Sensitivity and Specificity of Confocal Laser-Scanning Microscopy for In Vivo Diagnosis of Malignant Skin Tumors

Armin Gerger, MD1
Silvia Koller, MD2
Wolfgang Weger, MD2
Erika Richtig, MD2
Helmut Kerl, MD2
Hellmut Samonigg, MD1
Peter Krippel, MD1
Josef Smolle, MD3

1 Department of Internal Medicine, Division of Oncology, Medical University of Graz, Graz, Austria.
2 Department of Dermatology, Division of Analytic-Morphologic Dermatology, Medical University of Graz, Graz, Austria.
3 Institute for Medical Informatics, Statistics, and Documentation, Medical University of Graz, Graz, Austria.

BACKGROUND. Melanoma and nonmelanoma skin cancer are the most frequent malignant tumors by far among whites. Currently, early diagnosis is the most efficient method for preventing a fatal outcome. In vivo confocal laser-scanning microscopy (CLSM) is a recently developed potential diagnostic tool.

METHODS. One hundred seventeen melanocytic skin lesions and 45 nonmelanocytic skin lesions (90 benign nevi, 27 malignant melanomas, 15 basal cell carcinomas, and 30 seborrheic keratoses) were sampled consecutively and were examined using proprietary CLSM equipment. Stored images were rated by 4 independent observers.

RESULTS. Differentiation between melanoma and all other lesions based solely on CLSM examination was achieved with a positive predictive value of 94.22%. Malignant lesions (melanoma and basal cell carcinoma) as a group were diagnosed with a positive predictive value of 96.34%. Assessment of distinct CLSM features showed a strong interobserver correlation ($\kappa > 0.80$ for 11 of 13 criteria). Classification and regression tree analysis yielded a 3-step algorithm based on only 3 criteria, facilitating a correct classification in 96.30% of melanomas, 98.89% of benign nevi, and 100% of basal cell carcinomas and seborrheic keratoses.


KEYWORDS: confocal laser-scanning microscopy, sensitivity, specificity, malignant melanoma, basal cell carcinoma.

Early detection of malignant skin tumors is essential and is among the most challenging problems in clinical dermatoncology. Whereas surgical excision in early stages of tumor development almost always is curative, delayed recognition of skin malignancies puts the patient at risk for destructive growth and death from disease once the tumor has progressed to competence for metastasis.

However, the ability to make an early diagnosis of a malignant skin tumor has remained poor. Even in specialized centers, the accuracy of the clinical diagnosis for melanoma achieved with the unaided eye is only slightly better than 60%.

Technologic advancements have led to the development and investigation of imaging tools to provide information to the clinician that can improve the diagnostic performance for early diagnosis and assist in the management of cutaneous malignancies. Consequently, epiluminescence microscopy (dermoscopy) was introduced successfully in routine skin tumor screening. Dermoscopy is a noninvasive, in vivo examination with a microscope that uses incident light and oil.
immersion to make subsurface structures of the skin accessible to visual examination and therefore permits a more detailed inspection of skin tumors. In recent, systematic reviews of the diagnostic accuracy of dermoscopy in detecting malignant melanoma, an improvement of 49% and sensitivity and specificity values of approximately 86% and 89%, respectively (compared with inspection by the unaided eye), were reported. Despite several achievements, the relatively low magnification (usually ×10) in routinely applied instruments and the limited scope of observable structures restrict the usefulness of the method. Among novel noninvasive imaging techniques, confocal laser-scanning microscopy (CLSM) stands out because of its high resolution. CLSM facilitates in vivo examination of the skin at a level that allows the visualization of microanatomic structures and individual cells. High-contrast images are obtained by imaging a single, in-focus section and rejecting light reflected from out-of-focus portions of the object by means of a pinhole detector aperture that acts as a spatial filter. CLSM can image the epidermis and the dermis noninvasively with cellular-level resolution, producing impressive pictures of living tissues that represent horizontal planes of the skin.

After initial observations of healthy skin and various pathologic conditions, CLSM currently is being explored for diagnostic differentiation of skin tumors. Several diagnostic morphologic features of melanocytic and nonmelanocytic skin tumors determined by in vivo confocal microscopy have been investigated previously, and there is hope that established criteria may aid in improving its diagnostic accuracy. However, the numbers of studies and patients examined have been small and limited to melanocytic or nonmelanocytic skin tumors, which is why there is insufficient knowledge regarding the sensitivity and specificity of various morphologic criteria. In a first step toward exploring the potential use of CLSM in routine skin tumor practice, we conducted a prospective and observer-blinded study to systematically investigate the diagnostic impact and reliability of well-described morphologic features in a large series of melanocytic and nonmelanocytic skin tumors.

