Screening for dry eye disease using infrared ocular thermography

DOI:
10.1016/j.clae.2016.08.004

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Contact Lens and Anterior Eye

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Screening for dry eye disease using infrared ocular thermography

Li Li Tan* MMedSc (Optom), BOptom (Hons), FAMO, FIACLE
Srinivasan Sanjay§ MBBS, MRCS(Edin), MMed(Oph)(Sing), MS(Oph), DNB(Oph)
Philip B. Morgan† PhD, MCOptom, FAAO, FBCLA

*School of Chemical and Life Sciences, Singapore Polytechnic, Singapore
§Ophthalmology and Visual Sciences, Khoo Teck Puat Hospital, Singapore
§Yong Loo Lin School of Medicine, National University of Singapore, Singapore
†Eurolens Research, The University of Manchester, UK

Correspondence to:
Li Li Tan
500 Dover Road
School of Chemical and Life Sciences
Singapore Polytechnic
Singapore 139651
Fax: (65) 68704617
tanlili@sp.edu.sg

Number of tables: 4
Number of figures: 5
Introduction

Dry eye disease (DED) is a commonly encountered condition in clinical practice and affects up to 12.3% of the population in Singapore [1] with a world prevalence range of 5% to 38% [2]. The condition has remarkably impact on daily social and physical functioning, work place productivity and quality-of-life [3-6]. Diagnosing the disease can be a tedious and challenging task [7] and been hampered by the lack of objective tests with sufficient sensitivity and specificity, adequate repeatability, ease of performance, and suitability for the clinical practice setting particularly in early or mild cases [8]. Due to its multifactorial nature, DED potentially requires a broad spectrum of test measures in the monitoring of its diagnosis and treatment [9]. While there are many clinical tests for DED, the diagnostic values can be inconclusive [10-11] and may not be repeatable and/or reliable because of variable results, poor reproducibility and low sensitivity [12-15]. Determining the cause of dry eye when minimal clinical signs are present is difficult and the diagnosis is complicated further when there is a lack of correlation between its signs and symptoms [11,16-22].

Tear film stability is a key test in screening and diagnosing DED [8]. It has been reported that capturing ocular surface temperature (OST) changes using infrared (IR) ocular thermography reflects the nature of the tear film [23-24] and its stability [25] and can be used to screen DED [26, 27]. However, temperature metrics available from ocular thermography to screen DED were limited and remain unclear. Most studies on OST and dry eye have evaluated the temperature of the geometric center of the cornea [23,26,28-30]. A small number of studies evaluated other metrics such as the relative differences in temperature across the ocular surface [31,32], mean ocular surface temperature [32,33], temperature at the nasal and temporal conjunctiva [26] and temperature difference and compactness values of the OST [27]. Two reports on dry eye screening using ocular thermography were done using temperature of the geometric center of the cornea [26] and temperature difference and compactness values of the OST [27].

The current study was devised to evaluate the efficacy of IR ocular thermography as a diagnostic tool for DED and to determine the most effective temperature metrics, applied singly or in combination.

Methods

Subjects

The research protocol was approved by the Singapore National Health Group (NHG) Domain-Specific Review Board (DSRB) and the Singapore Polytechnic ethics review committee and the work adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from each subject at study enrolment. Sixty-two dry eye (48 ± 10 years; 14 Male and 48 Female) and 63 age- and sex-matched control subjects (46 ± 7 years; 16 Male and 47 Female) completed the study with the age matching range of ± 2 years. This study included mild or moderate dry eye subjects. They were classified based upon a composite disease severity index, derived from the Dry Eye Workshop severity scale [34]. The inclusion criteria for the dry eye subjects were as described
previously [35]: use of tear replacement therapy and had either a fluorescein tear break-up time of 10 seconds or less [36], or a Schirmer I test result of less than 10 mm in 5 min [31] along with presence of corneal or conjunctival staining. All dry eye patients were screened and diagnosed by an ophthalmologist at Khoo Tech Puat Hospital eye clinic. Control subjects were those not using tear replacement therapy or any topical medication and without signs or symptoms of dry eye. All subjects were required to be noncontact lens wearers for at least two years prior to enrolment. Subjects were excluded from control group if they had Schirmer I test result of less than 10 mm in 5 min or fluorescein tear break-up time of 10 seconds or less. Subjects with any anterior ocular anomalies (e.g. current ocular infection, allergy or ptosis), those undergone surgery or taking any medication that could affect the tear film or who were currently pregnant or breastfeeding were also excluded.

**Procedures**

The procedures were the same as described previously [35]. Subjects were asked to refrain from using their eye-drops or eye make-up on the day of measurement. Ocular thermography was performed in real time using an Infrared thermo-tracer (NEC TH9420) with resolution of 640 (H) x 480 (V) pixels, operational sensitivity of 0.06 °C and frequency of 30 frames per second, detecting infrared radiation between 8 and 14 µm. The emissivity of 0.98 was assumed [37]. A standard examination protocol as reported in the literatures [23,25,26,32] was adopted. All the measurements were performed from 9 am to 2 pm in the same room with controlled room temperature (24.06 ± 0.41 °C) and humidity (49.76 ± 2.61 %), with no air drifts and same brightness (380 lux). Subjects were adapted to the room for 20 minutes prior to ocular thermography as previous work has shown that this period was necessary to achieve ocular temperature stabilisation [38]. OST was recorded under the conditions described by Morgan and associates: the subjects blinked normally, closed for 3 s and the first image was recorded just after the eyes had opened [32,39]. 0 s was recorded as the time upon eye opening. 300 frames of real time thermal images reflecting OST changes at the ocular surface were captured over 10 s sustained eye opening. The measurement was done three times on right eye followed by left eye. At any time if subject blinked or changed fixation before 10 s, the measurement was discounted and repeated.

