Machine learning in high-frequency ultrasound skin imaging for cosmetics assessment

Massufero Vergilio, Mariane[1]; Flamini Kiihl, Samara[2]; Batista Florindo, João[3]; Soares de Freitas, Luan[2]; Ribeiro Duzzi, Matheus[2]; Martins, Dieine[2]; Moretti Aiello, Laura[4]; **Ricci Leonardi, Gislaine**[1,5*]

[1] School of Medical Sciences, University of Campinas (UNICAMP), SP, Brazil;[2] Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas (UNICAMP), SP, Brazil; [3] Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas (UNICAMP), SP, Brazil; [4] School of Pharmaceutical Sciences, University of Campinas (UNICAMP), SP, Brazil.

* Gislaine Ricci Leonardi, 200 Cândido Portinari st., +55 19 99203-0347, gislaine.leonardi@fcf.unicamp.br.

Abstract

Background: An effective way to assess the in vivo performance of anti-aging cosmetics is by using the high-frequency ultrasound (HFUS) skin image technique. Ultrasound-based skin measurements, such as echogenicity and thickness, are frequently used to evaluate skin conditions. Once manual measurements are operator-dependent and time-consuming, much research is being actively conducted on automated methods. However, the existing automated methods are still not specialized in measurements of the skin and its layers. The purpose of the study was to establish whether machine learning could be successfully applied to HFUS skin images and to develop new tools for cosmetic claims assessment.

Methods: To predict each target variable (echogenicity of dermis and epidermis, thickness of dermis and epidermis) using the ultrasound images, our approach considered supervised machine learning algorithms. The best-performing model was selected for each variable using an independent validation set. For echogenicity variables, Gradient Boosting Machine (GBM) algorithm and Principal Component Regression (PCR) were chosen for thickness variables. Using the proposed method, a labeled dataset containing 144 ultrasound skin images was used for training, and validation was performed with 40 ultrasound skin images.

Results: Our algorithms effectively predicted echogenicity and thickness variables, showing competitive performance. The dermis echogenicity variable stood out with a median absolute error (MAE) of 0.81 and a root mean squared error (RMSE) of 1.11 between the predicted and the expected value.

Conclusion: Machine learning algorithms are able to reduce the time and increase the quality of the skin ultrasound analysis method.

Keywords: Efficacy assessment; high-frequency ultrasound; machine learning; skin; supervised regression.

Introduction

For the evaluation and clinical efficacy of cosmetic products, the technique of skin image analysis by means of high frequency ultrasound (HFUS) stands out, which, in addition to being non-invasive, has shown great potential for application and characterization of changes in the skin [1]. The high frequency allows a detailed real-time visualization of the epidermis and dermis layers and subcutaneous tissue, allowing an in-depth characterization of several parameters [2]–[5].

This method provides quantitative tools such as thickness and acoustic impedance measurements of different skin layers; therefore, it can be used for non-invasive characterization and as a complementary and accurate instrument for the evaluation of skin diseases [2]–[5].

An HFUS imaging device uses the principle of transmitting short pulses of ultrasound energy to a given tissue and detecting the reflected waves (called echoes) resulting from the ultrasound wave-tissue interaction [6]. Echogenicity is a term used in radiology to describe how much a material allows the passage of or reflects ultrasound waves. In other words, echogenicity can be understood as the property of a given material, in this case the skin, to reflect ultrasound and originate echoes, which, in turn, generate echographic images, also called ultrasound [7].

The transducer is a very important component of the HFUS system, since it has a material with piezoelectric properties, it can convert electrical energy into vibrating mechanical energy, and vice versa, through the oscillatory expansion and contraction of the material [2], [8].

One of the main signal processing systems used to produce pulse-echo method recordings is the B-mode scanning [9]. In the B-mode scan processing system, the amplitude value is represented in shades of gray, ensuring the two-dimensional cross-sectional image

that resembles the cross-sectional images of tissues [9], [10]. B-mode scanning system displays tissue echotexture, allowing characterization of collagen fibers, keratin and dermal water content [9].

