Colorimetric Sensor Reading and Illumination Correction via Multi-Task Deep-Learning

Alejandra Castelblanco[1], Giusy Matzeu[2], Elisabetta Ruggeri[2], Fiorenzo G. Omenetto[2], Anne Hilgendorff [3,4], Julia A. Schnabel[1,5,6], Benjamin Schubert[1,5]

Abstract— **Colorimetric sensors represent an accessible and sensitive nanotechnology for rapid and accessible measurement of a substance's properties (e.g., analyte concentration) via color changes. Although colorimetric sensors are widely used in healthcare and laboratories, interpretation of their output is performed either by visual inspection or using cameras in highly controlled illumination set-ups, limiting their usage in end-user applications, with lower resolutions and altered light conditions. For that purpose, we implement a set of image processing and deep-learning (DL) methods that correct for non-uniform illumination alterations and accurately read the target variable from the color response of the sensor. Methods that perform both tasks independently vs. jointly in a multi-task model are evaluated. Video recordings of colorimetric sensors measuring temperature conditions were collected to build an experimental reference dataset. Sensor images were augmented with non-uniform color alterations. The best-performing DL architecture disentangles the luminance, chrominance, and noise via separate decoders and integrates a regression task in the latent space to predict the sensor readings, achieving a mean squared error (MSE) performance of 0.811±0.074[°C] and r^2=0.930±0.007, under strong color perturbations, resulting in an improvement of 1.26[°C] when compared to the MSE of the best performing method with independent denoising and regression tasks.**

Clinical Relevance— **The proposed methodology aims to improve the accuracy of colorimetric sensor reading and their large-scale accessibility as point-of-care diagnostic and continuous health monitoring devices, in altered illumination conditions.**

I. Introduction

Colorimetric sensors are a type of optical sensor that enable the rapid and cost-effective measurement of changes in physical parameters such as temperature and pH, as well as the detection of analyte levels in a substance. These sensing materials typically employ molecules or nanoparticles that switch their form in response to physical stimuli or to the degree of binding from a target analyte, thus changing the refractive index and the color reflected by the coating of the surface [1] [2]. In particular, colorimetric paper-based sensors have a large potential for point-care health diagnostic applications [3] [4], showing high sensitivity and specificity, both for binary analyte detection, such as the biosensors used for antigen SARS-COV-2 tests [5], or for measurement of parameters on a continuous scale, like the urine analysis used to identify glucose, red-blood-cells, and pH levels, among others, in point-of-care diagnostics [6]. The interpretation of the color output from these sensors is however not yet automated, relying on readings by end-point users with visual reference to a color scale, leading to a high degree of imprecision. As an alternative, smartphones, and image processing methods have been used for automated reading of colorimetric sensors [7] [8] [9], improving their accuracy and sensitivity. However, these approaches employ specialized set-ups with strictly controlled illumination, high-resolution cameras, and performance is tested only under subtle light variations, limiting the use of colorimetric sensors as point-of-care and wearable health devices in uncontrolled scenarios, where non-uniform illumination, low resolutions, and strong color alterations can add disruptive noise to the sensor's image and significantly reduce inference performance.

The colors that we observe on any digital image result from a combination of factors: the intrinsic surface reflectance of the objects, the color and arrangement of the light sources illuminating the scene, shadows cast over the objects, the camera sensor and white-balancing settings. Estimating the 'true' reflectance color of an object, under neutral or reference illumination, is an under-constrained problem and it, therefore, has been guided with references in the image, such as the color-checker charts, often used in photography. The color correction problem has been addressed, first, with global methods that find a transformation matrix between the observed color space and a reference color space for the whole image, such as the grayworld, white patch, and Finlayson polynomial methods [10]. These methods however do not account for strong non-linear gradients in the illumination of a scene. More recent approaches to the color constancy problem take cues from the local context of a scene, without color references, and employ DL methods in a semi- or unsupervised manner to transform the image into a predefined white-balanced representation, using convolutional neural networks (CNN) [11], integrating fast Fourier transformations [12], or k-neighbor inferences [13]. Still, we found that using such DL models to correct non-linear illuminants while considering the specific subspace of color variations and apriori knowledge of a colorimetric sensor is yet to be investigated.

