

Personal pdf file for
R. P. Kosilek, J. Schopohl, M. Grunke, M. Reincke,
C. Dimopoulou, G. K. Stalla, R. P. Würtz,
A. Lammert, M. Günther, H. J. Schneider

With compliments of Georg Thieme Verlag

www.thieme.de

Automatic Face Classification of Cushing's Syndrome in Women – A Novel Screening Approach

DOI 10.1055/s-0033-1349124
Exp Clin Endocrinol Diabetes

For personal use only.
No commercial use, no depositing in repositories.

Publisher and Copyright
© 2013 by
Georg Thieme Verlag KG
Rüdigerstraße 14
70469 Stuttgart
ISSN 0947-7349

Reprint with the
permission by
the publisher only

 **Thieme**

Automatic Face Classification of Cushing's Syndrome in Women – A Novel Screening Approach

Authors

R. P. Kosilek¹, J. Schopohl¹, M. Grunke¹, M. Reincke¹, C. Dimopoulou^{1,2}, G. K. Stalla², R. P. Würtz³, A. Lammert⁴, M. Günther⁵, H. J. Schneider¹

Affiliations

Affiliation addresses are listed at the end of the article

Key words

- pituitary adenoma
- cancer
- glucocorticoids
- hormones
- GH adrenal gland
- hormones

Abstract



Objective: Cushing's syndrome causes considerable harm to the body if left untreated, yet often remains undiagnosed for prolonged periods of time. In this study we aimed to test whether face classification software might help in discriminating patients with Cushing's syndrome from healthy controls.

Design: Diagnostic study.

Patients: Using a regular digital camera, we took frontal and profile pictures of 20 female patients with Cushing's syndrome and 40 sex- and age-matched controls.

Measurements: Semi-automatic analysis of the pictures was performed by comparing texture and geometry within a grid of nodes placed on the pictures. The leave-one-out cross-validation method was employed to classify subjects by the software.

Results: The software correctly classified 85.0% of patients and 95.0% of controls, resulting in a total classification accuracy of 91.7%.

Conclusions: In this preliminary analysis we found a good classification accuracy of Cushing's syndrome by face classification software. Testing accuracy is comparable to that of currently employed screening tests.

Introduction



Cushing's syndrome (CS) is characterized by the metabolic and phenotypical manifestations of endogenous or iatrogenic hypercortisolism. The typical metabolic complications (hypertension, diabetes, obesity, osteoporosis) and delayed diagnosis cause the increase in mortality and morbidity in CS patients [1]. Early diagnosis and treatment with consequent normalization of blood cortisol levels and metabolic parameters greatly improves disease outcome in CS patients [2]. Some clinical signs and symptoms of hypercortisolism are non-specific and can be seen in form of the metabolic syndrome in the general population, making it difficult to recognize [3]. Face classification software has been introduced as a new promising tool to screen for diseases that manifest with changes of facial features [4, 5].

We hypothesized that face classification software might aid in the screening for CS.

Subjects and Methods



Subjects

The study was approved by the ethics committee of LMU Munich and conforms to the declaration of Helsinki. All subjects gave written informed consent for participation and separate consent for publication of the pictures shown in **Fig. 1**. We recruited patients from the endocrine outpatient and inpatient clinics of the LMU Munich, the Mannheim University Hospital, and the Max Planck Institute of Psychiatry, Munich, from the rheumatologic outpatient clinic of the LMU and from an endocrinological patient information event. We included only female patients with a biochemically confirmed diagnosis of CS (established on the presence of histological proof if available or at least 2 abnormal test results among the following tests: midnight salivary cortisol, 1 mg dexamethasone suppression test, and 24-h urinary cortisol). Additionally, we included patients with iatrogenic CS as defined by oral glucocorticoid treatment greater than 7.5 mg Prednisone equivalent for longer than 3 months. We included 2 female control subjects per patient matched by age (2:1 matching). Exclusion criteria

received 11.03.2013
first decision 01.06.2013
accepted 13.06.2013

Bibliography

DOI <http://dx.doi.org/10.1055/s-0033-1349124>
Published online:
July 17, 2013
Exp Clin Endocrinol Diabetes
© J. A. Barth Verlag in
Georg Thieme Verlag KG
Stuttgart · New York
ISSN 0947-7349

Correspondence

H. J. Schneider

Medizinische Klinik und Poliklinik IV – Ludwig-Maximilians-University
Ziemssenstraße 1
80336 Munich
Germany
Tel.: +49/89/5160 2111
Fax: +49/89/5160 2194
harald.schneider@med.uni-muenchen.de