However, from this initiative a great challenge arises in the processing and evaluation of HFUS images, after all, within an image we can obtain several quantitative, semi-quantitative and qualitative parameters. Given the large amount of material to be evaluated, manual analysis performed by a trained person can bring bias, loss of reliability and low effectiveness [11], [12]. The trend identified is the increased use of the automatic image segmentation process through artificial intelligence technologies [13]–[16]. Manual and automated methods of measuring ultrasound parameters are different, as previously demonstrated [17].

Supervised machine learning can be implemented from ultrasound images, which can prevent wrong diagnostic decisions. The machine learning algorithms used to predict the quantitative parameters (dermis echogenicity, epidermis echogenicity, dermis thickness and epidermis thickness) from the ultrasound images were the Gradient Boosting Machine (GBM) algorithm by Friedman (2001) and Main Component Regression (PCR) [18], [19].

Thus, the objective of this study was to train and test/validate a model to automate obtaining the HFUS parameters of the skin through machine learning; and further contribute with the evaluation of the effectiveness of a cosmetic product, evaluating the skin of the individual before and after the use of a certain cosmetic.

Materials and Methods

Data preparation

The ultrasound image dataset was constructed from images in the untreated (control) region of participants in several studies carried out in the laboratory. These images were captured at different times and separated for further analysis in this dataset. The capture was performed on the ventral side of the forearm of 184 women, at a distance of 10 cm from the wrist. The region was previously demarcated according to the dimensions of the transducer, using a black hydrographic pen and a plastic mold (Figure 1).

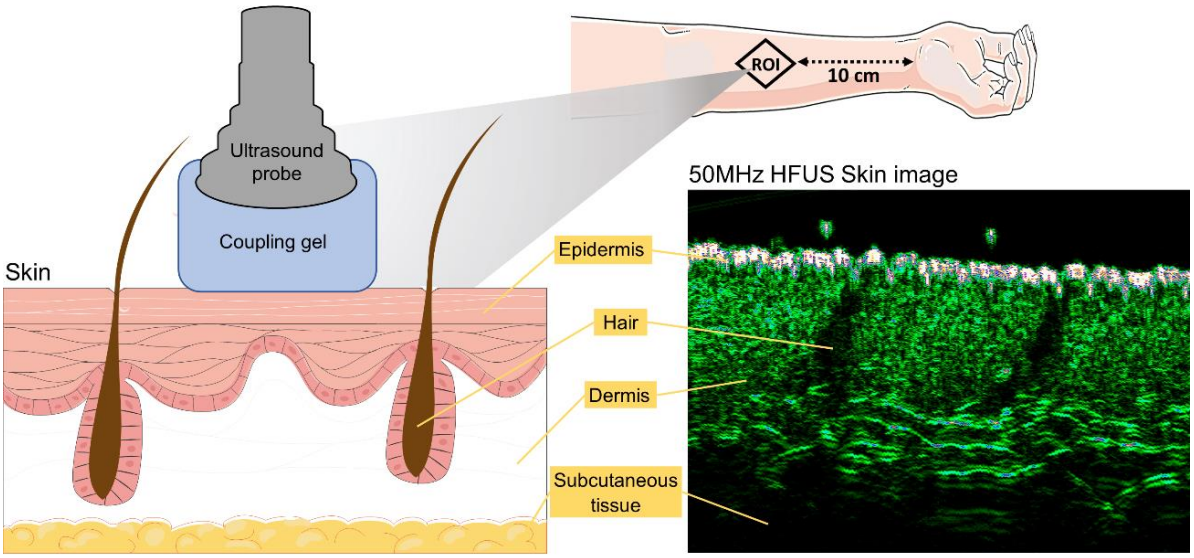

Figure 1. Scheme to obtain skin image using high-frequency ultrasound (HFUS). The tissue on the region of the interest (ROI) evaluated by the probe is reproduced as a real-time image in the visual display of the device (adapted from [12]).