In this work, we therefore, define the task of automated colorimetric sensor reading for uncontrolled scenarios as twofold: (1) we aim to correct for the illuminant of the scene, finding a local color homography between the current light condition and a reference illuminant scenario; (2) a regression

[1]Computational Health Center, Helmholtz Munich, Member of the German Center for Lung Research (DZL), Germany. [2]Silklab, Tufts University, Dept. of Biomedical Engineering, Medford, MA, USA. [3]Environmental Health Center, Helmholtz Munich, Member of the German Center for Lung Research (DZL), Germany. [4]Dr. von Haunersche Children's Hospital, Hospital of the Ludwig-Maximilians University, Member of the German Lung Research Center (DZL), Munich, Germany. [5]School of Computation, Information and Technology, Technical University of Munich, Germany. [6]School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK.

task is needed to map the color shown by the colorimetric sensor to the target variable. We hypothesize that integrating both the color correction and regression through multi-task deep-learning (DL) models, can improve the accuracy when compared to independent color correction and subsequent reading of the color response. We, therefore, tested a set of methods with both independent and combined denoising and regression tasks. In particular, we propose a multi-task autoencoder (AE) that disentangles the luminance, chrominance, and noise via separate decoders and performs regression in the latent space. To evaluate our approach, video recordings of temperature-responsive colorimetric sensors were acquired to build an experimental dataset. A synthetic data augmentation pipeline was implemented to add non-uniform light and color alterations to the sensor images. Our experiments show that colorimetric sensor reading using the multi-task AE with latent regression, improves prediction performance under strong color perturbation conditions when compared with other DL architectures and with independent color correction and regression methods.

II. Methods

We propose, implement and evaluate image processing methods and DL models for automated colorimetric sensor reading from images with non-uniform luminance and strong chrominance alterations.

A. Baseline Methods for Color Correction

The white patch (WP) and grayworld (GW) color constancy algorithms are used as global color correction reference baselines. In the WP method, each color channel C is rescaled by the maximum channel value within the selected white reference area (1), whereas for the GW method each channel is multiplied by the ratio between the global image average I_{avg} and the average of each channel (2).

$$C_{WP} = C \cdot \frac{255}{C_{max}} \quad (1); \qquad C_{GW} = C \cdot \frac{I_{avg}}{C_{avg}} \quad (2)$$

The Finlayson polynomial color correction method, implemented with the *colour 0.1.5* python library, considers the color-checker references from the ground truth (GT) vs. noisy sensor images to fit a color homography matrix via least squares.

A two-step baseline (*Lum-Fin*), which integrates prior knowledge of the sensor design is also proposed. First, a local luminance correction is performed by finding an image Δ'_Y with the pixel-wise grayscale differences between the noisy input and the GT background of the sensor, while masking other areas. Next, a 2D linear interpolation is calculated using Δ'_Y to estimate the luminance, on the previously masked sensing areas, caused by external light. The resulting interpolated image Δ_Y is scaled with a vector λ corresponding to the grayscale to RGB colorspace conversion, and subtracted from each RGB color channel. The second step is to apply the Finlayson color correction, with the rationale that, once the local luminance is corrected, the method can be applied globally with better performance (3).

$$I_{corrected} = Finlayson\,(I_{altered} - \lambda\,\Delta_Y) \quad (3)$$

B. Linear and Lasso Regression for Colorimetric Prediction

Performance for prediction of the temperature from the color response was first evaluated using simple Linear and Lasso regression models. The input features correspond to the average RGB color intensities from the pixels in each of the sensing areas.

C. Multilayer Perceptron for Colorimetric Prediction

We implement a multilayer perceptron (MLP) for direct prediction of the target variable from a sensor's image with standard orientation, using the mean-square error (MSE) as the metric for regression loss. Inclusion of an optional CNN module before the MLP was also tested. The model is trained and evaluated with masked or unmasked input images, in both the reference and color-altered conditions, as baselines for automated color-based prediction.

D. Multi-task decoder for luminance and chrominance with latent colorimetric prediction

Inspired by previous work on joint luminance and chrominance correction of underwater images [14], we propose a multi-task AE CNN model that integrates three denoising tasks and a regression task (Fig. 1). First, the AE model is trained to reconstruct the unperturbed images in the *YCrCb* colorspace, separating luminance *Y*, from chrominance *CrCb* channels, and to estimate the illumination gradient present in the image (light kernel). The aim is to build a latent space representation that disentangles the noise coming from the illumination of the scene vs the true color variations of the sensor's colorimetric response. From the latent representation, we use an MLP for prediction of the target variable from the color response of the sensor.

