Non-Invasive Detection of Diabetic Complications via Pattern Analysis of Temporal Facial Colour Variations

Majtner, Tomas; S. Nadimi, Esmaeil; Yderstræde, Knud Bonnet; Blanes-Vidal, Victoria

Published in:
Computer Methods and Programs in Biomedicine

DOI:
10.1016/j.cmpb.2020.105619

Publication date:
2020

Document version:
Accepted manuscript

Document license:
CC BY-NC-ND

Citation for published version (APA):

Go to publication entry in University of Southern Denmark’s Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:
- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 14. Sep. 2023
Non-Invasive Detection of Diabetic Complications via Pattern Analysis of Temporal Facial Colour Variations

Tomáš Majtner\textsuperscript{a,}\textsuperscript{*}, Esmaeil S. Nadimi\textsuperscript{a}, Knud B. Yderstræde\textsuperscript{b}, Victoria Blanes-Vidal\textsuperscript{a}

\textsuperscript{a}Group of Applied AI and Data Science, The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark
\textsuperscript{b}Steno Diabetes Center, Odense University Hospital, Odense, Denmark

Abstract

Background and Objective: Diabetes mellitus is a common disorder amounting to 400 million patients worldwide. It is often accompanied by a number of complications, including neuropathy, nephropathy, and cardiovascular diseases. For example, peripheral neuropathy is present among 20-30\% of diabetics before the diagnosis is substantiated. For this reason, a reliable detection method for diabetic complications is crucial and attracts a lot of research attention.

Methods: In this paper, we introduce a non-invasive detection framework for patients with diabetic complications that only requires short video recordings of faces from a standard commercial camera. We employed multiple image processing and pattern recognition techniques to process video frames, extract relevant information, and predict the health status. To evaluate our framework, we collected a dataset of 114 video files from diabetic patients, who were diagnosed with diabetes for years and 60 video files from the control group. Extracted features from videos were tested using two conceptually different classifiers.

Results: We found that our proposed framework correctly identifies patients with diabetic complications with 92.86\% accuracy, 100\% sensitivity, and 80\% specificity.

Conclusions: Our study brings a novel perspective on diagnosis procedures

\textsuperscript{*}Corresponding author
Email address: tomaj@mmmi.sdu.dk (Tomáš Majtner)
in this field. We used multiple techniques from image processing, pattern recognition, and machine learning to robustly process video frames and predict the health status of our subjects with high efficiency.

**Keywords:** pattern recognition; machine learning; skin patch analysis; feature extraction; skin redness

1. Introduction

Diabetes mellitus (DM) constitutes a group of metabolic disorders characterised by hyperglycemia resulting from defects in insulin secretion, insulin action, or both [1]. The chronic hyperglycemia of diabetes is associated with long-term damage, dysfunction, and failure of different organs, especially the eyes, kidneys, nerves, heart, and blood vessels. Especially type 2 diabetes is increasingly common, primarily because of increases in the prevalence of a sedentary lifestyle and obesity [2]. Globally, the number of people with diabetes mellitus has quadrupled in the past three decades, and diabetes mellitus is nowadays the ninth major cause of death [3].

Most patients with type 2 diabetes mellitus have at least one complication, and cardiovascular complications are the leading cause of morbidity and mortality in these patients. Cutaneous findings may be the first sign of metabolic disturbance related to undiagnosed diabetes, suboptimal management, or a prediabetic state [4]. Moreover, as it was shown by multiple studies, these findings may be used as external markers of internal complications in already diagnosed diabetic patients [5, 6].

One of the common skin manifestation associated with diabetes is called *rubeosis faciei* [7]. It is presented by flushing of blood to the face, and it has been associated with three non-cutaneous diabetic complications [8]. Typically, it is diagnosed during a clinical examination, when the intensity of red colouration in the diabetic patient is self-evident. However, the intensity of red colouration depends on the degree of vascular engorgement of the superficial venous plexus, and it may go unnoticed by physicians [9].
In our study, we examine the possibility of early and non-invasive detection of diabetic complications connected to rubeosis faciei based on changes in facial redness over a short time period. The aim of this study is to make a step towards a smartphone application for diabetic patients that will recognise early signs of diabetic complications using short video files captured at home. We employ a machine learning approach, where our proposed framework is based on short video recordings of faces. We localise the areas of interest in the face, which are subsequently used for the evaluation. Using a dataset of hundreds of video samples recorded in an environment with non-controlled lighting conditions and by applications of multiple data analysis techniques, we created a highly efficient framework, which works in real-world conditions.

