

High-Performance Detection of Corneal Ulceration Using Image Classification with Convolutional Neural Networks

Jan Gross
Aalen University,
Germany
jan.gross@hs-aalen.de

Johannes Breitenbach
Aalen University,
Germany
johannes.breitenbach@hs-aalen.de

Hermann Baumgartl
Aalen University,
Germany
hermann.baumgartl@hs-aalen.de

Ricardo Buettner
Aalen University,
Germany
ricardo.buettner@hs-aalen.de

Abstract

Corneal Ulcer, also known as keratitis, represents the most frequently appearing symptom among corneal diseases, the second leading cause of ocular morbidity worldwide. Consequences such as irreversible eyesight damage or blindness require an innovative approach that enables a distinction to be made between patterns of different ulcer stages to lower the global burden of visual disability. This paper describes a Convolutional Neural Network-based image classification approach that allows the identification of different types of Corneal Ulcers based on fluorescein staining images. With a balanced accuracy of 92.73 percent, our results set a benchmark in distinguishing between general ulcer patterns. Our proposed method is robust against light reflections and allows automated extraction of meaningful features, manifesting a strong practical and theoretical relevance. By identifying Corneal Ulcers at an early stage, we aid reduction of aggravation by preventively applying and consequently tracking the efficacy of adapted medical treatment, which contributes to IT-based healthcare.

1. Introduction

By potentially causing 1.5 to 2 million new cases of monocular blindness every year, ocular traumas and especially Corneal Ulcers (CUs) are the most severe causes of corneal blindness [1, 2]. The clinical picture of CUs has been referred to as a global challenge, for both the developing world and industrialized economies [1, 3, 4]. In the United States, about 75,000 cases of CUs have been reported annually [3].

Usually, the course of CUs disease is followed by corneal abrasions, which are often the direct cause of ocular injuries, amplifying the challenge of treating corneal blindness once it has occurred [5].

The success of state-of-the-art surgery concerning existing corneal blindness is very limited unless

experienced medical personnel and modern equipment are available [6].

Although human vision is afforded a high value, the identification of eyesight impairment is delayed in many cases due to the often slow and unnoticed course of the disease [7]. However, early intervention can immensely increase the healing process and lower aggravation, making cost-effective public health prevention programs highly crucial to industrialized and developing economies [1, 3]. Schwartz et al. [8] revealed that regarding the present levels of blindness in a Central African Republic's population that they investigated, 95.5 percent of all existing cases are preventable and treatable. While it is clear that prevention contributes to a cost-effective treatment of corneal blindness [1], manual analysis of the ocular surface is a widely used method for assessing patterns of CU [9].

However, manual assessment is highly time-consuming, and the capacities for treating the disease are very limited [1]. Furthermore, manual visual assessment of the widely used fluorescein staining images is considered based upon assumptions and clinical intuition rather than on an evidence-based analysis [10], making objective data analytics highly relevant [11].

The recent upswing within Information Systems research in healthcare sustainably contributes to a huge success in medicine, especially for the automated diagnosis and prevention of diseases [11-17].

Especially in image processing tasks, Convolutional Neural Networks (CNNs) have established themselves as a successful approach due to their superior performance [18-21], which is particularly important in the healthcare domain [11-16, 21]. The automated extraction of descriptive and discriminative features is among the major advantages of CNNs [18, 21].

Although several approaches have aimed to automate the identification of CUs using machine

learning techniques, they are either based on a small dataset, do not distinguish between different types of the disease, or do not manifest a sufficient level of accuracy to be implemented in the healthcare domain [9, 11, 21-29].

In this study, we aim to tackle these challenges and present a CNN-based classification approach that uses fluorescein staining images to differentiate between different patterns of CUs. Our five main contributions are:

1. We present a highly effective approach for distinguishing between different general ulcer patterns with a very good balanced accuracy of 92.73 percent.
2. Our CNN-based classification approach is accurate, robust, cost-efficient, and enables automated differentiation between different shapes and distribution characteristics of CU by only using a staining image of the ocular surface.
3. By presenting a model that is robust against light reflections on the ocular surface, we avoid loss of visual information, which is of high practical relevance.
4. We show that our model allows automated extraction of clinically meaningful features that underpin the predictive areas within the cornea, important for diagnosing CU, which is of high theoretical relevance.
5. By proposing an automated model that allows preventive intervention and monitoring therapeutic efficacy, we contribute to Information Systems research in healthcare [11, 30, 31].

The paper is organized as follows: First, we present our research background, including related work on automated CUs investigation, highlighting the need for our approach. Subsequently, we give an overview of our methodology, including our model architecture, evaluation data, and preprocessing. After that, we show the performance results of our CUs classification model. Next, we discuss the results before concluding with limitations and offering suggestions for future work.

2. Research Background

2.1 Clinical Picture of Corneal Ulcer

CU, also known as keratitis, is the most common symptom of corneal diseases. CUs can be caused by various risk factors and typically cause mild to severe eye discharge and reduced vision [1, 2, 32]. Late diagnosis or inappropriate treatment can lead to irreversible damage to human eyesight or even

blindness [1, 2, 7]. Medical diagnostics are crucial to identify corneal diseases and avoid visual impairment [1].

2.2 Assessment of Ocular Surface Disease

The most widely used diagnostic tool to investigate ocular surface diseases is staining with fluorescein [9]. The technique allows a determination to be made of the different types of CUs based on the chromatic and morphological properties [10].

Although manual analysis of the staining images is considered the most reliable method for assessing ulcer patterns, it is also highly time-consuming. While various factors contribute to the inconsistency of fluorescein's appearance, human factors such as bias, fatigue, and mindset contribute to the inconsistency and unreliability of manual visual assessments [9, 23, 33].

However, to avoid and reduce visual impairments, appropriate medical treatment has to be reliably applied at an early stage. Furthermore, it is essential to monitor clinical therapy efficacy to ensure that it leads to success [9, 34].

To overcome these challenges and meet the requirements of sophisticated CU management, automated techniques can be used since they are less influenced by human factors, allowing a consistent and objective investigation of CUs [9, 22, 27-29, 33].