A novel ‘diamond’ method was used to mark the ocular surface using a custom-designed OST Analysis V2 software (developed using Matlab Simulink 7.11.0, R2010b). The region of interest (ROI) formed by five anatomical points (labelled as 1 - 5) shaped like a diamond (Fig. 1). This method has the advantages of (1) overcome problems of truncated image by upper lids [40] and (2) minimize possible inconsistency in OST acquisition due to variation in palpebral aperture size and (3) enable study of the inferior zone of the ocular surface that reported to be a predictive area in detection of dry eye subtypes [41]. Each point marked represents an area of 3 × 3 pixels so that temperature was an average of nine pixels:-

1 - Temporal limbus (LT)
2 - Nasal limbus (LN)
3 - Extreme temporal conjunctiva (T1)
4 - Extreme nasal conjunctiva (T4)
5 - Most inferior point of the ocular surface

’Insert Fig. 1 here’

Once the marking of ROI was completed, OST acquisition and processing was performed automatically by double clicking the last point marked (point 5) to activate the OST Analysis V2 program and process all the 300 frames. All frame marking and data processing were undertaken by a single examiner (LL). Ten OST indices of the ocular surface were generated as shown in Table 1. In this study, GCC denotes the temperature of the geometric center of the cornea, obtained midway between LT and LN. The OST indices were selected to document the whole inferior zone of the exposed ocular surface within ROI and to include as far as possible, all the reported temperature metrics [23, 26-33]. All the ten OST indices extracted by the ‘diamond’ method has shown to be highly repeatable in assessing healthy and dry eyes [35] in terms of inter-image, inter-examiner and intra-examiner variability.

’Insert Table 1 here’

Data Analysis

Data on all 62 dry eye subjects and 63 controls were tabulated and analysed. The ten OST indices were studied in two aspects: static and dynamic measures. To prevent difficulties arising when non-independent data were collected from both eyes, only data obtained from right eye were used in the analysis [42].

Static measures were study of absolute OST at t = 0 s, 5 s and 10 s (3 static attributes) after eye opening. Data were obtained directly from the raw data. As we had ten OST indices, 30 static temperature metrics were generated. For example, GCC temperature at 0 s, 5 s and 10 s were labelled as GCC-0, GCC-5 and GCC-10 respectively.

Dynamic measures were study of temperature change over time. One phase exponential curve was fitted to the temperature vs. time data for each series of images using JMP 11 according to the following model:

\[ \text{Temperature} = a + b \times \text{Exp}(c \times \text{time}), \]

We adopted the terminology used by JMP (http://www.jmp.com; SAS Institute Inc., USA) as used the terms ‘asymptote’ (a), ‘scale’ (b) and ‘growth rate’ * [3 dynamic attributes] (c). Again, as we had
ten OST indices, in total 30 dynamic temperature metrics were generated. For example, asymptote, scale and growth rate for GCC temperature were labelled as GCC-A, GCC-S and GCC-GR respectively. The efficacy of ocular thermography in diagnosing DED was then evaluated in two phases: singly and in combination.

**Phase 1: Evaluating the efficacy of the 30 static- and 30 dynamic-metrics when applied singly**

Findings on dry eye subjects and their controls for each metric were compiled. Using GraphPrism 6 (www.graphpad.com; GraphPad Software Inc., USA), a range of testing threshold/criterion with their sensitivity, specificity, area under the ROC curve (AUC), predictive values [43,44] were then been derived. Receiver operating characteristics (ROC) [45,46] curves, which determine the static and 30 dynamic-metrics and AUC was extracted using trapezoidal numerical integration. The AUC (range: 50 to 100 in percentage) is a quantitative representation of overall test accuracy, where values from 50 to 70 represent low accuracy, values from 70 to 90 represent tests that are useful for some purposes/moderate accuracy, and values > 90 represent tests with high accuracy [48]. Metrics that has AUC of 70 or above were selected.

Although AUC is one of the main parameters of ROC, to better evaluate the performance of the technique as a detector, (i.e. the discrimination between dry eyes from those of normal subjects) it is important to determine the cutoff values, selected as the one that optimizes both sensitivity and specificity [43,49]. Sensitivity is the proportion of actual positives (i.e. dry eye subjects) that are correctly identified, while specificity is the proportion of actual negatives (i.e. normal subjects) that are correctly identified. To facilitate the detector performance comparison, we evaluated every tested metric by means of a set of statistical tools [50], namely the Youden’s index (\( \gamma \)) [51] and the discrimination power (DP) [50].

Youden’s index evaluates the algorithm’s ability to avoid failure and follows the expression

\[
\gamma = \text{sensitivity} + \text{specificity} - 100.
\]

Its value ranges from 0 to 100 in percentage, and has a zero value when a diagnostic test gives the same proportion of positive results for groups with and without the disease, i.e the test is useless. A value of 100 indicates that there are no false positives or false negatives, i.e. the test is perfect. Youden’s index is often used in conjunction with ROC analysis [52]. Graphically, \( \gamma \) is the maximum vertical distance between the ROC curve and the diagonal line. Cutoff values of the selected metrics can then be determined as the criterion that maximized the Youden index: max (sensitivity + specificity – 100), where \( c \) ranges over all possible criterion values [51]. The cutoff values that achieves this maximum is referred to as the optimal cutoff because it is the cutoff values that
optimizes the metric’s differentiating ability when equal weight is given to sensitivity and specificity [43,49,52-54].

DP is a measurement that summarizes sensitivity and specificity of the technique,

$$DP = \sqrt{3/\pi} \log X + \log Y,$$

Where X = sensitivity / (100-sensitivity) and Y = specificity / (100-specificity). Values of DP < 1 indicate poor discrimination performance, DP < 2 indicates limited performance, DP < 3 considered to be a fair discrimination, while values above 3 are classified as good.

