

Pigmented Dermatological Lesions Classification Using Convolutional Neural Networks Ensemble Mediated by Multilayer Perceptron Network

J. Lima, L. Araújo, F. Silva, and C. Figueiredo

Abstract—Skin cancer has the highest occurrence rate compared to other types of cancer. This paper presents the development of a Deep Learning model, trained from the Skin Cancer MNIST database (HAM10000). It is able to perform Classification of Pigmented Dermatological Lesions using Convolutional Neural Networks techniques by proposing an ensemble with Multilayer Perceptron Neural Networks. In order to evaluate the Convolutional Networks, the metrics Accuracy, Precision, Revocation and F1-Score were taken into consideration. The ensemble implementation was based on a Grid Search with Cross Validation and evaluated according to Accuracy. The obtained results show the relevance of the research and the consolidation of the techniques used in the development of Artificial Intelligence solutions applied to clinical images analysis. Accuracy and Recall reached 0.93 and Precision and F1-Score 0.92, which is superior performance to specialists and related researches.

Index Terms—Artificial Intelligence, Clinical Images Analysis, Deep Learning, Pigmented Dermatological Lesions.

I. INTRODUÇÃO

A EVOLUÇÃO da medicina demanda um alto desenvolvimento tecnológico, visando melhorar técnicas de tratamento de doenças e, principalmente, aumentar o índice de diagnósticos precoces, especialmente, de doenças que podem levar a óbito. A dermatologia, especificamente, possui inúmeros desafios no quesito diagnóstico precoce correto de lesões pigmentadas, uma vez que muitas delas são facilmente confundíveis ao olho humano.

O câncer de pele, por exemplo, é uma lesão pigmentada que corresponde a quase um terço dos casos de câncer no Brasil [14]. O tipo de câncer de pele mais danoso à saúde é o melanoma. Quando o melanoma é diagnosticado precocemente, este tem chance de cura de até 95%, enquanto, ao ser diagnosticado de forma tardia, pode haver redução para até 35% de chance de sobrevivência [15].

Este cenário pode parecer ainda pior quando analisada a taxa de acerto médico no diagnóstico do melanoma, que, em média, é menor que 70% [1]. Isso indica uma sublime oportunidade de aplicação de Inteligência Artificial na tentativa de aumentar o acerto no diagnóstico de lesões pigmentadas, trazendo assim benefícios à sociedade.

Em decorrência do contexto apresentado, é perceptível a necessidade de um método para, não somente identificar a existência de câncer de pele, mas classificar quanto a todas as categorias importantes de lesões pigmentadas. Este trabalho apresenta uma solução à problemática identificada utilizando técnicas de Aprendizado de Máquina Supervisionado.

A computação tem sido fundamental para o avanço da medicina, principalmente com o recente advento do Aprendizado de Máquina [2]. Em muitos destes casos, o aprendizado de máquina é aplicado a tarefas de processamento e análise de imagens. Neste sentido, técnicas de *Deep Learning* têm se apresentado eficientes por serem baseadas no sistema de percepção visual humano e por serem implementadas várias camadas de neurônios artificiais, aumentando assim a capacidade de aproximação de funções [3].

Este trabalho apresentará a utilização de Redes Neurais Convolucionais (CNN), uma abordagem de *Deep Learning*, para realização do diagnóstico automático de lesões pigmentadas. Na Seção II serão apresentados trabalhos relacionados ao contexto deste, em seguida, na Seção III, serão informados os materiais e métodos utilizados seguido da apresentação e análise dos resultados obtidos, na Seção IV. Por fim, na Seção V, serão expostas as considerações finais e perspectivas de trabalhos futuros.

II. TRABALHOS RELACIONADOS

Processamento e análise inteligente de imagens médicas é um tema que tem recebido bastante ênfase, principalmente, com uso de *Deep Learning* [3]. A principal aplicação para esta abordagem é obter diagnóstico através de classificação automática. A produção de trabalhos relacionados à Classificação de Lesões Pigmentadas tem sido muito frequente, trazendo abordagens semelhantes às apresentadas neste, empregando predominantemente Redes Neurais Convolucionais.

Uma abordagem comum, na tentativa de aumentar a eficiência das soluções, é reduzir a quantidade de doenças abrangidas pelo modelo, como acontece em [4] e [5], que visam

J. P. O. Lima, Universidade do Estado do Amazonas, Manaus, Amazonas, Brazil, jpdol.eng16@uea.edu.br.

L. C. S. Araújo, Universidade do Estado do Amazonas, Manaus, Amazonas, Brazil, lcsdafl.eng16@uea.edu.br.

F. S. Silva, Universidade do Estado do Amazonas, Manaus, Amazonas, Brazil, fssilva@uea.edu.br.

C. M. S. Figueiredo, Universidade do Estado do Amazonas, Manaus, Amazonas, Brazil, cfigueiredo@uea.edu.br.

a detecção de ocorrências de melanoma, e em [6] que apresenta um detector de carcinoma basocelular. Essa abordagem, apesar de altamente relevante, não consegue gerar um diagnóstico específico quanto a natureza da lesão, restringindo-se apenas à detecção da existência ou não de uma determinada lesão.