2.3 Automated Investigation of Corneal Ulcers

Several semi- and fully-automated approaches have been proposed in the field of CUs investigation [22-29].

Chun et al. [25] and Peterson et al. [26] have implemented a color extraction algorithm, while Rodriquez et al. [24] used digital image analysis to extract the CU area of interest. Furthermore, Deng et al. [23] proposed a semi-automated approach based on manual segmentation of the CU area.

Considering fully-automated approaches, Sun et al. [22] and Loo et al. [27] developed deep learning-based approaches that target the CU area segmentation. Furthermore, Akram and Debnath [29] used CNNs to detect and subsequently segment the ulcer area.

However, all these studies represent segmentation approaches of CUs and do not allow the classification of different types of CU patterns, which we propose and allow with our detection approach.

As one of the most recent approaches and closest in comparison to our study, Kuo et al. [28] developed a deep learning-based corneal photograph model for distinguishing between different types of microbial keratitis. Although their work shows promising results, the average diagnostic accuracy of 69.4 percent does not meet the high level of accuracy which is of vital

importance in the healthcare domain [21]. Furthermore, Kuo et al. [28] evaluated their diagnostic performance using a dataset with a very limited amount of 288 photographs, most likely possessing uncertainty in the model parameters and generalization capabilities [35].

All the presented approaches show promising results towards diagnosing and preventing the disease through automated CU investigation using corneal and fluorescein staining images. However, the proposed models are not suitable to efficiently and reliably realize the target of detecting and distinguishing between different general CU patterns. They either disregard the different types of CU patterns, include manual preprocessing, are only based on a small dataset size, or do not manifest a sufficient level of accuracy to be implemented in the healthcare domain [11, 21, 35]. All these limitations lead to vague generalization or feasibility capabilities in real-world applications.

To be able to provide clinically adapted therapy and track the therapeutic efficacy, a high level of accuracy is required. Investigations using larger sample sizes can reduce the epistemic uncertainty and improve generalization capabilities [35]. Furthermore, CUs must be diagnosed and assessed by their size, shape, and distribution characteristics [9].

3. Methodology

3.1 Model Architecture

In this study, we use a CNN-based approach [18] for the automated classification of different CU types. CNNs have been established as a widely used method in eye-related research [22, 33, 36, 37].

The combination of automated feature extraction and final classification represents one of the major advantages of CNNs [18, 38]. Regarding the objective of reducing negative human factors such as bias, fatigue, and mindset during the medical assessment process to achieve objective and reproducible results [9, 33], the automated feature extraction represents a fundamental backing for our approach.

Since the sample size manifests strong influences on the performance of CNNs [9], large and high-quality training datasets are required to ensure robust and good performance [39]. Even recent datasets in the eye-related image tasks show that the acquisition of data amounts needed to train a CNN from scratch represents an unfeasible process [9, 39-41].

To overcome the sampling challenge of quantitatively sufficient training data, we based our approach on a transfer learning strategy. Transfer learning greatly reduces the amount of required training data by transferring knowledge from other domains to the target domain. By removing the original and adding

a new individual classifier, the pre-trained model can be individually adapted to fit the purpose of classifying general CU patterns [42, 43].

For this study, we used the pre-trained CNN-based VGG16 architecture, proposed by Simonyan et al. [44], to realize the transfer learning approach. The VGG network represents a high-performance network for large-scale image classification and has been established as a representative network method for investigating CUs segmentation tasks [22]. The VGG network provides a lot of initial network knowledge that can be effectively generalized to fit the investigation of the ocular surface and many datasets (e.g., SUSTech-SYSU), which substantially fosters the performance and robustness of our model.

For the building of our model, we excluded the top layer of the VGG16 base model and applied global average pooling to transfer the 4D tensor to a 2D tensor using a layer with 512 units [45]. Subsequently, we added our classifier to the base model, consisting of two dense layers with 64 and 32 units, each followed by a dropout layer with a rate of 0.5 [46, 47]. We also used an appropriate L2 weight regularization with a weight decay of $1e-4$ (alpha), ensuring the preservation of the prior network knowledge and preventing our model from overfitting [38].

To achieve an efficient stochastic optimization, we used the AMSGrad variant of the Adam optimizer algorithm [48]. For our binary classification problem, we added the output layer in a final step. The output layer includes the sigmoid activation function, where the function always returns a value between 0 and 1 for the classification as either early or advanced stage [46]. To then train our new classifier, we froze the layers of the base model. For the fine-tuning of our model, we unfroze the last ten layers of the VGG16 base model and trained the layers with a learning rate of $1e-5$.

3.2 Evaluation Data and Preprocessing

We trained and evaluated our model on the SUSTech-SYSU dataset, which consists of 712 fluorescein staining images. It has been acquired at the Zhongshan Ophthalmic Center at Sun Yat-sen University. The study protocol was conducted in accordance with the ethical principles of the Declaration of Helsinki (2017KYPJ104) and approved by the Zhongshan Ophthalmic Centre ethics committee of Sun Yat-sen University. Also, informed consent was obtained from each patient. While no selection criteria for the patients (e.g., age, sex, ulcer cause) have been considered, the potential for sampling bias due to demographic factors was significantly reduced. The provided images capture the ocular surfaces from patients having various CU degrees [9]. In this study, we

used all 712 fluorescein staining images for our assessment.

Fluorescein was applied through drops of sterile saline to both eyes of the patients. Slit-beam lighting with the maximum width (30 mm) of the white light source, a diffusion lens at a 10 to 30° oblique angle, with the light source sitting at the midpoint between the pupil margin and limbus, a blue excitation filter, as well as a 10 or 16 magnification was used. A Haag-Streit BM 900 slit-lamp microscope (Haag-Streit AG, Bern, Switzerland) combined with a Canon EOS 20D digital camera (Canon, Tokyo, Japan) was utilized to capture the images. These were then saved with a resolution of 2,592×1,728 pixels in JPG format in 24-bit RGB color. Each image contains one cornea, fully presented in the image, and approximately centered in the visual field.