**MATERIALS AND METHODS**

**Patients**

One hundred nineteen patients (62 males and 57 females) were recruited prospectively from the Dermatooncology Clinic at the Department of Dermatology, Medical University of Graz (Graz, Austria) over 2 years. All patients provided informed consent for examination of their lesions by CLSM. All institutional rules governing clinical investigation of human subjects were followed strictly. We conformed to the Helsinki Declaration with respect to human subjects in biomedical research. Overall, 117 melanocytic skin lesions and 45 nonmelanocytic skin tumors, including malignant melanoma (MM), benign nevi (BN), basal cell carcinoma (BCC), and seborrheic keratosis (SK), were imaged consecutively by using a confocal microscope (Table 1). The tumors were not selected in any way for their CLSM features, nor was any tumor lacking particular CLSM features that were excluded from the study set. Seventy-two of 162 tumors (44%) that were included in the study were excised after clinical, dermoscopic, or confocal examination and were subjected to standard histopathologic assessment. The remaining 90 lesions were diagnosed on proven clinical and conventional dermoscopic criteria.

**In Vivo CLSM**

In vivo CLSM was performed with a commercially available, near infrared, reflectance confocal microscope (Vivascope 1000; Lucid Inc., Rochester, NY). The Vivascope 1000 uses a diode laser at 830 nm wavelength and a power of <35 mW at the tissue level. Because of the low power of the diode laser, no tissue damage occurs. A ×30 water-immersion objective lens with a numerical aperture of 0.9 is used with water (refractive index, 1.33) as an immersion medium. It images with a spatial resolution of from 0.5 μm to 1.0 μm in the lateral dimension and from 3.0 μm to 5.0 μm in the axial dimension, providing insight into cellular structures of the examined specimens in vivo. Usually, an examination depth of 350 μm can be reached, which corresponds to the papillary dermis. Placing the objective lens onto an adapter ring, which is fixed on the tumor, real-time images can be acquired in seconds “at the bedside.” All images that were obtained by CLSM in the study correspond to sections in the horizontal plane. At least 5 images,
comprising the stratum corneum, stratum granulorum, stratum spinosum, dermoepidermal junction, and papillary dermis, were recorded in each patient and were stored using a BMP file format.

**Diagnostic Morphologic CLSM Features**

Morphologic features of melanocytic skin tumors were assessed according to the results of our recently published investigations. The most striking finding was the identification of melanocytic cytomorphology and architecture, keratinocyte cell borders, and complex branching dendrites as highly diagnostic criteria. The set of confocal BCC features was selected based on qualitatively described criteria from previously published studies. Vascular architecture, tumor cells in a streaming pattern, and collagen fiber bundles were taken into account for diagnostic decisions and were discussed by the authors. In contrast, SK features were assessed solely based on well known, standard criteria used in conventional histopathology, because, until recently, data on the confocal examination of SK morphology were lacking.

**Training Data and Study Setting**

Four independent dermatoncologists without previous experience in CLSM received a standardized instruction about diagnostic CLSM features of MM, BN, BCC, and SK for 1 hour as a Power-Point presentation. Diagnostic criteria were explained, and 26 image examples were demonstrated for training purposes. For diagnostic assessment of the test set, 2 diagnostic images of each of the 162 skin tumors were shown on a computer screen using a macro procedure and were evaluated as either MM, BN, BCC, or SK by each of observer. In a second run, the presence or absence of each of the morphologic features was assessed by 2 observers irrespective of their assumed diagnosis. To ensure strict separation of learning and test set, none of the specimens presented in the training sample were used in the validation set. Furthermore, all of the experimenters were blinded with regard to the clinical and conventional dermoscopic or histopathologic diagnosis of the tumors.