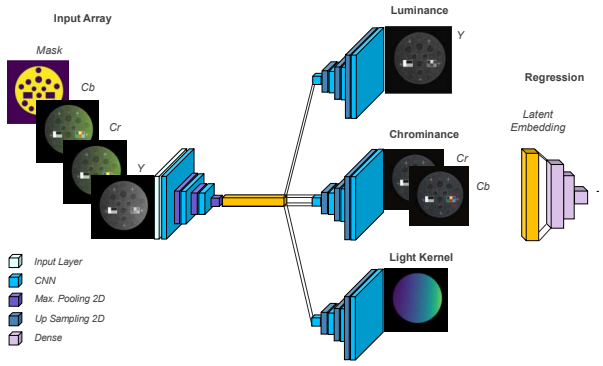

Figure 1. Multi-task CNN autoencoder model with a downstream regression task from the latent embedding representation.

A multi-task loss is therefore required for training the model (4), each loss is coupled with a factor θ that determines the scaling and importance of the loss. Three types of loss functions were evaluated for the denoising tasks: MSE, cross entropy loss, and structural similarity loss (SSIM). The regression task uses the MSE as the loss function.

$$L_{total} = \theta_{lum}\ L_{lum} + \theta_{chr}\ L_{chr} + \theta_{reg} L_{reg} + \theta_{reg} L_{reg} \quad (4)$$

E. Hyperparameter Optimization

For the proposed DL models, we optimized the architecture parameters (e.g., CNN depth, no. filters, MLP layers, batch

normalization), loss functions, loss weights θ, and the learning rate of the models using the framework *Optuna* version 3.0.3 [15]. Two sensors were used for training and validation, splitting the data with a 0.7/0.3 train-val ratio, leaving one sensor out for later-stage performance evaluation. The hyperparameter optimization was targeted to find the best combination with the lowest MSE average across 80 images. Training pipeline and model architectures are available in: https://github.com/SchubertLab/colorimetric_sensor_reading

III. Experiments and Dataset

A. Colorimetric Sensors

Our experiments were performed with colorimetric sensing material technology developed by Matzeu et al. [16] [17]. The paper-based sensor array integrates 7 types of temperature color-responsive dyes (*S0-S6*), each located in a specific sensing area (Fig. 2d), with a temperature sensing range of 31°C to 42°C, and resolution of 0.5°C.

B. Experimental dataset

An experimental set-up (Fig. 2a-b) was created to record videos (30fps, 8 MP) of the continuous color response of three sensors for gradually increasing temperatures between 31°C to 42°C. The sensors were placed inside an impermeable layer, attached to a glass container. The temperature of the water surrounding the sensors was increased 1°C every two minutes using a closed-loop control water heater, while recording the color variations for temperatures in between.

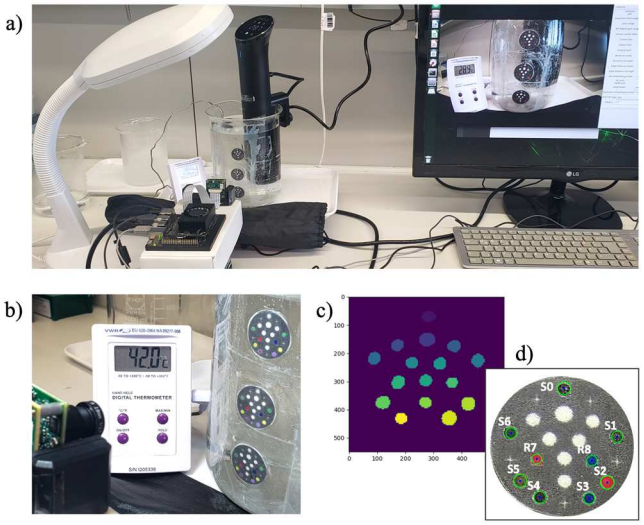

Figure 2. a) Experimental set-up for recording the sensor's continuous color response. b) Close-up of the experimental set-up and sensors at 42°C. c) Sample of U-Net sensing area detection for color data extraction. d) Sensor Design with Sensing area labels (*S0-S6*) and color-reference spots (*R7, R8*).

A pipeline for digitizing the temperature values and the color changes from the experimental videos was developed. Segmentation of the sensing areas of the sensors in every sampled frame of the video was done via a fine-tuned U-Net model [18] (Fig. 2c).