Linked to this process, a tabular data file was manually created with age, echogenicity, skin thickness and layers per participant. Thus, each of these characteristics form the following variables:

- Age: variable of the study filled in after questioning by the researcher to each of the study participants and will try to be predicted using the images and covariates stored in the tabular data file.

- Epidermis and dermis echogenicity: variables that record the echogenicity measure in different skin layers, the first one being the sum of the values found for epidermis and dermis [12].

- Thickness of epidermis and dermis: variables that record the distance between the beginning and end of a layer. This measurement is indicated by the researcher herself and the first measurement reflects the sum of the thickness of the epidermis and dermis [12].

Data modeling and evaluation

To predict each target variable (echogenicity of dermis and epidermis, thickness of dermis and epidermis) using the ultrasound images, our approach considered supervised machine learning algorithms. Several methods were tested, but here we describe the

algorithms with the best performance: the Gradient Boosting Machine (GBM) algorithm by [18] and Principal Component Regression (PCR) [19]. These methods were combined with feature extraction techniques from the ultrasound images. To extract the information from the images, Local Binary Patterns (LBP) histograms [20] were used. Then, principal component analysis (PCA) was applied to reduce the dimension and correlation of the LBP features. Hyperparameters tuning of the GBM algorithm was performed using leave-one-out cross validation (LOOCV). To tune, implement and validate the GBM and the PCR algorithms the R package caret [21] was used. LBP feature extraction from grayscale converted ultrasound images was performed using the wvtool R package [22]. GBM had the best performance to predict dermis_echogenicity and epidermis_echogenicity. To predict dermis_thickness and epidermis_thickness, PCR had the best performance. Metrics used for evaluation of the models: median absolute error (MAE) and root mean squared error (RMSE). The sample (n=184) was randomly divided into two: training (n=144) and testing (n=40) datasets respectively to train the models and test their prediction performance.

Results

Descriptive data (mean, SD, median, min and max) of age and of the echogenicity variables of the dermis and epidermis described in arbitrary units (AU); and thickness of the dermis and epidermis described in μm, measured manually, in the 184 ultrasound images of different participants, are presented in Table 1. The models developed here were based on the data in the table.

Table 1. Descriptive data from the ultrasound image collection of participants with age, echogenicity and skin thickness data**.**

Variables	Age group (years)	Number of values	Mean	SD	Median	Min	Max
Age							
	10 to 19	1	19.0	0.0	19.0	19.0	19.0
	20 to 29	8	24.9	2.1	25.0	22.0	29.0
	30 to 39	20	35.9	3.1	37.0	30.0	39.0

	40 to 49	73	45.2	2.9	45.0	40.0	49.0
	50 to 59	70	53.5	2.7	53.5	50.0	59.0
	60 to 69	12	62.5	2.5	62.0	60.0	68.0
Epidermis echogenicity (AU)							
	10 to 19	1	155.6	0.0	155.6	155.6	155.6
	20 to 29	8	107.4	30.2	98.4	67.8	144.7
	30 to 39	20	102.5	25.9	104.8	39.7	150.0
	40 to 49	73	95.7	28.7	95.9	13.0	157.4
	50 to 59	70	99.9	20.2	97.9	60.9	165.0
	60 to 69	12	123.2	27.2	121.6	86.6	172.2
Edipermis thickness (µm)							
	10 to 19	1	82.0	0.0	82.0	82.0	82.0
	20 to 29	8	73.3	15.0	76.0	51.0	94.0
	30 to 39	20	71.0	17.4	71.0	11.0	93.0
	40 to 49	73	77.6	17.6	78.0	12.0	106.0
	50 to 59	70	79.2	17.3	82.0	23.0	128.0
	60 to 69	12	92.8	12.5	94.0	66.0	109.0
Dermis echogenicity (AU)							
	10 to 19	1	11.2	0.0	11.2	11.2	11.2
	20 to 29	8	16.4	5.9	13.8	11.2	26.5
	30 to 39	20	17.8	4.8	17.5	10.7	27.5
	40 to 49	73	15.8	4.5	15.0	7.8	29.3
	50 to 59	70	14.7	4.1	14.3	8.2	30.2
	60 to 69	12	19.0	7.3	18.8	11.0	32.3
Dermis thickness (µm)							
	10 to 19	1	945.0	0.0	945.0	945.0	945.0
	20 to 29	8	1045.0	261.7	1002.0	622.0	1484.0
	30 to 39	20	981.0	180.5	986.5	645.0	1383.0
	40 to 49	73	1013.0	175.4	1008.0	668.0	1527.0
	50 to 59	70	1040.0	190.1	1020.0	691.0	1539.0
	60 to 69	12	1095.0	286.8	1037.0	672.0	1688.0