1.1. Related Work

The topic of diabetes mellitus attracts a lot of research attention. In particular, regarding diabetic complications, researchers have studied the recent progress in pathogenesis of diabetic neuropathy [10], or they focused on the treatment of neuropathic pain [11]. The current research progress in diabetic neuropathy was summarized by Feldman et al. [12].

In 2018, Dagliati et al. [13] created a data mining pipeline to derive a set of predictive models of complications in type 2 diabetes mellitus based on electronic health record data. Their models, tailored in accordance with the complications, provided an accuracy up to 83.8\%. Other machine learning methods for diabetes detection are based on the measurement of volatile organic compounds in biological samples such as urine [14]. For classification of diabetes vs control, they report only the area under ROC curve of 0.825. In other studies, computer–aided systems were used to identify complex combinations of treatment effect modifiers [15] and predict blood sugar level [16].

Our study is inspired by previous studies aimed at detecting diabetes mellitus using a facial key block analysis [17]. In their work, the authors can distinguish between diabetes and control classes with an average accuracy of 97.54\%. The high performance is achieved by employing a combination of the facial blocks and
specialised recording devices under laboratory conditions. Their approach was further extended using Gabor filters [18] and using probabilistic collaborative representation-based classifier [19]. The original motivation for this concept comes from traditional Chinese medicine where it is believed that the status of the internal organs can be determined by studying different regions of the face [20][21].

In current practice, the two main techniques to diagnose diabetic complications like neuropathy and retinopathy rely on measuring conductance of nerves [22] and the devices used in ophthalmology. These devices are very accurate but also very complex, when compared to a digital camera. Therefore, we decided to extend the facial key block analysis method in the temporal domain where changes of facial redness follow the heart pulsation. Our methodology has been recognised as a promising marker of diabetic complications [23].

The main difference of our work compared to the previous published methods is the utilisation of video recordings from a standard commercial camera. We simulate a real-world scenario by choosing not to control lighting conditions during the recording. This decision is crucial for a subsequent practical application of our framework outside laboratory conditions. Various advanced image processing methods are used to preprocess and normalise our samples before we employ classifiers to test the reliability of the proposed framework.

2. Materials and Methods

2.1. Study Participants

In total, thirty subjects were enrolled in our study and they were divided into two groups. The first one consists of twenty insulin-treated diabetic patients with an average age of 63±10 years (fourteen males and six females). We refer to this group as group DM. The second group has ten control subjects with an average age of 60±6 years (seven males and three females), and this group will be called group C. The two groups did not differ in average age and gender proportion. For the age comparison, a two-sample two-tailed t-test was used
Table 1: Specification of diabetic complications for each patient (DM subject).

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuropathy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Retinopathy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nephropathy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>P11</th>
<th>P12</th>
<th>P13</th>
<th>P14</th>
<th>P15</th>
<th>P16</th>
<th>P17</th>
<th>P18</th>
<th>P19</th>
<th>P20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuropathy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Retinopathy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nephropathy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

and for the proportions of gender, we used a Fisher exact test (no statistically significant difference was found, \( p > 0.10 \) in both cases). Eleven subjects from the DM group were diagnosed with type 1 diabetes with the average time since diagnosis 28\( \pm \)13 years. Nine patients were diagnosed with type 2 diabetes with the mean time from diagnosis 22\( \pm \)5 years.

Out of twenty DM subjects, nineteen had been diagnosed with peripheral neuropathy, seven with diabetic nephropathy, and nineteen with retinopathy (see Table 1). Facial erythema was not apparent in any of the participants, except for one DM subject with a visible facial red colouration. None of the subjects from DM group had exhibited cranial neuropathies such as facial nerves palsy, optic neuropathy, or auditory neuropathy. Control subjects had not been diagnosed with DM, peripheral arterial disease, or any other disease affecting the nervous system.

Two different types of factors can affect facial erythema during video acquisition and potentially introduce errors. The first type are individual-specific factors that mostly affect facial colour in a homogeneous manner, equally affecting all regions of the face (e.g. individual-specific score in the Fitzpatrick scale [24], sun exposure if bilateral). The second type are circumstantial factors that cause colour differences among different facial regions (e.g. physical activity, emotional state, and certain skin conditions).
We address the first type by incorporating the difference of redness between two different face regions, as it will be explained in detail later in our proposed method. Heterogeneous colour variations caused by second factors were minimised by using the same protocol for all participants. None of participants engaged in any physical activity during the previous twenty minutes prior to the beginning of experimental recordings.

