

ESTEPHANY NICOLE DA SILVA

JOSEPH DE MOURA FREITAS

THIAGO MORALES RIBEIRO

RecoSkin AI: Plataforma de classificação de dermatites atópicas por meio de deep learning

Projeto de Pesquisa realizado sob orientação do Prof. Robson Ferreira Lopes, co-orientação do Prof. Ricardo Agostinho de Rezende Júnior e da pesquisadora biomédica Bruna Amorim dos Santos.

Guarulhos

2021

Resumo

O projeto RecoSkin AI foi inspirado em uma pesquisa desenvolvida por estudantes e pesquisadores da Universidade Estadual de Campinas (Unicamp), que estão desenvolvendo um software capaz de emitir o diagnóstico do câncer de pele do tipo melanoma. Por utilizarem uma inteligência artificial como objeto central de análise para o diagnóstico, consideramos muito válido seguir um rumo semelhante; no entanto, nosso projeto visa abordar outra espécie de doença dermatológica, a dermatite Atópica. Esse tipo de doença consiste basicamente em uma inflamação na pele, que pode causar vermelhidão, coceira, descamação e pequenas bolhas. Todo o desenvolvimento do projeto prioriza auxiliar cada vez mais o diagnóstico médico e o acompanhamento de pessoas que sofrem com esse tipo de doença, a qual, se tratada no início, dificilmente se agrava. A ferramenta em desenvolvimento será capaz de acelerar o processo de diagnósticos médicos em relação a essa doença epitelial. O RecoSkin, no momento, está apenas no início, ou seja, necessita de aperfeiçoamento e pesquisa. Utilizaremos a linguagem de programação aliada ao deep learning na área de visão computacional para tornar viável o desenvolvimento da plataforma. O objetivo principal do projeto é proporcionar ao dermatologista um instrumento capaz de facilitar seu ofício, o acompanhando na difícil tarefa de diagnosticar um indivíduo com essa doença de pele. A partir de informações presentes em artigos, monografias e até mesmo sites especializados, a coletânea de estudos foi montada, esses recursos nos possibilitam aplicar de forma técnica, os dados da área da saúde e o deep learning em nosso algoritmo, ou seja, a metodologia do projeto se resume à união das pesquisas bibliográficas aplicadas de modo descritivo. Quanto aos resultados obtidos ao longo da pesquisa para o desenvolvimento da ferramenta, podemos citar: o bom funcionamento da rede convolucional, uma baixa acurácia no classificador e a capacidade de manipular a imagem.

Palavras-chave: Dermatite; Inteligência Artificial; Ferramenta; Diagnóstico; Classificação.

1. Introdução

O presente projeto busca enfatizar e trabalhar especificamente sobre uma doença dermatológica que pode ser confundida com outras, a dermatite Atópica, muito semelhante a doenças como a Psoríase e o Eczema. Essa incerteza se deve ao fato de que sua causa ainda é considerada incerta entre os médicos. Esse contexto possibilita que a dermatite seja tratada de forma errada ou duvidosa. Por esse motivo o RecoSkin buscará auxiliar no pré-diagnóstico do médico, analisando uma imagem que apresente a doença manifestada, acompanhada de uma mensagem relatando a porcentagem da classificação da doença, ou seja, o valor retornado será um número que permitirá ao usuário saber se sua doença realmente se trata de uma dermatite Atópica ou possível variante. É necessário ressaltar que esse projeto é destinado a pessoas que estejam sofrendo com a variante da dermatite de tipo Atópica.

Este trabalho aborda um tema conhecido como “Pré-diagnóstico”, uma espécie de análise ampla, que será feita através de uma imagem fornecida pela câmera do usuário. Como o próprio nome informa, o termo “pré-diagnóstico” consiste em uma informação passada ao paciente a respeito da doença que ele está enfrentando e, posteriormente, a situação da pessoa é avaliada por um médico especializado na área. Esse tipo de diagnóstico prévio permite um conforto maior ao paciente, já que ele terá uma breve noção do que será tratado e, dessa maneira, a tendência é que o avanço da doença seja retardado.

Obviamente, abordando o tema dessa maneira tão ampla, há uma grande dificuldade de entender a dermatite em sua completude, por esse motivo decidiu-se focar em uma das variantes dessa doença, a dermatite Atópica. Essa doença é mais comum nas dobras dos braços e nas partes de trás dos joelhos, ou seja, ela se localiza em uma região que apresenta tendência a ficar úmida, por conta do suor do indivíduo (fator que gera a irritação na pele). Seu agravamento pode estar atrelado a doenças crônicas respiratórias como a asma e a rinite alérgica, o que explica o motivo de a dermatite ser considerada uma doença de reação alérgica.

A doença abordada no projeto apresenta muitos sintomas visualmente perceptíveis (inflamação, vermelhidão, erupção, descascamento, fissura e secura). Em virtude dessa

informação, torna-se clara a necessidade de uma identificação por meio de uma imagem, que receberá o tratamento e manipulação específica para que dessa forma a máquina possa compreender que tipo de informação está sendo enviada a ela.

Essa doença epitelial não é capaz de provocar a morte de um indivíduo, porém seus sintomas em estados mais graves são capazes de afetar o sono, relações sociais, sexuais, trabalho, esporte, lazer e imagem, segundo Ainara Rodríguez (2019), responsável médica de Dermatite Atópica de Sanofi (trata-se de uma empresa farmacêutica de Gentilly, França). A coceira intensa e contínua acaba afetando o psicológico do portador dessa doença, podendo levá-lo a desenvolver tendências suicidas, muitas das quais surgem por conta da carga emocional atrelada a essa doença. Todo o estresse acumulado, unido à irritação da pele e ao hábito de se coçar, faz com que a pessoa opte pela morte ao invés de continuar sofrendo ao longo dos dias. Pesquisas desenvolvidas pela Associação de Apoio à Dermatite Atópica (AADA, 2010) apontam que cerca de 10% a 15% da população brasileira sofre com essa doença, sendo necessário o apoio familiar para que sintomas psicológicos não venham a se agravar e, nesse sentido, grupos de apoio e campanhas de conscientização se tornam mais do que necessárias nesse momento.

A ferramenta produzida pelo RecoSkin surge com o intuito de mudar esse cenário, permitindo que os diagnósticos sejam mais certos, o que favorece um tratamento mais rápido da doença, já que não haverá muita demora para sua identificação. A dermatite Atópica infelizmente não possui cura até o presente momento, mas há diversos tratamentos que amenizam o sofrimento. Quanto mais pacientes tiverem a sua patologia identificada, menos pessoas sofrerão, é isso o que o projeto RecoSkin visa fazer, contribuir para a saúde com o apoio de profissionais da área.

