

## **ACURÁCIA E REPRODUTIBILIDADE DA IDENTIFICAÇÃO DE PONTOS CEFALOMÉTRICOS REALIZADA POR DOIS SOFTWARES COM INTELIGÊNCIA ARTIFICIAL EM IMAGENS DE PACIENTES BRASILEIROS E COREANOS: A ORIGEM POPULACIONAL INFLUENCIA A PERFORMANCE DOS SOFTWARES?**

**Palavras-chave:** Inteligência artificial; Identificação anatômica; Aprendizagem profunda; População.

**Autores(as):**

**Giovanna Sachs Puntigam [FOP – UNICAMP];**

**Tháisa Pinheiro Silva [FOP – UNICAMP];**

**Prof. Dr. Deborah Queiroz de Freitas (orientadora) [FOP – UNICAMP].**

---

### **INTRODUÇÃO:**

O uso de IA na identificação de pontos cefalométricos tem mostrado avanços significativos em acurácia, tornando sua aplicação na prática clínica cada vez mais viável. Estudos recentes destacam a eficácia de alguns softwares nessa tarefa (Arik, Ibragimov e Xing, 2017; Park et al., 2019; Hwang et al., 2020; Bulatova et al., 2021). Entre eles, o CefBot, um software brasileiro, demonstrou excelente acurácia na identificação de pontos cefalométricos, mas suas avaliações foram limitadas a imagens de pacientes brasileiros (Silva et al., 2021). Por outro lado, o WebCeph, um software coreano, foi relatado como menos preciso em estudos que não especificaram a origem populacional das imagens utilizadas (Yassir et al., 2021). O estudo indicando sua acurácia utilizou radiografias de uma população não coreana, sugerindo que diferenças populacionais entre conjuntos de treinamento e avaliação podem impactar os resultados.

Se um software foi treinado em uma população diferente daquela na qual foi avaliado, isso pode levar a discrepâncias na sua performance. Diferenças nas características faciais e cranianas entre populações podem impactar a capacidade do modelo de generalizar efetivamente seu padrão de reconhecimento para novos casos. Portanto, considerando que as imagens selecionadas para o treinamento do algoritmo podem influenciar fortemente o desempenho final do software de IA, o objetivo deste estudo foi avaliar o desempenho da identificação de pontos cefalométricos realizada por dois programas de software com IA em imagens de diferentes origens populacionais (brasileira e coreana), com base na comparação das coordenadas de cada ponto com a identificação humana.

### **OBJETIVO:**

Avaliar o desempenho da identificação de pontos cefalométricos realizada por dois programas de software com IA em imagens de diferentes origens populacionais (brasileira e coreana).

## **MATERIAIS E MÉTODOS:**

Após a aprovação do Comitê de Ética em Pesquisa (CEP) da Universidade Estadual de Campinas, foram selecionadas 30 telerradiografias cefalométricas laterais de indivíduos brasileiros e 30 de indivíduos coreanos. As imagens brasileiras foram obtidas dos arquivos da Área de Radiologia da Faculdade de Odontologia de Piracicaba/Unicamp, utilizando o aparelho OP300 Maxio, módulo de Telerradiografia Cefalométrica Lateral. As imagens coreanas foram adquiridas de arquivos de acesso aberto do International Symposium on Biomedical Imaging: From Nano to Macro, de 2015. Foram selecionados indivíduos com idade superior a 18 anos e classe esquelética I ( $ANB \leq 0^\circ$ ). Foram escolhidos vinte pontos cefalométricos para avaliação nas 60 imagens: Básio (Ba), Espinha Nasal Posterior (ENP), Espinha Nasal Anterior (ENA), Mentoniano (Me), Násio (Na), Orbital (Or), Pogônio (Pog), Pório (Po), Subespinhal (A), Supramentoniano (B), Násio Linha (Na'), Pogônio Linha (Pog'), Supramentoniano Linha (B'), Lábio Inferior (LI), Lábio Superior (LS), Sela (S), Subnasal (Sn), Mentoniano Linha (Me'), Glabela (Gla) e Ponta do Nariz (Pn).