Calculating the Youden’s index and DP could give clearer evidence of a test performance [49]. In this report, AUC was used as an indicator for test accuracy [48] and where the AUC does not show much difference, Youden’s index and DP were used as indicators for test performance [49].

**Phase 2: Evaluating the best combined temperature metrics in screening DED**

The diagnostic power of the metrics can be possibly maximized by combining them. In this part of the study, we sought to evaluate if AUC can be further maximised by a factor analysis model using principal component analysis [55-56] through Excel’s ‘solver’ function (Microsoft Excel 2013, USA). The analysis was developed for each dataset in order to reduce the dimensionality of the variables down to one or two factors combining these variables. This will help to determine the best detectors for dry eye. Again, data was first tabulated on the 30 static- and dynamic-metrics. A range of testing threshold/criterion was encapsulated in 0.1 intervals. Using Excel spreadsheet, sensitivity, specificity and AUC were derived for all the metrics. After which, a tool in the Excel spreadsheet ‘solver’ was used to test on all possible combinations within the 30 static- and dynamic-metrics that could possibly maximising the AUC and determine the best detectors.

**Results**

**Phase 1: Evaluating the efficacy of the 30 static- and 30 dynamic-metrics when applied singly**

Figure 2 shows the ROC curves for the 30 static metrics, grouped in three different time points (0 s, 5 s and 10 s). Figure 3 shows the ROC curves for the 30 dynamic metrics, grouped into its three attributes.

Based on Figure 2, the best AUC for static metrics were T4-5 and T4-10 with AUC of 72% (95% CI: 63 to 81%) and 73% (64 to 82%) respectively, indicating the tests were useful for some purpose/moderate accuracy [48]. AUC for the rest of the static metrics lied below 70% indicating low accuracy. For example, AUC for T4-0 was 69% (60 to 79%). Based on Figure 3, the AUC for dynamic metrics ranged from 50 to 64, indicating that all dynamic metrics were of low accuracy as their AUC lied below 70% [48].
Table 2 shows a summary of AUC, sensitivity, specificity, Youden's index, DP and the selected cutoff values for the 30 static metrics. The best results were, again, obtained for T4-5 and T4-10 metrics with DP of 1.07 and 1.05 respectively indicating limited performance. DP for the rest of the static metrics were less than 1 indicating poor discrimination performance including T4-0 [50]. Youden's index for T4-5 and T4-10 was also found to be highest of all (37.9 and 39.5 respectively). T4-0 had lower DP of 0.79 and Youden's index of 34.4.

When the cutoff values for T4-5 was set at 34.7 °C (i.e., values < 34.8 °C; Table 2), sensitivity and specificity was 87.1% (76.2 to 94.3%) and 50.8% (37.9 to 63.6%) respectively and when the cutoff values for T4-10 was set at 34.5 °C (i.e., values < 34.6 °C; Table 2), sensitivity and specificity was 77.6% (64.7 to 87.5%) and 61.9% (48.8 to 73.9%) respectively.

Table 3 shows a summary of AUC, sensitivity, specificity, Youden's index, DP and the selected cutoff values for the 30 dynamic metrics. All dynamic metrics were shown to be of low accuracy, with AUC below 70% [48]. The DP for all dynamic metrics ranged from 0.20 to 0.69 indicating poor discrimination performance as they were less than 1 [50]. Youden's index was also found to be low for all dynamic metrics ranging 8.6 to 27.9.

Phase 2: Evaluating the best combined temperature metrics in screening DED

AUC for static metrics at 0 s can be increased to 71% by a combined metrics in the expression of 0.95GCC + 0.19MOST + 0.26MinT + 0.50MaxT + 0.09T1 + 0.94T4 + 0.04CT + 0.18LT + 0.07LN + 0.37CN.

AUC for static metrics at 5 s can be increased to 72% by a combined metrics in the expression of 0.02GCC + 0.15MOST + 0.04MinT + 0.65MaxT + 0.27T1 + 0.90T4 + 0.26CT + 0.12LT + 0.11LN + 0.19CN.

AUC for static metrics at 10 s can be increased to 73% by a combined metrics in the expression of 0.03T1 + 0.98T4 + 0.06CT + 0.02LT + 0.02LN + 0.05CN.

It was shown that T4 was the main contributor to AUC for static metrics at all three time points. However, by just looking at the AUC at respective metric (without combining them), the AUC for T4 were shown to be very similar: 69% at 0 s, 72% at 5 s and 73% at 10 s (Table 2). It was therefore concluded that combining metrics was not able to meaningfully improve AUC for static measures.
On the other hand, AUC for dynamic metrics for asymptote can be increased to 53% by a combined metrics in the expression of

\[ 0.31 \text{GCC} + 0.34 \text{MOST} + 0.55 \text{MinT} + 0.23 \text{MaxT} + 0.28 \text{T1} + 0.87 \text{T4} + 0.84 \text{CT} + 0.25 \text{LT} + 0.02 \text{LN} + 0.68 \text{CN}. \]

AUC for dynamic metrics for scale can be increased to 56% by a combined metrics in the expression of

\[ 0.21 \text{GCC} + 0.06 \text{MOST} + 0.04 \text{MinT} + 0.06 \text{MaxT} + 0.92 \text{T4}. \]

AUC for dynamic metrics for growth rate can be increased to 51% by \( 1.00 \text{T4} \). It was again shown that, T4 was the main contributor to AUC for dynamic metrics at all three attributes. Similarly, by just looking at the AUC at respective metric (without combining them), the AUC for T4 were shown to be very similar: 59% for asymptote, 61% for scale and 50% for growth rate (Table 3). It was again concluded that combining metrics was not able to meaningfully improve the AUC for dynamic measures.