C. Synthetic Data Generation

To generate a synthetic dataset, we sample the experimental data every 25 frames. For every frame and every sensor, we sample the temperature value and the average color of each of the six sensing areas. A synthetic image of the sensor is generated using the sampled color and adding random noise to simulate the texture of the material. The reference image dataset (N=12357 images) has identical color checkers and a standard orientation (Fig. 3). Original images are generated in a [250×250 pixels] resolution, however, the resolution of the image generators for model training was downsized to [72x72 pixels] to test our approach with challenging low resolutions that can be found in less ideal scenarios.

D. Noise Model

An image augmentation pipeline was implemented to add non-uniform light and color alterations to the sensor images (Fig. 3). We define a noise model where a light kernel K is added or subtracted to each color channel matrix (R,G,B) of the original image, with independent chroma intensity factors α, β, γ (5).

$$I_{altered} = [\, R \pm \alpha K\,,\; G \pm \beta K\,, B \pm \gamma K\,] \quad (5)$$

Chroma intensity factors α, β, γ were randomly sampled in the range from [0-0.5]· 255 with (0.01 steps). The resulting image I is clipped within [0, 255].

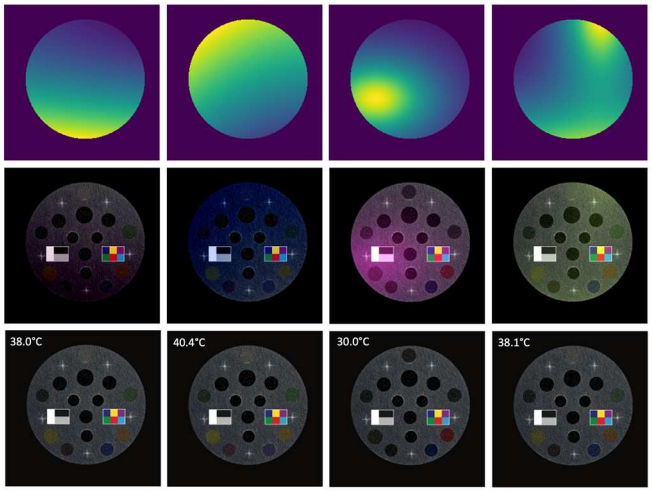

Figure 3. Example of synthetic image augmentations. Top row: noise kernels used for color augmentations (2 polynomial, elliptical and combined kernel). Middle row: sensor images with color perturbations. Bottom row: ground truth sensor images with the corresponding temperature.

Light kernels (K) were defined with a kernel function $f(x, y)$ that determines the local luminance distribution. We define three types of luminance functions: polynomial (6) with degrees $n \in [1,3]$, elliptical (7), and a sum of two random polynomial and elliptical kernels.

$$f(x,y) = (m_x \cdot x^n + o_x) + (m_y \cdot y^n + o_y) \quad (6)$$

$$f(x,y) = \frac{i}{d(x,y)^2}\;;\; d(x,y)^2 = s_x\left(\frac{x-c_x}{N}\right)^2 + s_y\left(\frac{y-c_y}{N}\right)^2 + 1 \quad (7)$$

The parameters offset o, slope m, and center of the ellipse c_x and c_y, are sampled in proportional to the image dimensions (NxN), intensity $i \in [-100,100]$ and scale $s_x, s_y \in [-1,20]$.

E. Model Evaluation in Altered Color Conditions

The best models from the hyperparameter optimization were tested in a leave-one-sensor-out validation scheme. We evaluated two test scenarios, one with in-distribution noise which uses polynomial and elliptical kernels, and one alternative noise-distribution (excluded from model training and validation) with polynomial+elliptical kernels that can be seen as modeling two overlapping light distributions.

IV. Results and Discussion

A. Sensor Colorimetric Response

The color response to continuous temperature increase showed a distinctive progression of colors for each of the seven color-responsive dyes (Fig. 4); altogether they define a sub-colorspace for temperature inference.

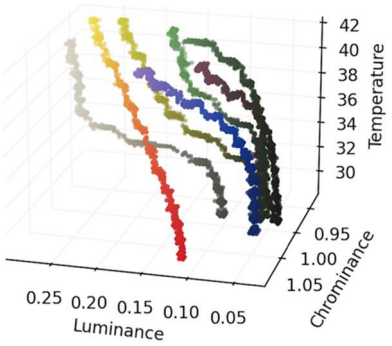

Figure 4. Experimental response of a colorimetric sensor with seven sensing dyes responsive to temperature variations. The horizontal axis shows Luminance and Chrominance as (Cr+Cb) dimensions. Scattered points show the color of the sensing area in the RGB colorspace.