Participants did not show signs of skin conditions causing heterogeneous redness such as psoriasis, rosacea, lupus, or allergic reactions. However, other skin conditions that cause more subtle colour changes may have occurred. Video recordings were mostly taken in the morning, and participants had not consumed alcohol prior to the experiment.

Finally, the emotional state during the experiment may vary from person to person. Bodily responses are an integral component of emotional experience, and it is certainly not possible to control the emotional state of the study participants during the experiment for standardisation purposes. However, regarding the potential stress caused by the fact of undergoing the experiment itself, it could well be assumed that the stress level may decrease during the course of the video recording, and therefore, randomisation of the video files was used to minimise nuisance effects.

2.2. Dataset

For the data acquisition, we prepared a recording setup consisting of a laptop, an external camera (Canon EOS 1300D), and a tripod. The camera was placed behind the laptop using the tripod, and it was capturing the entire face of tested subjects (see Figure 1 for illustration). On the laptop screen, we showed participants a short video sequence of a white point randomly moving across the black screen. The video was part of a larger experiment, where we also tracked the eyes motion [25]. The task was to follow the white point at the laptop screen while keeping the head steady. With each subject, we recorded three video sequences of different lengths during a single recording session. Their length was 4850 frames (97 s), 5900 frames (118 s), and 6250 frames (125 s).
Each subject completed two recording sessions, except for two subjects from DM group, where due to technical problems only one session was recorded. This resulted in six separate video files for each subject (in case of two DM subjects it was three video files). In total, we acquired 114 video files from group DM and 60 video files from group C. The camera was recording at 50 frames per second (fps) with a resolution of $1280 \times 720$ pixels.

Before we started to analyse recordings, we first randomly split all videos into two halves, where the first one was for training and the other one for testing. Data from a single person were either in training or in testing set. We employed five-fold cross-validation approach to find optimal parameters for training dataset but a holdout testing approach to report the final performance of our system. Some studies choose a cross-validation approach over entire dataset but from the statistical point of view, the reported performance tends to drop significantly when the algorithm is applied to new, previously unseen data [20]. The training part of the dataset consisted of 60 DM videos and 30 C videos, where all three video lengths are equally present. For the testing part of the dataset, we allocated 54 DM videos and 30 C videos.
2.3. Preprocessing

For a high-level abstraction of our proposed method, we show a simplified workflow scenario in Figure 2.

We intended our recordings to simulate real-world conditions, where we do not control the environment around our subjects. Therefore, to compensate for the varying lighting conditions, we employed a colour constancy method known as gray world [27, 28]. This algorithm is based on the assumption that the average reflectance of surfaces in the world is achromatic. It was previously successfully applied for skin colour normalisation for melanoma recognition [29] and for colour-based skin classification [30].

Despite the fact that our subjects were trying to keep their heads still during the experiment, there was a minor head movement between frames. It was typically caused by a very slow falling of the head of our elderly subjects while they were watching the video. To compensate for this minor head movement, Kanade–Lucas–Tomasi (KLT) feature tracker [31] was employed. Detected key points in each video frame were subsequently used to estimate the geometric transformation between two consecutive frames to track the movement of the head, which was typically only one or two pixels between consecutive frames. Both the colour constancy algorithm and KLT feature tracker enabled us to compensate for the more relaxed recording conditions.

One of the features of our presented framework is the analysis of temporal facial colour variations. We call it facial redness, and it occurs when the blood flows in and out of the face during the blood pulse wave. It is hard to see it
with the naked eye, but Eulerian video magnification (EVM) [32] helps us to visualise these small changes. The algorithm takes a video sequence as input and applies spatial decomposition followed by temporal filtering to the frames. To make the facial redness variation perceptible, the resulting signal is then amplified by a factor $\alpha$, as it is shown in Figure 3.

We conducted experiments where we employed spatial filtering using a Gaussian blur together with temporal filtering using an ideal bandpass filter, as it was suggested by the original study [32]. However, the magnification process does not bring any new information to video, it only magnifies the already existing signal. Moreover, in some cases it introduces unwanted artefacts, especially at the beginning and the end of video sequence. Since the results were not outperforming our method without EVM, we are using it here only for the visualisation purposes to demonstrate the presence of desired facial redness signal in the original video recording.