Dois examinadores identificaram os vinte pontos cefalométricos nas imagens usando o software digital RadioCef Studio 3 (RadioMemory, MG, Brasil). Para assegurar a calibração dos examinadores, foi calculado o Coeficiente de Correlação Intraclasse (ICC) das coordenadas X e Y das dez primeiras imagens, resultando em 1,0 e 0,999, respectivamente, indicando excelente concordância (Koo and Li, 2016). A média dos valores obtidos pelos dois examinadores foi utilizada como referência para avaliar a performance dos sistemas de IA.

As mesmas imagens foram então submetidas à identificação automatizada de pontos cefalométricos por dois softwares com IA: Cefbot (RadioMemory, MG, Brasil) e Webceph (AssembleCircle, República da Coreia). Os pontos identificados foram tabulados para a determinação das coordenadas X (abscissa) e Y (ordenada) utilizando o software ImageJ, e os dados foram registrados no Microsoft Office Excel 2007 (Microsoft Corporation, Redmond, WA, EUA). Para análise de reprodutibilidade, 30% das imagens foram reavaliadas 30 dias após a primeira análise. Foi calculada a média da diferença entre as coordenadas X e Y estipuladas pelos avaliadores e softwares, de acordo com os pontos e nacionalidade. Os dados obtidos foram submetidos à análise de acordo com o teste Shapiro-Wilk referente à normalidade. A distribuição foi atribuída como normal e então a Análise de Variância unidirecional (ANOVA) e teste post-hoc de Tukey foram usados para a comparação dos valores encontrados pelos softwares com IA com o padrão de referência, que foi determinado pelas duas examinadoras.

## **RESULTADOS:**

O Cefbot, programa brasileiro, apresentou, na coordenada X, boa performance na identificação dos pontos em imagens brasileiras, exceto nos pontos Glabela e Mentoniano. Nas imagens coreanas, houve discrepância na identificação dos pontos Glabela, Mento, Básio e Orbital. Na coordenada Y, o Cefbot apresentou irregularidades na marcação dos pontos Glabela e Mentoniano Linha tanto nas imagens brasileiras quanto nas coreanas.

O Webceph, programa coreano, apresentou, na coordenada X, diferenças significativas na determinação do Básio, Mentoniano Linha e Pório nas imagens brasileiras. Porém, houve discrepância apenas no Mentoniano Linha nas imagens coreanas. Na coordenada Y, o Webceph diferiu no Mentoniano Linha e Násio Linha em imagens brasileiras e na Glabela, Mentoniano Linha e Násio Linha nas imagens coreanas.

Tabela 1 – Média da diferença entre as coordenadas estipuladas pelos avaliadores e pelos softwares, de acordo com os pontos e nacionalidade

Landmark	Software program	X <sup>#</sup>		Y <sup>s</sup>	
		Brazilian	Korean	Brazilian	Korean
Basio	Cefbot	-0.25	-1.88*	-0.05	-1.90
	Webceph	-0.37*	-0.67	0.68	-1.19
ENA	Cefbot	-0.05	-0.65	-0.01	0.14
	Webceph	0.72	-1.83	0.66	-0.92
ENP	Cefbot	-0.15	-1.80	0.01	0.61
	Webceph	0.26	-1.32	0.63	-0.97
Glabela	Cefbot	13.30*	92.23*	15.65*	88.11*
	Webceph	0.94	0.22	1.24	4.84*
Labio_Inferior	Cefbot	0.06	0.32	0.02	0.03
	Webceph	0.97	0.07	0.35	-3.15
Labio_Superior	Cefbot	0.13	-0.89	-0.01	-0.28
	Webceph	1.00	-1.12	0.63	-1.40
Mentoniano	Cefbot	-0.02	-0.35	-0.03	-0.29
	Webceph	1.16	1.89	0.35	-3.55
Mentoniano_L	Cefbot	13.08*	87.76*	3.21*	17.04*