Um dos possíveis problemas que o RecoSkin enfrentará está inteiramente relacionado à “ignorância” ou desaprovação às pesquisas científicas, pois, de acordo com os estudos realizados pela equipe do RecoSkin, foi possível concluir que grande parcela da população que sofre com alguma doença, qualquer que seja, busca se informar e se autodiagnosticar através da internet; no entanto, segundo sites como “Pebmed” (2019), “Bonde” (2019) e “BemParaná” (2019), 7 em cada 10 brasileiros recebem diagnósticos falsos e 40% dos

brasileiros se auto diagnosticam, como aponta os sites “R7” (2018) e “Exame” (2018). Foi possível perceber que grande parte das pessoas que visualizam qualquer tipo de sintomas em suas peles, logo os associam ao câncer, o que é um terrível engano, já que há muitas doenças que se manifestam através da pele, como a própria dermatite, que possui sintomas semelhantes aos do câncer de pele, entre eles, o descamamento. Esse falso diagnóstico pode prejudicar a saúde das pessoas e até mesmo matá-las, caso as doenças não sejam tratadas da maneira correta.

A maior motivação do projeto de pesquisa RecoSkin se deu por conta de estudos realizados na UFRGS (Universidade Federal do Rio Grande do Sul) em conjunto com o IEEE (Instituto de Engenheiros Eletricistas e Eletrônicos), que desenvolveram um projeto cujo objetivo era facilitar o conhecimento em relação ao câncer de pele e prover o diagnóstico da doença, já que anteriormente o diagnóstico só poderia ser emitido através de uma biópsia ou dermatoscópio, que são maneiras muito caras e lentas de se conseguir um resultado. De acordo com o Instituto Nacional do Câncer (INCA, 2018), todos os anos são registrados mais de 180 mil casos de câncer de pele e com toda essa lentidão e dificuldade na disponibilidade de diagnósticos não há maneira de conseguir tratar uma boa parcela de pessoas com a doença. O projeto desenvolvido em parceria pela UFRGS e IEEE utiliza recursos muito simples, como uma câmera fotográfica, uma rede neural é um software que informa se a pele tem algum tipo de lesão, diagnosticando a pessoa com 90% de certeza, segundo o professor Jacob Scharcanski (2019), membro do IEEE.

O projeto de pesquisa apresentado neste relatório inspirou-se na abordagem da UFRGS e IEEE e pretende dar continuidade às ideias expandindo seu alcance para outro ramo de doenças que apresentam sintomas na pele, permitindo que um especialista possa pré-diagnosticar com maior velocidade a doença que está afligindo o paciente, lembrando que o objetivo visado pelo programa não é substituir o trabalho de um profissional da medicina, mas apenas auxiliar e aumentar a chance de combater a dermatite atópica, diminuindo os casos de autodiagnóstico errôneos. Por ser um pré-diagnóstico, irá permitir o tratamento da doença enquanto não estiver espalhada pelo corpo, ou seja, ainda estará em uma fase inicial. Sabe-se que, atualmente, a falta de informação sobre diversos tipos de doenças é muito menor, quando comparada a tempos anteriores, e a tecnologia é uma das grandes responsáveis por mudar essa situação. Podemos,

então, imaginar que com a implementação da Inteligência Artificial no cotidiano e na área da saúde, um número maior de pessoas será capaz de lidar com os males da doença, não havendo preocupações com o avanço da dermatite Atópica.

Para o projeto ter um bom andamento foi necessário estabelecer uma questão norteadora, um modelo de pergunta responsável por guiar a pesquisa até o seu objetivo principal. A plataforma (algoritmo) desenvolvida pelo projeto de pesquisa abarca tanto o meio social quanto a área da saúde, visto que visa ao bem-estar da população. A questão que vem guiando o projeto é **“Em que nível o modelo de pré-diagnóstico fornecido pelo projeto RecoSkin será capaz de facilitar a vida de um profissional da saúde e de um portador da Dermatite Atópica?”** Quando o projeto se iniciou, não se sabia ao certo como seria a ferramenta, porém a pergunta estabelecida propunha que deveria ser um software simples, que permitisse o manuseio por um profissional (para diagnóstico) e por um leigo (para acompanhamento). Infelizmente esse instrumento ainda não se encontra disponível para uso e quando for disponibilizado buscará atingir primeiramente a população brasileira que convive com a dermatite atópica. Como já foi mencionado ao longo do relatório, a dermatite é uma doença autoimune (patologia que afeta o sistema imunológico, fazendo com que o organismo provoque ataques a si mesmo), que acaba afetando o ser humano tanto emocionalmente quanto fisicamente. Com todos esses sintomas afetando um organismo, é possível que ele desenvolva transtornos depressivos, que muitas vezes são responsáveis pela formação de tendências suicidas. Ainda não há como comprovar a eficácia da ferramenta em desenvolvimento em uma situação real de análise, mas ela está apresentando avanços e, em um futuro próximo, facilitará os estudos dermatológicos brasileiros.

O objetivo geral da pesquisa consiste, basicamente, em auxiliar médicos e pessoas que sofrem com uma variante específica da dermatite. Um meio de compreender nosso desígnio de uma maneira mais específica seria da seguinte forma: através de um pré-diagnóstico, o intuito é auxiliar pessoas que, por algum motivo de desinformação ou “ignorância”, acabam confundindo diferentes tipos de doenças que se manifestam na pele com a doença abordada pelo projeto (Dermatite atópica). O pré-diagnóstico visa facilitar a relação entre o paciente e o especialista, já que ambos terão uma maior precisão para tratar dessa doença dermatológica de reação alérgica.

2. Fundamentação teórica

O projeto de pesquisa trata-se de um trabalho voltado para o ramo da saúde, mais especificamente para a área da dermatologia. O RecoSkin visa apoiar o diagnóstico feito pelo profissional, tendo como uma possível consequência a diminuição de pré-diagnósticos realizados pela internet.

De acordo com as informações apresentadas durante a introdução, é possível concluir que o projeto é voltado para o estudo da dermatite Atópica por meio de uma aplicação relacionada com *deep learning*. Dentre os temas abordados no projeto de pesquisa estão: a própria dermatite Atópica, Aprendizado profundo e a influência da internet em um pré-diagnóstico.

