROC (*Area Under Curve*, ou AUC) seja próxima a 1, sendo que quando este valor é próximo a 0.5, diz-se que o classificador não é capaz de diferenciar os dois grupos. Para se construir o gráfico ROC plota-se a especificidade no eixo das ordenadas e a sensibilidade no eixo das abscissas (Prati et al.; 2008). Sensibilidade e especificidade são métricas que fornecem a probabilidade de que um modelo forneça resultados corretos ao classificar positivamente ou negativamente uma entrada, sendo estas duas medidas definidas pelas equações 2 e 3, respectivamente. (Castanho et al.; 2004).

$$sensibilidade = \frac{VP}{VP + FN} \quad (2)$$

$$especificidade = \frac{VN}{VN + FP} \quad (3)$$

Ante o exposto, os autores optaram por avaliar o desempenho dos classificadores treinados através de curvas ROC.

6 Resultados

Esta seção descreve os resultados obtidos. A subseção 5.1 detém-se a analisar o desempenho dos modelos de ConvNets propostos na seção de Metodologia e a subseção 5.2 apresenta um parecer dos autores sobre a viabilidade do protótipo desenvolvido com base no modelo Melampus.

6.1 Desempenho dos Modelos de ConvNets Avaliados

Na Tabela 3 são apresentados os valores de AUC referentes aos cinco modelos de ConvNets propostos. Em cada linha destacam-se em verde e vermelho os melhores e os piores resultados obtidos para cada item, respectivamente.

Para Swets et al. (2000), valores de AUC superiores a 0.9 são considerados resultados excelentes, enquanto que valores compreendidos na faixa entre 0.9 e 0.8 podem ser interpretados como bons resultados. Para os mesmos autores, valores entre 0.8 e 0.7 são apenas razoáveis, e valores abaixo de 0.7 representam baixa performance. Porém, Youngstrom (2014) pondera que apesar destes valores provavelmente serem adequados para os campos da engenharia e biomedicina, estes são menos representativos no contexto do diagnóstico

psicológico. Na prática, segundo Youngstrom (2014), muitos dos melhores instrumentos de diagnóstico possuem AUC entre 0.7 e 0.8 em condições clínicas realistas. Considerando-se então esta perspectiva, observa-se que para cinco itens da escala foi possível obter classificadores com boa AUC.

Com exceção do item "Adaptação do Cabelo", os melhores resultados foram apresentados pelos modelos 3, 4 e 5, sendo estes modelos de ConvNets mais complexos quando comparados aos modelos 1 e 2. Apesar disto, não é possível afirmar que houve uma relação direta entre a complexidade do modelo e sua performance nos cenários propostos, uma vez que o modelo de ConvNet que demonstrou melhor performance em um maior número de cenários não foi necessariamente o modelo com mais camadas.

Um dos fatores que possivelmente teve influência na baixa performance dos modelos para certos itens foi o tamanho da base de dados disponível. Segundo Sermanet et al. (2014), a principal desvantagem das ConvNets é a grande quantidade de dados geralmente requerida para o treinamento dos modelos. Estudos que obtiveram bons resultados utilizando-se de ConvNets para a geração de classificadores de imagens tem utilizado grandes bases, tais como a CIFAR10 (60.000 imagens), MNIST (70.000 imagens) e ImageNet 2012 (1.2 milhões de imagens) (Ciresan et al.; 2012; Sermanet et al.; 2014). Segundo Goodfellow et al. (2016), uma das soluções mais efetivas para aumentar a performance de uma ConvNet é coletar mais dados e treiná-la novamente.