'Insert Fig. 4 and 5 here'

ROC curves on T4 metrics alone were then plotted and shown in Figure 4 and 5 with a summary of their AUC. It was clear that T4 metrics (static) were good detectors for DED as points were far above the diagonal line in ROC curve \([45]\) (Fig. 4) as compared to T4 metrics (dynamic) where points were on / fairly above the diagonal line (Fig. 5) indicating poor detectors for DED \([45]\).

Discussion

This is the first study to demonstrate that measuring temperature at the extreme nasal conjunctiva was able to discriminate mild to moderate dry eye from non-DED patients. The test was comparable to other well established methods in testing tear stability based on different principles, such as tear break-up time (BUT) \([21,57,58]\) and non-invasive tear break-up time (NIBUT) \([58]\) (Table 4). A cutoff value of 34.7 °C for T4-5 temperature has given higher sensitivity as compared to wet BUT and dry BUT \([57]\), BUT \([21,58]\) and NIBUT \([58]\). On the other hand, a cutoff value of 34.5 °C for T4-10 temperature had lower sensitivity in general when compared with above-mentioned tests but its specificity was better than wet BUT \([57]\), BUT \([21,58]\) and NIBUT \([58]\) (Table 4).

'Select Table 4 here'

Static metrics were found to have better efficacy in diagnosing DED as compared to dynamic metrics. It was sufficient to perform the test singly as combined metrics was not successful to increase its performance. The limited discrimination performance of any single method highlights the complexity of diagnosing DED. The 30 tested dynamic metrics were shown to have low efficacy in diagnosing DED (singly and in combination) and deserved no further discussion.
It has been reported that temperature at the conjunctiva was higher than that of the central corneal [37,59]. Although the reasons remains unclear, the temperature of the nasal conjunctiva was reported to be higher than that of the temporal conjunctiva because of the influence of greater blood flow and vascularization in the nasal conjunctiva [26]. More large vessels (eg. the dorsal nasal artery and the angular artery) are situated at the nasal side of the eye. In addition, the medial rectus muscle has two anterior ciliary arteries, whereas the lateral rectus muscle has only one artery. All these anatomical factors has caused a more vascularised nasal conjunctiva with higher blood flow [26] and hence causing a higher evaporation rate / less stable tear film at T4. Further studies would be required to confirm these postulation. Nevertheless, the temperature changes at nasal conjunctiva may also be contributed by other factors such as allergy rather than dry eye and warranted further investigations.

Our study was in agreement with past studies [26,27] i.e., measuring OST can be a good diagnostic tool for dry eye. Sensitivity and specificity of Tomey IR thermographer tested at GCC was reported to be 83% and 80% respectively using a cutoff value of 0.13 °C but reduced to 80% and 73% respectively using a cutoff value of 0.11 °C [26]. The values derived in Kamao et al’s [26] study was not defined clearly but believed was a decrease in dynamic temperature over 10 seconds (for cutoff value of 0.13 °C) and 5 seconds (for cutoff value of 0.11 °C) respectively. Using a custom-designed IR thermal image system, sensitivity and specificity of a combined temperature metrics (temperature difference and compactness values) were 84% and 83% respectively with unclear cutoff value [27]. Limited temperature metrics were included in the above mentioned studies as compared to sixty temperature metrics in the current report. Different in findings as compared to the current report could be due to different in methodology and subjects recruited. Ocular thermographers with lower resolution of 400 (H) x 240 (V) pixels and 320 (H) x 240 (V) pixels in Kamao et al’s [26] and Su et al’s [27] studies respectively may have hindered the visibility of the anatomical structures of the ocular surface and created errors in OST acquisition. Dry eye subjects recruited by both studies were much older (mean age 52.9 years for dry eye and 42.7 years for control subjects) [26] and mean age of 49 years for dry eye and 34 years for controls [27]) and the two groups were not age-matched. This is important because OST was reported to decrease with age at a rate of - 0.010 °C / year [39]. Differences in OST between groups may be age-related rather than to the disease itself.

Other tests evaluating DED such as tear osmolarity and tear evaporation rate measurement have been reported to have better diagnostic ability. Using a threshold value of 312 mOsms/L, tear osmolarity had sensitivity of 73% and specificity of 92% as compared to other tests (54% for corneal staining, 60% for conjunctival staining and 61% for meibomian gland grading on sensitivity; 45% for tear break-up time and 51% for Schirmer test on specificity [21]. On the other hand, values of 41 g/m(-2)/h for tear evaporation rate has found to give sensitivity of 96% and specificity of 67% [47].
This paper presented a comprehensive study of the diagnostic ability for IR ocular thermography in screening DED. All dry eye patients were mild to moderate with no inflamed meibomian glands. We acknowledged that many disease severity criteria are confounded by complex disease subtypes and a lack of standardisation, and the selection of single criteria for assessment of disease severity is therefore fraught with difficulties [8,60,61].

The results of this study suggest that IR ocular thermography is a suitable test to be incorporated in the non-invasive diagnostic assessment of dry eye. It is repeatable, rapid, and easy to use and in this study, shown to give good sensitivity of specificity in diagnosing DED. Future studies on dry eye screening using ocular thermography should include temperature of the conjunctiva. T4-5 and T4-10 metrics are simple, static measures and we speculate that they could be used in combination with other conventional tests to further refine diagnostic discrimination for DED.

Since this was a preliminary study, we attempted to standardize every aspect of our testing protocol (e.g., temperature and humidity controlled room, 20 minutes of room acclimation) in order to reduce testing variability as much as possible. While this is a common research strategy, additional work is needed to determine if our findings hold up in a normal clinic environment. Nevertheless, our overall findings do indicate that ocular thermography may be useful for understanding dry eye in clinical practice.