2.1 Dermatite Atópica

Essa doença pode ser definida como uma espécie de inflamação localizada na região cutânea. Além de ser conhecida como dermatite Atópica também é conhecida, entre o meio da medicina, como eczema atópico. Ela costuma formar placas ou pequenos caroços, que costumam coçar bastante, também é importante ressaltar que não há idade para o aparecimento dessa doença, porém é comum se manifestar até os cinco anos de idade de um indivíduo (Viana, Aleksana, Dermatologista, 2021).

Por mais que não haja cura para essa variação da doença, existem tratamentos eficazes como a utilização de anti-inflamatórios, pomadas, comprimidos ou hidratação da pele.

Segundo o artigo disponibilizado pela professora adjunta do departamento de Dermatologia da UNIFESP, Silmara C. P. Cestari, a dermatite Atópica é uma doença que costuma variar muito, portanto é considerada de difícil definição. A localização da pessoa, tempo, histórico familiar e pessoal são questões analisadas pelo profissional antes de dar um veredito a respeito da doença. É necessário ressaltar que o diagnóstico é totalmente clínico, que é uma espécie de diagnóstico realizado quando há doenças semelhantes, mas existe um sintoma muito

específico capaz de auxiliar o médico nesse processo de análise, chama-se prurido e faz referência ao desconforto causado pelo ato de se coçar (Cestari, Silmara. 2010).

Existem relatos que buscam enfatizar como o desequilíbrio emocional pode ocasionar um avanço no agravamento da doença. A publicitária Simone Bispo comenta um pouco sobre essa situação em uma matéria produzida pelo site VivaBemUol (2020). Ela conta que teve sua primeira experiência com a dermatite sendo agravada pela emoção em 2018. A mulher cita que o estresse gerado por discussões políticas começou a se manifestar em sua pele, o que gerou sua primeira crise severa e acabou afetando sua autoestima, visto que seu corpo ficou repleto de manchas causadas pela enfermidade.

2.2 Aprendizado profundo

A unidade básica de uma rede neural artificial é um neurônio matemático, também denominado nó, baseado no neurônio biológico. As ligações entre esses neurônios matemáticos estão relacionadas com àquelas dos cérebros biológicos, e principalmente no modo como essas associações se desenvolvem ao longo do tempo, denominado “treinamento”. (Chagas, Edgar. 2019)

O Deep Learning, também denominado aprendizagem profunda, é uma parte do Machine Learning ou aprendizagem de Máquina, que aplica algoritmos para processar dados e reproduzir o processamento realizado pelo cérebro humano. (Chagas, Edgar. 2019)

Algumas das principais aplicações de Deep Learning utilizadas atualmente são: compreensão do comportamento do cliente realizado por sites de e-commerce robustos (Amazon e eBay). Durante todo o tempo no site, os dados do consumidor são coletados. Quanto maior a experiência no local, maiores serão as chances de efetivar a compra. Assim, o site é otimizado a cada acesso por meio dos dados gerados, para que a experiência destinada ao cliente seja mais cativante;

Reconhecimento facial é uma aplicação já utilizada em sites de redes sociais, como o Facebook, para identificar quais são os potenciais usuários a serem marcados em uma determinada foto. Porém, atualmente, tem sido bastante utilizada para os sistemas de segurança. No carnaval de 2019, inclusive, o sistema auxiliou a Polícia Militar da Bahia a prender um homem foragido, reconhecido por meio desse tipo de sistema;

Suporte técnico personalizado, de forma que não é necessário, em muitos casos, atendentes humanos para prestar o serviço de assistência remota a um cliente. Alguns exemplos são a Clara, o Howdy e o GridSpace Sift;

Classificação de doenças. Por exemplo, há hoje tecnologias que auxiliam oftalmologistas em exames médicos para identificar, por exemplo, retinas afetadas pela diabetes, por meio da análise de imagens;

Redução no erro do diagnóstico de câncer, auxiliando na identificação de células cancerígenas por meio das imagens dos linfonodos;

Carros autônomos, que funcionam por meio de um Computador Neural Diferenciável (DNC), ao ser apresentado a mapas, linhas e paradas, o carro consegue se conduzir para uma rota mais curta, facilitando as viagens. (Gaea)

2.3 Influência da Internet em um Autodiagnóstico

O conceito de Internet remonta de um período relativamente próximo ao atual, ou seja, é uma invenção muito recente se comparada a criações como a máquina a vapor ou a máquina fotográfica. A internet é basicamente um conjunto de redes de computadores espalhadas por todo o mundo, ou seja, ela cria uma conexão global, que permite o compartilhamento de dados entre dispositivos variados como, *notebooks*, *smartphones*, impressoras, televisões etc. A internet surgiu no período de Guerra Fria (1947-1991), sendo desenvolvida por norte-americanos, que visavam transformá-la em um instrumento de comunicação interno. Dessa maneira eles poderiam utilizá-la caso os meios de comunicação da época fossem destruídos ou inutilizados (UOL, 2001).

Em uma matéria escrita pelo jornalista Felipe Lourenço no blog Iclinic são discutidos os principais motivos que levam um paciente a optar por um diagnóstico feito pela internet ao invés de consultar um profissional que estudou anos para exercer sua profissão. As consultas feitas via navegador de internet têm se tornado cada vez mais comuns. Cabe ressaltar que a prática do auto diagnóstico pode gerar riscos severos à saúde de um indivíduo, já que pode criar um pânico desnecessário ou omissão de uma doença severa. Muitos pacientes apenas buscam o médico para obter uma confirmação, pois acreditam que já se diagnosticaram por meio dessa grande rede. O autodiagnóstico é uma questão ainda muito recente, pois ela passa a se consolidar mediante ao avanço global da internet. Com toda essa desconfiança do paciente que passa a acreditar em informações duvidosas, o médico, mais do que nunca, passou a vender confiança, porque necessita conquistar a pessoa antes de examiná-la (Lourenço, Felipe. 2014).

3. Metodologia de pesquisa

Para o desenvolvimento do projeto foi necessário iniciá-lo por meio de pesquisas bibliográficas e estudos sobre o tema. Ambas as questões abordadas pela ferramenta são de difícil compreensão, já que ela abrange conhecimentos da área da saúde aliados ao aprendizado de máquina. Por conta do momento de pandemia em que o mundo se encontra desde o ano de 2020, não houve a possibilidade de ser realizada uma pesquisa de campo, ou seja, o processo de criação da plataforma RecoSkin se limitou à revisão bibliográfica disponibilizada por cursos, documentos acadêmicos e sites especializados no assunto. Felizmente há muito conteúdo proporcionado pela internet, logo não houve uma carência de documentação para o projeto, muito menos falta de meios de pesquisa.

