IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 February 2024, accepted 18 April 2024, date of publication 29 April 2024, date of current version 14 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3395063

RESEARCH ARTICLE

A Systematic Investigation of Image Pre-Processing on Image Classification

PEGAH DEHBOZORGI[1,2], OLEG RYABCHYKOV[1,2], AND THOMAS BOCKLITZ[1,2,3]

[1]Institute of Physical Chemistry (IPC) and Abbe Center of Photonics (ACP), Friedrich Schiller University Jena, Leibniz Centre for Photonics in Infection Research (LPI), 07743 Jena, Germany
[2]Leibniz Institute of Photonic Technology, Leibniz Health Technologies, Leibniz Centre for Photonics in Infection Research (LPI), 07745 Jena, Germany
[3]Institute of Computer Science, Faculty of Mathematics, Physics and Computer Science, University Bayreuth, 95447 Bayreuth, Germany

Corresponding author: Thomas Bocklitz (Thomas.bocklitz@uni-jena.de)

This work was supported in part by Bundesministerium für Bildung und Forschung (BMBF) through Photonics Research Germany under Grant 13N15710 (LPI-BT3); in part by the Leibniz Center for Photonics in Infection Research (LPI); in part by the LPI initiated by Leibniz-IPHT, Leibniz-HKI, Friedrich Schiller University Jena, and Jena University Hospital funded by the BMBF National Roadmap for Research Infrastructures; in part by European Research Council (ERC) through the project STAIN-IT under Grant 101088997; in part by the German Research Foundation Projekt-Nr. 512648189; and in part by the Open Access Publication Fund of the Thueringer Universitaets-und Landesbibliothek Jena.

ABSTRACT AI-powered image analysis is a transformative technology with immense potential to enhance diagnostics and patient care. Accurate medical image assessment plays a crucial role in disease detection and treatment planning, yet challenges arise due to noise and visual variations in medical imaging. Image pre-processing is a key solution to address these challenges, and while widely used, there is a lack of studies on its effectiveness. Recognizing this gap, our research aims to contribute insights to this scientific scope. This research specifically delves into the impact of pre-processing on the binary classification model performance, rather than model and hyperparameter optimization. We deliberately selected a limited yet comprehensive subset of methods and datasets; H&E-stained tissue, chest X-ray, and retina OCT images were chosen to ensure the generalizability of our findings. Analysis revealed that implementing a pre-processing significantly improved mean sensitivity in the binary classification models: from 0.87 to 0.97 for H&E-stained tissue, 0.92 to 0.96 for chest X-rays, and 0.96 to 0.99 for Retina OCT images. Two different sequences for applying pre-processing steps were explored, with minimal effect observed in the altered sequences, indicating consistent improvement regardless of the chosen sequence. We investigated the pre-processing steps employed in the 40 of the best-performing and worst-performing models, determined by the higher and lower mean sensitivities. We have uncovered that the pre-processing steps of the best-performing models displayed only minimal similarities, except for the pooling mode. This observation also applied to the worst-performing models with lower sensitivity.

INDEX TERMS Image pre-processing, image classification, machine learning (ML) method, transfer learning (TL).

I. INTRODUCTION

Medical science generates a wide range of data, rapidly growing in volume and diversity. Advances in medical image scanning and digital storage capabilities have led to the development of automated analysis systems. On the other hand, Artificial Intelligence (AI) has emerged as a prominent tool in medicine and healthcare with continuous growth driven by researchers, businesses, and individuals striving to harness the AI technologies for an automated analysis. The primary objectives of AI in medicine are to expedite diagnostic process, support therapeutic decision-making, and facilitate clinical research. The data required for the diagnosis and treatment can be derived from various sources,

The associate editor coordinating the review of this manuscript and approving it for publication was Mouloud Denai.

including clinical notes, laboratory tests, pharmacy records, and medical imaging. Medical images hold a critical role in clinical practice, as they enable visualization of internal tissues within organisms. Different imaging protocols yield diverse modalities of medical images, each characterized by unique visualization attributes. Consequently, the integration of AI-based technologies into medical practices is expected to bring significant transformations across various aspects of the medicine and healthcare [1], [2].

Medical images are susceptible to different issues that can compromise their visual interpretation. Inherent noises can emerge during the image acquisition process, while factors such as uneven illumination, image blurriness, inadequate contrast, and improper exposure further contribute to these challenges. These factors as such, significantly affect the quality of medical images, emphasizing the importance of employing image pre-processing techniques to mitigate these issues. Data pre-processing transforms raw data into a clean and refined format, eliminating unwanted variations such as instrumental and experimental artifacts while enhancing the overall quality of the images. In a comprehensive data analysis procedure, multiple pre-processing methods are typically utilized. It is important to note that an improper choice of pre-processing approach can significantly undermine the predictive capabilities of AI models. Thus, making the appropriate pre-processing choices is crucial to ensure more reliable, accurate, and optimal outcomes in AI-driven analysis [3].