AU= arbitrary units; SD=Standard deviation; Min= Minimum; Max=Maximum.

After being developed, the models were trained and validated. The accuracy assumed by MAE and RMSE of the models are provided in Table 2. The values indicated better performance for the echogenicity variables (MAE values between 0.81 and 10.41 AU) than for the thickness variables (MAE values between 8.13 at 104.41 µm). The dermis echogenicity variable stood out with a MAE value of 0.81 and RMSE of 1.11.

Table 2. Performance of the models for training and testing datasets.

Target Variable	Best Model	Datasets	MAE	RMSE
Dermis echogenicity	GBM	Training (n=144)	0.81	1.11
Dermis echogenicity	GBM	Testing (n=40)	1.94	2.57
Epidermis echogenicity	GBM	Training (n=144)	6.25	8.54
Epidermis echogenicity	GBM	Testing (n=40)	10.41	12.98
Dermis thickness	PCR	Training (n=144)	102.68	127.53
Dermis thickness	PCR	Testing (n=40)	104.41	143.21
Edipermis thickness	PCR	Training (n=144)	8.86	12.37
Edipermis thickness	PCR	Testing (n=40)	8.13	10.77

AU= arbitrary units; GBM= Gradient Boosting Machine; PCR= Principal Component Regression; MAE= median absolute error; RMSE= root mean squared error.

To better characterize this good performance, Figure 2 shows the proximity of the observed values to the predicted values for each variable, as well as the respective values of the Spearman coefficient index (α=0.05).

Figures 2a and 2b show the predicted values of dermis echogenicity by the GBM algorithm and the actual values observed in the test sample and in the training sample. The diagonal line represents the strong and positive (R=0.96) and strong and positive (R=0.81) prediction for the training sample and the test sample, respectively. Figures 2c and 2d show the predicted values of epidermis echogenicity by the GBM algorithm and the actual values observed in the test sample and in the training sample. The diagonal line represents the strong and positive (R=0.97) and strong and positive (R=0.81) prediction for the training sample and the test sample, respectively. Figures 2e and 2f show the predicted values of dermis thickness by the PCR algorithm and the actual values observed in the test sample and in the training sample. diagonal line represents the moderate and positive (R=0.77) and moderate and positive (R=0.70) prediction for the training sample and the test sample, respectively. Figures 2g and 2h show the predicted values of epidermis thickness by the PCR algorithm and the actual values observed in the test sample and in the training sample. The diagonal line represents the moderate and positive (R=0.71) and moderate and positive (R=0.77) prediction for the training sample and the test sample, respectively. These data indicate that the performance was superior for the echogenicity variables when compared to the thickness variables.