**Statistical Analysis**

Statistical analyses (sensitivity, specificity, positive predictive value [PPV], negative predictive value [NPV], median value, mean value, standard deviation, and \( \kappa \) statistic) were performed by using SPSS statistical software package for Windows (version 12.0; SPSS Inc., Chicago, IL). Reliability data (interobserver agreement) were produced in the form of the \( \kappa \) statistic. Kappa \( \kappa \) takes a value between 0 (no agreement) and 1 (perfect agreement); therefore, it was assumed that reliability was highly specific when \( \kappa \) was >0.8, excellent when \( \kappa \) was >0.6, moderate when \( \kappa \) was >0.4, and poor when \( \kappa \) was ≤0.4. For classification purposes, we used the Classification and Regression Tree (CART) software (version 4.0; Salford Systems, San Diego, CA). In CART analysis, classification trees are applied to learning sets in the search for optimal split criteria, which facilitate an optimal classification of all elements to the particular class.

**RESULTS**

**General Observations**

The test set comprised 162 skin tumors, including 27 MM lesions and 15 BCC lesions (all histologically verified), 90 BN lesions (30 histologically verified), and 30 SK lesions (none histologically verified). Among the MM lesions, 5 were in situ, 8 showed regression structures (fibrosis with melanophages), 3 showed ulceration, 8 were the nodular type, and none were amelanotic. The median and mean Breslow thickness were 1.2 mm and 1.75 mm, respectively (standard deviation, 1.51 mm; range, from in situ to 6 mm). Among the BCC lesions, 8 were nodular, 2 were the superficial type, and 5 were sclerodermatous (morphea type). Three of the BCC lesions showed ulceration. Within the set of BN lesions, there were 52 junctional, compound, or dermal nevi (6 histologically verified) and 29 dysplastic nevi (20 histologically verified) identified. The remaining 9 BN lesions included congenital, inkspot, hypermelanotic, blue, and Reed nevi (4 histologically verified). Thus, the test set comprised a variety of subtypes of the examined skin tumors.

**Qualitative Description of CLSM Criteria**

In general, melanocyte tumor cells could be delineated clearly from nonmelanocytic cytomorphology. In BN, round to oval, bright, and monomorphic cells were found (Fig. 1), whereas MM tended to present polymorphic and irregular cells that were seen frequently with complex branching dendrites (Fig. 2). More or less clearly defined, junctional and dermal nevus cell nests were found in BN (Fig. 1), whereas MM showed disarray of the melanocytic cell architecture (Fig. 2). Keratinocyte cell borders could be detected readily or showed focal absence in BN (Fig. 1) and were poorly defined or absent in MM (Fig. 2). In BCC, an increase in the number and dimension of blood vessels with horizontally orientated vessels in parallel arrangement and loss of vascular architecture was found (Fig. 3A). Furthermore, large, elongated cells with dark nuclei in a streaming pattern could be detected clearly (Fig. 3B). In addition, greater numbers of bright collagen fiber bundles stood out (Fig. 3C). In SK, disarray of the stratum corneum architecture was
recognized (Fig. 4A), and greater amounts of melanin in keratinocytes were found (Fig. 4B). Occasionally, cystic inclusions could be detected (Fig. 4C).

**Sensitivity and Specificity**

Diagnostic differentiation of MM from BCC, BN, and SK reached sensitivity and specificity values of 85.19% and 98.52% (Observer 1), 92.59% and 98.52% (Observer 2), and 92.59% and 99.26% (Observers 3 and 4), respectively with the following overall performance: sensitivity, 90.74%; specificity, 98.89%; PPV, 94.22%; and NPV, 98.17%. Correlations between the assumed CLSM diagnoses and the assessed pathologic or clinical diagnoses are shown in Table 1. When the sepa-

**FIGURE 1.** In benign nevi, round-to-oval and bright tumor cells are seen in well defined nests. Note the indicated honeycomb pattern between the tumor nests that represents keratinocytes.

**FIGURE 2.** In malignant melanoma, there is disarray of large and polymorphic tumor cells with complex branching dendrites and a loss of keratinocyte cell borders.