3.1 Método de abordagem

O método de abordagem capaz de descrever o processo de criação dessa plataforma é o hipotético dedutivo. Trata-se basicamente de um método que impõe diversas verificações às hipóteses estabelecidas no início do projeto, verificando se elas se tornarão verdadeiras ou falsas. A ideia central desse processo metodológico consiste em levantar hipóteses, porém elas deverão ser minimamente viáveis para que a pesquisa não sofra atrasos no momento de responder os problemas em questão (POPPER, 1935).

3.2 Métodos de procedimentos

Esses tipos de método compreendem todas as formas pelas quais as pesquisas e estudos foram realizados. As pesquisas estavam voltadas para análises de documentos descritivos sobre o tema em questão. Para que fosse possível compreender o tema em sua completude optou-se por utilizar o método de estudo de caso. Através desse método foi possível dividir os conteúdos a serem estudados em áreas de maior relevância e de menor relevância. Com os temas divididos em ciências dermatológicas e visão computacional, cada um dos integrantes recebeu um tópico para ser estudado e, graças a esse procedimento, o projeto passou a se desenvolver rapidamente, chegando a uma fase em que pode ser avaliado pelos critérios impostos.

3.3 Técnicas de Pesquisa

A técnica utilizada ao longo da pesquisa foi a descritiva. Nessa metodologia é realizada uma análise extremamente focada no objeto de pesquisa, buscando relatar todos os detalhes possíveis para que não haja espaços para possíveis erros desnecessários. Esse modelo de pesquisa visa ao estudo aprofundado, ou seja, é necessário que o estudante possua um conhecimento prévio para que seja possível dar continuidade a esse método. É necessário ressaltar que esse conjunto de métodos não interfere na análise de dados. Optou-se por essa técnica de pesquisa, por conta da divisão do projeto em duas fases, sendo a primeira focada em aprendizagem e domínio do tema e a segunda voltada à análise de dados que seriam aplicados na ferramenta em desenvolvimento.

4. Materiais e Métodos

4.1 Pesquisa Bibliográfica

Para que seja possível compreender todo o caminho percorrido é necessário que sejam citados os recursos bibliográficos que serviram de referência para o projeto. Ao longo do projeto foi utilizada uma vasta gama de conteúdos que trabalham, como questão principal, com a disseminação do conhecimento de redes neurais, lógica de programação e desenvolvimento de IA, que são fontes constituídas por materiais já elaborados. Também foram utilizados materiais que citam a Rede Neural Convolucional para análise de doenças dermatológicas. Todas essas informações puderam ser adquiridas através de várias monografias, teses, dissertações, livros, artigos e matérias de websites de tecnologia. Os cursos adquiridos no decorrer do projeto foram de extrema valia para a compreensão de termos técnicos e utilização de algoritmos específicos nos âmbitos de desenvolvimento de Machine Learning e Deep Learning. Extensas pesquisas foram realizadas em busca de um algoritmo “open-source” capaz de realizar a classificação de imagens e fornecer um retorno decimal. Por meio dos documentos, monografias e livros utilizados foi possível compreender como a máquina trabalha com imagens, ou seja, como ela vê a imagem fornecida a ela. Com todos esses dados fornecidos pelas pesquisas realizadas, iniciou-se o processo de adequação do algoritmo para o objetivo almejado, que, felizmente, está apresentando bons resultados.

O conhecimento prático sobre problemas de saúde voltados à pele foi fornecido aos membros do projeto pela pesquisadora biomédica Bruna Amorim dos Santos, formada em biomedicina pela Universidade Nove de Julho (UNINOVE) desde 2019. Atualmente ela exerce sua profissão no centro de diagnósticos Sanitas, um laboratório de análises clínicas localizado em Guarulhos. O contato entre o profissional e o grupo foi mantido por meio de aplicativos de conversa, que possibilitam um contato mais próximo entre as pessoas que os utilizam. Ao longo dos diálogos, a biomédica buscou sanar dúvidas sobre estudos e aplicações laboratoriais da

dermatite Atópica. Artigos sobre estudos com especialização em Dermatologia também tiveram seus dados utilizados. Essas informações contribuíram para expandir o repertório intelectual de cada uma das fases do projeto: análise, preparação e aplicação.

4.2 Ferramentas

Como principal ferramenta utilizada para a construção e manipulação do algoritmo pode-se mencionar a linguagem de programação denominada de Python. Trata-se de uma linguagem Open-Source de alto nível, comumente aplicada em áreas de data science, machine learning, desenvolvimento web, desenvolvimento de aplicativos, automação de scripts e muito mais. Essa linguagem foi criada por Guido Van Rossum, em 1991. O fato de ser uma linguagem que se aproxima muito mais da nossa língua do que da linguagem de máquina é o que permite ao usuário trabalhar com mais facilidade e melhor compreensão do código. Sua escolha se deu por conta de diversos fatores, mas o principal foi a disponibilidade de recursos voltados ao aprendizado de máquina, pois provavelmente seria muito difícil encontrar o mesmo ambiente favorável em outra linguagem de programação.

Para a construção do modelo buscou-se focar na utilização do Keras, uma API (“conjunto de rotinas e padrões de programação para acesso a um aplicativo de software ou plataforma baseado na Web; a sigla **API** provém do termo inglês, Application Programming Interface, que em uma tradução literal significa, Interface de programação de aplicativos. Essa biblioteca surge com o intuito de reduzir o tempo de execução de redes neurais profundas, por cumprir esse requisito, profissionais da área a avaliam como uma ferramenta excepcional. Por ser uma aplicação escrita em python, torna-se muito fácil uni-la à área de ciência de dados abordada nas pesquisas mencionadas. Ela é capaz de rodar a partir de bibliotecas de código aberto como o TensorFlow, Microsoft Cognitive Toolkit, R, entre outras. Keras facilita o processo de experimentação de redes neurais convolucionais, visto que foi construída para trabalhar com o aprendizado profundo (deep learning). Em união ao Keras houve a necessidade de trabalhar com uma biblioteca focada na área da visão computacional. OpenCV foi a biblioteca escolhida por ser recorrente em trabalhos de segmentação e classificação de imagem. Sua amplitude em recursos

permite que até mesmo um leigo no assunto possa desenvolver aplicações voltadas para o reconhecimento facial ou de um objeto.

Bibliotecas como Scikit-learn, TensorFlow e Matplotlib permitiram que o modelo do algoritmo tomasse forma. Scikit-learn disponibilizou as funções de classificação de modelo, TensorFlow permitiu que a rede neural convolucional fosse treinada através de um loop de épocas que continham um número específico de seções. Já a biblioteca MatPlot foi a responsável pela criação de imagens ao longo do algoritmo, muitas das quais mostram a eficácia do modelo por meio de um gráfico de linhas que acompanha todos os resultados obtidos nos mais diferentes testes de acurácia.