Por outro lado, [1] e [7] apresentam modelos baseados em arquiteturas tradicionais de Redes Neurais Convolucionais, fazendo-se uso das camadas de convolução para extração de características e utilizando as camadas *fully-connected* para realizar a classificação final. Este trabalho faz uso de CNNs, no entanto, se diferencia por usar uma arquitetura não trivial, a InceptionV3 [17].

Na Abordagem não trivial proposta e experimentada neste trabalho, são utilizadas várias CNNs para realizar a análise de imagens e, por fim, obter um resultado a partir da saída de todas as CNNs por meio de uma rede Perceptron de Multicamadas (MLP). Trabalhos que apresentam a utilização de *ensemble*, como acontece em [8], possuem sistema de votação comum e não fazem uso de CNNs específicas de determinada classe para realização de classificação binária, tampouco utilizam técnicas de busca para definição dos hiperparâmetros do sistema usado para o *ensemble*.

Os principais pontos que diferem este trabalho de outros relacionados são: a implementação de um *ensemble* utilizando rede MLP para analisar saídas advindas de outras redes; a técnica de escolha de classes específicas para implementação de CNNs binárias; e a técnica para escolha da arquitetura da MLP, baseada em busca em grade com validação cruzada [9].

Assim, nota-se, neste trabalho, a existência de abordagens interessantes não encontradas em artigos científicos relacionados. A evolução nos resultados obtidos, que serão expostos na Seção 4, em comparação com trabalhos que fazem uso da mesma base de dados, é justificada pelas abordagens coerentes e incomuns utilizadas.

III. MATERIAIS E MÉTODOS

Esta seção descreve os materiais e métodos utilizados para o desenvolvimento e análise dos experimentos conduzidos. As subseções abrangem os seguintes tópicos: Dados Experimentais, que apresenta e analisa a base de dados utilizada; Descrição da Tarefa de Aprendizado de Máquina, que expõe como o Aprendizado de Máquina será utilizado no contexto deste trabalho em face à base de dados apresentada; Proposição de Modelos, que indica os tipos de modelos utilizados bem como a sua construção e emprego.

A. Dados Experimentais

O conjunto de dados de doenças dermatológicas utilizado para o desenvolvimento deste trabalho é denominado *Skin Cancer MNIST* (HAM10000), que contém 10015 imagens dermatoscópicas de diferentes populações [16]. Além das imagens, a base de dados conta com os seguintes atributos preditores: sexo e idade do paciente e local no corpo de onde a imagem fora capturada.

Os exemplos contidos na base representam todas as categorias diagnósticas importantes no domínio das lesões pigmentadas divididas em 7 classes: queratoses actínicas e carcinoma intraepitelial (Akiec), carcinoma basocelular (Bcc), lesões semelhantes a ceratose benigna (Bkl), dermatofibroma

(Df), melanoma (Mel), nevos melanocíticos (Nv) e lesões vasculares (Vasc), como ilustra a Fig. 1.

É possível notar, a partir da Fig. 1, que as classes possuem algumas semelhanças entre si, aumentando as chances de erro no diagnóstico médico. Existe, no entanto, características atenuantes para a distinção entre elas. Um especialista, ao avaliar um melanoma, por exemplo, leva em consideração os seguintes fatores: assimetria, bordas, cor variável, diâmetro e evolução da lesão [18]. Entre esses, o único que não pode ser identificado através de uma foto é a evolução, ou seja, pela natureza de uma rede neural convolucional, de buscar padrões visuais como cantos, arredondamentos, contraste de cores, entre outros [3], é provável que esta abordagem seja eficiente na classificação de lesões dermatológicas pigmentadas.

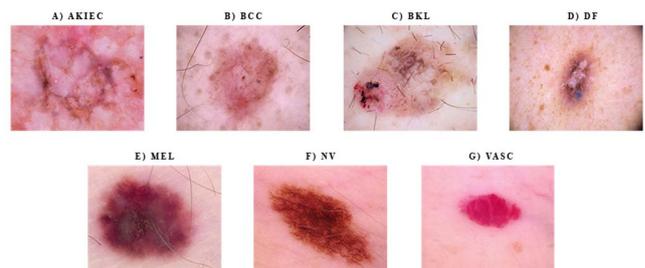


Fig. 1. Amostras de Imagens Dermatoscópicas do *Skin Cancer MNIST*.