Acknowledgements

This study was funded by Singapore ToteBoard Organisation no. LS/CLS/TM/2009/007. The authors wish to thank Dr. Cai Zhi Qiang from School of Electronic and Electrical Engineering in writing the OST analysis V2 program using MatLab Simulink 7.11.0 (R2010b) and Dr. Robert Straughan from School of Mathematics and Science for his valuable statistical advice.

Financial disclosures

The authors have no financial interest in any of the materials used in this study.

Conflict of interest statement

No conflicting relationship exists for any author.

References


Table 1
The ten OST indices studied.

<table>
<thead>
<tr>
<th>OST indices</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCC</td>
<td>Geometric center of the cornea (midway between LT and LN)</td>
</tr>
<tr>
<td>MOST</td>
<td>Mean OST of the ROI</td>
</tr>
<tr>
<td>T1</td>
<td>Extreme temporal conjunctiva</td>
</tr>
<tr>
<td>T4</td>
<td>Extreme nasal conjunctiva</td>
</tr>
<tr>
<td>CT</td>
<td>Mid-temporal conjunctiva (midway between T1 and LT)</td>
</tr>
<tr>
<td>LT</td>
<td>Temporal limbus</td>
</tr>
<tr>
<td>LN</td>
<td>Nasal limbus</td>
</tr>
<tr>
<td>CN</td>
<td>Mid-nasal conjunctiva (midway between T4 and LN)</td>
</tr>
<tr>
<td>MinT</td>
<td>Minimum temperature of ROI</td>
</tr>
<tr>
<td>MaxT</td>
<td>Maximum temperature of ROI</td>
</tr>
</tbody>
</table>
Table 2
Test effectiveness of static metrics at 0, 5 and 10 s. AUC, sensitivity, specificity, Youden's index (Y), discrimination power (DP) and the selected cutoff values are shown.