**FIGURE 3.** (A) In basal cell carcinoma (BCC), loss of vascular architecture is observed with an increase in the dimension of a blood vessel. Because of trafficking blood cells in real-time observation, vessels can be detected easily. (B) In this BCC, elongated tumor cells are seen with dark nuclei. In contrast to the honeycomb appearance of keratinocytes, BCC cells show a streaming pattern. (C) Bright collagen fiber bundles are seen in this BCC.
RATION OF BENIGN VERSUS MALIGNANT SKIN TUMORS WAS EVALUATED, SLIGHTLY HIGHER DIAGNOSTIC PERFORMANCE, WITH SENSITIVITY AND SPECIFICITY OF 90.48% AND 98.33% (OBSERVER 1), 95.24% AND 98.33% (OBSERVER 3), AND 95.24% AND 99.17% (OBSERVERS 2 AND 4), RESPECTIVELY, COULD BE ACHIEVED, WITH THE FOLLOWING OVERALL PERFORMANCE: SENSITIVITY, 94.05%; SPECIFICITY, 98.75%; PPV, 96.34%; AND NPV, 97.94%. TAKING INTO ACCOUNT ONLY THE BIOPSY DOCUMENTED LESIONS (72 OF 162 SPECIMENS, INCLUDING ALL MM AND BCC SPECIMENS AND 30 BN SPECIMENS) FOR THE BENIGN VERSUS MALIGNANT CLASSIFICATION, SENSITIVITY AND SPECIFICITY VALUES OF 90.48% AND 96.67% (OBSERVER 1), 95.24% AND 100% (OBSERVER 2), 95.24% AND 96.67% (OBSERVER 3), AND 97.62% AND 100% (OBSERVER 4), RESPECTIVELY, WERE FOUND, WITH THE FOLLOWING OVERALL PERFORMANCE: SENSITIVITY, 94.65%; SPECIFICITY, 96.67%; PPV, 97.50%; AND NPV, 92.99%.

DIAGNOSTIC IMPACT AND RELIABILITY OF MORPHOLOGIC FEATURES

WHEN THE PRESENCE OR ABSENCE OF MORPHOLOGIC FEATURES WAS ASSESSED BY 2 OBSERVERS (TABLE 2), CLASSIFICA-

FIGURE 4. (A) IN SEBORRHEIC KERATOSIS (SK), THE STRATUM CORNEUM PRODUCED THE FIRST IMAGE OF THE TOP SURFACE OF THE SKIN. ACCORDING TO THE CLINICAL APPEARANCE, DISARRAY OF THE ARCHITECTURE WAS FOUND. (B) A GREATER AMOUNT OF MELANIN IN KERATINOCYTES WAS FOUND FREQUENTLY IN SK, WHICH IS DISTINGUISHABLE CLEARLY FROM THE MELANOCYTIC TUMOR CELLS. (C) THIS SK SHOWS A CYSTIC INCLUSION THAT REPRESENTS A HORN CYST.

TABLE 2
MEAN FREQUENCY (IN %) OF THE OCCURRENCE OF EACH INDIVIDUAL FEATURE IN ALL CONFOCAL TUMOR IMAGES EVALUATED BY TWO OBSERVERS IRRESPECTIVE OF THEIR DIAGNOSIS

<table>
<thead>
<tr>
<th>Feature</th>
<th>MM (n = 27)</th>
<th>BN (n = 90)</th>
<th>BCC (n = 15)</th>
<th>SK (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monomorphous melanocytic cells</td>
<td>0.0</td>
<td>98.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Melanocytic cell nests</td>
<td>0.0</td>
<td>93.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Readily detected keratinocyte cell borders</td>
<td>3.7</td>
<td>91.7</td>
<td>0.0</td>
<td>1.7</td>
</tr>
<tr>
<td>Polymorphous melanocytic cells</td>
<td>98.1</td>
<td>1.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Disarray of melanocytic architecture</td>
<td>98.1</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Poorly defined or absent keratinocyte cell borders</td>
<td>92.6</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Complex branching dendrites</td>
<td>92.6</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Loss of vascular architecture</td>
<td>0.0</td>
<td>0.0</td>
<td>86.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Elongated shaped cells in streaming architecture</td>
<td>0.0</td>
<td>0.0</td>
<td>50</td>
<td>0.0</td>
</tr>
<tr>
<td>Bright collagen fiber bundles</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>0.0</td>
</tr>
<tr>
<td>Disarray of stratum corneum architecture</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>96.7</td>
</tr>
<tr>
<td>Cystic inclusions</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>56.7</td>
</tr>
<tr>
<td>Greater amount of melanin in keratinocytes</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>93.3</td>
</tr>
</tbody>
</table>

MM INDICATES MALIGNANT MELANOMA, BN, BENIGN NEVI; BCC, BASAL CELL CARCINOMA; AND SK, SEBORRHEIC KERATOSIS.