Three experienced ophthalmologists from the Zhongshan Ophthalmic Centre at Sun Yat-sen University, responsible for clinical work and ophthalmic research, performed the labeling of the ocular staining images. While two ophthalmologists classified each staining image independently, the third ophthalmologist re-classified the images with inconsistent classification results from the other two ophthalmologists. Although the results of subjective assessments (i.e., labeling of ocular staining images) can differ due to human factors (e.g., bias, fatigue, mindset), the re-classification approach using three experienced ophthalmologists has significantly reduced the aleatoric uncertainty and noise in the labels of the dataset [9, 35]. Although this procedure is highly time-consuming and would be unfeasible in a daily clinical routine, it fosters the validity and reliability of the provided labels. As the performance of our classification approach as a type of supervised learning is strongly bound to the quality of the underlying reference values [18], the labels will significantly support the validity of our study results.

Concerning the shape, size, and distribution of the ulcer, the staining images have been classified into the three categories point-like (358 images), point-flaky mixed (263 images), and flaky (91 images), according to general ulcer patterns [9].

Point-like CU patterns are characterized by a relatively mild clinical picture. Identifying this state of the disease is of high relevance due to its preliminary appearance. At this particular time, appropriate and effective treatments can significantly increase the chance of a full recovery and prevent further progression of the disease. Furthermore, *point-like* patterns represent CUs that are almost cured and can, therefore, be used to validate the applied treatment's success [9].

Point-flaky mixed ulcer patterns describe an irregularly distributed shape containing *point-like* and flaky ulcers within the cornea. This condition usually

refers to a degree of severity that lies between the two types of CUs. Due to the range of this stage of the disease, it can also refer to either an advanced *point-like* state or provide a clinical picture that shows initial success in treating *flaky* ulcer patterns [9].

Flaky patterns represent the most severe state of CU patterns. This state may leave scars on the ocular surface, noticeably affecting the patient's eyesight [9].

In this study, we propose a binary model approach (see Figure 1). This approach is based on the aggravation and healing progress of CUs, leading to the need for a preliminary detection and continuous monitoring of the applied therapy [9]. Therefore, we combined the provided ulcer categories into two classes, where we define *point-like* ulcers as an early stage, while we refer to *point-flaky mixed* and *flaky* as advanced stages. The consolidation into two classes might generally lead to an ambiguity of the point-flaky-mixed-related characteristics. However, based on our specific use case of identifying CUs at an early stage, our approach aims to identify point-flaky mixed patterns as a premature advanced stage. Since most cases are preventable and the potential to fully recover and prevent further progression is better the earlier the disease is diagnosed, our approach allows experts to prematurely allocate an adapted medical treatment [1, 8, 9].

1. Early Stage: Monitoring therapeutic efficacy and identifying potential risks of an aggravated course of the disease.
2. Advanced Stage: Identifying clinical pictures that require appropriate medical treatment.

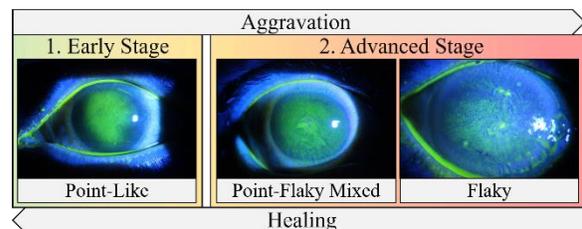


Figure 1. Our preventive model approach.

Through our approach, we achieved a nearly balanced dataset of 358 staining images belonging to the negative class (early stage) and 354 staining images belonging to the positive class (advanced stage).

Considering the image data acquisition in real-world scenarios, reflections are inevitable, which can strongly harm many computer vision tasks [49, 50]. Due to the slight difference between the chromatic (i.e., green color emitted from the dye) appearance of the ulcer patterns and potential light reflection spots within the dataset, we used thresholding (global binarization) to avoid loss of visual information by identifying and reducing the reflection spots in the staining images.

To improve the computational efficiency, all images were resized to a fixed input size of 324x216 pixels. Furthermore, we rescaled the input RGB channels to [-1,1] and applied data augmentation to the training data by rotating the images by 25 degrees, zoom in and out on within a range of 80 to 120 percent, and changing the brightness of the images within a range of 70 to 130 percent. These preprocessing steps cover most of the varying conditions that might occur during data acquisition and further improved our model’s robustness and classification accuracy [46].

To foster our approach’s robustness and internal validity, we used hold-out cross-validation to evaluate the performance of our classification model on unseen image data. This method splits the entire dataset into subsets for both training and testing [38, 51]. For this study, we applied a randomized split to separate the dataset into 80 percent training data (569 images) and 20 percent testing data (143 images) using a five-fold approach. For each of the five folds, the training data was divided into five splits, while a model was trained with four splits and validated with one split. The final performance of the model was then evaluated using the completely unseen testing data.

Based on the nearly balanced dataset through our binary approach, the testing data of 143 images contained 72 images of early-stage CUs (negative class) and 71 images of advanced-stage CUs (positive class).

4. Results

For the building and training of our CNN-based model, we used Python 3.6.9 with the Keras 2.3.1 package [52] and TensorFlow 1.15.2 as backend [53], running on an NVIDIA Tesla T4 16GB graphics processing unit. We trained our model, over all five folds of the hold-out cross-validation [51], for 50 epochs with a batch size of 64 images.

To achieve the final results of our classifier over all five folds of the hold-out cross-validation, we used the arithmetic mean to average the values of each fold. Following this approach, the proposed CNN provides a balanced accuracy of 92.73 percent with a very low standard deviation of 0.94 percent in differentiating between early and advanced CU staining images. Cohen’s-Kappa score is 85.46 percent, and the AUC-ROC is 0.98. Both the very good classification accuracy and the low standard deviation of all the performance indicators underpin our model’s very good results and robustness. Table 1. provides the arithmetic mean values of our considered performance indicators of the CNN model over all five folds. The prevalence of advanced CUs does not manifest a standard deviation since one unseen test-split is used for all five folds.