4.3 Explicação do conceito de redes neurais

A Inteligência Artificial é uma programação que consegue exercer de forma complexa uma função humana. Não são funções “simples” como um cálculo, mas na verdade são funções complexas, como por exemplo: escolher se vai para a praia a partir da observação do clima.

Para a IA ser capaz de exercer uma função que seja relacionada às escolhas humanas, é necessária uma base de programação que separe as funções em etapas, uma rede neural. Conforme está descrito no Anexo C, cada camada possui vários “inputs” e “outputs”, que podem ter diferentes instruções e caminhos dependendo do objetivo. Quando um ou mais dos inputs da primeira camada identificarem uma função que os ativa, eles irão enviar um valor numérico, que será multiplicado por outro valor determinado no caminho até as entradas da segunda camada. Na segunda camada os valores que chegam são somados dentro de cada entrada e se os valores dentro de cada entrada corresponderem, os valores presentes prosseguirão para a próxima camada, sendo multiplicados por um valor específico novamente. A terceira etapa, que possui vários “outputs” conectados às entradas da etapa anterior, é onde a IA irá fazer escolhas. Se o

valor recebido em algum de seus “outputs” estiver dentro do período estimado, a IA executará a função escolhida para aquele “output”.

4.4 Processo de criação da Inteligência Artificial

Toda a estrutura dessa ferramenta é baseada no modelo de classificação elaborado pelo cientista de dados e pesquisador de aprendizado profundo, Toni Esteves (2020). O processo de criação do modelo está descrito de forma teórica e referenciada no site Medium, cujo objetivo é abrir espaços para novos desenvolvedores e possibilitar a visualização de diferentes projetos. Por meio das explicações e partes de código elaboradas pelo cientista foi possível construir o algoritmo utilizado pelo RecoSkin.

O primeiro passo é criar uma base de dados composta por imagens, gerenciadas em dois grupos distintos: o grupo de treino e o grupo de teste. O grupo de treino é usado para treinar e ensinar a rede neural, e o grupo de teste é usado para testar a eficiência da rede neural após o treino. Após a criação do “dataset” é necessário importá-lo de alguma forma para o algoritmo, a forma utilizada no projeto importa os dados de um repositório presente no google drive, através da função: `from google.colab import drive drive.mount('/content/drive')`. Mediante ao comando “pip” foram instaladas as seguintes bibliotecas: “OpenCV”, “Keras.Utilities”, “Scikit.Learn.Classification.Report” e “MatplotlibLibrary”.

O segundo passo é o pré-processamento: com ele a rede neural é capaz de trabalhar com dados padronizados, criando uma oportunidade para o sistema interpretar a imagem com maior facilidade, desempenho e velocidade. Existem muitas formas de realizar o pré-processamento de uma imagem, e a que será usada no projeto pertence à biblioteca “keras”, que é capaz de manipular a largura, o comprimento, dar zoom, alterar a escala, rotacionar etc. Essa forma de manipular a imagem se chama Image Data Generator, e encontra-se na repartição de processamento de imagens do Keras. Realizando essa aplicação, a base de dados sofre uma alteração de maneira indireta, visto que a partir desse momento cada uma das épocas de testes da rede neural receberá uma imagem transformada, com o intuito de capacitar a rede para o processo

de classificação. Encerrando esse processo de manipulação da imagem, foi estabelecida uma proporção fixa para as imagens, (150, 150, 3), dessa forma não haverá necessidade de alterar parâmetros futuros da rede neural convolucional (CNN).

O terceiro passo é a criação da rede neural convolucional, que irá realizar a função principal do programa: construir um modelo que fornecerá percentuais das épocas para serem utilizados na análise de uma pele infectada com a doença e distinguir se ela é uma dermatite Atópica ou não. A estrutura dessa CNN é baseada em um modelo sequencial fornecido pelo Keras, junto do apoio de funções, cujos objetivos são compreender as camadas dos dados. O modelo sequencial permite a inserção de camadas de uma rede neural em série, por meio desse método é possível utilizar o “output” da primeira camada como o “input” da segunda camada, e assim por diante. Com a construção do modelo finalizada, deve-se adicionar uma lista de instâncias de camada para a função construtora, a utilizada no algoritmo chama-se Dense, ela calcula uma função de ativação em relação aos dados de entrada e peso. Dessa maneira o modelo está finalizado, basta compilá-lo e adicionar um otimizador capaz de extrair a capacidade máxima de aprendizado de uma rede neural profunda.

A rede neural convolucional é capaz de dividir os dados em múltiplos “pedaços” de tamanhos diferentes, a mais comum consegue analisar a menor divisão de dados e depois os utiliza para analisar divisões maiores de dados, que são também utilizados para analisar outra divisão ainda maior de dados, e assim sucessivamente, até todos os processos acabarem. Este tipo de processamento pode ser muito útil para a análise de imagens, conforme o anexo - D, em que a imagem pré-processada é analisada por uma camada convolucional, que guarda dados para serem usados posteriormente e cria três novas “imagens” que possuem uma maior quantidade de informação, cada uma representando uma cor do espectro de cores, vermelho, verde e azul, também conhecido como RGB. A imagem criada pela camada convolucional é processada pelo pooling, que diminui a resolução da imagem, deixando-a menos massiva para a próxima análise com outra camada convolucional, de modo que os processos são repetidos até as últimas camadas. Estas são importantes para realizar o julgamento dos dados recebidos da imagem atual e comparar com a informação adquirida na fase de treino. Os dados que coincidirem irão

acrescentar valores a uma porcentagem que define a probabilidade de a imagem ser um barco, um gato, um cachorro ou um pássaro. Para a classificação de uma imagem é recomendado dividir a imagem em um “array”, por estar fracionado em diversas partes a probabilidade de predição aumenta e a perda de valores da imagem se reduz. Com a probabilidade definida, basta informá-la para o usuário.