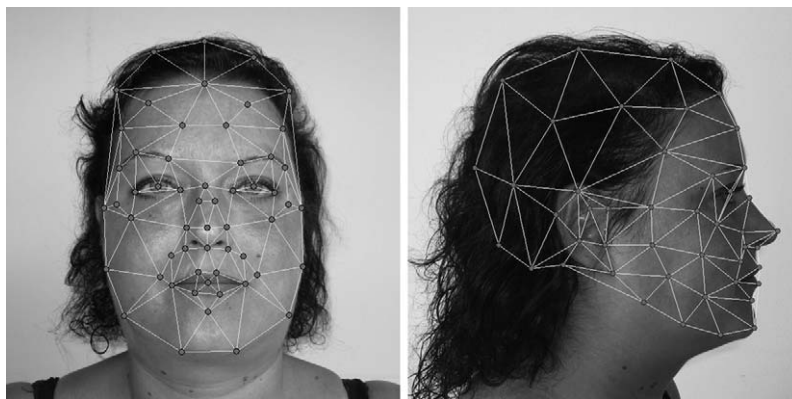


Fig. 1 Representative example of the placement of nodes used for analyses. Nodes are used to calculate functions for the texture around them and distances between them. Diagnosis is based on averages of these readings.

were any diseases that might cause facial changes (particularly CS, acromegaly, Graves disease, or consuming diseases). Controls were recruited from coworkers and patients from the endocrine and rheumatologic outpatient clinics of the LMU.

Picture acquisition & processing

We took frontal and profile pictures with a regular digital camera in a standardized way (homogenous lighting, neutral facial expression and background). The camera models used were Canon Ixus 115 HS (n=18), Canon EOS 450d (n=40), and Casio EX-Z75 (n=2).

Pictures were then processed in a standardized way, which included cropping, resizing, renaming for pseudonymization, and import into the software FIDA (Facial Image Diagnostic Aid, Würtz/Günther, Ruhr-University Bochum).

◉ **Fig. 1** shows examples of patient photographs.

Automatic face classification

After preparation as described above, pictures were semi-automatically 'labeled' with a grid of nodes (face graph) covering critical defined landmarks, i.e., cheeks for plethora, facial margins for width (◉ **Fig. 1**). The positions of the landmarks are standardized and identical for both groups. The labeling procedure was performed with knowledge of the diagnosis. The software then automatically tested the subjects.

Classification into either category was based on comparing texture and geometry of the images. The analysis of texture relies on the calculation of similarity functions for Gabor jets at the nodes. A Gabor jet is a vector that describes the image's texture information around the node [6]. The analysis of geometry is based on the comparison of the distances between nodes. We used the functions P (Gabor jets) and L (geometry) for face classification as described previously [4]. Please see the provided addendum for a more detailed description of the mathematical functions.

The data was evaluated using the 'leave-one-out-cross-validation' method, in which every subject is once excluded from the training set and then tested, thus giving a valid reading for every testing subject.

Statistical analyses

We calculated means and standard deviations (SD) and compared means using Student's t-test. A two-sided p of 0.05 was considered statistically significant. Sensitivity and specificity were calculated to assess test accuracy.

Table 1 Software classification accuracy given in percent – summary of study results. Results are based on analysis of frontal and profile photographs in terms of geometry and texture.

Classification Accuracy (%)	n	FIDA ^a
overall	60	91.7
patients/ <i>sensitivity</i>	20	85.0
endogenous CS	12	75.0
- central CS	8	62.5
- adrenal CS	4	100
iatrogenic CS	8	100
controls/ <i>specificity</i>	40	95.0

^a Facial Image Diagnostic Aid, Würtz/Günther, Ruhr-University Bochum. Trained with a set of 20 female patients and 40 age- and sex-matched controls

Results

▼ A total of 20 female patients were included in the study (mean age ± SD 50.3 ± 15.3 years) (n=8 steroid therapy, n=8 central CS, n=4 adrenal CS). 7 out of initially 27 patients were excluded because of insufficient image standardization. The control group consisted of 40 female subjects (50.2 ± 13.9 years, p=0.98 vs. patients). All patients and controls were Caucasian. We only have data on the BMI for a part of the subject group.

◉ **Table 1** summarizes the classification results. The software correctly classified 85.0% of patients and 95.0% of controls using frontal- and profile photographs, resulting in a total classification accuracy of 91.7%. All cases of iatrogenic and adrenal CS, and 5 out of 8 cases of central CS were correctly identified.

Discussion and Conclusions

▼ In this preliminary study we hypothesized that CS can be detected by face classification software. Using a small sample of subjects, we achieved 85% and 95% classification accuracy for patients and controls, respectively, supporting our hypothesis.

In a recent meta-analysis the sensitivity and specificity of late-night salivary cortisol measurement, a currently recommended first-line screening test, was 92% and 96%, respectively [7,8]. Interestingly, the performance of automatic face classification was almost as high in our study.