<table>
<thead>
<tr>
<th>Static metrics at 0 s</th>
<th>AUC, % (95% CI)</th>
<th>Cutoff values, °C</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Y</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCC-0</td>
<td>69 (60 to 78)</td>
<td>&lt; 34.3</td>
<td>72.6 (59.8 to 83.2)</td>
<td>60.3 (47.2 to 72.4)</td>
<td>32.9</td>
<td>0.77</td>
</tr>
<tr>
<td>MOST-0</td>
<td>68 (58 to 77)</td>
<td>&lt; 34.7</td>
<td>79.0 (66.8 to 88.3)</td>
<td>54.0 (40.9 to 66.6)</td>
<td>33.0</td>
<td>0.82</td>
</tr>
<tr>
<td>MinT-0</td>
<td>67 (58 to 77)</td>
<td>&lt; 34.0</td>
<td>79.0 (66.8 to 88.3)</td>
<td>57.1 (44.1 to 69.5)</td>
<td>36.2</td>
<td>0.89</td>
</tr>
<tr>
<td>MaxT-0</td>
<td>67 (57 to 76)</td>
<td>&lt; 35.7</td>
<td>82.3 (70.5 to 90.8)</td>
<td>47.6 (34.9 to 60.6)</td>
<td>29.9</td>
<td>0.79</td>
</tr>
<tr>
<td>T1-0</td>
<td>66 (57 to 76)</td>
<td>&lt; 35.2</td>
<td>61.3 (48.1 to 73.4)</td>
<td>71.4 (58.7 to 82.1)</td>
<td>32.7</td>
<td>0.76</td>
</tr>
<tr>
<td>T4-0</td>
<td>69 (60 to 79)</td>
<td>&lt; 34.5</td>
<td>69.4 (56.4 to 80.4)</td>
<td>65.1 (52.0 to 76.7)</td>
<td>34.4</td>
<td>0.79</td>
</tr>
<tr>
<td>CT-0</td>
<td>68 (58 to 77)</td>
<td>&lt; 34.6</td>
<td>56.5 (43.3 to 69.0)</td>
<td>74.6 (62.1 to 84.7)</td>
<td>31.1</td>
<td>0.74</td>
</tr>
<tr>
<td>LT-0</td>
<td>67 (57 to 77)</td>
<td>&lt; 34.5</td>
<td>79.0 (66.8 to 88.3)</td>
<td>54.0 (40.9 to 66.6)</td>
<td>33.0</td>
<td>0.82</td>
</tr>
<tr>
<td>LN-0</td>
<td>65 (55 to 75)</td>
<td>&lt; 34.7</td>
<td>64.5 (51.3 to 78.3)</td>
<td>63.5 (50.4 to 75.3)</td>
<td>28.0</td>
<td>0.63</td>
</tr>
<tr>
<td>CN-0</td>
<td>68 (57 to 77)</td>
<td>&lt; 35.4</td>
<td>83.9 (72.3 to 92.0)</td>
<td>52.4 (39.4 to 65.1)</td>
<td>36.3</td>
<td>0.96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Static metrics at 5 s</th>
<th>AUC, % (95% CI)</th>
<th>Cutoff values, °C</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Y</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCC-5</td>
<td>68 (58 to 77)</td>
<td>&lt; 33.9</td>
<td>67.7 (54.7 to 79.1)</td>
<td>65.1 (52.0 to 76.7)</td>
<td>32.8</td>
<td>0.75</td>
</tr>
<tr>
<td>MOST-5</td>
<td>66 (57 to 76)</td>
<td>&lt; 33.9</td>
<td>46.8 (34.0 to 59.9)</td>
<td>84.1 (72.7 to 92.1)</td>
<td>30.9</td>
<td>0.85</td>
</tr>
<tr>
<td>MinT-5</td>
<td>63 (53 to 73)</td>
<td>&lt; 32.9</td>
<td>41.9 (29.5 to 55.2)</td>
<td>82.5 (70.9 to 91.0)</td>
<td>24.5</td>
<td>0.68</td>
</tr>
<tr>
<td>MaxT-5</td>
<td>68 (59 to 77)</td>
<td>&lt; 35.6</td>
<td>77.4 (65.0 to 87.1)</td>
<td>54.0 (40.9 to 66.6)</td>
<td>31.4</td>
<td>0.77</td>
</tr>
<tr>
<td>T1-5</td>
<td>66 (56 to 75)</td>
<td>&lt; 35.1</td>
<td>59.7 (46.5 to 72.0)</td>
<td>71.4 (58.7 to 82.1)</td>
<td>31.1</td>
<td>0.72</td>
</tr>
<tr>
<td>T4-5</td>
<td>72 (63 to 81)</td>
<td>&lt; 34.8</td>
<td>87.1 (76.2 to 94.3)</td>
<td>50.8 (37.9 to 63.6)</td>
<td>37.9</td>
<td>1.07</td>
</tr>
<tr>
<td>CT-5</td>
<td>66 (56 to 75)</td>
<td>&lt; 34.7</td>
<td>75.8 (63.3 to 85.8)</td>
<td>52.4 (39.4 to 65.1)</td>
<td>28.2</td>
<td>0.68</td>
</tr>
<tr>
<td>LT-5</td>
<td>64 (54 to 74)</td>
<td>&lt; 34.1</td>
<td>72.6 (59.8 to 83.2)</td>
<td>52.4 (39.4 to 65.1)</td>
<td>25.0</td>
<td>0.59</td>
</tr>
<tr>
<td>LN-5</td>
<td>63 (54 to 73)</td>
<td>&lt; 34.4</td>
<td>59.7 (46.5 to 72.0)</td>
<td>71.4 (58.7 to 82.1)</td>
<td>31.1</td>
<td>0.72</td>
</tr>
<tr>
<td>CN-5</td>
<td>67 (57 to 76)</td>
<td>&lt; 35.3</td>
<td>85.5 (74.2 to 93.1)</td>
<td>47.6 (34.9 to 60.6)</td>
<td>33.1</td>
<td>0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Static metrics at 10 s</th>
<th>AUC, % (95% CI)</th>
<th>Cutoff values, °C</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Y</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCC-10</td>
<td>65 (55 to 74)</td>
<td>&lt; 33.9</td>
<td>77.6 (64.7 to 87.5)</td>
<td>49.2 (36.4 to 62.1)</td>
<td>26.8</td>
<td>0.67</td>
</tr>
<tr>
<td>MOST-10</td>
<td>64 (54 to 74)</td>
<td>&lt; 33.9</td>
<td>46.6 (33.3 to 60.1)</td>
<td>82.5 (70.9 to 91.0)</td>
<td>29.1</td>
<td>0.78</td>
</tr>
<tr>
<td>MinT-10</td>
<td>60 (50 to 70)</td>
<td>&lt; 32.7</td>
<td>34.5 (22.5 to 48.1)</td>
<td>82.5 (70.9 to 91.0)</td>
<td>17.0</td>
<td>0.50</td>
</tr>
<tr>
<td>MaxT-10</td>
<td>69 (59 to 78)</td>
<td>&lt; 35.5</td>
<td>75.9 (62.8 to 86.1)</td>
<td>55.6 (42.5 to 68.1)</td>
<td>31.4</td>
<td>0.75</td>
</tr>
<tr>
<td>T1-10</td>
<td>66 (56 to 75)</td>
<td>&lt; 34.8</td>
<td>44.8 (31.7 to 58.5)</td>
<td>82.5 (70.9 to 91.0)</td>
<td>27.4</td>
<td>0.74</td>
</tr>
<tr>
<td>T4-10</td>
<td>73 (64 to 82)</td>
<td>&lt; 34.6</td>
<td>77.6 (64.7 to 87.5)</td>
<td>61.9 (48.8 to 73.9)</td>
<td>39.5</td>
<td>1.05</td>
</tr>
<tr>
<td>CT-10</td>
<td>64 (54 to 74)</td>
<td>&lt; 34.2</td>
<td>48.3 (35.0 to 61.8)</td>
<td>81.0 (69.1 to 89.8)</td>
<td>29.2</td>
<td>0.76</td>
</tr>
<tr>
<td>LT-10</td>
<td>62 (52 to 72)</td>
<td>&lt; 34.1</td>
<td>74.1 (61.0 to 84.7)</td>
<td>47.6 (34.9 to 60.6)</td>
<td>21.8</td>
<td>0.53</td>
</tr>
<tr>
<td>LN-10</td>
<td>62 (52 to 72)</td>
<td>&lt; 34.3</td>
<td>63.8 (50.1 to 76.0)</td>
<td>63.5 (50.4 to 75.3)</td>
<td>27.3</td>
<td>0.62</td>
</tr>
<tr>
<td>CN-10</td>
<td>67 (57 to 76)</td>
<td>&lt; 35.2</td>
<td>82.8 (70.6 to 91.4)</td>
<td>52.4 (39.4 to 65.1)</td>
<td>35.1</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Table 3: Test effectiveness for the dynamic metrics (A-asymptote, S-scale, GR-growth rate). Sensitivity, specificity, Youden's index (Y), discrimination power (DP) and the selected cutoff values are shown.