Figure 3: Illustration of the temporal facial colour variations. The first row shows specified frames from the original video, while the second row depicts results on the very same frames after the magnification with $\alpha = 100$. 

Frame 1  Frame 20  Frame 39  Frame 58
Figure 4: The top left image shows the location of regular patches P1 and P2, the top right image shows the location of neutral patches N1 and N2, the bottom left image shows the location of neutral patch N3, and the bottom right image shows the location of neutral patches N4 and N5. The generated face image was obtained from a beautycheck project [33].

2.4. Extraction of Skin Patches

In the next step, we extract skin patches for further analysis from all the recorded frames. The position of patches is marked by blue squares in Figure 4. This approach is inspired by the previous work on diabetes detection through facial specific regions [17, 18, 19, 21].

We distinguish between two different types of patches. The first type is located in cheeks, where the increase in number and diameter of superficial venules of diabetic patients has been described histologically [34]. These patches are referred to as regular patches P1 and P2. However, if we look only at these areas, our data are dependent on a particular subject and recording conditions. Therefore, we need some normalisation, which is performed within each frame. This enables us to measure differences of signals rather than subject-specific values. The normalisation is performed by neutral areas that form the second type of so-called neutral patches. Because there are more potential neutral areas, we experimentally evaluate multiple locations. After the normalisation, changes in redness of specified regions are comparable between different recordings.
The first parallel thread

Frame 1

\[ \text{sum of R values } \rightarrow v_1 \]
\[ \text{sum of R values } \rightarrow v_2 \]

(v1 – v2)\text{t}_1 = \text{final redness}

Frame 2

\[ \text{sum of R values } \rightarrow v_1 \]
\[ \text{sum of R values } \rightarrow v_2 \]

(v1 – v2)\text{t}_2 = \text{final redness}

Result: 1D signal

---

The second parallel thread

Frame 1

\[ \text{sum of R values } \rightarrow v_1 \]
\[ \text{sum of R values } \rightarrow v_2 \]

(v1 – v2)\text{h}_1 = \text{final redness}

Frame 2

\[ \text{sum of R values } \rightarrow v_1 \]
\[ \text{sum of R values } \rightarrow v_2 \]

(v1 – v2)\text{h}_2 = \text{final redness}

Result: 1D signal

---

Figure 5: Illustration of skin patches processing for a single video. \textit{R values} correspond to intensity values of the red channel of the RGB image. As a result, two 1D signals are extracted. The first parallel thread

The size of skin patch was set to 31 × 31 pixels, which was determined by the area between the nose and the mouth used for normalisation. The precise location of patches is defined by the position of eyes and nose, which were detected using a real-time detection algorithm \cite{35} implemented in MATLAB R2019b. The patches have constant predefined distance from the centre of detected eyes and nose, respectively, as shown in Figure 4. Their potential misplacement during the video due to the detection failure is controlled and evaluated every frame.

2.5. Processing of Skin Patches

Since we are interested in changes of facial redness, after the extraction of all the skin patches from all the videos, we looked at the red channel of their RGB colour representation. We derived the sum of all intensity pixels from the red channel inside each patch. It represents the \textit{redness value} or simply the \textit{R value} and we stored it for each frame of each video. Multiple options for the normalisation of R values are considered in this study: P1N1&P2N2 (P1 normalised by N1 & P2 normalised by N2), P1N2&P2N1, P1N3&P2N3, P1N4&P2N5, and P1N5&P2N4. In each case, two combinations of patches are
processed separately. We refer to them as two parallel threads. As a result, we have two 1D signals capturing the changes of redness inside each frame of the video (see Figure 5).

To illustrate the process described above, we provide figures of redness values for a diabetic (Figure 6) and a control (Figure 7) subject. In this example, each figure consists of three graphs capturing redness values of a regular patch P1

Figure 6: Three signals capturing the redness values in a single video of a diabetic subject.

Figure 7: Three signals capturing the redness values in a single video of a control subject.
Table 2: Specification of how many frames were cut out at the beginning and end of each video plus indication of the number of FV forming the remaining part of the video (in bold). For example, for FV length of 150 frames (150 f.) and video of 4850 frames, we cut out 150 frames at the beginning and 200 frames at the end of the video. The remaining frames of this video were split into 30 FV of length 150.