Data pre-processing generally involves multiple steps, each targeting specific artifacts within the data. A sequential application of individual pre-processing methods is required to address all the artifacts. The choice of the most optimal pre-processing method, or a combination of methods, relies on several factors: such as data properties and the specific objectives of the data analysis. Careful consideration of these factors enables researchers to enhance the quality and reliability of their data analysis, establishing a solid foundation for subsequent analysis and generating valuable insights. By effectively conducting data pre-processing, not only can the accuracy of results be improved, but also potential biases and errors that could impact the validity of conclusions can be minimized. This preparatory step enables researchers to derive meaningful and robust insights from their data analysis [4].

The Rachana et al., suggested that the adaptive histogram equalization can produce better results for image analysis tasks [5]. In contrast, the Pradhan et al., and Ali et al., took a different approach by combining the median smoothing and the contrast adjusting techniques to improve the quality of image data and increase the performance of analysis [6], [7]. The Giełczyk et al. asserted that the utilization of adaptive masking, histogram equalization, and Gaussian blur as the chosen pre-processing techniques yielded a notable improvement in the F1-score, elevating it from 93% to well beyond 97% [8]. To set up the classification task, the Pradhan et al. improved the quality of their image datasets by employing methods such as resizing, cropping, Global Pixel Normalization (GPN) and Gaussian filtering (GF) the single pre-processing steps. This ensures that the predictive models utilized in this study are supplied with high-quality, meaningful data, contributing to more accurate and reliable classification results [9]. The Farhan et al. employed a combination of optimal filtering and contrast adjustment not only to enhance the quality of chest X-ray images but also to boost the feature extraction and, accordingly, improve the classification performance [10]. The Ahmad Hameed et al. introduced a stacked CNN model to improve skin lesion diagnosis, tackling challenges in distinguishing skin injuries from cancer due to visual similarities. They ensured model robustness with image augmentation for dataset balance and a comprehensive pre-processing phase involving resizing, edge detection, color space conversion (BGR to YUV), and histogram equalization. This pre-processing procedure enhanced image quality, leading to improved stacked CNN performance and achieving an accuracy of 95.2% [11]. The Hewan Shrestha et al. conducted research on segmentation and classification algorithms for early skin cancer detection. They employed image resizing and normalization as essential steps to enhance the quality of the image data, ultimately contributing to the overall accuracy improvement of the implemented models [12]. The Shumoos Al-Fahdawi et al. focused on identifying multiple ocular diseases in fundus images through the development of an automated deep-learning classification system. They implemented five critical image pre-processing steps, encompassing circular border cropping, image resizing, contrast enhancement, noise reduction, and data augmentation, to not only enhance the quality of the image data but also ensure notable improvements in the overall performance of the model [13]. Despite numerous studies on the pre-processing techniques, finding the best method or the combination of methods to eliminate all the artifacts in a dataset remains challenging. It is essential to search for the most effective pre-processing method actively, considering its influence on subsequent data analysis and results.

Depending on the nature of the problem, researchers may choose to apply a single pre-processing method or a combination of techniques to address measurement errors and artifacts in the data [14]. Here we conduct the systematic study, focusing on the image data, with the primary objective of better understanding the effect of the different pre-processing strategies on the overall quality of images and the subsequent analysis results.

To the best of our knowledge, a systematic study of the impact of image pre-processing on model performance has never been conducted before. Our study represents a pioneering effort, addressing numerous questions related to the impact of pre-processing on the final analysis results. We have undertaken an extensive exploration, incorporating a diverse range of techniques, exploring all the possible

combinations, and considering various medical image modalities. This systematic approach aims to fill the existing gap in understanding the role of pre-processing in image analysis while ensuring the generalizability of the experiments.

The workflow of this study is as follows: the first step is involved in organizing all the suitable and most common pre-processing combinations and determining their respective order of application. Two distinct order sequences, sequence 1 and sequence 2, along with the pre-processing combinations, are chosen and thoroughly examined to assess the influence of each sequence. Following the pre-processing phase, key features have to be extracted from the pre-processed images using the VGG16 [15], a well-known pre-trained network widely recognized for extracting high-level features. Implementing the pre-trained network as a feature extractor ensures that the extracted features contain meaningful information. In the final phase, the classification task has to be performed. Therefore, a series of models must be designed and trained using the extracted features. The models have to be evaluated and compared to assess their performance when fed with different pre-processing combinations as the input. The ultimate goal is to identify the pre-processing combination that yields the most effective results, ensuring accurate classification and optimal analysis outcomes. By conducting this comprehensive analysis, we are aiming to determine the optimal approach that maximizes image classification accuracy.