Available testing methods for screening and confirmation of diagnosis are an ongoing topic of research. Given the long delay between onset of symptoms and final diagnosis of the disease it is beyond doubt that improvements in screening and early diagnosis are needed.

The strength of this new screening approach lies in its simplicity. Our method merely requires 2 photographs, taken according to simple instructions with a standard digital camera. In the future, we hope to establish a system that will give the physician an immediate classification result upon importing the pictures into the software. Similar to current standard testing procedure, combining this method with other screening tests will further improve accuracy of diagnosis.

The small sample of only female subjects is the key limitation of this study. We do not know how this method performs in men. To eliminate gender-specific facial features as a confounder, analyses must be performed separately by gender. Due to the epidemiology of the disease we could not recruit enough male patients for reliable analyses. We have also tried to minimize ethnicity as a source of error. All patients and controls were Caucasian and we have tried to standardize the setting of the picture acquisition as much as possible.

The fact that 3 different cameras were used for collection of data is a potential source of error. However, we have previously shown that the use of different cameras has minimum impact on classification accuracy as long as standardization instructions are respected [4].

This method shares one major limitation with clinically employed tests. Good performance was only demonstrated for pre-selected patient populations. This is also true for our method, as all patients included had confirmed, active CS or received relevant doses of oral steroid treatment. Generalizability remains to be evaluated in further studies.

The fact that some symptoms of CS are represented in the general population by the rising prevalence of the 'metabolic syndrome' is another shared concern. This issue needs to be addressed by additional studies. While our method might be susceptible to subsequent facial changes, biochemical changes of the metabolic syndrome also inhibit the performance of other methods [9,10]. This study should serve as a proof of concept. We intentionally did not match controls by BMI or include an obese control group, as this represents an advanced problem.

To ensure comparability we have used the same settings for analysis as in our previous study, including the landmarks on the photographs. We are aware that different facial features are relevant to diagnosis in CS and acromegaly. In future analyses we plan to address this issue by changing the placement of landmarks to better represent CS features. Improving the model for face recognition and analysis is an extensive process that first requires building a significantly larger database of subjects. Even though the labeling procedure was done in knowledge of the diagnosis, we expect this to have minimum impact on classification results because of the standardization of landmarks.

We are currently working on simplifying the method with the end-user in mind.

It is of note that only patients with central hypercortisolism were misclassified. We can only speculate on potential reasons for this. This might be a chance finding. On the other hand, slight hyperandrogenism as a typical consequence of ACTH-dependent hypercortisolism may have changed the facial geometry and influenced the classification result.

The implications of developing a technology that reliably diagnoses diseases based on simple photographs must be considered. Beside those concerns that commonly apply to diagnostic

tests, it is important to remember that analysis relies on pictures of the face, clearly identifying the patients. Secure data handling must be ensured before the method can be tested on a larger scale.

We have shown that face classification software can be successfully employed to detect CS on a small scale. These results warrant further research by recruiting more subjects and customizing the method to better match the facial characteristics of the disease.

Author Contributions

▼
RP Kosilek had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Schneider, Schopohl, Kosilek, Günther, Würtz, Reincke.

Acquisition of data: Kosilek, Schneider, Grünke, Dimopoulou, Stalla, Lammert, Reincke.

Development of computer software: Günther, Würtz.

Analysis and interpretation of data: Kosilek, Schneider, Günther, Würtz.

Drafting of the manuscript: Kosilek.

Critical revision of the manuscript for important intellectual content and final approval of the manuscript: Kosilek, Schneider, Schopohl, Grünke, Dimopoulou, Stalla, Lammert, Reincke, Günther, Würtz.

Statistical analysis: Kosilek.

Funding: This work has been partially funded by a German Research Foundation (DFG) grant WU 314/6-2 to R.P. Würtz. The funding source was not involved in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Declaration of interest: RPK and AL report no conflict of interest. HJS and CD received speaker fees and travel grants from Novartis. GKS and MGr received speaker fees from Novartis. JS received speaker fees, travel grants, and research grants from Ipsen and Novartis. MGü and RW received research grants from the DFG related to this work and hold copyrights on the software FIDA. MR is scientific advisory board member for Novartis.

Supplemental Information

▼
Online Addendum: Description of similarity functions.