<table>
<thead>
<tr>
<th>Dynamic metrics-A</th>
<th>AUC, % (95% CI)</th>
<th>Cutoff values, °C</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Y</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCC-A</td>
<td>51 (41 to 61)</td>
<td>&gt; -0.2861</td>
<td>40.3 (28.1 to 53.6)</td>
<td>68.3 (55.3 to 79.4)</td>
<td>8.6</td>
<td>0.21</td>
</tr>
<tr>
<td>MOST-A</td>
<td>56 (46 to 66)</td>
<td>&lt; -0.5155</td>
<td>45.2 (32.5 to 58.3)</td>
<td>73.0 (60.4 to 83.4)</td>
<td>18.2</td>
<td>0.44</td>
</tr>
<tr>
<td>MinT-A</td>
<td>54 (44 to 64)</td>
<td>&lt; -0.2918</td>
<td>71.0 (58.1 to 81.8)</td>
<td>42.9 (30.5 to 56.0)</td>
<td>13.8</td>
<td>0.33</td>
</tr>
<tr>
<td>MaxT-A</td>
<td>62 (52 to 72)</td>
<td>&lt; -0.1750</td>
<td>50.0 (37.0 to 63.0)</td>
<td>76.2 (63.8 to 86.0)</td>
<td>26.2</td>
<td>0.64</td>
</tr>
<tr>
<td>T1-A</td>
<td>58 (48 to 68)</td>
<td>&lt; -0.1371</td>
<td>45.2 (32.5 to 58.3)</td>
<td>81.0 (69.1 to 89.8)</td>
<td>26.1</td>
<td>0.69</td>
</tr>
<tr>
<td>T4-A</td>
<td>59 (49 to 69)</td>
<td>&lt; -0.0921</td>
<td>54.8 (41.7 to 67.5)</td>
<td>73.0 (60.4 to 83.4)</td>
<td>27.9</td>
<td>0.66</td>
</tr>
<tr>
<td>CT-A</td>
<td>51 (41 to 61)</td>
<td>&lt; -0.6576</td>
<td>17.7 (9.2 to 29.5)</td>
<td>92.1 (82.4 to 97.4)</td>
<td>9.8</td>
<td>0.51</td>
</tr>
<tr>
<td>LT-A</td>
<td>52 (41 to 62)</td>
<td>&lt; -0.3737</td>
<td>66.1 (53.0 to 77.7)</td>
<td>46.0 (33.4 to 59.1)</td>
<td>12.2</td>
<td>0.28</td>
</tr>
<tr>
<td>LN-A</td>
<td>60 (50 to 70)</td>
<td>&lt; -0.0571</td>
<td>82.3 (70.5 to 90.8)</td>
<td>34.9 (23.3 to 48.0)</td>
<td>17.2</td>
<td>0.50</td>
</tr>
<tr>
<td>CN-A</td>
<td>57 (47 to 67)</td>
<td>&lt; -0.4083</td>
<td>40.3 (28.1 to 53.6)</td>
<td>74.6 (62.1 to 84.7)</td>
<td>14.9</td>
<td>0.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dynamic metrics-S</th>
<th>AUC, % (95% CI)</th>
<th>Cutoff values, °C</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Y</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCC-S</td>
<td>51 (41 to 61)</td>
<td>&lt; 0.3310</td>
<td>45.2 (32.5 to 58.3)</td>
<td>63.5 (50.4 to 75.3)</td>
<td>8.7</td>
<td>0.20</td>
</tr>
<tr>
<td>MOST-S</td>
<td>57 (46 to 67)</td>
<td>&gt; 0.5252</td>
<td>45.2 (32.5 to 58.3)</td>
<td>74.6 (62.1 to 84.7)</td>
<td>19.8</td>
<td>0.49</td>
</tr>
<tr>
<td>MinT-S</td>
<td>55 (44 to 65)</td>
<td>&gt; 0.3363</td>
<td>67.7 (54.7 to 79.1)</td>
<td>46.0 (33.4 to 59.1)</td>
<td>13.8</td>
<td>0.32</td>
</tr>
<tr>
<td>MaxT-S</td>
<td>64 (55 to 74)</td>
<td>&gt; 0.1571</td>
<td>54.8 (41.7 to 67.5)</td>
<td>68.3 (55.3 to 79.4)</td>
<td>23.1</td>
<td>0.53</td>
</tr>
<tr>
<td>T1-S</td>
<td>59 (48 to 69)</td>
<td>&gt; 0.1531</td>
<td>48.4 (35.5 to 61.4)</td>
<td>77.8 (65.5 to 87.3)</td>
<td>26.2</td>
<td>0.66</td>
</tr>
<tr>
<td>T4-S</td>
<td>61 (51 to 71)</td>
<td>&gt; 0.0004</td>
<td>62.9 (49.7 to 74.8)</td>
<td>58.7 (45.6 to 71.0)</td>
<td>21.6</td>
<td>0.49</td>
</tr>
<tr>
<td>CT-S</td>
<td>53 (43 to 63)</td>
<td>&gt; 0.0101</td>
<td>79.0 (66.8 to 88.3)</td>
<td>33.3 (22.0 to 46.3)</td>
<td>12.4</td>
<td>0.35</td>
</tr>
<tr>
<td>LT-S</td>
<td>52 (42 to 62)</td>
<td>&gt; 0.3985</td>
<td>59.7 (46.5 to 72.0)</td>
<td>49.2 (36.4 to 62.1)</td>
<td>8.9</td>
<td>0.20</td>
</tr>
<tr>
<td>LN-S</td>
<td>61 (51 to 71)</td>
<td>&gt; 0.3398</td>
<td>64.5 (51.3 to 76.3)</td>
<td>57.1 (44.1 to 69.5)</td>
<td>21.7</td>
<td>0.49</td>
</tr>
<tr>
<td>CN-S</td>
<td>58 (48 to 68)</td>
<td>&gt; 0.3175</td>
<td>48.4 (35.5 to 61.4)</td>
<td>68.3 (55.3 to 79.4)</td>
<td>16.6</td>
<td>0.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dynamic metrics-GR</th>
<th>AUC, % (95% CI)</th>
<th>Cutoff values, °C</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Y</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCC-GR</td>
<td>54 (44 to 64)</td>
<td>&gt; -0.1771</td>
<td>61.3 (48.1 to 73.4)</td>
<td>52.4 (39.4 to 65.1)</td>
<td>13.7</td>
<td>0.31</td>
</tr>
<tr>
<td>MOST-GR</td>
<td>52 (42 to 63)</td>
<td>&lt; -0.2436</td>
<td>37.1 (25.2 to 50.3)</td>
<td>74.6 (62.1 to 84.7)</td>
<td>11.7</td>
<td>0.30</td>
</tr>
<tr>
<td>MinT-GR</td>
<td>55 (45 to 65)</td>
<td>&lt; -0.0322</td>
<td>83.9 (72.3 to 92.0)</td>
<td>30.2 (19.2 to 43.0)</td>
<td>14.0</td>
<td>0.45</td>
</tr>
<tr>
<td>MaxT-GR</td>
<td>53 (43 to 64)</td>
<td>&gt; -0.3078</td>
<td>85.5 (74.2 to 93.1)</td>
<td>33.3 (22.0 to 46.3)</td>
<td>18.8</td>
<td>0.60</td>
</tr>
<tr>
<td>T1-GR</td>
<td>55 (44 to 65)</td>
<td>&gt; -0.1719</td>
<td>71.0 (58.1 to 81.8)</td>
<td>46.0 (33.4 to 59.1)</td>
<td>17.0</td>
<td>0.41</td>
</tr>
<tr>
<td>T4-GR</td>
<td>50 (40 to 61)</td>
<td>&gt; -0.6860</td>
<td>90.3 (80.1 to 96.4)</td>
<td>17.5 (9.1 to 29.1)</td>
<td>7.8</td>
<td>0.37</td>
</tr>
<tr>
<td>CT-GR</td>
<td>56 (46 to 66)</td>
<td>&lt; -0.4405</td>
<td>33.9 (22.3 to 47.0)</td>
<td>79.4 (67.3 to 88.5)</td>
<td>13.2</td>
<td>0.37</td>
</tr>
<tr>
<td>LT-GR</td>
<td>51 (41 to 61)</td>
<td>&gt; -0.1985</td>
<td>58.1 (44.9 to 70.5)</td>
<td>52.4 (39.4 to 65.1)</td>
<td>10.4</td>
<td>0.23</td>
</tr>
<tr>
<td>LN-GR</td>
<td>55 (45 to 65)</td>
<td>&lt; -0.3620</td>
<td>27.4 (16.9 to 40.2)</td>
<td>87.3 (76.5 to 94.4)</td>
<td>14.7</td>
<td>0.53</td>
</tr>
<tr>
<td>CN-GR</td>
<td>56 (46 to 66)</td>
<td>&gt; -0.4496</td>
<td>85.3 (73.8 to 93.0)</td>
<td>28.6 (17.9 to 41.4)</td>
<td>13.8</td>
<td>0.46</td>
</tr>
</tbody>
</table>
Table 4
A comparison of ocular thermography to similar tests found in the literature.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Tests</th>
<th>Cutoff values</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>AUC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current study</td>
<td>T4-5 temperature</td>
<td>34.7 °C</td>
<td>87.1</td>
<td>50.8</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>T4-10 temperature</td>
<td>34.5 °C</td>
<td>77.6</td>
<td>61.9</td>
<td>73</td>
</tr>
<tr>
<td>Kim et al., 2015</td>
<td>Wet BUT</td>
<td>4.48 s</td>
<td>79.0</td>
<td>54.8</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Dry BUT</td>
<td>3.50 s</td>
<td>72.6</td>
<td>69.4</td>
<td>72</td>
</tr>
<tr>
<td>Lemp et al., 2011</td>
<td>TBUT</td>
<td>&lt; 10 s</td>
<td>84.4</td>
<td>45.3</td>
<td>-</td>
</tr>
<tr>
<td>Yeh et al., 2015</td>
<td>NIBUT</td>
<td>-</td>
<td>72.0</td>
<td>52.0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>TBUT</td>
<td>-</td>
<td>68.0</td>
<td>57.0</td>
<td>-</td>
</tr>
</tbody>
</table>
Fig. 1 The ‘diamond’ method in marking the ocular surface and OST acquisition. ROI = region of interest.