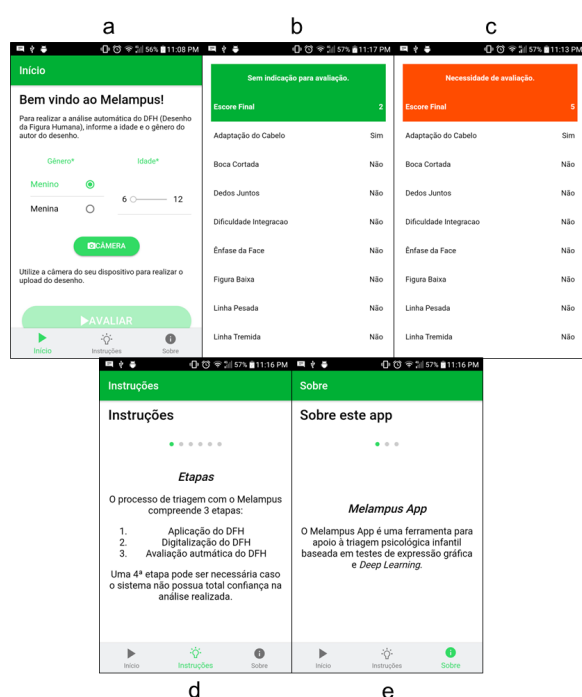
O grande desbalanceamento da base de dados também pode ter afetado o desempenho dos modelos. Apesar de ter-se utilizado as técnicas de *weighting* e *data augmentation*, ainda há a possibilidade de que este fator tenha influenciado nos resultados obtidos, principalmente para itens onde o desbalanceamento foi mais acentuado, como "Transparência", "Adaptação do Cabelo" e "Dificuldade de Integração". Em se tratando especificamente do item "Transparência", sendo este o item com maior desbalanceamento nos dados (para cada exemplo positivo existem cerca de 56 exemplos negativos), observa-se que um dos modelos testados apresentou AUC de aproximadamente 0.89 enquanto os demais modelos apresentaram desempenho muito ruim. Acredita-se que o pequeno número de exemplos positivos no conjunto de teste possa ter distorcido o valor de AUC, e devido a isto sugere-se que a ser necessário avaliar este modelo com maior número de exemplos positivos no conjunto de teste.

Além dos fatores previamente citados, os autores pensam ser relevante para o entendimento dos resultados levar em consideração a complexidade dos padrões de imagem que representam cada um dos itens da escala clínica. Dentre os itens que compõe as escalas clínicas propostas por de Oliveira (2013), entende-se que existem aqueles podem ser interpretados como padrões de maior ou menor complexidade computacional. Itens como "Figura Pequena" e "Linha Pesada" são conceitos computacionalmente menos complexos do que itens como "Adaptação do Cabelo" e "Penteado", uma vez que estes últimos representam conceitos mais abstratos e podem ser expressos por um número muito maior de variações gráficas.

Tabela 3: Resultados obtidos após treinamento das Convnets propostas

Item	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Penteado	0,504106	0,609649	0,529586	0,640442	0,640895
Problemas de Representação de Roupa	0,456624	0,657801	0,700038	0,361955	0,544453
Figura Baixa	0,810425	0,892019	0,949811	0,919847	0,912343
Linha Tremida	0,431795	0,464615	0,552821	0,336923	0,470769
Dedos Juntos	0,514909	0,494545	0,677091	0,493818	0,503636
Transparência	0,500000	0,101266	0,126582	0,886076	0,126582
Adaptação do Cabelo	0,500000	0,500000	0,365385	0,500000	0,275641
Boca Cortada	0,536000	0,500000	0,562667	0,557333	0,640000
Dificuldade de Integração	0,414474	0,457237	0,618421	0,552632	0,588816
Ênfase da Face	0,470395	0,473684	0,592105	0,500000	0,884868
Linha Pesada	0,781333	0,408000	0,528000	0,816000	0,372000

Fonte: Elaborado pelos autores.



(a) Tela inicial: permite a execução a avaliação do desenho digitalizado através da câmera. (b) Resultado positivo de uma avaliação. (c) Resultado negativo de uma avaliação. (d) Tela de instruções sobre uso do aplicativo. (e) Tela de informações sobre o aplicativo.

Figura 8: Principais telas do aplicativo

6.2 Protótipo Desenvolvido

Em relação ao protótipo construído, confirmou-se a viabilidade técnica da implementação do Melampus como um aplicativo para *smartphones*. Na Figura 8 são apresentadas as principais telas do aplicativo desenvolvido.

Através do protótipo, foi possível fazer a integração dos dados coletados através do aplicativo com o servidor responsável pelas tarefas de computação intensiva das imagens através das ConvNets e cálculo do escore final da triagem. Pontua-se também que os autores entendem que a possibilidade de uso da câmera do dispositivo como meio de digitalização dos

desenhos facilita o processo de *upload* das imagens e agiliza o processo. Porém, percebe-se que a utilização da câmera requer alguns cuidados para a captura da imagem, tais como a disponibilidade de boas condições de iluminação e o correto enquadramento da figura.