	Webceph	2.06*	6.86*	0.65*	-1.81*
Nasio	Cefbot	-0.02	0.09	-0.02	-0.44
	Webceph	0.93	0.21	1.16	0.61
Nasio_L	Cefbot	-0.02	0.30	0.01	0.01
	Webceph	1.01	1.03	1.53*	5.10*
Orbital	Cefbot	-0.42	-2.17*	-0.08	-0.48
	Webceph	0.24	-1.69	0.67	0.17
Pogonio	Cefbot	0.01	-0.38	0.05	0.82
	Webceph	0.79	-0.35	0.38	-2.38
Pogonio_L	Cefbot	0.08	0.31	0.25	1.15
	Webceph	0.96	0.21	0.29	-3.07
Ponta_Nariz	Cefbot	0.28	0.23	-0.09	-0.54
	Webceph	1.10	0.12	0.69	-0.58
Porio	Cefbot	-0.30	-1.14	-0.13	-0.84
	Webceph	-0.37*	-0.82	0.92	-0.70
Sela	Cefbot	-0.09	-0.14	0.09	0.37
	Webceph	0.16	0.21	0.88	0.64
Subspinal	Cefbot	0.12	0.57	-0.01	0.49
	Webceph	0.79	-1.36	0.83	-0.31
Subnasal	Cefbot	0.00	-0.95	0.04	-0.44
	Webceph	0.94	-0.15	0.59	-2.12
Supramentoniano	Cefbot	0.02	-0.02	-0.26	-1.17
	Webceph	0.84	0.02	0.28	-3.33
Supramentoniano_L	Cefbot	-0.07	-0.44	-0.07	0.05
	Webceph	0.89	0.08	0.41	-3.40

\* indicam que a diferença é estatisticamente significativa, de acordo com ANOVA e teste post-hoc de Tukey ( $p < 0.05$ ).

# Medidas dos brasileiros e coreanos diferiram em todos os pontos para o eixo X, independente do software

§ Medidas dos brasileiros e coreanos não diferiram no eixo Y

## DISCUSSÃO:

No geral, ambos os softwares tiveram boa performance na determinação dos pontos cefalométricos. O software brasileiro apresentou, nas imagens brasileiras, acurácia de 90%, exceto nas regiões de Mentoniano Linha (Me') e Glabela (Gla). Nas imagens coreanas, a acurácia foi de 80%, indicando potenciais dificuldades na aplicação do programa em outras populações.

O software coreano apresentou irregularidades na identificação de certos pontos em imagens brasileiras, tendo acurácia de 85%. Em imagens coreanas, a acurácia foi de 95%, indicando ótima aplicabilidade na população nativa.

O Coeficiente de Correlação Intraclasse (ICC) entre a primeira avaliação por IA pelos softwares e da reavaliação 30 dias depois foi de 1,0. Assim, infere-se que os softwares garantiram reprodutibilidade na marcação dos pontos cefalométricos.

Cada software apresentou melhor desempenho ao analisar imagens da própria população, com leve queda na acurácia quando avaliaram imagens de outra população.

Os achados da presente pesquisa destacam a importância de considerar variações específicas a cada população ao utilizar a IA em programas para a identificação de pontos.

## **CONCLUSÃO:**

No geral, o software brasileiro e o coreano demonstraram acertos comparáveis na identificação da maioria dos pontos cefalométricos analisados neste estudo, independentemente da origem populacional das imagens. No entanto, os pontos cefalométricos Glabela e o Mentoniano L apresentaram diferenças significativas para o software brasileiro, e para os softwares brasileiro e coreano, respectivamente.

## **REFERÊNCIAS:**

- ARIK, S. Ö.; IBRAGIMOV, B.; XING, L.; Fully automated quantitative cephalometry using convolutional neural networks. *J. Med. Imag.* 4(1), 014501; 2017 doi:10.1117/1.JMI.4.1.014501.
- BULATOVA, G.; KUSNOTO, B.; GRACE, V.; TSAY, T.P.; AVENETTI, D.M.; SANCHEZ, F.J.C. Assessment of automatic cephalometric landmark identification using artificial intelligence. *Orthod Craniofac Res.* 2021; 2437–42. <https://doi.org/10.1111/ocr.12542>
- HWANG H.W.; PARK J.H.; MOON J.H. Automated identification of cephalometric landmarks: Part 2-Might it be better than human? *Angle Orthod.* 2020; 90(1):69-76.
- PARK J-H.; HWANG H-W.; MOON J-H. Automated identification of cephalometric landmarks: Part 1-Comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthod.* 2019; 89(6):903-909.
- SILVA, T. P.; HUGHES, M. M.; MENEZES, L. DOS S.; MELO, M. DE F. B. DE; TAKESHITA, W. M.; YASSIR, Y. A.; SALMAN, A. R.; NABBAT, S. A. The accuracy and reliability of WebCeph for cephalometric analysis. *Journal of Taibah University Medical Sciences*, p. S1658361221001645, set. 2021.