Fig. 2 ROC curves for static metrics at three different time points. AUC are shown at the legend.

Fig. 3 ROC curves for dynamic metrics at three different attributes. AUC are shown at the legend.

Fig. 4 ROC curves for T4 metrics (static). AUC are shown at the legend.

Fig. 5 ROC curves for T4 metrics (dynamic). AUC are shown at the legend.
Fig. 1
at 0 s

GCC (AUC = 69)
MOST (AUC = 68)
MinT (AUC = 67)
MaxT (AUC = 67)
T1 (AUC = 66)
T4 (AUC = 69)
CT (AUC = 68)
LT (AUC = 67)
LN (AUC = 65)
CN (AUC = 68)

at 5 s

GCC (AUC = 68)
MOST (AUC = 66)
MinT (AUC = 63)
MaxT (AUC = 68)
T1 (AUC = 66)
T4 (AUC = 72)
CT (AUC = 66)
LT (AUC = 64)
LN (AUC = 63)
CN (AUC = 67)

at 10 s

GCC (AUC = 65)
MOST (AUC = 64)
MinT (AUC = 60)
MaxT (AUC = 69)
T1 (AUC = 66)
T4 (AUC = 73)
CT (AUC = 64)
LT (AUC = 62)
LN (AUC = 62)
CN (AUC = 67)

Fig. 2
**asymptote**

![Asymptote Diagram]

**scale**

![Scale Diagram]

**growth rate**

![Growth Rate Diagram]

Fig. 3
Fig. 4

Fig. 5