Conforme mostrado na Fig. 2, a quantidade de dados estava extremamente heterogênea. Por conta disso, foi feito um balanceamento dos dados, a fim de evitar *overfitting* e assim obter um melhor treinamento, diminuindo consideravelmente a quantidade de dados disponíveis. Primeiramente, foram removidos todos os exemplos que continham dados faltantes, obtendo-se, assim, 102 exemplos para a classe com o menor número de exemplos, a classe Df. Em seguida, foram excluídos, aleatoriamente, exemplos das demais classes até que, por fim, restassem 714 exemplos igualmente distribuídos entre as 7 classes.

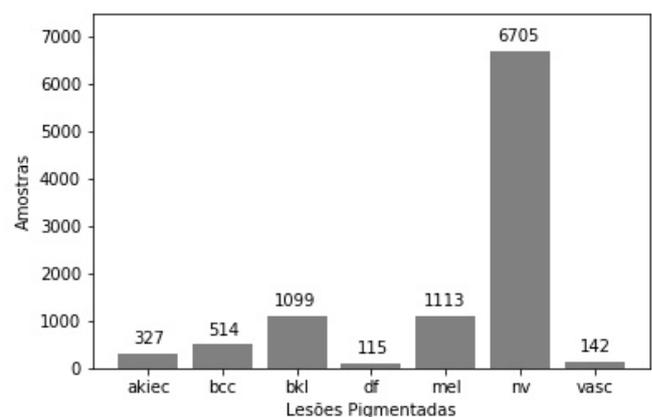


Fig. 2. Distribuição de frequência para as doenças na base de dados HAM10000.

B. Descrição da Tarefa de Aprendizado de Máquina

A base de dados será utilizada para realização de um aprendizado de máquina supervisionado com a tarefa de classificação multiclasse. Tal tarefa será executada em duas etapas: fase de treino e fase de validação.

Na fase de treino, um determinado conjunto de exemplos contido na base de dados será submetido a uma CNN [19] a fim de que ocorram os ajustes de seus parâmetros e, portanto, a geração do primeiro modelo, ou seja, o seu aprendizado inicial. Para garantir o poder de generalização do modelo, este passará pela fase de validação, que consiste em submeter atributos preditores de exemplos não contidos no conjunto de treino e avaliar seu desempenho com relação ao atributo alvo desejado para cada exemplo.

Para a dinâmica de treino e validação das redes implementadas foi realizado, para cada uma delas, uma validação cruzada com 10 *splits* de teste, isto é, a base de dados usada para cada rede foi dividida em 10 partes, onde, a cada iteração, uma das partes é usada como conjunto de validação do modelo, enquanto as outras são usadas como conjunto de treino. Isso se repete 10 vezes, até que todos os subconjuntos tenham sido utilizados como conjunto de teste.

C. Proposição de Modelos

Primeiramente, a abordagem adotada foi a de realizar o treinamento, por completo, de uma rede neural convolucional com a mesma arquitetura da InceptionV3 [17] para a classificação quanto aos 7 grupos de diagnósticos.

A fim de melhorar os resultados obtidos após a validação do primeiro modelo, decidiu-se implementar outras 7 CNNs, uma para cada uma das 7 classes da base de dados. As novas redes também empregam a arquitetura InceptionV3 e foram treinadas para a tarefa de classificação binária, isto é, cada rede obtém a probabilidade de um caso pertencer ou não a uma classe específica. Dessa forma, cada rede binária treinada é capaz de determinar se uma imagem corresponde ou não a uma das classes da base de dados. Para o treino de cada rede binária utilizou-se a base inicial, desbalanceada, apenas com os exemplos com dados faltantes removidos. Isso foi realizado para que cada CNN utilizasse o máximo de imagens disponíveis de uma classe, aumentando assim o poder de generalização da rede e, no entanto, mantendo os dados de treino de cada rede balanceados.

Em seguida, foi desenvolvido um *ensemble* em que as probabilidades de saída de cada classe para cada uma das redes binárias treinadas anteriormente foram dadas como entrada para uma MLP, além dos dados fornecidos pelo conjunto de dados para cada imagem: idade, sexo e local da lesão. À rede Perceptron de multicamadas cabe a tarefa de mediar as probabilidades de saída das redes anteriores e decidir a qual classe o caso de entrada pertence. A arquitetura do *ensemble* é ilustrada pela Fig. 3.

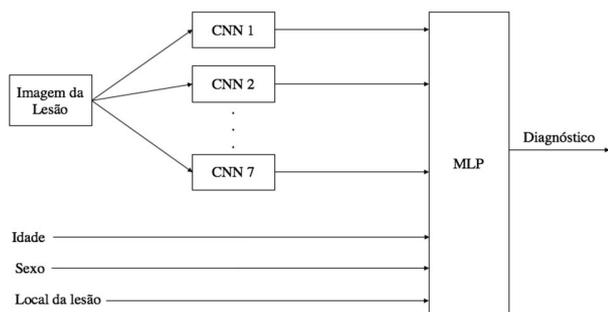


Fig. 3. Arquitetura do *ensemble* das 7 CNNs binárias.