<table>
<thead>
<tr>
<th>FV length of 50 f.</th>
<th>FV length of 150 f.</th>
<th>FV length of 250 f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4850 frames</td>
<td>150+91*50+150</td>
<td>150+30*150+200</td>
</tr>
<tr>
<td>5900 frames</td>
<td>150+112*50+150</td>
<td>150+37*150+200</td>
</tr>
<tr>
<td>6250 frames</td>
<td>150+119*50+150</td>
<td>100+40*150+150</td>
</tr>
</tbody>
</table>

(a red signal), a neutral patch N2 (a blue signal), and their difference (a green signal). The observed oscillations correspond to the pulse signal and their variations for a diabetic subject are much higher than for a control subject. This was also the conclusion of analysis published earlier [23]. The captured signal is not perfect and some artefacts could be present in the video. However, as we will demonstrate in the evaluation part, diabetic patients are still recognised with high accuracy.

2.6. Extraction of Feature Vectors

The last step of our method consists of splitting the extracted 1D signals to so-called feature vectors. The feature vector (FV) in our study is a short chunk of signal that is directly used as an input to a classifier. We want the FV to be as short as possible but at the same time to be long enough to capture changes in blood circulation. Therefore, multiple FV lengths are considered and compared in this work, namely FV with a length of 50 frames (1 s), 150 frames (3 s), 250 frames (5 s), 350 frames (7 s), and 500 frames (10 s).

Before extracting FV, we cut out some frames from the beginning and the
Table 3: Specification of data distribution. Each cell specifies the number of FV used for training (the number of independent test samples is in brackets).

<table>
<thead>
<tr>
<th></th>
<th>FV length of 50 f.</th>
<th>FV length of 150 f.</th>
<th>FV length of 250 f.</th>
<th>FV length of 350 f.</th>
<th>FV length of 500 f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>group_DM</td>
<td>6,440 (5,796)</td>
<td>2,140 (1,926)</td>
<td>1,280 (1,152)</td>
<td>920 (828)</td>
<td>640 (576)</td>
</tr>
<tr>
<td>group_C</td>
<td>3,220 (3,220)</td>
<td>1,070 (1,070)</td>
<td>644 (640)</td>
<td>460 (460)</td>
<td>320 (320)</td>
</tr>
</tbody>
</table>

end of each video. It is our boundary condition to prevent potential inclusion of boundary artefacts. Table 2 specifies how we split each video length to FV and how many frames were cut out. The total number of samples used for training and for testing is specified in Table 3.

2.7. Classification

We use two conceptually different classifiers, namely a linear discriminant analysis (LDA) and a support vector machine (SVM). The LDA uses a generative approach and has an analytical solution, where it assumes that the data points have the same covariance. On the other hand, the SVM is a discriminative method, which makes use of kernels to transform the parameter space resulting in a non-linear classifier. We employed radial basis function kernel for SVM that is typically used to classify data that are not linearly separable. Other kernels like polynomial might lead to overfitting to training data.

Both LDA and SVM classifiers are very popular and were chosen because of their differences. For hyper-parameter settings, we performed an automatic optimisation to minimise five-fold cross-validation loss over the training dataset. It automatically sets optimal parameters for each dataset separately without any manual intervention.
3. Results

In this section, we address the question if it is possible to distinguish reliably between recorded diabetic and control samples. For this purpose, our framework is evaluated on two levels: a feature vector (FV) level and a video level. The illustration for these two levels of classification is provided in Figure 8.

The first set of tests was conducted on the FV level. We merged all FV inside each parallel thread together. As a result, for FV length of 50 (denoted as \( fv_{level.50f} \)), we had in each parallel thread in total 6440 diabetic plus 3220 control FV for training and 5976 diabetic plus 3220 control FV for testing. For different FV lengths, exact numbers could be analogously derived from Table 2.

The feature vector classification was performed on each thread independently, which implies that we had separate classification output for both combinations of patches. In the end, we assigned to each testing FV a label with the highest posterior probability among both threads.

The second set of tests was conducted on the video level, where we grouped classification outputs belonging to the same video. As a result, we classify the entire video, and we make a prediction about the health status based on the majority voting. In this scenario, we accept some parts of the video (some FV) to be incorrectly classified, which increases the robustness of our system.