7 Conclusão

O objetivo deste trabalho foi apresentar o modelo Melampus, uma ferramenta de apoio ao processo de triagem psicológica infantil que faz uso de *deep learning* para a avaliação automática de testes de expressão gráfica.

O modelo foi avaliado através da implementação de um protótipo, fazendo uso de um sistema de escalas para a avaliação do teste do Desenho da Figura Humana. Para a geração dos classificadores de imagem, utilizou-se Redes Neurais Convolucionais treinadas em uma base composta por desenhos previamente coletados.

Os resultados mostram que houveram modelos de ConvNet treinados que obtiveram bom desempenho no conjunto de testes para certos itens da escala clínica. Porém, os autores reconhecem que mais esforços se fazem necessários em função de obter melhores resultados onde o desempenho foi inesperado.

Além do tamanho relativamente pequeno da base de dados disponível, outra limitação deste trabalho foi o fato de que, devido a questões éticas e legais, o protótipo desenvolvido não pode ser avaliado em um contexto real de triagem psicológica. Cabe ressaltar também que a viabilidade do Melampus como ferramenta de triagem psicológica está diretamente relacionada ao instrumento de avaliação psicológica selecionado para compor a sua implementação. Sendo assim, o Melampus não poderá ser reconhecido como um instrumento de triagem válido caso o método de avaliação do teste de expressão selecionado não possuir validade clínica adequada.

Os autores entendem que o Melampus se mostra um modelo viável, porém mais estudos serão necessários para que as implementações baseadas neste recebam parecer favorável das entidades reguladoras de testes psicológicos.

Sugere-se como continuidade desta pesquisa: aumentar o banco de imagens utilizado através da

coleta de novos desenhos ou fusão com outras bases preexistentes; treinar e avaliar outros modelos de ConvNet; identificar demais sistemas de triagem psicológica aderentes a proposta do Melampus; realizar testes de aceitação do modelo através do protótipo desenvolvido e de instrumentos como o TAM (*Technology Acceptance Model*).

Referências

- Anxiety and Depression Association of America (2017). Anxiety and depression in children. Disponível em: <https://www.adaa.org/living-with-anxiety/children/anxiety-and-depression>. Acesso em: 21 fevereiro 2017.
- Arteche, A. X. (2006). Indicadores emocionais do desenho da figura humana : construção e validação de uma escala infantil. 164f. Tese de Doutorado. Programa de Pós-Graduação em Psicologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS. Disponível em: <http://www.lume.ufrgs.br/handle/10183/8730>. Acesso em: 21 maio 2017.
- Baldi, P. and Sadowski, P. J. (2013). Understanding dropout, in C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger (eds), *Advances in Neural Information Processing Systems* 26, Curran Associates, Inc., pp. 2814–2822. URL: <http://papers.nips.cc/paper/4878-understanding-dropout.pdf>
- Bandeira, D. R., Costa, A. and Arteche, A. (2012). The flynn effect in brazil: Examining generational changes in the draw-a-person and in the raven's coloured progressive matrices, *Revista Latinoamericana de Psicologia* 44(3).
- Borsa, J. C. and Bauermann, M. (2013). O desenho da figura humana na avaliação da agressividade infantil, *Avaliação Psicológica* 12: 273 – 274.
- Carpenter, K. L. H., Sprechmann, P., Calderbank, R., Sapiro, G. and Egger, H. L. (2016). Quantifying risk for anxiety disorders in preschool children: A machine learning approach, *PLOS ONE* 11(11): e0165524.
- Castanho, M., Barros, L., Vendite, L. and Yamakami, A. (2004). Avaliação de um teste em medicina usando uma curva roc fuzzy, *Biomatematica* 14: 19–28.
- Ciresan, D., Meier, U., Masci, J. and Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification, *Neural Networks* 32: 333–338.
- Curry, E. and Grace, P. (2008). Flexible self-management using the model-view-controller pattern, *IEEE software* 25(3).
- Davis, J. and Goadrich, M. (2006). The relationship between precision-recall and roc curves, *Proceedings of the 23rd international conference on Machine learning*, ACM, pp. 233–240.
- de Oliveira, S. E. S. (2013). Construção de escalas clínicas do desenho da figura humana para crianças de 6 a 12 anos: Normas e evidências de validade. 75 f. Monografia apresentada como exigência parcial do Curso de Especialização em Psicologia Clínica – Instituto de Psicologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 2013. Disponível em: <http://www.lume.ufrgs.br/bitstream/handle/10183/95473/000917471.pdf?sequence=1>. Acesso em: 12 maio 2017.
- Falissard, B. (2016). Early detection of child and adolescent mental disorders: some elements of a necessary debate, *European Child & Adolescent Psychiatry* 25(10): 1041–1043.
- Foland-Ross, L. C., Sacchet, M. D., Prasad, G., Gilbert, B., Thompson, P. M. and Gotlib, I. H. (2015). Cortical thickness predicts the first onset of major depression in adolescence, *International Journal of Developmental Neuroscience* 46: 125–131.
- Gabriels, R. L., Wamboldt, M. Z., McCormick, D. R., Adams, T. L. and McTaggart, S. R. (2000). Children's illness drawings and asthma symptom awareness., *The Journal of asthma : official journal of the Association for the Care of Asthma* 37: 565–574.
- Gauy, F. V. and Guimarães, S. S. (2006). Triagem em saúde mental infantil, *Psicologia: Teoria e Pesquisa* 22(1): 5–16.
- Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*, MIT Press. <http://www.deeplearningbook.org>.
- Hammond, L. C. and Gantt, L. (1998). Using art in counseling: Ethical considerations, *Journal of Counseling & Development* 76(3): 271–276.
- Hilbert, K., Lueken, U., Muehlhan, M. and Beesdo-Baum, K. (2017). Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: A multimodal machine learning study, *Brain and Behavior* 7(3): e00633.
- Hosseini, B., Moradi, M. H. and Rostami, R. (2013). Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal, *Computer Methods and Programs in Biomedicine* 109(3): 339–345.
- Hutz, C. S. and Bandeira, D. R. (1995). Avaliação psicológica com o desenho da figura humana: técnica ou intuição?, *Temas em Psicologia* 3: 35 – 41.
- in Kim, S., Han, J. and Oh, Y.-J. (2012). A computer art assessment system for the evaluation of space usage in drawings with application to the analysis of its relationship to level of dementia, *New Ideas in Psychology* 30(3): 300–307.
- in Kim, S., Kang, H.-S., Chung, S. and Joo Hong, E. (2012). A statistical approach to comparing the effectiveness of several art therapy tools in estimating the level of a psychological state, *The Arts in Psychotherapy* 39(5): 397–403.