Por fim, foi desenvolvido um segundo *ensemble*, o qual, ao contrário do primeiro, recebe as probabilidades de saída da CNN multiclasse, as probabilidades de saída das 3 CNNs binárias que, após a fase de validação da CNN multiclasse, obtiveram as piores métricas, além dos atributos: sexo; idade e local da lesão. A arquitetura do Segundo *ensemble* é ilustrada pela Fig. 4.

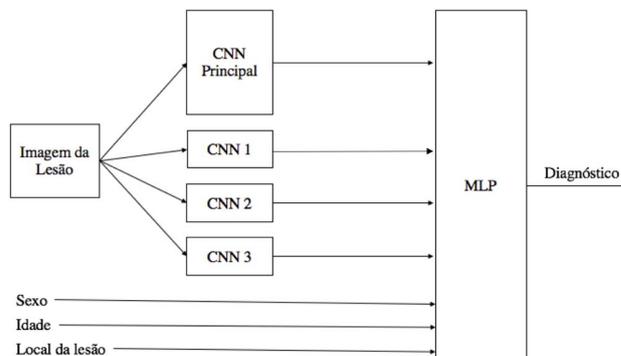


Fig. 4. Arquitetura do *ensemble* da CNN principal e das 3 CNNs com piores métricas.

As arquiteturas das redes Perceptron de multicamadas foram estabelecidas através de uma busca em grade, conforme descrito a seguir, para otimizar a busca por seus melhores parâmetros e hiperparâmetros. A busca em grade consiste no treino de todas as possíveis combinações de parâmetros e hiperparâmetros previamente estabelecidos [9]. Para isso, uma base de dados supervisionada secundária foi criada, contendo os dados de entrada e as saídas desejadas das redes MLP. A métrica final utilizada para descrever a melhor rede MLP foi a acurácia média dos 10 *splits* de teste da validação cruzada.

Os parâmetros e hiperparâmetros pré-definidos para a busca em grade foram: número de camadas ocultas, número de neurônios por camada oculta, tamanho de *batch*, taxa de aprendizagem, função de ativação e *solver*. O número de camadas ocultas foi fixado em 2, visto que, com duas camadas ocultas, uma rede neural pode implementar qualquer função [10]. O número de neurônios ocultos foi obtido pela Regra da Pirâmide Geométrica (RPG) [11], dada por:

$$Nh = \lceil \alpha \cdot \sqrt{Ni \cdot No} \rceil \quad (1)$$

em que Nh é o número total de neurônios ocultos, Ni o número de neurônios na camada de entrada, que corresponde ao número de atributos preditores, No o número de neurônio na camada de saída, que corresponde ao número de classes, e α uma constante arbitrária. Foi levado em consideração todos as combinações, de duas camadas, formadas pelo total de neurônios ocultos obtidos pela RPG quando $\alpha = 1; 1,5; 3$ e $3,8$. O tamanho de *batch* foi de 16; 24 e 32. As taxas de aprendizagem determinadas foram 0,0001 e 0,0005. Enquanto os valores de *solver* foram '*adam*', '*sgd*' e '*lbfgs*' e funções de ativação *relu*, *sigmoidal*, *tangente hiperbólica* e *identidade*.

IV. RESULTADOS E DISCUSSÕES

Nesta seção são descritos os resultados obtidos a partir da metodologia proposta em classificar automaticamente lesões

pigmentadas. As subseções abrangem os seguintes tópicos: Desempenho da CNN multiclasse, que apresenta e analisa os resultados da validação cruzada da classificação multiclasse de uma única rede neural convolucional; desempenho das CNNs binárias, que analisa os resultados obtidos para as redes treinadas individualmente para reconhecimento de um tipo de lesão; desempenho do Ensemble das 7 CNNs binárias, que analisa os resultados obtidos pelo *ensemble* realizado por uma rede MLP com as 7 CNNs binárias, além dos resultados obtidos pela busca em grade que definiu os parâmetros e hiperparâmetros da rede MLP; e desempenho do *Ensemble* com multiclasse e redes binárias das classes com as 3 piores métricas, que analisa os resultados obtidos pelo *ensemble* realizado por uma rede MLP com as 3 CNNs binárias que possuíam piores métricas na classificação binária, além dos resultados obtidos pela busca em grade que definiu os parâmetros e hiperparâmetros da rede MLP.

A. Desempenho da CNN multiclasse

O primeiro modelo treinado foi a CNN com a tarefa de classificação multiclasse. Foram utilizadas 4 métricas de desempenho para avaliar os resultados obtidos pela validação cruzada do modelo. São elas as médias aritméticas da: Acurácia; Precisão; Revocação e F1-Score para os 10 *splits* de teste. A Acurácia obtida pela validação foi de 0,71. Os demais resultados podem ser ilustrados pela Tabela I.