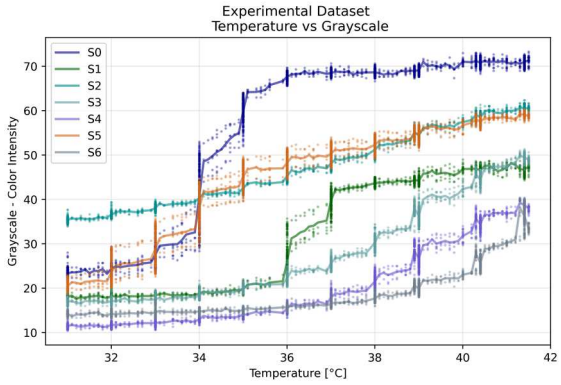

Figure 5. Grayscale intensity vs temperature, for all sensing dyes (*S0-S6*). Scatter points show the individual measurements for 3 sensors after normalization; the average response per frame is shown with a solid line.

Inter-sensor variability was analyzed for the experimental data after grayscale min-max scaling (Fig. 5). The highest variations were found for *S0* and *S5*, in the temperature range of 33-34°C, with standard deviations (SD) of 3.21°C and 2.93°C respectively. The ranges between 33-34°C and 34-35°C had the highest variability across all dyes (avg. SD: 1.28°C and 1.17°C), therefore adding experimental variability to the predictions within that range.

B. MLP reaches performance close to the experimental boundary with unperturbed low-resolution images

The MLP prediction of temperature from low-resolution sensor images (72x72 pixels) with no illumination perturbations (Table 1, Fig.7a), was comparable (MSE 0.398°C) to the performance of the Linear and Lasso regression models using the experimental features, with a maximum difference in average performance of 0.062°C.

TABLE I. Temperature Regression from Experimental Data

	MSE [°C]	Coef. of Det. r^2
Multiple-Linear Reg. (0.5 train/test, 10 rep.)	0.336 ± 0.005	0.977 ± 5.6e-4
Lasso Reg. ($\alpha = 0.01$) (0.5 train/test, 10 rep.)	0.349 ± 0.006	0.976 ± 5.9e-4
MLP (leave-one-out)	0.398 ± 0.136	0.981 ± 0.004

C. Baseline Methods for Color Correction

Baseline color correction methods were applied to the perturbed dataset (N test=400 images per sensor) (Fig. 6). The average pixel-wise MSE between the sensing areas of the resulting denoised images and the respective GT images was: *Lum-Fin:*85.06±14.24, *GW*:97.16±21.72, *WP*: 102.57±58.35. With the *Lum-Fin* method showing higher denoising performance, closely followed by the GW correction.

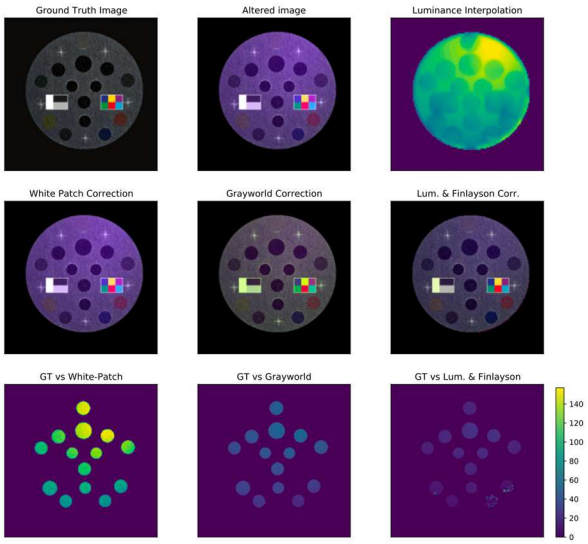

Figure 6. Sample of performance of baseline methods for color correction. Lum. = Local luminance correction via interpolation. GT=Ground Truth.