- Jiang, H., Popov, T., Jylänki, P., Bi, K., Yao, Z., Lu, Q., Jensen, O. and van Gerven, M. (2016). Predictability of depression severity based on posterior alpha oscillations, *Clinical Neurophysiology* **127**(4): 2108–2114.
- Jiménez-Serrano, S., Tortajada, S. and García-Gómez, J. M. (2015). A mobile health application to predict postpartum depression based on machine learning, *Telemedicine and e-Health* **21**(7): 567–574.
- Kim, E. Y., Lee, M. Y., Kim, S. H., Ha, K., Kim, K. P. and Ahn, Y. M. (2017). Diagnosis of major depressive disorder by combining multimodal information from heart rate dynamics and serum proteomics using machine-learning algorithm, *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **76**: 65–71.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, in F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger (eds), *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., pp. 1097–1105.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning, *Nature* **521**(7553): 436–444.
- LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE* **86**(11): 2278–2324.
- LeCun, Y., Kavukcuoglu, K. and Farabet, C. (2010). Convolutional networks and applications in vision, *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, IEEE, pp. 253–256.
- Li, X., Hu, B., Sun, S. and Cai, H. (2016). EEG-based mild depressive detection using feature selection methods and classifiers, *Computer Methods and Programs in Biomedicine* **136**: 151–161.
- Luxton, D. D. (2014). Artificial intelligence in psychological practice: Current and future applications and implications., *Professional Psychology: Research and Practice* **45**(5): 332–339.
- Mental Health Foundation (2016). Fundamental facts about mental health 2016, *Technical report*, Mental Health Foundation, Londres.
- Muller, B. E. and Erford, B. T. (2012). Choosing assessment instruments for depression outcome research with school-age youth, *Journal of Counseling & Development* **90**(2): 208–220.
- Murphy, M. and Fonagy, P. (2013). Mental health problems in children and young people, *Annual Report of the Chief Medical Officer 2012, Our Children Deserve Better: Prevention Pays*, Departamento de Saúde do Reino Unido, Reino Unido.
- Noda, K., Yamaguchi, Y., Nakadai, K., Okuno, H. G. and Ogata, T. (2014). Audio-visual speech recognition using deep learning, *Applied Intelligence* **42**(4): 722–737.
- Patel, M. J., Andreescu, C., Price, J. C., Edelman, K. L., Reynolds, C. F. and Aizenstein, H. J. (2015). Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction, *International Journal of Geriatric Psychiatry* **30**(10): 1056–1067.
- Popi, J. M., de Sá Riechi, T. I. J. and Hamdan, A. C. (2016). Triagem infantil neuropsicológica computadorizada para indicação do TDAH em crianças e adolescentes, *Psicologia Argumento* **34**(84): 39–50.
- Prati, R., Batista, G. and Monard, M. (2008). Curvas roc para avaliação de classificadores, *Revista IEEE América Latina* **6**(2): 215–222.
- Provost, F. J., Fawcett, T. and Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms., *ICML*, Vol. 98, pp. 445–453.
- Rice, F., Eyre, O., Riglin, L. and Potter, R. (2017). Adolescent depression and the treatment gap, *The Lancet Psychiatry* **4**(2): 86–87.
- Rollins, J. A. (2005). Tell me about it: Drawing as a communication tool for children with cancer, *Journal of Pediatric Oncology Nursing* **22**(4): 203–221.
- Saur, A. M., Pasian, S. R. and Loureiro, S. R. (2010). Desenho da figura humana e a avaliação da imagem corporal, *Psicologia em Estudo* **15**(3): 497–507.
- Seixas, F. L., Zadrozny, B., Laks, J., Conci, A. and Saade, D. C. M. (2014). A bayesian network decision model for supporting the diagnosis of dementia, alzheimer's disease and mild cognitive impairment, *Computers in Biology and Medicine* **51**: 140–158.
- Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and Lecun, Y. (2014). Overfeat: Integrated recognition, localization and detection using convolutional networks, *International Conference on Learning Representations (ICLR2014)*, CBLS, April 2014.
- Shim, M., Hwang, H.-J., Kim, D.-W., Lee, S.-H. and Im, C.-H. (2016). Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features, *Schizophrenia Research* **176**(2–3): 314–319.
- Sokolova, M., Japkowicz, N. and Szpakowicz, S. (2006). Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation, *Australasian Joint Conference on Artificial Intelligence*, Springer, pp. 1015–1021.
- Swets, J. A., Dawes, R. M. and Monahan, J. (2000). Psychological science can improve diagnostic decisions, *Psychological Science in the Public Interest* **1**(1): 1–26. PMID: 26151979.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2015). Going deeper with convolutions, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Vargas, A. C. G., Paes, A. and Vasconcelos, C. N. (2016). Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres, *SIBGRAPI - Conference on Graphics, Patterns and Images*, Sociedade Brasileira de Computação, Sociedade Brasileira de Computação, São José dos