Table 1. Performance indicators.

Performance Indicator	Mean Value	STD
Balanced Accuracy	92.73%	0.94%
True positive rate	93.52%	1.26%
True negative rate	91.94%	2.67%
Positive predictive value	92.03%	2.41%
Negative predictive value	93.53%	1.12%
Kappa	85.46%	1.89%
AUC-ROC	0.9798	0.0018
Prevalence	49.65%	-

The confusion matrix shown in Table 2 provides an overview of our classifier’s results, including the amount and type of errors that occurred during evaluation. The shown values are averaged over all five folds of the hold-out cross-validation using the arithmetic mean. Since the confusion matrix shows the model’s evaluation results, the individual values add up and refer to the testing data of 143 images. While, on average, 66.4 of the 71 referenced advanced CU images within the testing data were correctly classified, our model achieved a sensitivity (true positive rate) of 93.52 percent. Of the 72 images referenced as early-stage CUs, on average, 66.2 were correctly classified as images showing early-stage ulcers, leading to a specificity (true negative rate) of 91.94 percent.

Table 2. Confusion matrix: mean values based on unseen testing data.

		Reference	
		Early	Advanced
Prediction	Early	46.29% (66.2)	3.22% (4.6)
	Advanced	4.06% (5.8)	46.43% (66.4)

While the evaluation reveals that our CNN-based model is able to correctly classify CUs of the early (true negative) as well as advanced (true positive) stage in equal measure, we used the Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm to visually explain the predictive areas of our classification model by transferring the gradients of our target concept into the final convolutional layer [54]. We applied a pseudo/false-color to the heatmap, where the areas highlighted in red/orange represent important areas, whereas less relevant regions are marked in blue/turquoise. We highlight the fact that due to up- and downscaling of the images for the algorithm, the provided heatmaps are not pixel-accurate and mainly manifest a reliable trend towards the predictive areas.

Figure 2 shows the preprocessed staining images (A) and the corresponding heatmaps with predictive

areas (B) for the early and advanced stages. Images labeled as (C) show the preprocessed image with the overlaid heatmap (B) with 70 percent opacity. The red circular markings highlight the respective cornea overlay, which has been provided as part of the dataset, and which we superposed for evaluation purposes.

We enlarged the provided cornea overlay by manually increasing the line weight (thickness), following the provided path for better visualization. The white circular markings focus on the glare spot artifacts that we reduced as part of our preprocessing.

As shown in images (C), our model pays special attention to the corneal region, which is very important for providing a clinically valid diagnosis, by representing meaningful features for the CUs differentiation. While the two staining images (A) appear to manifest very similar ulcer pattern distributions, the well-balanced classification performance and predictive feature of our model foster our binary approach and the functionality of our model for a variety of pattern types.

5. Discussion

Our results reveal that our model extracts meaningful features and correctly classifies the two CU stages based on our binary approach.

Based on the images (B) and (C) shown in Figure 2, it can be seen that our CNN model mainly focuses on the corneal area, referring to a heatmap focus within the cornea overlay. While the CU severity degree is closely related to the location in the cornea area [9], the region that is most highly and commonly stained in corneal lesions, our model fosters the theoretical relevance of our work, providing meaningful features for the diagnosis of CUs [9, 34, 55].

Moreover, image (B) reveals that our model is robust against the influence of light reflection spots. Reflections are inevitable in real-world settings and can cause loss of essential visual information. However, the predictive areas (B) reveal that our model is not affected by the reduced light reflection artifacts, which is of high practical relevance. This insight fosters our preprocessing step and shows that the glare reduction, resulting in a reduced contrast between the reflection and the fluorescein dye's appearance, contributes to our model's performance, allowing meaningful features to be extracted. Although the raw staining images provide a wide variety of ulcer patterns within each of the initially provided classes (*point-like*, *point-flaky mixed*, and *flaky*), leading to potentially overlapping appearances across the classes, our model can extract meaningful features fully-automated, based on the corneal size, shape, and distribution, which allows the

fine-graded distinction of early and advanced stage ulcer patterns.

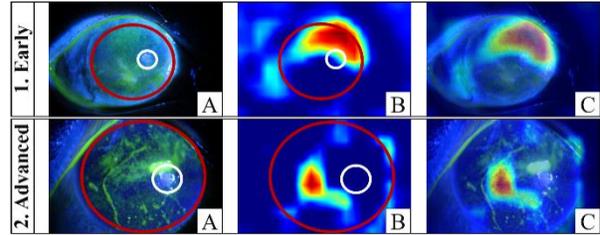


Figure 2. Post-hoc analysis of correctly classified images using Grad-CAM heatmaps.

Based on the true positive rate, our model reveals considerable similarities between the dataset-referenced classes *point-flaky mixed* and *flaky* regarding shape, size, and distribution of the ulcer patterns. Furthermore, the dominance of the false-positive values (4.06 percent) compared to the false-negative values (3.22 percent) shows that the *point-flaky mixed* class recordings significantly need to be considered a premature advanced stage. Our results may also stimulate future research concerning the consideration of the *point-flaky mixed* as an intermediate stage of the disease. These outcomes of our model foster the idea of our approach. Furthermore, we contend that considering the *point-flaky mixed* as an intermediate stage might lead to a trivialization of preliminary *flaky* patterns, potentially causing the loss of valuable time in treating CUs [1, 7, 9]. To foster the architecture of our model, we trained and compared the classification results of other pre-trained networks (i.e., VGG19, Xception, ResNet50V2) on the 712 staining images of the SUSTech-SYSU dataset [9]. We used the same binary sampling method and the identical train-test split among the studies to allow a standardized comparison of the classification results. Table 3 provides an overview of the classification performance of the other pre-trained models, averaged overall five folds of the hold-out cross-validation using the arithmetic mean. While none of the pre-trained models we used as baseline architecture were able to outperform our VGG16-based architecture in terms of classification accuracy on the SUSTech-SYSU dataset [9], our model also manifests by far the lowest standard deviation with 0.94 percent, underlining the robustness of our model.

Table 3. Performance indicators of other pre-trained networks.