TABELA I
RESULTADOS OBTIDOS PELA CNN MULTICLASSE PARA TODAS AS 7 CLASSES ANTES DO ENSEMBLE

Classe	Precisão	Revocação	F1-Score
Akiek	0,51	0,44	0,47
Bcc	0,69	0,63	0,66
Bkl	0,47	0,75	0,58
Df	0,83	0,96	0,89
Mel	0,71	0,53	0,61
Nv	0,80	0,71	0,74
Vasc	0,98	0,95	0,96
Total	0,71	0,71	0,71

De acordo com os resultados obtidos, a rede tem desempenho geral razoável, com uma Acurácia e F1-Score total de 0,71. No entanto, no cenário da problemática, é necessário que haja uma taxa de acerto maior. A classe ‘Mel’, por exemplo, que representa o diagnóstico de melanoma, o tipo de câncer de pele mais nocivo, obteve F1-Score de 61%. Outras duas classes obtiveram F1-Score abaixo de 60%, ‘Akiec’ e ‘Bkl’, cujos desempenhos foram de 47% e 58% respectivamente. Assim, identifica-se que estes resultados não foram suficientes, sobretudo no contexto médico, onde a taxa de acerto deve ser bastante elevada. A solução encontrada para isso foi treinar outras CNNs, desta vez binárias, para classificar o diagnóstico específico das classes.

B. Desempenho das CNNs Binárias

Com objetivo de melhorar as métricas de desempenho, decidiu-se dividir as tarefas de classificação entre redes binárias para cada classe. As redes binárias são responsáveis pela verificação da existência da lesão pigmentada nos dados de

entrada. Ou seja, cada rede binária é capaz de dar o diagnóstico de ‘positivo’ ou ‘negativo’ da lesão que esta corresponde.

Para o treinamento e validação das 7 redes, foram selecionados todos os casos de saída desejada positiva contidos na base de dados, isto é, todos os exemplos da classe que a rede representa. Para compor o conjunto de saída desejada negativa, foram selecionados, de forma aleatória e proporcional, exemplos das demais classes que não correspondem à rede binária em questão, de tal forma que fosse obtido, para cada rede binária, um subconjunto balanceado de dados para treino e validação.

Os resultados médios obtidos pela validação cruzada dos modelos de classificação binária são apresentados na Tabela II. Em comparação aos resultados obtidos na primeira rede, é perceptível a melhoria significativa de desempenho, principalmente para as classes BCC e Mel, onde os valores de F1-Score aumentaram em 13% e 19% respectivamente.

TABELA II
RESULTADOS OBTIDOS PELAS CNNs DE CLASSIFICAÇÃO BINÁRIA

Classe	Precisão	Revocação	F1-Score
Akiek	0,87	0,47	0,61
Bcc	0,81	0,78	0,79
Bkl	0,62	0,88	0,73
Df	0,89	0,95	0,92
Mel	0,76	0,84	0,80
Nv	0,83	0,83	0,83
Vasc	0,98	0,97	0,97
Total	0,82	0,82	0,82

C. Desempenho do Ensemble das 7 CNNs Binárias

Agora com todas as redes convolucionais treinadas, foi realizada uma busca em grade com validação cruzada para definir os parâmetros e hiperparâmetros da rede MLP, responsável pelo *ensemble*. As arquiteturas das 5 melhores redes resultantes da busca em grade são apresentadas na Tabela III. Ao fim da busca, foi selecionada a rede com maior acurácia média em relação aos 10 *splits* de teste realizados. A rede encontrada possui os seguintes parâmetros: 2 camadas ocultas com 10 e 28 neurônios respectivamente, tamanho de *batch* de 16, função identidade como função de ativação, taxa de aprendizado constante de 0,0005 e o *solver* ‘adam’.

TABELA III
MELHORES ARQUITETURAS DE REDES OBTIDAS COMO RESULTADO DA PRIMEIRA BUSCA EM GRADE

Camadas Ocultas	Ativação	Solver	Taxa de Aprend.	Tam. <i>batch</i>	Acurácia
(10, 28)	Ident.	adam	0,0005	16	0,78
(26, 12)	tanh	adam	0,0005	24	0,77
(26, 12)	relu	lbfgs	0,0005	16	0,76
(25, 13)	tanh	adam	0,0005	16	0,76
(10, 28)	tanh	adam	0,0005	24	0,75

Após a obtenção da melhor rede MLP para o *ensemble*, como resultado da busca em grade, foram analisadas as métricas Acurácia, Precisão, Revocação e F1-Score para a abordagem

em questão. A acurácia obtida foi de 0,80. Os demais resultados são informados pela Tabela IV.