D. AE with integrated color correction and regression yields the best performance with perturbed images

Prediction of temperatures under strong color perturbations was evaluated for two subgroups of methods: (1) first, a set of two-step methods that perform color correction of the images via (*GW, WP* and *Lum-Fin*) and subsequent temperature prediction with an MLP, and (2) DL models combining the color correction and inference tasks (MLP model trained directly on noisy data and the AE latent regression model). The AE latent regression model showed better performance when compared to other DL models and independent task methods (Table 2). The best overall performance (MSE 0.811°C) was found with the AE latent regression model using the cross-entropy loss for denoising. A significant improvement of 1.26°C MSE was found over the best performing two-step method *Grayworld+MLP*. Statistical analysis was performed first testing for normality and applying the Kruskal–Wallis and Mann–Whitney U tests.

TABLE II. REGRESSION PERFORMANCE WITH STRONG COLOR PERTURBATIONS

Leave-one-sensor-out N test = 1200	MSE [°C]	Coef. of Det. r^2
Lum-Fin + MLP†	2.894 ± 0.362 ***	$0.870 \pm 3.18\times10^{-2}$
White Patch + MLP†	2.726 ± 2.012 ***	$0.899 \pm 5.20\times10^{-2}$
Grayworld + MLP†	2.072 ± 0.149 ***	$0.897 \pm 2.62\times10^{-2}$
MLP (trained with noisy images)	2.906 ± 2.363	$0.793 \pm 1.16\times10^{-1}$
AE Latent Reg. (MSE)	1.266 ± 0.239 ***	$0.917 \pm 1.25\times10^{-2}$
AE Latent Reg. (SSIM)	0.929 ± 0.110	$0.915 \pm 9.70\times10^{-3}$
AE Latent Reg. (Cross. Entr.)	**0.811 ± 0.074**	**$0.930 \pm 6.63\times10^{-3}$**

† MLP trained with images after color correction with respective method. Significance for the Mann–Whitney U test, MSE comparisons with the best model (AE latent Reg. Cross. Entr.) with (***=p-val≤ 0.001, **=p-val≤0.01, *=p-val≤0.05).

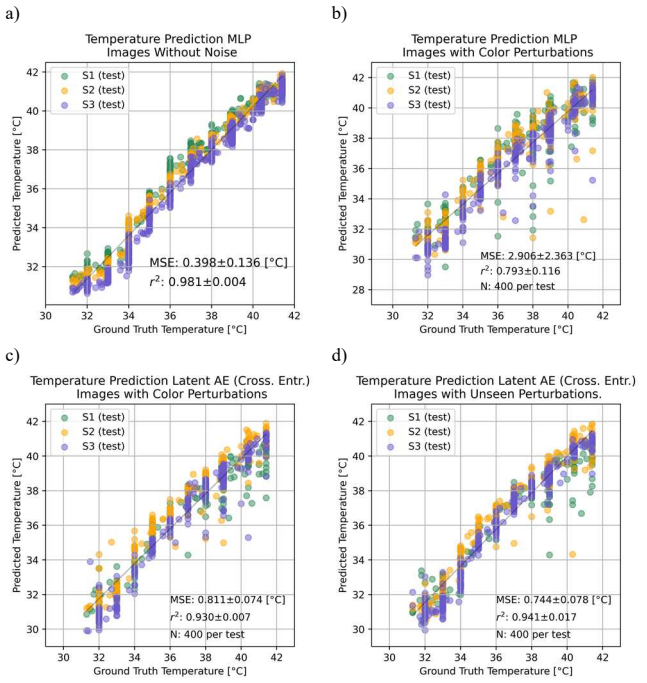

Figure 7. Temperature prediction with colorimetric sensor images a) MLP with ground truth images. b) MLP trained and tested with color perturbations. c) Autoencoder (AE) with latent regression and cross entropy loss with color perturbations. d) AE with latent regression, performance with color alterations for unseen noise kernels (polynomial + elliptical). Models trained and tested in a leave-one-out scheme for each sensor (S1, S2, S3).

The MLP model trained with noisy images exhibited a large prediction variability (SD 2.363, Fig. 7b), while the best AE model had the lowest prediction error variability (SD 0.074, Fig. 7c). Outliers in performance can be associated with large image information losses, since perturbations of up to ±50% in color intensity and image clipping were applied. Performance of the AE latent regression model with unseen noise distributions (polynomial + elliptical kernels, Fig 7d) was satisfactory with a MSE 0.74±0.08°C and r^2 of 0.94±0.017, indicating transferability to a different noise distribution.