Pre-Trained Network	Balanced Accuracy	STD
VGG16 [44]	92.73%	0.94%
VGG19 [44]	92.47%	2.64%
Xception [56]	67.31%	1.44%
ResNet50V2 [57]	73.45%	3.31%

6. Conclusion

In this study, we have proposed and validated a highly effective and automated classification method for CU patterns. Based on a transfer learning strategy, we were able to greatly reduce the amount of necessary training data [35]. Alongside the implementation of data augmentation, the application of thresholding to remove light reflections represents a substantial step to improve the performance and robustness of our approach for real-world applications. Since we designed our model so that it requires a small input size of just 324x216 pixels, we were also able to improve the computational efficiency and lower the requirements regarding the quality of the input images.

Utilizing a total of 712 fluorescein staining images, we have demonstrated the performance of the proposed method in terms of classification accuracy for different shape and distribution characteristics of CUs. Related studies either disregard different types of CU patterns, include manual preprocessing, are only based on a small dataset size or do not manifest a sufficient level of accuracy. However, a method is required that allows for a distinction to be made between different types of CUs to adapt treatments and track their efficacy. Our approach is an example of such a method and achieves a balanced accuracy of 92.73 percent based on completely unseen testing image data.

According to the assumption of our approach for this specific use case, the advanced types of CUs should be considered as a holistic advanced stage through a binary strategy, requiring increased attention concerning the treatment. Especially regarding the potential of preventable and treatable corneal blindness [8], the challenge is to counteract the often slow and unnoticed development of eyesight impairments [1, 7].

While surgical interventions, required due to a delayed identification, are not cost-effective concerning the elimination of corneal blindness, prevention is more cost-effective. It leads to higher success in decreasing the prevalence of blindness [1]. With our model, particularly in the era of big data [11, 12], we contribute to the IT-based healthcare domain by strongly improving the efficiency of identifying CU.

6.1 Limitations

The sample size of the SUSTech-SYSU dataset represents one of the main limitations [9, 40, 41]. While the internal validation is very high due to the hold-out cross-validation, the external validity needs to be further investigated. Tests involving a wider variety of ulcer patterns are required to fully evaluate our model's robustness and its generalization capabilities for large-scale medical applications. Although our model can

achieve a good classification performance, another validation step for our future work will be to evaluate how the model performs in real-world scenarios, where the image quality may vary due to different conditions (e.g., equipment, lighting, camera angle).

Another limitation is related to the hardware we used for training and performance evaluation. Calculations were made on a hosted graphics processing unit that might exceed some computers' hardware capabilities in the healthcare environment. Therefore, tests on computers, used in medical environments, need to be performed.

We highlight that all the presented limitations represent general challenges of medical image analysis models, particularly CU detection approaches.

6.2 Future Work

Future work includes a re-evaluation of our model using more extensive datasets. This includes the patients' demographic factors (e.g., age, sex) and a wider variety of point-like CU patterns to improve the preventive detection. In particular, further investigating point-like CUs at an early stage will help to increase the chances of a full recovery [8, 9]. Furthermore, we will examine the performance and runtime of our approach for different embedded platforms.

Future research also involves the investigation, how our approach supports the CU diagnosis at an early stage. A case-control study could help to determine which value our approach adds and how it performs compared to the manual diagnosis of the disease.

The presence of reflections in real-world data represents an unavoidable challenge, potentially causing a significant loss of visual information. The absence of clear eye visibility, in particular, can affect the performance of computer vision tasks [49].

The severity degree of CUs strongly relies on the patterns' location in the cornea relative to the central optical zone [9]. Therefore, reflections in the data can potentially affect automated CUs investigation models. While we have already successfully implemented thresholding to reduce light reflections as part of our preprocessing, we will further investigate the influence of reflections in eye-related research areas.

Furthermore, we will pursue to investigate how the integration of our automated CU approach into real-world Human-Computer-Interaction in multi-agent-settings [58, 59] might influence the cognitive workload [60-62] and other user-oriented concepts [63-65] of the health personnel (e.g., surgeons and nurses). We aim to analyze this interaction using physiological sensor data (e.g., electroencephalographic data [66-72] and spectra [73-78], electrocardiographic data [79, 80], electrodermal activity [81], eye fixation [82, 83], eye

pupil diameter [84, 85], facial expressions [86]). We also plan to investigate levels of trust [87, 88] and technology acceptance [89-92] of both clinical experts and patients towards the practical implementation of our model and confirm if the automated approach improves the coordination of doctors and treatments more efficiently.

Acknowledgments

This research is funded by the Carl Zeiss Foundation and the German Federal Ministry of Education and Research (13FH4E03IA, 13FH4E07IA, 13FH176PX8, 13FH4E05IA).