All sets of tests were performed using a multiclass condition and a binary condition. In the multiclass condition, we split diabetes group into two subgroups based on number of complications they encounter (see Table 1). The
first subgroup consisted of people with all three complications (Diab3), while people in the second subgroup were diagnosed with one or two complications (Diab1or2). There is also a control group with zero complications. For this condition, we provide accuracy results in Figure 9 and confusion matrices for the best performing variant, which is SVM classifier with FV length of 350 in Figure 10. Please note that classes are not balanced here (see Table 1).

In the binary condition, we merged all diabetic patients with complications to a single group and we provide overall accuracy results in Figure 11, sensitivity and specificity results in Figure 12, and detailed results for the best performing variant in Table 4.
Figure 10: Confusion matrices of multiclass classification. Rows and columns correspond to predicted and actual/true classes, respectively. Values are in %.

4. Discussion

All presented figures depict a clear pattern, where P1N2&P2N1 combination dominates over all other variants. In multiclass classification, P1N2&P2N1 showed a significantly higher accuracy (p<0.05) than all other normalization approaches (i.e. P1N1&P2N2, P1N3&P2N3, P1N4&P2N5, P1N5&P2N4). In binary classification, P1N2&P2N1 showed higher accuracy than all other normalization approaches, but this difference in accuracy was statistically significant (p<0.05) only when compared to P1N1&P2N2, P1N4&P2N5, P1N5&P2N4, and not statistically significant (p>0.05) for P1N3&P2N3. The paired-samples t-tests were used for determining statistical significance here.

The P1N2&P2N1 variant normalises values from the right-hand side cheek with the left-hand side area under the nose and vice versa. We see the reason in the potential uneven illumination of facial area, when the light source is not directly in the centre of the face. Therefore, the cross-over normalisation of patches makes more sense than using the same side of the head. Alternatively, the central normalisation patch on the nose could be also considered in the future, since normalisation using N3 patches had the second best results.

Areas in the forehead (patches N4 and N5) had the lowest accuracy. We see the explanation in Figure 3 where the forehead is clearly changing colours during the heart beat. These areas in forehead were previously not recognised with the increase in number or diameter of superficial venules in diabetic patients, therefore we tested them for normalisation purposes.
Figure 11: Overall accuracy results of binary classification.

Figure 12: Overall sensitivity and specificity results of binary classification.
Table 4: Results for P1N2&P2N1 on two different levels and for five different FV lengths.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SVM</td>
<td>LDA</td>
<td>SVM</td>
</tr>
<tr>
<td><strong>fv_level_50f.</strong></td>
<td>92.78 %</td>
<td>86.30 %</td>
<td>100 %</td>
</tr>
<tr>
<td></td>
<td>79.97 %</td>
<td>79.94 %</td>
<td></td>
</tr>
<tr>
<td><strong>video_level_50f.</strong></td>
<td>92.86 %</td>
<td>86.90 %</td>
<td>100 %</td>
</tr>
<tr>
<td></td>
<td>79.97 %</td>
<td>79.94 %</td>
<td></td>
</tr>
<tr>
<td><strong>fv_level_150f.</strong></td>
<td>78.50 %</td>
<td>86.35 %</td>
<td>86.86 %</td>
</tr>
<tr>
<td></td>
<td>86.35 %</td>
<td>86.86 %</td>
<td>89.88 %</td>
</tr>
<tr>
<td></td>
<td>63.46 %</td>
<td>80.00 %</td>
<td></td>
</tr>
<tr>
<td><strong>video_level_150f.</strong></td>
<td>78.57 %</td>
<td>86.90 %</td>
<td>87.04 %</td>
</tr>
<tr>
<td></td>
<td>97.74 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>63.33 %</td>
<td>80.00 %</td>
<td></td>
</tr>
<tr>
<td><strong>fv_level_250f.</strong></td>
<td>92.35 %</td>
<td>85.60 %</td>
<td>99.22 %</td>
</tr>
<tr>
<td></td>
<td>88.72 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80.00 %</td>
<td>80.00 %</td>
<td></td>
</tr>
<tr>
<td><strong>video_level_250f.</strong></td>
<td>92.86 %</td>
<td>85.71 %</td>
<td>100 %</td>
</tr>
<tr>
<td></td>
<td>88.89 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80.00 %</td>
<td>80.00 %</td>
<td></td>
</tr>
<tr>
<td><strong>fv_level_350f.</strong></td>
<td>90.22 %</td>
<td>86.65 %</td>
<td>95.53 %</td>
</tr>
<tr>
<td></td>
<td>90.34 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80.65 %</td>
<td>80.00 %</td>
<td></td>
</tr>
<tr>
<td><strong>video_level_350f.</strong></td>
<td>92.86 %</td>
<td>86.90 %</td>
<td>100 %</td>
</tr>
<tr>
<td></td>
<td>90.74 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80.00 %</td>
<td>80.00 %</td>
<td></td>
</tr>
<tr>
<td><strong>fv_level_500f.</strong></td>
<td>87.61 %</td>
<td>75.33 %</td>
<td>90.80 %</td>
</tr>
<tr>
<td></td>
<td>81.77 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>81.87 %</td>
<td>63.75 %</td>
<td></td>
</tr>
<tr>
<td><strong>video_level_500f.</strong></td>
<td>92.86 %</td>
<td>75.00 %</td>
<td>100 %</td>
</tr>
<tr>
<td></td>
<td>81.48 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80.00 %</td>
<td>63.33 %</td>
<td></td>
</tr>
</tbody>
</table>