5. Resultados e Discussão

Após as pesquisas para compormos o conjunto de materiais e métodos, buscamos compreender como funcionava a interpretação de uma imagem de modo computacional feita por uma Inteligência Artificial. Por esse motivo houve a necessidade de desenvolver um teste com uma abordagem bem simples, se comparada à do projeto idealizado pelo grupo RecoSkin AI: um Machine Learning, que diferencia cachorros de gatos. Analisando o problema em questão, criamos um dataset com valores variados, que representassem características de um gato ou de um cachorro, assim como uma imagem, estipulamos as “perguntas” que diferenciam um gato de um cachorro, fizemos a máquina compreender nosso raciocínio através de dados numéricos, estipulamos variáveis e escolhemos um método técnico de Machine Learning, o Naive Bayes, que, baseado no teorema de Bayes, calcula a probabilidade e classifica a amostra determinada. Para compreendermos melhor o funcionamento desse algoritmo, utilizaremos um exemplo: o algoritmo de Bayes, basicamente, nos informa sobre a probabilidade posterior, sendo que para chegar a esse resultado, é necessário considerar a probabilidade multiplicada pela probabilidade original da classe, dividida pelo preditor da possibilidade posterior, cf. Anexo A.

Supondo que temos uma tabela com dias de jogos e seus respectivos climas e queremos saber qual a chance de jogar em um dia chuvoso, para isso realizamos um cálculo evidenciando a probabilidade posterior, que seria igual à uma probabilidade (JOGOS NO PERÍODO CHUVOSO tal que REALMENTE ACONTECERAM) multiplicada por outra probabilidade total de jogos que realmente aconteceram (REALMENTE ACONTECERAM) dividida pelo meu preditor, que seria a razão de jogos que ocorreram e não ocorreram em período chuvoso em relação a todos os jogos. Dessa maneira chegaríamos à predição da porcentagem de o jogo ser realizado no período chuvoso, excelente para trabalhar com probabilidades. No caso, em nosso exemplo, buscamos apenas uma confirmação de que nossa imagem é realmente um gato e não um cachorro, estabelecendo as figuras dos animais que serão analisados pela máquina e estipulando as perguntas através de valores numéricos, sendo 0 para uma resposta falsa e 1 para uma verdadeira. Agora que já temos os animais, colocamos todos em uma única variável e chamamos uma função com base no nosso algoritmo de Bayes. No primeiro momento, a inteligência artificial

obviamente não acertaria todos os animais, afinal tínhamos uma base de dados pequena. Nosso algoritmo acertou duas imagens, sendo que havíamos colocado na variável duas imagens de gato e uma de cachorro, ou seja, o algoritmo supõe que todos eram gatos, nos revelando uma acurácia de aproximadamente 66%. Todos esses resultados foram obtidos graças a um exemplo de teste com um passo a passo anexado, fornecido pelo site de tecnologia Medium. Com esses resultados obtidos, descartamos parcialmente o algoritmo de Bayes, já que trabalharemos com uma grande quantidade de valores presentes em matrizes (imagens) e com um modelo convolucional; por isso, será improvável trabalhar apenas com esse algoritmo, já que nesse caso existem processadores e classificadores melhores, como o Keras.Processing e a biblioteca Numpy, capaz de suportar “arrays” multidimensionais.

Por conta da necessidade de realmente identificar a imagem no futuro, realizamos uma pesquisa por métodos de segmentação, buscamos algoritmos que cumprissem essa tarefa e fossem capazes de nos fornecer bons resultados, cf. Anexo B. Os algoritmos que foram capazes de cumprir tais requisitos foram o “Thresholding”, “Algorithm with Histograms” e um novo algoritmo que atualmente está sendo muito discutido por profissionais, chama-se “Phase Stretch Transform”. O Thresholding analisa o valor de cada pixel de uma imagem, com base no valor de intensidade presente em pontos específicos. Se a intensidade do pixel supera o valor devido, ele terá valor branco, caso ele seja menor, ele passa a ser preto, dando origem a uma imagem em preto e branco. Já os algoritmos baseados em histogramas apresentam uma eficiência melhor, se comparados com outros métodos de segmentação, pois eles realizam apenas uma passagem pelos pixels das imagens, classificando os pixels escuros e os claros de modo muito mais rápido. O algoritmo “Phase Stretch Transform” é capaz de analisar as bordas de uma imagem extraindo informações. Outro fator que nos cativou muito é a sua capacidade de melhorar imagens e realizar análises de texturas.

Diante dos resultados, podemos evidenciar algumas questões, que podem acabar sendo realizadas com maior facilidade. Nosso principal problema com a parte prática do projeto sempre esteve atrelado à qualidade da imagem que nos será fornecida, já que pretendemos futuramente expandir o nosso projeto para uma biblioteca de aplicativos para celular. Uma solução viável

pode ser utilizar o algoritmo “Phrase Stretch Transform”. Com relação ao nosso problema em questão geral, a solução será mais complexa, pois mudar o pensamento da maioria da população tem que ser feito com uma estratégia de marketing muito eficiente, com comerciais, vídeos exemplificando, apoio governamental, entre outros. Com toda certeza, caso optemos por uma expansão do projeto, será necessário um patrocínio para continuar as pesquisas e desenvolvimentos mais complexos para amplificar o número de usuários e pessoas que podem ser afetadas de maneira benéfica pelo aplicativo. A utilização da biblioteca “Anaconda” será essencialmente útil, já que poderemos gerenciar e implantar pacotes com mais facilidade, no entanto existem alguns pacotes que não são disponibilizados por esta biblioteca; por esse motivo trabalharemos com uma mescla de bibliotecas incluindo o “Keras” e principalmente o “TensorFlow”, um dos frameworks mais utilizados para a aprendizagem profunda. Após os nossos resultados de testes iniciais, observamos que temos um longo caminho para trilhar, por esse motivo visamos primeiramente treinar nossa Inteligência Artificial, criando uma rede neural bem estruturada, que dificilmente apresente erros. Posteriormente iremos ampliar o projeto para um desenvolvimento mobile, algo que demanda tempo de estudo, mas abrirá muitas portas para a aliança entre a medicina e a tecnologia.

Infelizmente os resultados obtidos nessa fase do projeto ainda não permitem um avanço para mais testes e aplicações em dispositivos móveis. O algoritmo atual foi capaz de classificar a dermatite Atópica com apenas 46% de acurácia, ou seja, trata-se de um valor muito baixo, já que sobram 54% de incerteza do modelo. A conclusão que se pode tirar com o resultado é que o código de aprendizado profundo necessita de aperfeiçoamento e neste momento não é capaz de solucionar uma situação real, em que o profissional necessita de uma ferramenta de pré-diagnóstico. O resultado obtido está totalmente atrelado à escassez de imagens de dermatite Atópica presentes na web, muitas das imagens não são da dermatite em si, logo fica muito difícil de treinar um modelo com imagens que não fornecem o que realmente se deseja abordar ao longo do algoritmo. Ainda que estivesse presente no modelo uma função capaz de gerar mais imagens com pequenos detalhes, essa quantidade não pode ser considerada suficiente para o modelo de

treinamento, também é necessário ressaltar que caso não fosse utilizado um otimizador em nosso modelo sequencial, o valor da acurácia seria ainda menor. Adam, é o nome do algoritmo de otimização utilizado, ele é capaz de aplicar cálculos nas taxas de aprendizagem individuais para cada um dos parâmetros. Todos os recursos à disposição do grupo foram utilizados para melhorar a CNN, no entanto não surtiram o efeito imaginado, certamente em uma das fases futuras haverá a necessidade de adquirir um repositório de imagens maior, que contenha tanto imagens de peles saudáveis quanto de peles devastadas pela dermatite Atópica.