TABELA IV

RESULTADOS OBTIDOS PELA REDE PERCEPTRON DE MULTICAMADAS PARA TODAS AS 7 CLASSES APÓS O ENSEMBLE ENTRE TODAS AS REDES BINÁRIAS

Classe	Precisão	Revocação	F1-Score
Akiek	0,89	0,57	0,69
Bcc	0,87	0,98	0,92
Bkl	0,68	0,65	0,66
Df	0,97	0,92	0,94
Mel	0,71	0,68	0,69
Nv	0,80	0,87	0,83
Vasc	0,97	0,82	0,89
Total	0,84	0,78	0,81

D. Desempenho do Ensemble com Multiclasse e Redes Binárias das Classes com as 3 Piores Métricas

Uma outra abordagem utilizada foi de escolher as classes com pior desempenho na classificação multiclasse para implementar as redes de classificação binária e realizar um *ensemble* dessas redes juntamente com a CNN multiclasse, além de receber como entrada os dados de sexo e idade do paciente e local da lesão. O critério de escolha das piores classes foi a métrica Revocação, visto que a revocação é uma métrica crucial no contexto de diagnósticos médicos, por indicar, quando realmente um diagnóstico pertence a uma determinada classe, o quão frequente este é classificado corretamente. Portanto, ficou definido que as classes com Revocação inferior à Revocação Média da classificação multiclasse teriam suas redes neurais específicas integrantes do comitê de CNNs desta abordagem. Como a Revocação Média obtida pela rede foi de 0,71, as classes escolhidas foram ‘Akiek’, ‘Bcc’ e ‘Mel’, com, respectivamente, 0,44; 0,63 e 0,53 de Revocação.

Novamente, foi realizada uma busca em grade com validação cruzada para a obtenção dos melhores parâmetros e hiperparâmetros da MLP responsável pelo *ensemble*. Ao fim da busca, foi selecionada a rede com maior acurácia média em relação aos 10 *splits* de teste realizados. A Tabela V apresenta as melhores redes encontradas a partir da busca em grade. A melhor rede encontrada possui os seguintes parâmetros: 2 camadas ocultas com 12 e 28 neurônios respectivamente, tamanho de *batch* de 16, função identidade como função de ativação, taxa de aprendizado constante de 0,0005 e o *solver* ‘adam’.

TABELA V

MELHORES ARQUITETURAS DE REDES OBTIDAS COMO RESULTADO DA SEGUNDA BUSCA EM GRADE

Camadas Ocultas	Ativação	<i>Solver</i>	Taxa de Aprend.	Tam. <i>batch</i>	Acurácia
(12, 28)	Ident.	adam	0,0005	16	0,91
(11, 27)	Relu	lbfgs	0,0001	24	0,90
(24, 14)	tanh	adam	0,0005	16	0,89
(25, 13)	tanh	lbfgs	0,0001	16	0,89
(10, 28)	tanh	adam	0,0005	24	0,86

Após a obtenção da melhor rede MLP para o *ensemble*, como resultado da busca em grade, foram analisadas as métricas

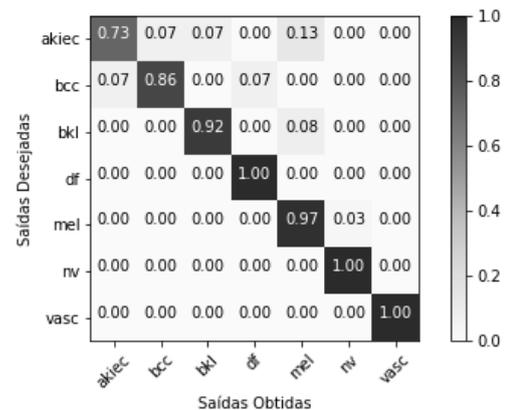
Acurácia, Precisão, Revocação e F1-Score para a abordagem em questão. A acurácia obtida foi de 0,93. Os demais resultados são informados pela Tabela VI.

TABELA VI

RESULTADOS OBTIDOS PELA REDE PERCEPTRON DE MULTICAMADAS PARA TODAS AS 7 CLASSES APÓS O ENSEMBLE DAS 3 PIORES CLASSES COM A CNN MULTICLASSE

Classe	Precisão	Revocação	F1-Score
Akiek	0,78	0,99	0,87
Bcc	1,00	0,90	0,95
Bkl	0,99	0,85	0,91
Df	0,90	0,98	0,96
Mel	0,95	0,84	0,89
Nv	0,82	0,98	0,89
Vasc	1,00	0,94	0,97
Total	0,92	0,93	0,92

Percebe-se, através da comparação entre as Tabelas I e VI, a grande evolução, em termos de desempenho, que os modelos propostos tiveram no decorrer deste trabalho. Tendo o F1-Score, por exemplo, aumentado de 0,71, do modelo com uma única CNN, para 0,92, no modelo com *ensemble* com apenas 3 redes binárias e CNN multiclasse. Pode-se assim verificar a eficiência da abordagem para diagnóstico de lesões pigmentadas. Uma análise mais detalhada dos resultados pode ser ilustrada pela matriz de confusão (Fig. 5).