Campos.

URL: <http://urlib.net/8IMKD3MGPAW/3ME3L2P>

Viapiana, V. F., Bandeira, C. M. and Giacomoni, C. H. (2016). Bem-estar subjetivo infantil: avaliação por meio do desenho da figura humana, *Revista Avaliação Psicológica* **15**(1): 49-59.

Wang, L., Qiao, Y. and Tang, X. (2015). Action recognition with trajectory-pooled deep-convolutional descriptors, *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*.

Wazlawick, R. S. (2014). *Metodologia de Pesquisa Para Ciência da Computação*, CAMPUS - GRUPO ELSEVIER.

Witten, I. H. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*, Morgan Kaufmann.

World Health Organization (2013). Mental health action plan 2013 - 2020, *Technical report*, World Health Organization.

World Health Organization (2017). Child and adolescent mental health. Disponível em: http://www.who.int/mental_health/maternal-child/child_adolescent/en/. Acesso em: 22 fevereiro 2017.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to roc, *Journal of Pediatric Psychology* **39**: 204-221.

Zheng, H., Zheng, P., Zhao, L., Jia, J., Tang, S., Xu, P., Xie, P. and Gao, H. (2017). Predictive diagnosis of major depression using NMR-based metabolomics and least-squares support vector machine, *Clinica Chimica Acta* **464**: 223-227.

Zhu, Y., Shang, Y., Shao, Z. and Guo, G. (2017). Automated depression diagnosis based on deep networks to encode facial appearance and dynamics, *IEEE Transactions on Affective Computing* pp. 1-1.