Evaluation of additional sensor designs and inclusion of more complex noise models, are limitations that we would like to address in the future. Moreover, the complementary task of detecting and aligning the sensor orientation must be accounted for in order to have a complete evaluation of the robustness of automated colorimetric sensor reading systems in non-standardized scenarios.

V. CONCLUSION

This work aims to explore and validate methods for automated digital reading of colorimetric sensors in settings with perturbed illumination conditions and low image-resolutions. The use of a multi-task DL approach resulted in an improvement of the prediction performance over independent two-step color correction and prediction models. Further development of these methods can enable the application of colorimetric sensors as point-of-care diagnostic devices and accessible continuous health monitoring devices in remote settings.

ACKNOWLEDGMENT

A.C. is supported by the Helmholtz joined graduate school Munich School for Data Science – MUDS. This work is supported by the German Lung Research Center (DZL).

REFERENCES

[1] Ajay V.S et al., "Colorimetric sensors for rapid detection of various analytes," Materials Science and Engineering: C, vol. 78, Sep. 2017.
[2] I. I. Ebralidze, N. O. Laschuk, J. Poisson, and O. V. Zenkina, "Colorimetric Sensors and Sensor Arrays," Nanomaterials Design for Sensing Applications. pp. 1–39, 2019. . pp. 1–39, 2019.
[3] S. S. Gambhir, T. Jessie Ge, O. Vermesh, R. Spitler, and G. E. Gold, "Continuous health monitoring: An opportunity for precision health," Science Translational Medicine, vol. 13, no. 597. 2021.
[4] S. Krishnan et al. "Colorimetric Visual Sensors for Point-of-needs Testing," Sensors and Actuators Reports, vol. 4, p. 100078, Nov. 2022.
[5] M. Pohanka, "Progress in Biosensors for the Point-of-Care Diagnosis of COVID-19," Sensors, vol. 22, no. 19, p. 7423, Sep. 2022.
[6] M. Ra et al. "Lab-in-a-Cup (LiC): An autonomous fluidic device for daily urinalysis using smartphone," Sens. Actuators B Chem., vol. 355, p. 131336, Mar. 2022.
[7] S. D. Kim, Y. Koo, and Y. Yun, "A Smartphone-Based Automatic Measurement Method for Colorimetric pH Detection Using a Color Adaptation Algorithm," Sensors, vol. 17, no. 7, Jul. 2017.
[8] M. N. Sakinah et al. "Color Correction Technique using an Artificial Color Board and Root-polynomial Color Correction for Smartphone-Based Urinalysis.", International Conference on QiR. 2021.
[9] M. Ra et al. "Smartphone-Based Point-of-Care Urinalysis Under Variable Illumination.", IEEE J.Transl. Eng. in Health and Med. 2017.
[10] G. Finlayson, H. Gong, and R. B. Fisher, "Color Homography: Theory and Applications," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 1, pp. 20–33, Jan. 2019.
[11] S. Bianco, C. Cusano, and R. Schettini, "Single and Multiple Illuminant Estimation Using Convolutional Neural Networks," IEEE Trans. Image Process., vol. 26, no. 9, pp. 4347–4362, Sep. 2017.
[12] J. T. Barron and Y.-T. Tsai, "Fast Fourier Color Constancy," in 2017 IEEE CVPR, Honolulu, HI, Jul. 2017.
[13] M. Afifi, B. Price, S. Cohen, and M. S. Brown, "When color constancy goes wrong: Correcting improperly white-balanced images," in 2019 IEEE/CVF (CVPR), Long Beach, CA, USA, Jun. 2019.
[14] X. Xue, Z. Hao, L. Ma, Y. Wang, and R. Liu, "Joint Luminance and Chrominance Learning for Underwater Image Enhancement," IEEE Signal Process. Lett., vol. 28, pp. 818–822, 2021.
[15] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," Jul. 2019.
[16] G. Matzeu et al., "Large-Scale Patterning of Reactive Surfaces for Wearable and Environmentally Deployable Sensors," Adv. Mater., vol. 32, no. 28, p. e2001258, Jul. 2020.
[17] G. Matzeu et al., "Functionalized Mouth-Conformable Interfaces for pH Evaluation of the Oral Cavity," Adv. Sci., vol. 8, no. 12, Jun. 2021.
[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Notes in Computer Science. pp. 234–241, 2015. doi: 10.1007/978-3-319-24574-4_28.