6. Considerações Finais

Testes estão sendo feitos com a IA através de imagens, para que antes de tomar qualquer rumo, possamos definir como será o aprendizado dela. O dataset de imagens utilizado no projeto está sendo montado por nós mesmos, pois não existem repositórios de pesquisas oficiais para essa variante da dermatite disponibilizada na web. A pequena variação de fotos é um problema para nós, visto que o algoritmo passa a ter um processo de aprendizado mais lento. Temos 22 imagens contendo a dermatite atópica e 21 imagens que apresentam uma pele supostamente sadia. Testes estão sendo executados para verificar a acurácia do algoritmo, pois como a doença apresenta diversas características visuais, podem ocorrer erros na otimização da visão computacional. Para que essa situação não venha ocorrer em grande escala, temos que tornar a IA capaz de classificar a imagem entre uma pele saudável e uma contaminada. Futuramente, o que vai nos ajudar a ter uma maior eficiência nos resultados será implementar uma nova classe, responsável por armazenar imagens de doenças variadas, que possuem uma certa semelhança com a dermatite Atópica. Para que o algoritmo saiba diferenciar a dermatite de outro tipo de doença, a nossa IA terá que ser capaz de classificar a imagem em 3 resultados: pele saudável, pele contaminada com dermatite Atópica e pele contaminada com outra doença.

Dentro de um determinado tempo, a eficiência do sistema vai progressivamente aumentar, tornando-o suficientemente útil para o uso de pré-diagnósticos. Relembrando que o diagnóstico ainda será feito por um profissional na área da medicina dermatológica, a IA apenas será usada como um auxílio para diminuir possíveis gastos de dinheiro, tempo e principalmente visando o bem-estar do paciente. Acreditamos que o RecoSkin será uma inovação para o ramo da saúde e para o meio social. Com relação a projetos futuros, visamos transformar essa plataforma em um software disponível para aparelhos mobile, buscando favorecer grande parte da população brasileira usuária de “smartphones”. O aplicativo criado será de caráter totalmente seguro, contendo login, senha e informações do paciente. A partir do resultado obtido por meio do pré-diagnóstico realizado pelo profissional, o “app” fornecerá uma mensagem automática, em que estará localizada uma porcentagem que definirá o quanto aquela infecção se assemelha a uma dermatite Atópica. De acordo com os dados fornecidos pelo aplicativo, o dermatologista ou

alergista irá indicar os exames necessários para ter certeza se o paciente está ou não com a doença e, posteriormente, diagnosticar o paciente, portanto o tempo de análise será encurtado, pois o médico irá se basear na porcentagem que o aplicativo o conceder.

7. Referência

BENITO, Emilio de. **A doença (ainda sem cura) que obriga a pessoa a se coçar o tempo todo.** Disponível em: https://brasil.elpais.com/brasil/2017/11/20/ciencia/1511196768_287794.html#:~:text=/> Acesso em: 05 jun. 2020.

CESTARI, Silmara. **Dermatite atópica diagnóstico e tratamento.** Disponível em: <http://www.luzimarteixeira.com.br/wp-content/uploads/2010/05/dermatite-atopica-diagnostico.pdf> Acesso em: 27 ago. 2021.

Gaea. **Afinal, o que é deep learning.** Disponível em: <https://gaea.com.br/afinal-o-que-e-deep-learning/> Acesso em: 27 ago. 2021.

CHAGAS, Edgar Thiago De Oliveira. **Deep Learning e suas aplicações na atualidade.** Disponível em: https://www.nucleodoconhecimento.com.br/administracao/deep-learning#_ftn1 Acesso em: 27 ago. 2021.

ACHÔA, Yara. **MARCAS ALÉM DA PELE** Disponível em: <https://www.uol.com.br/vivabem/reportagens-especiais/dermatite-atopica-os-impactos-fisicos-e-mentais-da-doenca-de-pele/#cover> Acesso em: 26 ago. 2021.

LOURENÇO, Felipe. **Dr. Google: como lidar com pacientes que confiam demais no diagnóstico da internet** Disponível em: <https://blog.iclinic.com.br/dr-google-como-lidar-com-pacientes-que-confiam-demais-no-diagnostico-da-internet/> Acesso em: 26 ago. 2021.

Editorial Team. **Association Between Atopic Dermatitis and Suicidality**. Disponível em: <https://atopicdermatitis.net/clinical/eczema-mental-health-suicidality?utm_source=weekly&utm_medium=email&utm_campaign=ff7d0db7-d3d7-4524-821a-30885bf24a97&utm_confid=sovj7m9z3&aGVhbHRoIHVuaW9uIGJsYWg=8e258a283071ec72caf27a214fb89135e09f3b0c56f8c97cf8c576f9cc975e02/> Acesso em: 05 jun. 2020.

GÓIS, Cerqueira Alexis. **Já é possível detectar câncer de pele com uma câmera de celular**. Disponível em: <<https://www.tecmundo.com.br/ciencia/148433-possivel-detectar-cancer-pele-camera-celular.htm>> Acesso em: 23 set. 2020.

Redação Bonde com Assessoria de imprensa. **Diagnóstico médico na internet: 7 em 10 pacientes recebem informações falsas**. Disponível em: <<https://www.bonde.com.br/saude/noticias/diagnostico-medico-na-internet-7-em-10-pacientes-recebem-informacoes-falsas-499982.html>> Acesso em: 23 set. 2020.

VINICIUS. **Modelo sequencial do Keras**. Disponível em: <<https://www.monolitonimbus.com.br/modelo-sequencial-do-keras/#:~:text=O%20Keras%20é%20uma%20API,como%20TensorFlow%2C%20CNTK%20ou%20Theano/>> Acesso em: 18 jun. 2021.

COLL, Liana. **Pesquisa identifica câncer de pele com câmeras de celular**. Disponível em: <<https://valedosulicio.com.br/noticias/noticia/pesquisa-identifica-cancer-de-pele-com-cameras-de-celular>> Acesso em: 21 jun. 2021.