7. References

- [1] J. P. Whitcher, M. Srinivasan, and M. P. Upadhyay, "Corneal blindness: a global perspective," *Bull. World Health Organ.*, vol. 79, no. 3, pp. 214–221, 2001.
- [2] K. H. Cheng, S. L. Leung, H. W. Hoekman, W. H. Beekhuis, P. Mulder, A. Geerards, and A. Kijlstra, "Incidence of contact-lens-associated microbial keratitis and its related morbidity," *Lancet*, vol. 354, no. 9174, pp. 174–175, 1999.
- [3] G. Amescua, D. Miller, and E. C. Alfonso, "What is causing the corneal ulcer? management strategies for unresponsive corneal ulceration," *Eye*, vol. 26, pp. 228–236, 2011.
- [4] J. P. Whitcher and M. Srinivasan, "Corneal ulceration in the developing world—a silent epidemic," *Brit. J. Ophthalmol.*, vol. 81, no. 8, pp. 622–623, 1997.
- [5] M. P. Upadhyay, P. C. Karmacharya, S. Koirala, D. N. Shah, S. Shakya, J. K. Shrestha, H. Bajracharya, C. K. Gurung, and J. P. Whitcher, "The Bhaktapur eye study: ocular trauma and antibiotic prophylaxis for the prevention of corneal ulceration in Nepal," *Brit. J. Ophthalmol.*, vol. 85, no. 4, pp. 388–392, 2001.
- [6] D. Yorston, M. Wood, and A. Foster, "Penetrating keratoplasty in africa: graft survival and visual outcome," *Brit. J. Ophthalmol.*, vol. 80, pp. 890–894, 1996.
- [7] C. Knauer and N. Pfeiffer, "The value of vision," *Graefes Arch. Clin. Exp. Ophthalmol.*, vol. 246, pp. 477–482, 2008.
- [8] E. C. Schwartz, R. Huss, A. Hopkins, B. Dadjim, P. Madjitouloum, C. Hénault, and V. Klauss, "Blindness and visual impairment in a region endemic for onchocerciasis in the Central African Republic," *Brit. J. Ophthalmol.*, vol. 81, no. 6, pp. 443–447, 1997.
- [9] L. Deng, J. Lyu, H. Huang, Y. Deng, J. Yuan, and X. Tang, "The SUSTech-SYSU dataset for automatically segmenting and classifying corneal ulcers," *Sci. Data*, vol. 7, no. 23, pp. 1–7, 2020.
- [10] P. B. Morgan and C. Maldonado-Codina, "Corneal staining: do we really understand what we are seeing?," *Cont. Lens Anterior Eye*, vol. 32, no. 2, pp. 48–54, 2009.
- [11] K. Tsoi, S. Poon, and P. Hung, "Introduction to the minitrack on big data on healthcare application," in *HICSS-53 Proc.*, p. 3215, 2020.
- [12] C. Poon, N. Gorji, M. Latt, K. Tsoi, B. Choi, C. Loy, and S. Poon, "Derivation and Analysis of Dynamic Handwriting Features as Clinical Markers of Parkinson's Disease," in *HICSS-52 Proc.*, pp. 3721–3730, 2019.
- [13] K. Tsoi, M. Lam, C. Chu, M. Wong, and H. Meng, "Machine Learning on Drawing Behavior for Dementia Screening," in *DH '18 Proc.*, pp. 131–132, 2018.
- [14] S. Zhang, S. Poon, K. Vuong, A. Sneddon, and C. Loy, "A Deep Learning-Based Approach for Gait Analysis in Huntington Disease," *Stud. Health Technol. Inform.*, vol. 264, pp. 477–481, 2019.
- [15] R. Buettner, M. Buechele, B. Grimmeisen, and P. Ulrich, "Machine Learning Based Diagnostics of Developmental Coordination Disorder using Electroencephalographic Data," in *HICSS-54 Proc.*, 2021, in press.
- [16] J. Frick, T. Rieg, and R. Buettner, "Detection of schizophrenia: a machine learning algorithm for potential early detection and prevention based on event-related potentials," in *HICSS-54 Proc.*, 2021, in press.
- [17] D. Chiu, P. Hung, V. Cheng, and E. Kafeza, "Protecting the Exchange of Medical Images in Healthcare Process Integration with Web Services," in *HICSS-40 Proc.*, pp. 131–140, 2007.
- [18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [19] H. Baumgartl, J. Tomas, R. Buettner, and M. Merkel, "A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring," *Prog. Addit. Manuf.*, pp. 277–285, 2020.
- [20] R. Buettner and H. Baumgartl, "A highly effective deep learning based escape route recognition module for autonomous robots in crisis and emergency situations," in *HICSS-52 Proc.*, pp. 659–666, 2019.
- [21] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical Image Classification with Convolutional Neural Network," in *ICARCV '14 Proc.*, pp. 844–848, 2014.
- [22] Q. Sun, L. Deng, J. Liu, H. Huang, J. Yuan, and X. Tang, "Patch-Based Deep Convolutional Neural Network for Corneal Ulcer Area Segmentation," in *Fetal, Infant and Ophthalmic Medical Image Analysis*, pp. 101–108, Springer, 2017.
- [23] L. Deng, H. Huan, J. Yuan, and X. Tang, "Automatic segmentation of corneal ulcer area based on ocular staining images," *Biomedical Applications in Molecular Structural and Functional Imaging*, vol. 10578, 2018.
- [24] J. D. Rodriguez, K. J. Lane, G. W. Ousler, E. Angjeli, L. M. Smith, and M. B. Abelson, "Automated Grading System for Evaluation of Superficial Punctate Keratitis Associated With Dry Eye," *Investig. Ophthalmol. Vis. Sci.*, vol. 56, no. 4, pp. 2340–2347, 2015.
- [25] Y. S. Chun, W. B. Yoon, K. G. Kim, and I. K. Park, "Objective Assessment of Corneal Staining Using Digital Image Analysis," *Investig. Ophthalmol. Vis. Sci.*, vol. 55, no. 12, pp. 7896–7903, 2014.
- [26] R. Peterson and J. Wolffsohn, "Objective grading of the anterior eye," *Optom. Vis. Sci.*, vol. 86, pp. 273–278, 2009.
- [27] J. Loo et al., "Open-source automatic segmentation of ocular structures and biomarkers of microbial keratitis on slit-lamp photography images using deep learning," *IEEE J. Biomed. Health Inform.*, 2020, in press.