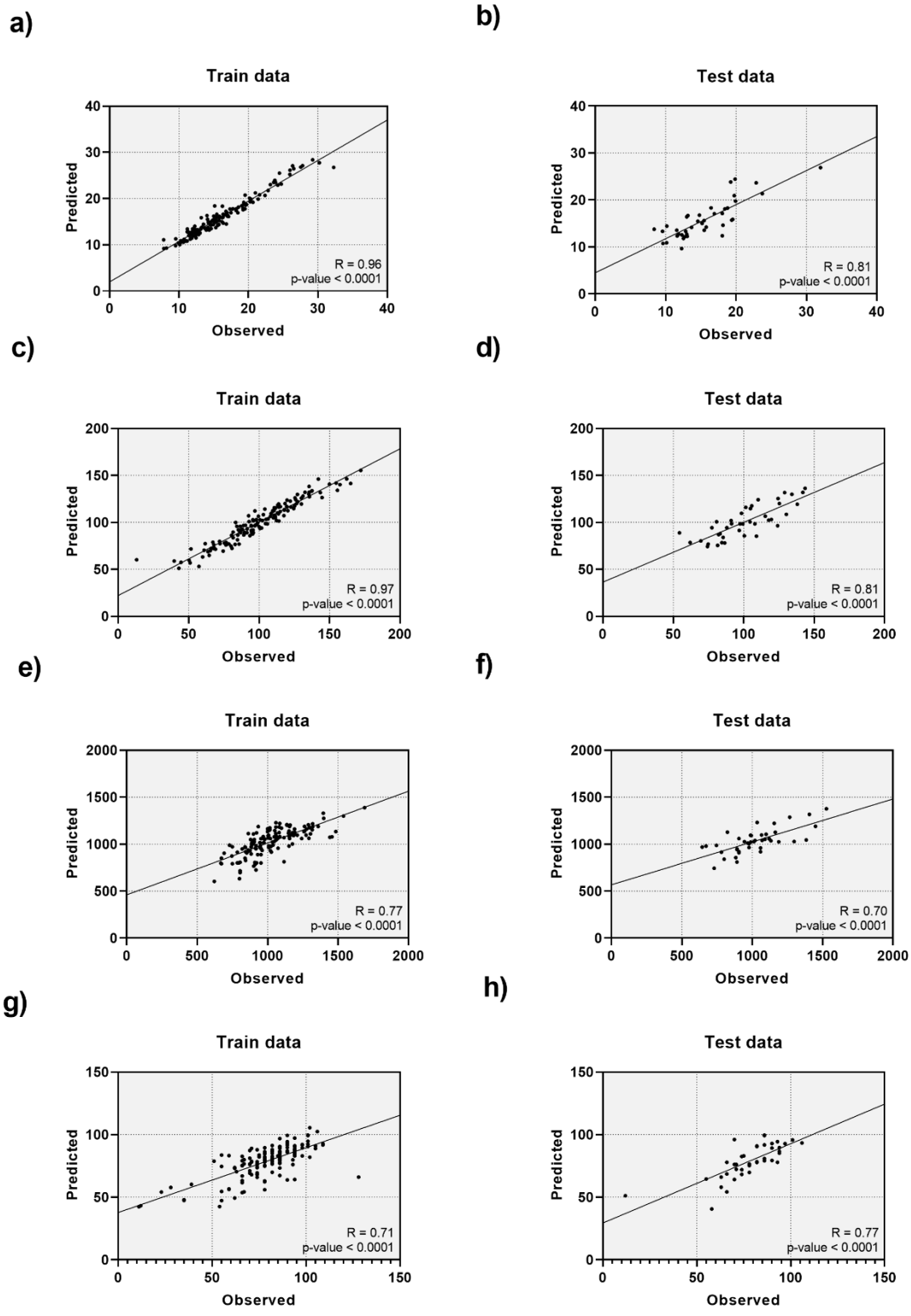

Figure 2: Prediction results of (A) training and (B) testing data from GBM algorithm for ecogenicidade da derme values; (C) training and (D) testing data from GBM algorithm for

ecogenicidade da epiderme values; (E) training and (F) testing data from PCR algorithm for espessura da epiderme values; and (G) training and (H) testing data from PCR algorithm for espessura da derme values by HFUS. The x and y axis represent observed and predicted values respectively.