MERKER, Júlia. **Software pode diagnosticar câncer de pele com precisão de 86%**. Disponível em: <<https://www.unicamp.br/unicamp/ju/noticias/2020/01/20/software-pode-diagnosticar-cancer-de-pele-com-precisao-de-86>> Acesso em: 21 jun. 2021.

R7. Mais de 40% dos brasileiros fazem autodiagnóstico pela internet. Disponível em: <<https://noticias.r7.com/saude/mais-de-40-dos-brasileiros-fazem-autodiagnostico-pela-internet-09082018>> Acesso em: 21 jun. 2021.

NEVES, Úrsula. **Sete em cada dez pacientes recebem informações falsas de diagnósticos na internet.** Disponível em: <<https://pebmed.com.br/sete-em-cada-dez-pacientes-recebem-informacoes-falsas-de-diagnosticos-na-internet/>> Acesso em: 21 jun. 2021.

ESTEVES, Toni. **Agrupando conceitos e classificando imagens com Deep Learning.** Disponível em: <<https://estevestoni.medium.com/agrupando-conceitos-e-classificando-imagens-com-deep-learning-5b2674f99539>> Acesso em: 21 jun. 2021.

POPPER, Karl. **A Lógica da Pesquisa Científica.** São Paulo: Cultrix; 2013.

8. Anexos

Imagens Anexadas:

ANEXO-A

Sabendo que : $P(A \cap B) = P(B|A) \cdot P(A)$

Podemos fazer :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

E assim obtemos o Teorema de Bayes :

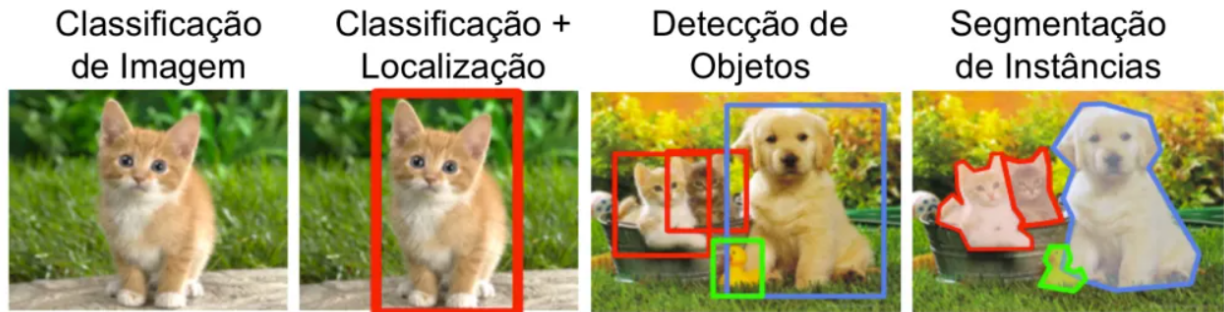
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Teoria Básica de probabilidade por trás do Teorema de Bayes.

Disponível em: <https://miro.medium.com/max/1544/1*n7WYOozwdieetW6jEpxCbA.png> Acesso em: 31 Jan.

2021

ANEXO-B



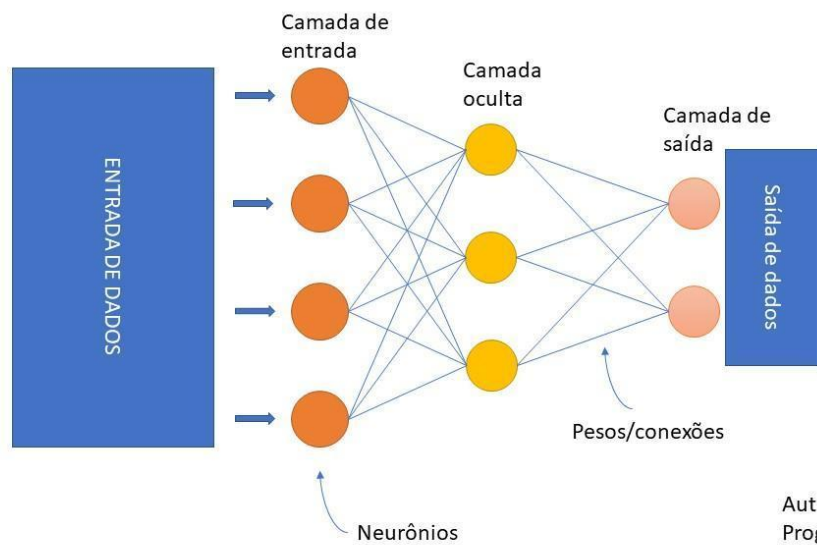
**Representação de um algoritmo segmentando os objetos necessários de uma imagem
(Atividades na visão computacional).**

Disponível em: <<http://datascienceacademy.com.br/blog/segmentacao-de-imagens-medicas-com-deep-learning/>>

Acesso em dia: 29 Jan. 2021

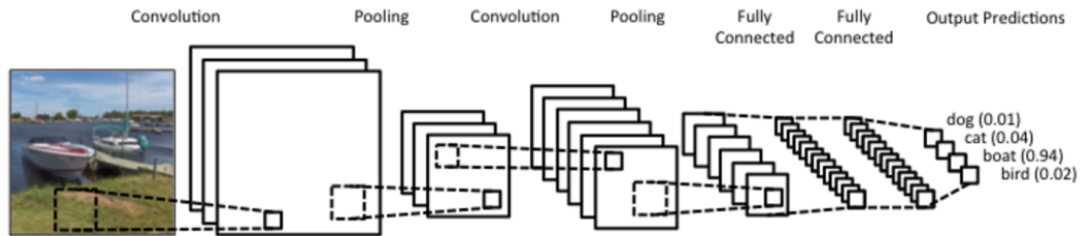
ANEXO-C

Redes Neurais



Autor: Joseph de Moura Freitas
Programa: PowerPoint
Data de Realização: 31/01/2021

ANEXO-D



Resolvendo CAPTCHAs com Redes Neurais Convolucionais

Disponível em: <<https://matheusfacure.github.io/img/tutorial/convnet.png>> Acesso em: 05 Jun. 2021

ANEXO-E

**Algoritmo desenvolvido pela equipe do RecoSkin, com base nas instruções do pesquisador
Toni Esteves.**

Disponível em: <https://colab.research.google.com/drive/1e_w9sRVdRpTmMJtNeW-fbR-qLk3wPbp7?usp=sharing>

Acesso em: 20 Jun. 2021