- [28] M. Kuo et al., "A deep learning approach in diagnosing fungal keratitis based on corneal photographs," *Sci. Rep.*, vol. 10, pp. 1–8, 2020, Article 14424.
- [29] A. Akram and R. Debnath, "An Efficient Automated Corneal Ulcer Detection Method using Convolutional Neural Network," in *ICCIT '19 Proc.*, pp. 1–6, 2019.
- [30] R. Kohli and S. S.-L. Tan, "Electronic Health Records: How Can IS Researchers Contribute to Transforming Healthcare?," *MIS Quarterly*, vol. 40, no. 3, pp. 553–573, 2016.
- [31] D. Romanow, S. Cho, and D. Straub, "Riding the Wave: Past Trends and Future Directions for Health IT Research," *MIS Quarterly*, vol. 36, no. 3, pp. iii–x, 2012.
- [32] E. J. Cohen, P. R. Laibson, J. J. Arentsen, and C. S. Clemons, "Corneal Ulcers Associated with Cosmetic Extended Wear Soft Contact Lenses," *Ophthalmology*, vol. 94, no. 2, pp. 109–114, 1987.
- [33] S. Schmitz-Valckenberg, A. P. Goebel, S. C. Saur, J. J. Steinberger, S. Thiele, C. Wojek, C. Russmann, and F. G. Holz, "Automated Retinal Image Analysis for Evaluation of Focal Hyperpigmentary Changes in Intermediate Age-Related Macular Degeneration," *Transl. Vis. Sci. Technol.*, vol. 5, no. 2, pp. 1–9, 2016, Article 3.
- [34] P. Joyce, "Corneal vital staining," *Ir. J. Med. Sci.*, vol. 42, pp. 359–367, 1967.
- [35] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *NIPS '17 Proc.*, pp. 5574–5584, 2017.
- [36] V. Gulshan, L. Peng, M. Coram, M. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. Nelson, J. Mega, and D. Webster, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [37] R. Srivastava, J. Cheng, D. Wong, and J. Liu, "Using deep learning for robustness to parapapillary atrophy in optic disc segmentation," in *'15 Proc. Int. Symp. Biomed. Imag.*, pp. 768–771, 2015.
- [38] H. Baumgartl and R. Buettner, "Development of a highly precise place recognition module for effective human-robot interactions in changing lighting and viewpoint conditions," in *HICSS-53 Proc.*, pp. 563–572, 2020.
- [39] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, 2016.
- [40] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabudhe, and F. Meriaudeau, "Indian Diabetic Retinopathy Image Dataset (IDriD): A Database for Diabetic Retinopathy Screening Research," *Data*, vol. 3, no. 3, 2018.
- [41] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-Based Vessel Segmentation in Color Images of the Retina," *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
- [42] X. Li, Y. Grandvalet, and F. Davoine, "Explicit Inductive Bias for Transfer Learning with Convolutional Networks," in *PMLR '18 Proc.*, pp. 2825–2834, 2018.
- [43] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE T. Knowl. Data. En.*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [44] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *ICLR '15 Proc.*, pp. 163–171, 2015.
- [45] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *CVPR '16 Proc.*, pp. 2921–2929, 2016.
- [46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *NIPS '12 Proc.*, pp. 1097–1105, 2012.
- [47] H. Wu and X. Gu, "Towards dropout training for convolutional neural networks," *Neural Networks*, vol. 71, pp. 1–10, 2015.
- [48] S. J. Reddi, S. Kale, and S. Kumar, "On the Convergence of Adam and Beyond," in *ICLR '18 Proc.*, pp. 1–23, 2018.
- [49] T. Sandhan and J. Y. Choi, "Anti-Glare: Tightly Constrained Optimization for Eyeglass Reflection Removal," in *CVPR '17 Proc.*, pp. 1675–1684, 2017.
- [50] T. D'Orazio, M. Leo, G. Cicirelli, and A. Distanti, "An algorithm for real time eye detection in face images," in *ICPR '04 Proc.*, vol. 3, pp. 278–281, 2004.
- [51] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat. Surv.*, vol. 4, pp. 40–79, 2010.
- [52] F. Chollet et al., "Keras," <https://keras.io>, 2015.
- [53] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: A System for Large-scale Machine Learning," in *USENIX-OSDI '16 Proc.*, pp. 265–283, 2016.
- [54] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization," in *ICCS '17 Proc.*, pp. 618–626, 2017.
- [55] D. M. Maurice, "The structure and transparency of the cornea," *J. Physiol.*, vol. 136, no. 2, pp. 263–286, 1957.
- [56] F. Chollet, "Xception: Deep Learning With Depthwise Separable Convolutions," in *CVPR '17 Proc.*, pp. 1251–1258, 2017.
- [57] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *ECCV '16 Proc.*, pp. 630–645, 2016.
- [58] J. Landes and R. Buettner, "Argumentation-Based Negotiation? Negotiation-Based Argumentation!," in *EC-Web '12 Proc.*, vol. 123, pp. 149–162, 2012.
- [59] R. Buettner, "Cooperation in Hunting and Food-sharing: A Two-Player Bio-inspired Trust Model," in *BIONETICS '09 Proc.*, pp. 1–10, 2009.
- [60] R. Buettner, "The relationship between visual website complexity and a user's mental workload: A NeuroIS perspective," in *Information Systems and Neuro Science*, vol. 16, pp. 107–113, Springer, 2016.
- [61] R. Buettner, "A user's cognitive workload perspective in negotiation support systems: An eye-tracking experiment," in *PACIS '16 Proc.*, 2016.
- [62] R. Buettner, I. J. Timm, I. F. Scheuermann, C. Koot, and M. Roessle, "Stationarity of a user's pupil size signal as a precondition of pupillary-based mental workload