Affiliations

¹ Med. Klinik und Poliklinik IV, Ludwig-Maximilians-University, Munich, Germany

² Neuroendocrinology Group, Max Planck Institute of Psychiatry, Munich, Germany

³ Institute for Neural Computation, Ruhr-University, Bochum, Germany

⁴ V. Medizinische Klinik, Mannheim University Hospital, Mannheim, Germany

⁵ Idiap Research Institute, Martigny, Switzerland

References

- 1 *Psaras T, Milian M, Hattermann V et al.* Demographic factors and the presence of comorbidities do not promote early detection of Cushing's disease and acromegaly. *Exp Clin Endocrinol Diabetes* 2011; 119: 21–25
- 2 *Biller BM, Grossman AB, Stewart PM et al.* Treatment of adrenocorticotropin-dependent Cushing's syndrome: a consensus statement. *J Clin Endocrinol Metab* 2008; 93: 2454–2462
- 3 *Schneider HJ, Dimopoulou C, Stalla GK et al.* Discriminatory value of signs and symptoms in Cushing's syndrome revisited: what has changed in 30 years? *Clin Endocrinol (Oxf)* 2012
- 4 *Schneider HJ, Kosilek RP, Gunther M et al.* A novel approach to the detection of acromegaly: accuracy of diagnosis by automatic face classification. *J Clin Endocrinol Metab* 2011; 96: 2074–2080
- 5 *Learned-Miller E, Lu Q, Paisley A et al.* Detecting acromegaly: screening for disease with a morphable model. *Med Image Comput Comput Assist Interv* 2006; 9: 495–503
- 6 *Günther M.* Statistical Gabor Graph Based Techniques for the Detection, Recognition, Classification, and Visualization of Human Faces. Shaker Verlag, Aachen: 2012
- 7 *Nieman LK, Biller BM, Findling JW et al.* The diagnosis of Cushing's syndrome: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2008; 93: 1526–1540
- 8 *Carroll T, Raff H, Findling JW.* Late-night salivary cortisol for the diagnosis of Cushing syndrome: a meta-analysis. *Endocr Pract* 2009; 15: 335–342
- 9 *Ellis E, Chin PK, Hunt PJ et al.* Is late-night salivary cortisol a better screening test for possible cortisol excess than standard screening tests in obese patients with Type 2 diabetes? *N Z Med J* 2012; 125: 47–58
- 10 *Baid SK, Rubino D, Sinaii N et al.* Specificity of screening tests for Cushing's syndrome in an overweight and obese population. *J Clin Endocrinol Metab* 2009; 94: 3857–3864



Supplemental Information



Description of similarity functions

Gabor jet-based functions compared texture around the nodes by calculating similarities of the Gabor jets J and J' at the node positions in the corresponding facial images. Geometry-based functions compared distances between edges and nodes. The 5 different Gabor jet-based functions we used were:

- ▶ A (scalar product):

$$S_A(J, J') = \sum_{j=1}^{40} a_j a'_j$$

- ▶ P (scalar product including phase):

$$S_P(J, J') = \sum_{j=1}^{40} a_j a'_j \cos(\phi_j - \phi'_j)$$

- ▶ C (Canberra similarity):

$$S_C(J, J') = 1 - \frac{1}{40} \sum_{j=1}^{40} \frac{|a_j - a'_j|}{a_j + a'_j}$$

- ▶ M (modified Manhattan similarity):

$$S_M(J, J') = 1 - \frac{\sum_{j=1}^{40} |a_j - a'_j|}{(\sum_{j=1}^{40} |a_j|) + (\sum_{j=1}^{40} |a'_j|)}$$

- ▶ D (disparity similarity):

$$S_D(J, J') = \sum_{j=1}^{40} a_j a'_j \cos(\phi_j - \phi'_j - \vec{d}^T \vec{k}_j)$$

The pair (a_j, ϕ_j) represents the result of convolving the picture at a node position with the Gabor wavelet with the parameter vector \vec{k}_j ($j=1, \dots, 40$) with a_j representing the absolute part and ϕ_j representing the phase. The absolute part contains information about the strength of a particular spatial frequency at the image location, while the phase contains information about the relative location. The Gabor jet is then the aggregation of the responses of all 40 wavelet responses. All Gabor jets were normalized to unit length: $\sum_{j=1}^{40} a_j^2 = 1$. D (disparity) estimates the

disparity vector \vec{d} from the given jets J and J' for compensation of the phase difference.

Geometry-based functions compared distances between edges and nodes. The 4 functions based on geometry are:

- ▶ E (Edge Length Difference): $S_E(E, E') = -|\vec{x}^2 - \vec{x}'^2|$
- ▶ L (Edge Difference Length): $S_L(E, E') = -\|\vec{x} - \vec{x}'\|^2$
- ▶ H (Edge Vector Difference): $S_H(E, E') = -|x_1 - x'_1|$ and $-|x_2 - x'_2|$
- ▶ N (Node Position Distance): $S_N(N, N') = -\|N - N'\|$

\vec{x} is the edge vector, representing the directed difference of connected nodes with the horizontal and vertical components x_1 and x_2 , respectively. The functions E and L used one single similarity value per edge, whereas H used 2 similarity values for horizontal and vertical distance separately. For function N, the Euclidean distance between node positions N and N' was calculated.