* NOTE THE CLEAR-CUT SEPARATION WITHIN THE NONMELANOCYTIC TUMORS AND THE NEARLY CLEAR-CUT SEPARATION BETWEEN MELANOCYTIC AND NONMELANOCYTIC TUMORS. THIS IS REFLECTED IN THE CLASSIFICATION AND REGRESSION TREE ANALYSIS RESULTS.
tion tree software (CART; Salford Systems) was applied on the data set to search for optimal split features, which facilitate an optimal classification of all tumors. Furthermore, CART automatically performed a ranking of all features that depended on their diagnostic value. In the order of the analysis ranking, the results indicated out that mainly monomorphic melanocytic cells, bright collagen fiber bundles, and disarray of the melanocytic architecture followed by polymorphic melanocytic cells, melanocytic cell nests, and greater amounts of melanin in keratinocytes were taken into account for diagnostic classification by the software. Readily detected or focal absence and poorly defined keratinocyte cell borders, complex branching dendrites, and disarray of the stratum corneum and vascular architecture also were identified as good working features. In contrast, elongated cells with dark nuclei in a streaming pattern and cystic inclusions had less to no diagnostic importance. Overall, using only the assessed presence or absence of the top 3 confocal features, the CART software (Fig. 5) correctly classified 96.30% of MM lesions, 98.89% of BN lesions, and 100% of BCC and SK lesions. When each feature was measured for its reliability (interobserver agreement) by using the statistic, the results showed that most of the diagnostic criteria were highly reliable, indicating good definitions of the morphologic features (Fig. 6).

DISCUSSION
CLSM is a novel, high-resolution imaging technique that opens a window into living tissue. It has been deemed a safe procedure with no evidence of tissue damage from the low-level energy laser beams. Many different sites and skin tumors can be examined during the same office visit, and the same tumor can be examined at different times. Images can be stored electronically and can be shared quickly for consultations. The time requirement for examining skin lesions is comparable to that required for routinely practiced digital dermoscopy. In contrast to dermoscopy, which is limited by the low magnification (usually ×10) in routinely applied instruments, CLSM offers the unique opportunity to analyze skin structures noninvasively at a "quasihistopathologic" resolution.

The current results demonstrate the application of CLSM to the diagnostic classification of benign and malignant melanocytic and nonmelanocytic skin tumors. A large number of tumors comprising a variety of subtypes were imaged by using CLSM and were evaluated in an observer-blinded manner to determine sensitivity and specificity of the diagnosis. The analysis was sterile and artificial, in that no clinical diagnosis or dermoscopic appearance was taken into account for diagnostic decisions.

Four independent observers received a standardized 1-hour instructional presentation about diagnostic CLSM features and evaluated 162 skin tumors. None of the observers in this study had been trained formally in CLSM or had any previous experience with this method. It is important to note that the confocal morphologic features that were used for evaluating the test set are easy to learn and use, and this was reflected in the interobserver agreement. In contrast, to become proficient in dermoscopy, formal training of at least 9 hours is required. Moreover, the diagnostic accuracy is no better with dermoscopy applied by nonexperts than with the unaided eye. This finding underlines the necessity of intensive training for the application of dermoscopy.

When all morphologic features were taken into account for diagnostic decisions by the 4 observers, significantly higher diagnostic accuracy than has been reported for dermoscopic examination could be achieved for the diagnosis of melanoma. In evaluating benign versus malignant skin tumors, which is often the main objective in routine screening practice, further increases in sensitivity and specificity were found, as expected. In contrast, taking into account only the biopsy-documented lesions, slightly lower specificity value was reported. Moreover, all observer showed similar sensitivity and specificity values, indicating a stable diagnostic performance.

To evaluate the diagnostic power of each individual feature in an objective manner, classification tree software was applied. Highly diagnostic features could be identified by the software and were incorporated automatically into the classification process. It is noteworthy that, in the CART analysis, the evaluation of only 3 features, namely, monomorphic melanocytic...
cells, bright collagen fiber bundles, and disarray of the melanocytic architecture, was required to reach sensitivity and specificity values superior to those achieved by any of the observers, taking into account all morphologic features for their diagnostic decisions.

One limitation in the current state of technologic CLSM development that has to be addressed is that the assessment of microanatomic structures can be done only to a depth of 350 μm, which corresponds to the papillary dermis. Therefore, processes in the reticular dermis and tumor invasion depth cannot be evaluated reliably. Furthermore, we could not conclude from the results of 4 independent observers that similar classification results would be achieved by the majority of dermatooncologists in everyday practice.

The cumulative experience with CLSM by different investigators clearly holds promise for this technology in the future. The results of the current study underline the finding that CLSM merits application for use as a screening tool in skin oncology.

REFERENCES