- evaluation,” in *Information Systems and Neuro Science*, vol. 25, pp. 195–200, Springer, 2017.
- [63] R. Buettner, “Predicting user behavior in electronic markets based on personality-mining in large online social networks,” *Electron. Mark.*, vol. 27, no. 3, pp. 247–265, 2017.
- [64] R. Buettner, “Getting a job via career-oriented social networkingmarkets,” *Electron. Mark.*, vol. 27, no. 4, pp. 371–385, 2017.
- [65] R. Buettner, S. Sauer, C. Maier, and A. Eckhardt, “Real-time Prediction of User Performance based on Pupillary Assessment via Eye Tracking,” *AIS Trans. Hum.-Comput. Interact.*, vol. 10, no. 1, pp. 26–56, 2018.
- [66] R. Buettner, M. Hirschmiller, K. Schlosser, M. Roessle, M. Fernandes, and I. J. Timm, “High-performance exclusion of schizophrenia using a novel machine learning method on EEG data,” in *IEEE Healthcom '19 Proc.*, pp. 1–6, 2019.
- [67] R. Buettner, J. Fuhrmann, and L. Kolb, “Towards high-performance differentiation between Narcolepsy and Idiopathic Hypersomnia in 10 minute EEG recordings using a Novel Machine Learning Approach,” in *IEEE Healthcom '19 Proc.*, pp. 1–7, 2019.
- [68] R. Buettner, D. Beil, S. Scholtz, and A. Djemai, “Development of a machine learning based algorithm to accurately detect schizophrenia based on one-minute EEG recordings,” in *HICSS-53 Proc.*, pp. 3216–3225, 2020.
- [69] R. Buettner, A. Grimmeisen, and A. Gotschlich, “High-performance Diagnosis of Sleep Disorders: A Novel, Accurate and Fast Machine Learning Approach Using Electroencephalographic Data,” in *HICSS-53 Proc.*, pp. 3246–3255, 2020.
- [70] T. Rieg, J. Frick, M. Hitzler, and R. Buettner, “High-performance detection of alcoholism by unfolding the amalgamated EEG spectra using the Random Forests method,” in *HICSS-52 Proc.*, pp. 3769–3777, 2019.
- [71] R. Buettner, T. Rieg, and J. Frick, “Machine Learning based Diagnosis of Diseases Using the Unfolded EEG Spectra: Towards an Intelligent Software Sensor,” in *Information Systems and Neuroscience*, vol. 32, pp. 165–172, 2019.
- [72] R. Buettner, T. Rieg, and J. Frick, “High-performance detection of epilepsy in seizure-free EEG recordings: A novel machine learning approach using very specific epileptic EEG sub-bands,” in *ICIS '19 Proc.*, 2019.
- [73] J. Gross, H. Baumgartl, and R. Buettner, “A Novel Machine Learning Approach for High-Performance Diagnosis of Premature Internet Addiction Using the Unfolded EEG Spectra,” in *AMCIS '20 Proc.*, 2020.
- [74] J. Breitenbach, H. Baumgartl, and R. Buettner, “Detection of Excessive Daytime Sleepiness in Resting-State EEG Recordings: A Novel Machine Learning Approach Using Specific EEG Sub-Bands and Channels,” in *AMCIS '20 Proc.*, 2020.
- [75] H. Baumgartl, E. Fezer, and R. Buettner, “Two-Level Classification of Chronic Stress Using Machine Learning on Resting-State EEG Recordings,” in *AMCIS '20 Proc.*, 2020.
- [76] H. Baumgartl, P. Roessler, D. Sauter, and R. Buettner, R. “Measuring Social Desirability using a Novel Machine Learning Approach based on EEG Data,” in *PACIS '20 Proc.*, 2020.
- [77] D. Raab, H. Baumgartl, and R. Buettner, “Machine Learning based Diagnosis of Binge Eating Disorder using EEG recordings,” in *PACIS '20 Proc.*, 2020.
- [78] H. Baumgartl, F. Dikici, D. Sauter, and R. Buettner, “Detecting Antisocial Personality Disorder using a Novel Machine Learning Algorithm based on Electroencephalographic Data,” in *PACIS '20 Proc.*, 2020.
- [79] R. Buettner, L. Bachus, L. Konzmann, and S. Prohaska, “Asking Both the User’s Heart and Its Owner: Empirical Evidence for Substance Dualism,” in *Information Systems and Neuroscience*, vol. 29, pp. 251–257, 2018.
- [80] R. Buettner and M. Schunter, “Efficient machine learning based detection of heart disease,” in *IEEE Healthcom '19 Proc.*, pp. 1–6, 2019.
- [81] A. Eckhardt, C. Maier, and R. Buettner, “The Influence of Pressure to Perform and Experience on Changing Perceptions and User Performance: A Multi-Method Experimental Analysis,” in *ICIS '12 Proc.*, 2012.
- [82] R. Buettner, “Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology,” in *KI '13 Proc.*, vol. 8077, pp. 37–48, 2013.
- [83] A. Eckhardt, C. Maier, J. P.-A. Hsieh, T. Chuk, A. B. Chan, A. B. Hsiao, and R. Buettner, “Objective measures of IS usage behavior under conditions of experience and pressure using eye fixation data,” in *ICIS '13 Proc.*, 2013.
- [84] R. Buettner, B. Daxenberger, A. Eckhardt, and C. Maier, “Cognitive Workload Induced by Information Systems: Introducing an Objective Way of Measuring based on Pupillary Diameter Responses,” in *SIGHCI '13 Proc.*, 2013.
- [85] R. Buettner, S. Sauer, C. Maier, and A. Eckhardt, “Towards ex ante Prediction of User Performance: A novel NeuroIS Methodology based on Real-Time Measurement of Mental Effort,” in *HICSS-48 Proc.*, pp. 533–542, 2015.
- [86] R. Buettner, “Robust user identification based on facial action units unaffected by users’ emotions,” in *HICSS-51 Proc.*, pp. 265–273, 2018.
- [87] R. Buettner, “The Impact of Trust in Consumer Protection on Internet Shopping Behavior: An Empirical Study using a Large Official Dataset from the European Union,” in *'20 IEEE BDS-Proc.*, pp. 66–69, 2020.
- [88] S. C. Rodermund, R. Buettner, and I. J. Timm, “Towards Simulation-based Preplanning for Experimental Analysis of Nudging,” *WI-2020 Proc.*, pp. 1219–1233, GITO, 2020.
- [89] R. Buettner, B. Daxenberger, and C. Woesle, “User acceptance in different electronic negotiation systems – a comparative approach,” in *ICEBE '13 Proc.*, 2013.
- [90] R. Buettner, “Towards a New Personal Information Technology Acceptance Model: Conceptualization and Empirical Evidence from a Bring Your Own Device Dataset,” in *AMCIS '15 Proc.*, 2015.
- [91] R. Buettner, “Analyzing the Problem of Employee Internal Social Network Site Avoidance: Are Users Resistant due to their Privacy Concerns?,” in *HICSS-48 Proc.*, pp. 1819–1828, 2015.
- [92] R. Buettner, “Getting a Job via Career-oriented Social Networking Sites: The Weakness of Ties,” in *HICSS-49 Proc.*, pp. 2156–2165, 2016.