Fig. 5. Matriz de Confusão normalizada da validação final do *ensemble*.

A matriz de confusão normalizada mostra explicitamente as relações entre saídas desejadas e saídas obtidas, onde as colunas representam as saídas obtidas pelo modelo, enquanto as linhas indicam as saídas desejadas. Nota-se, pela coloração mais escura, a alta densidade de valores na diagonal principal, o que evidencia o alto índice de classificação correta em face dos triângulos superior e inferior da matriz que, por sua vez, possuem seus elementos iguais ou próximos a zero, indicando assim baixa taxa de erro.

Uma análise muito importante, e que não pode deixar de ser feita, é a comparação entre o modelo desenvolvido e um especialista. Visto que, em média, um dermatologista realiza a análise clínica para classificação de lesões pigmentadas com acurácia de 73,2% [12], tem-se mais uma comprovação da eficiência e aplicabilidade do modelo na classificação de lesões pigmentadas, uma vez que demonstrou acurácia superior em quase 20%.

Pesquisas semelhantes mostram ainda o desempenho de redes neurais não convolucionais na classificação de lesões pigmentadas, chegando a atingir a acurácia de 0,87 [13]. Redes Neurais Convolucionais utilizadas para a classificação em questão e treinadas com a mesma base de dados atingem, em média, F1-Score de 0,87 [1] e acurácia de 0,91 [7]. Resultados estes inferiores ao desempenho do modelo apresentado neste trabalho, que atingiu F1-Score de 0,92 e Acurácia de 0,93, indicando, assim, a importância do trabalho e avanço da pesquisa no cenário de classificação de imagens clínicas para o apoio ao diagnóstico de lesões dermatológicas pigmentadas. Estes dados podem ser verificados com mais detalhes para fins comparativos na Tabela VII, que apresenta os valores de Precisão e F1-Score obtidos pelos modelos propostos em [1] e pelo modelo com *ensemble* de 3 CNNs binárias e uma CNN multiclasse proposto neste trabalho.

TABELA VII
RESULTADOS OBTIDOS POR [1] E PELA METODOLOGIA PROPOSTA NESTE
TRABALHO COM A MESMA BASE DE DADOS

Classificador	Precisão	F1-Score
DenseNet201	0,87	0,87
ResNet152	0,85	0,85
Inception v3	0,83	0,82
InceptionResNet v2	0,85	0,83
Modelo Proposto	0,92	0,93

V. CONSIDERAÇÕES FINAIS

O *Ensemble* de Redes Neurais Convolucionais com arquitetura InceptionV3 mediado por uma Rede Neural do tipo Perceptron de Multicamadas apresentou ótimos resultados na tarefa de classificação de imagens dermatoscópicas quanto às 7 classes da base de dados *Skin Cancer MNIST* (HAM10000). O modelo se apresentou bastante consistente quanto a todas as métricas utilizadas para sua avaliação, com resultados superiores ao desempenho de especialistas e de modelos desenvolvidos em trabalhos similares, demonstrando-se assim estado-da-arte.

O avanço em pesquisas do mesmo segmento deve ocorrer através de testes com outras combinações de classes para o *ensemble* e testes de diferentes arquiteturas para as CNNs. A busca mais eficiente por hiperparâmetros da Rede Perceptron de Multicamadas, que media o *ensemble*, e, ainda, a proposta de um outro modelo para realização deste também deve apontar as possíveis evoluções para a tarefa proposta. Uma vez alcançada a melhoria das métricas avaliadas, têm-se um novo desafio: tornar a classificação das lesões pigmentadas mais específica, isto é, aumentar o número de classes da tarefa. O modelo apresenta o diagnóstico automático quanto a classificação dentre os principais grupos de lesões. Assim, com o aumento das classes, obter-se-á um diagnóstico ainda mais objetivo.

REFERÊNCIAS

[1] A. Rezvantalab, H. Safigholi and S. Karimijeshni, "Dermatologist Level Dermoscopy Skin Cancer Classification Using Different Deep Learning

Convolutional Neural Networks Algorithms," Oct. 21, 2018. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1810/1810.10348.pdf>

[2] L.C. Lobos, "Inteligência Artificial e Medicina" *Rev. bras. educ. med.* vol.41, n.2, pp.185-193. 2017. DOI: 10.1590/1981-52712015v41n2esp

[3] Y. LeCun, Y. Bengio and G. Hinton. "Deep Learning," *Nature*, vol. 521, pp. 436-444, May, 2015.