Discussion

It is known that with advancing age the echogenicity of the skin tends to decrease, since it is related to the content and organization of collagen fibers in the skin; and also changes in skin thickness [12]. To develop the model, the ventral region of the forearm was chosen for the collection of HFUS data. The forearms are one of the areas of the body that stand out for assessing skin aging. [23]. In addition, at HFUS, site differences are also closely related to the fact that these areas are more or less exposed to the sun [24], [25]. Furthermore, in HFUS, the differences in sites are also closely related to the fact that these areas are more or less exposed to the sun, since the relative reduction in skin thickness caused by aging is smaller in the areas exposed to the sun compared to with the protected areas [25], [26]. In addition, photo-protected regions present greater echogenicity of the dermis [24], [25], [27], due to the degradation of the dermal extracellular matrix, caused by UV [25].

As a result of this study, it was observed that our algorithms effectively detect and measure echogenicity and skin layer thickness variables. Clinically, to support cosmetic claims, the error values of the models are within the expected range, since the expected values vary around 10% from the actual values [28]. We consider this to be highly valuable for future cosmetic development, as HFUS analyses can be performed without invasive sampling that can cause physical and mental stress in the subjects. Sample accessibility has been a key challenge with conventional methods to expand and perform more general analyses.

Interestingly, the most successful variable, dermal echogenicity, is also the most studied variable in ultrasound imaging for the assessment of skin aging [14], [23]. It is suggested that both the prediction models and the dermis echogenicity variable itself is the most reliable parameter for skin assessment by HFUS.

As a future perspective, we aim to evaluate the ability of algorithms to detect the benefits of a skin treatment, by replicating the results obtained in a clinical study, evaluating echogenicity and thickness before and after a given treatment. In addition, we would like to test the skin image analysis tool on a larger set of skin image data, on participants of different ages, genders and ethnicities, as well as different skin aging conditions.

Conclusion

The study developed here showed the possibility of automatic prediction of fundamental dermatological variables for the development and evaluation of cosmetics. In particular, the use of machine learning algorithms such as GBM and PCR showed promise especially in the prediction of dermis echogenicity, reaching a MAE of 0.81 and RMSE of 1.11. Through mathematical algorithms, the possibility arises of reducing these evaluations to minutes and further accelerating the development of new strategies for evaluating cosmetic claims. The developed system brought very promising data, and could be improved and later coupled to high-frequency ultrasound equipment, allowing an easier use, with less bias, and therefore more effective in evaluating the effectiveness of prescribed dermocosmetics. The use of machine learning algorithms has the potential to revolutionize the reality of clinical dermatology, research, as well as companies that sell products to improve skin conditions.

Acknowledgments

Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2018/06973-4; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: 0001.

Conflict of Interest Statement

None.

References

1. Vergilio MM, Monteiro e Silva SA, Vasques LI, Leonardi GR. Characterization of women’s young and mature skin by the 50 MHz high-frequency ultrasound technique. J Cosmet Dermatol. 2021;20(11):3695–7.