In the multiclass classification option, we created a very challenging setting. Our aim was to test if we can recognise samples from patients with various number of diagnosed complications and capture differences between them. However, based on presented confusion matrices, we were not very successful in this task. The majority of video samples with all three complications were classified to the other diabetic category. We see multiple reasons for this. One of them is relatively low number of samples and their uneven distribution. The Diab1or2 class represents a majority class and attracts a lot of test samples from Diab3 class. The other reasons we see in technical limitations, which resulted in small extracted patches.

However, confusion matrices of multiclass classification also reveal that control group is recognised with much higher accuracy. Moreover, the main miss-classification is with Diab1or2 class and there is almost no miss-classification between control class and Diab3 class. This indicates that there is a measurable difference between samples from different categories. Despite all the drawbacks, our proposed framework had classified accuracy of 67.31% on FV level and
71.43% on video level in this challenging multiclass classification setting.

In the binary classification option, we aimed to distinguish between control samples and samples from people with diabetic complications. Since the classification accuracy itself is not a sufficient metric and could be misleading, we provide measures of sensitivity and specificity, which should be considered together as complementary metrics. For the SVM classifier, the results are generally higher than for LDA classifier and most importantly, the sensitivity on video level is 100% for almost all FV lengths. It shows that we correctly recognise all patients with diabetic complications, which is the most important tasks. These results are supported by the specificity of 80%.

Because of their nature and specific conditions, these results can not be directly compared with other frameworks. As we mentioned in related work, there are works that report 97.54% average accuracy [17] in laboratory conditions using specific equipment designed for this task. Moreover, they are working with static images only, while our framework works in temporal domain. Another machine learning model for diabetic complications was reporting 83.8% accuracy [13] but they were working with electronic health data.

To the best of our knowledge, we had the very first successful study on the usage of videos from the conventional camera for this type of medical prediction [23]. Our study confirms the hypothesis that it is possible to discriminate between samples recorded from patients with diabetic complications and control subjects. We combined multiple techniques from image processing, pattern recognition, and machine learning to process video frames and predict the health status of our subjects with high efficiency. The presented results are very promising and encouraging, especially because of 100% sensitivity on video level tests.

At the same time, we should state here that this is a pilot study. One of its limitations is the relatively small number of participants resulting in a small number of samples. Therefore, the current results, while encouraging, should be interpreted carefully and considered as the proof of concept for a potential novel marker for diabetic complications.
In the future, we plan to extend the study by involving more patients with more degrees of diabetic complications, from absent to severe. The potential confirmation of our results would provide the patients and with a feasible, non-invasive, and cost-effective telemedicine framework for automatic detection of early signs of diabetic complications.

Declaration of Competing Interest

We have no conflict of interest to declare.

Study Protocol

The study protocol was reviewed and approved by the scientific ethical committee for Region Southern Denmark. All participants were informed about the study, and they signed written informed consents.

Acknowledgements

This study was funded by the University of Southern Denmark (Strategic Research Focus Areas 2016 - SMILE project - A mobile Health solution for early detection of complications in diabetic patients) and Region Southern Denmark (OUHs Innovationspulje 2017 – Project: Spatiotemporal video magnification in patients with diabetic complications).

References