[4] M.H. Jafari, *et al.* (2016, december). Skin Lesion segmentation in clinical images using deep learning. Presented at 23rd ICPR. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7899656>

[5] J. Premaladha and K.S. Ravichandran, "Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms," *J. Med. Syst.*, pp. 40-96, Feb, 2016. DOI: 10.1007/s10916-016-0460-2, [Online] Available: <https://doi.org/10.1007/s10916-016-0460-2>

[6] A.A. Crus-Roa, J.E. Arevalo Ovalle, A. Madabhushi and F.A. González Osorio. (2013). A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection. Presented at MICAI 2013. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-40763-5_50

[7] N.D. Reddy. "Classification of Dermoscopy Images using Deep Learning," Aug. 5, 2018. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1808/1808.01607.pdf>

[8] G. Schaefer, *et al* "An ensemble classification approach for melanoma diagnosis," *Memetic Computing*, vol. 6, no. 4, pp. 233-240, Dec, 2014. DOI: 10.1007/s12293-014-0144-8, [Online] Available: <https://doi.org/10.1007/s12293-014-0144-8>

[9] H. Brink, J. Richards and M. Fetherolf, *Real-World Machine Learning* 1st ed. USA: Manning Publications, 2016.

[10] A.P. Braga. *Redes Neurais Artificiais: Teorias e aplicações*, 2nd ed. BR: LTC. 2016

[11] T. Masters. *Advanced Algorithms for Neural Network: A C++ Sourcebook*, 1st ed. New York, NY, USA: Wiley, 1995.

[12] C. Dolianitis *et al.* "Comparative performance of 4 dermoscopic algorithms by nonexperts for the diagnosis of melanocytic lesions," PubMed – US National Library of Medicine. USA. September 2005. [Online] Available: https://www.researchgate.net/publication/7657968_Comparative_Performance_of_4_Dermoscopic_Algorithms_by_Nonexperts_for_the_Diagnosis_of_Melanocytic_Lesions

[13] J.L.G. Arroyo and B.G. Zapirain. "Detection of pigment network in dermoscopy images using supervised machine learning and structural analysis," *Computers in Biology and Medicine*, vol. 44, no. C, pp. 144-157, Jan., 2014. DOI: 10.1016/j.combiomed.2013.11.002. [Online] Available: <https://www.sciencedirect.com/science/article/pii/S001048251300320X?via%3Dihub>

[14] Instituto Nacional de Câncer. Câncer de pele não melanoma. Brazil. Nov. 2018. [Online]. Available: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-pele-nao-melanoma>

[15] Instituto Nacional de Câncer. Câncer de pele melanoma – Versão para Profissionais de Saúde. Brazil. Nov. 2018. [Online]. Available: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-pele-melanoma/profissional-de-saude>

[16] Kaggle. Skin Cancer MNIST: HAM10000, a large collection of multi-source dermatoscopic images of pigmented lesions. 2018. [Online] Available: <https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000>.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," Dec. 11, 2015. [Online] Available: <https://arxiv.org/abs/1512.00567>

[18] World Health Organization. Prevention. Cancer control: knowledge into action: WHO guide for effective programs. 2017. [Online] Available: http://www.who.int/cancer/publications/cancer_control_detection/en/.

[19] T. Liu *et al.* "Implementation of Training Convolutional Neural Networks," Jun. 4, 2015. [Online] Available: <https://arxiv.org/abs/1506.01195>



Jean Phelipe de Oliveira Lima

Graduando em Engenharia de Computação pela Escola Superior de Tecnologia da Universidade do Estado do Amazonas (EST/UEA). Possui experiência em sistemas embarcados, inteligência artificial, aprendizado de máquina, visão computacional, *deep learning* e tem interesse nos temas de eficiência energética, inteligência das coisas, indústria 4.0 e computação aplicada a medicina.



Luiz Carlos Silva de Araújo Filho

Graduando em Engenharia de Computação pela Escola Superior de Tecnologia da Universidade do Estado do Amazonas (EST/UEA). Pesquisador do Laboratório de Sistemas Inteligentes (LSI). Possui experiência em sistemas embarcados, inteligência artificial, aprendizado de máquina, visão computacional, *deep learning* e tem interesse nos temas de redes generativas, processamento de sinais e imagens e computação aplicada a medicina.



Carlos Maurício Seródio Figueiredo

Engenheiro Eletricista pela Universidade Federal do Amazonas, mestre e doutor em Ciências da Computação pela Universidade Federal de Minas Gerais. É professor adjunto da Escola Superior de Tecnologia da Universidade do Estado do Amazonas (EST/UEA) e professor colaborador do Programa de Pós-graduação em Informática da Universidade Federal do Amazonas. Tem experiência na área de Ciência da Computação, com ênfase em Sistemas Embarcados e Móveis, atuando principalmente nos seguintes temas de Internet das Coisas e Sistemas Inteligentes.



Fábio Santos da Silva

Doutor em Ciências com ênfase em Sistemas Digitais pela Escola Politécnica da Universidade de São Paulo. Atualmente, é professor da Escola Superior de Tecnologia da Universidade do Estado do Amazonas (EST/UEA) e pesquisador do Laboratório de Sistemas Inteligentes (LSI). Atua nas áreas: Computação Ubíqua, Inteligência Artificial, Recuperação de Informação.