2. Barcaui EDO, Carvalho PaCP, Piñeiro-Maceira J, Valiante PM, Barcaui CB. Ultrassonografia de alta frequência (22MHz) na avaliação de neoplasias cutâneas malignas. Surg Cosmet Dermatol. 2014;6(2):105–11.
3. Kleinerman R, Whang TB, Bard RL, Marmur ES. Ultrasound in dermatology: Principles and applications. J Am Acad Dermatol. 2012;67(3):478–87.
4. Naredo E, Pascau J, Damjanov N, Lepri G, Gordaliza PM, Janta I, et al. Performance of ultra-high-frequency ultrasound in the evaluation of skin involvement in systemic sclerosis: a preliminary report. Rheumatology. 2020;59(7):1671–8.
5. Mlosek RK, Malinowska S. Ultrasound image of the skin, apparatus and imaging basics. J Ultrason. 2013;13(53):212–21.
6. Hwang JH. Principles of Ultrasound. In: Endosonography. 4th editio. Philadelphia, PA 19103-2899: Elsevier; 2019. p. 2–14.
7. Waller JM, Maibach HI. Age and skin structure and function, a quantitative approach (I): blood flow, pH, thickness, and ultrasound echogenicity. Ski Res Technol. 2005;11(4):221–35.
8. Foster FS, Pavlin CJ, Harasiewicz KA, Christopher DA, Turnbull DH. Advances in ultrasound biomicroscopy. Ultrasound Med Biol. 2000 Jan;26(1):1–27.
9. Rallan D, Harland CC. Ultrasound in dermatology - basic principles and applications. Clin Exp Dermatol. 2003;28(6):632–8.
10. Schmid-Wendtner M-H, Burgdorf W. Ultrasound Scanning in Dermatology. Arch Dermatol. 2005;141(2):217–24.
11. Burk RS, Parker A, Sievers L, Rooney MB, Pepperl A, Schubert CM, et al. Effects of position and operator on high-frequency ultrasound scan quality. Intensive Crit Care Nurs. 2015;31(3):148–54.
12. Vergilio MM, Monteiro e Silva SA, Jales RM, Leonardi GR. High-frequency ultrasound as a scientific tool for skin imaging analysis. Exp Dermatol. 2021 Jul 7;30(7):897–910.
13. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. Ultrasound Obstet Gynecol. 2020 Jun 12;uog.22122.
14. Sciolla B, Le Digabel J, Josse G, Dambry T, Guibert B, Delachartre P. Joint

segmentation and characterization of the dermis in 50 MHz ultrasound 2D and 3D images of the skin. Comput Biol Med. 2018 Dec;103:277–86.

15. Frontin JB, Anthony BW. Quantifying Dermatology: Method and Device for User-Independent Ultrasound Measurement of Skin Thickness. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2019.
16. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019 Dec 19;19(1):64.
17. Alsing KK, Serup J. High-frequency ultrasound skin thickness: Comparison of manual reading and automatic border detection includes assessment of interobserver variation of measurement. Ski Res Technol. 2020 Jun 14;(00):srt.12884.
18. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat. 2001 Oct 1;29(5). Available from: https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full
19. Nordhausen K. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition by Trevor Hastie, Robert Tibshirani, Jerome Friedman. Int Stat Rev. 2009;77(3).
20. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on feature distributions. Pattern Recognit. 1996;29(1).
21. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A. caret: Classification and Regression Training. R package version 6.0-90. R package version 6.0-90. 2021. Available from: https://cran.r-project.org/package=caret
22. Sugiyama J, Kobayashi K. wvtool: Image Tools for Automated Wood Identification. R package version 1.0. 2016. Available from: https://cran.r-project.org/package=wvtool
23. Vergilio MM, Vasques LI, Leonardi GR. Characterization of skin aging through high-frequency ultrasound imaging as a technique for evaluating the effectiveness of anti-aging products and procedures: A review. Ski Res Technol. 2021;27(5):966–73.
24. de Rigal J, Escoffier C, Querleux B, Faivre B, Agache P, Lévêque J-L. Assessment of Aging of the Human Skin by In Vivo Ultrasonic Imaging. J Invest Dermatol.

1989;93(5):621–5.

25. Carvalho PRS, Sumita JM, Soares JLM, Sanudo A, Bagatin E. Forearm skin aging: characterization by instrumental measurements. Int J Cosmet Sci. 2017;39(5):564–71.
26. Lasagni C, Seidenari S. Echographic assessment of age-dependent variations of skin thickness. Ski Res Technol. 1995;1(2):81–5.
27. Sandby-moller J, Wulf HC. Ultrasonographic subepidermal low-echogenic band, dependence of age and body site. Ski Res Technol. 2004;10(1):57–63.
28. Lévêque JL. Skin Capacitance Imaging. In: Non Invasive Diagnostic Techniques in Clinical Dermatology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 177–86.