The analysis has been established by investigating three datasets. The initial dataset comprised image patches that have been extracted from the histological images of human colorectal cancer (CRC) and the normal tissues, stained with the hematoxylin & eosin (H&E) [16]. The second dataset consists of chest X-ray images of humans categorized as normal and pneumonia. Lastly, the study has incorporated the retinal Optical Coherence Tomography (OCT) scans as the third set of data that have been obtained from adult patients, containing two distinct classes: the choroidal neovascularization (CNV) and the normal retina [17]. H&E-stained, OCT, and chest X-ray images are examples of the most used data in pathology, ophthalmology, and radiology domains and were chosen for their clinical relevance. H&E staining is the preferred choice for routine diagnostics among pathologists, providing exceptional clarity in visualizing cellular and tissue structures. OCT, widely adopted in healthcare, is crucial for diagnosing ocular conditions like glaucoma and macular degeneration. X-ray imaging, including chest X-rays, is instrumental in identifying fractures, abnormalities, and respiratory conditions such as pneumonia, tuberculosis, and lung cancer, offering essential insights for pulmonologists. The decision to focus on these modalities was driven by the need for a detailed analysis. The selected modalities provided a rich subset of cases, allowing us to explore the generalizability of our findings.

The study significantly advances image analysis by offering valuable insights into the optimizing pre-processing techniques for the image classification. Through this systematic analysis of various pre-processing combinations across the different image modalities, effective approaches for the binary classification models are distinguished. These findings serve as a practical guide for practitioners, aiding in an informed pre-processing technique selection.

In the following sections, we will conduct a thorough exploration of the datasets utilized in the ''MATERIALS'' section, providing comprehensive insights into their properties. Subsequently, the employed models will be carefully explained in the ''METHODS'' section, offering a detailed understanding of their functionalities. Our focus will then shift to the presentation of results, wherein we attempt to address various questions covering the influence of pre-processing on the analysis outcomes. In conclusion, we will present our findings in a concrete summary, including the key takeaways from our research. This structured approach ensures a systematic and insightful exploration of the materials, methods, results, and conclusions, enhancing the clarity and coherence of our study.

II. MATERIALS

To enhance the overall applicability of our analysis, three different types of image modalities are thoroughly examined. By considering the multiple modalities, the goal is to derive more comprehensive and reliable conclusions from our research. In the following sections, a detailed overview of the specifications for the three datasets is provided.

A. HEMATOXYLIN AND EOSIN STAIN (H&E STAIN) DATASET

The first data set consists of the image patches derived from the human colorectal cancer (CRC) and the normal tissues stained with the hematoxylin & eosin (H&E) technique [16], [18], [19]. All the patches are of the size of 224 × 224 pixels and into the two classes, the normal and colorectal adenocarcinoma epithelium tumor (TUM) tissues. The histological analysis of the stained human tissue samples stands as the gold standard for evaluating many diseases, forming the fundamental basis for the pathologic assessments. This analysis involves the careful examination of the histologically stained tissue samples affixed on the glass slides, utilizing either traditional microscopes or digitized versions of the histologic images obtained through advanced Whole Slide Image (WSI) scanners. Within the pathology workflow, the histological staining step is critical as it plays a pivotal role in providing contrast and color to a tissue, enabling a clear differentiation among various tissue components. The most employed stain, known as the routine stain, is the hematoxylin and eosin (H&E) stain, which finds application in nearly all clinical cases, encompassing approximately 80% of global histologic staining [20], [21].

B. X-RAY DATASET

C. As another dataset, the chest X-ray images are used to extend the analysis beyond the H&E images [17], [22]. The

dataset contains the normal chest X-ray images depicting clear lungs and the images with the sign of pneumonia (PNE) in lungs all of the size of 224 × 224 pixels.

X-Ray is an electromagnetic radiation similar to visible light and infrared. However, it has higher energy levels and can effectively penetrate most objects, including the human body.

In medicine, X-ray technology is utilized to capture detailed images of tissues and structures within the body. As X-rays travel through the body, the absorption is different depending on the tissue and the density. Simultaneously, they pass through an X-ray detector positioned on the other side of the patient, forming an image that represents the ''shadows'' created by the organs within the body. Medical X-ray images can be considered as an essential diagnostic tool, enabling healthcare professionals to visualize and assess the structure within the body. For further information on this topic, readers are encouraged to refer to the [23].

C. OPTICAL COHERENCE TOMOGRAPHY (OCT) DATASET

The final image modality that is incorporated within the study is the optical coherence tomography (OCT) images [24] of the adult patients including two classes, the normal OCT images showing normal retina in absence of any retinal fluid/edema and the images with sign of choroidal neovascularization (CNV) all with the same pixel size of 224 × 224. The optical coherence tomography (OCT) is an advanced optical imaging modality that enables a high-resolution, cross-sectional tomographic imaging of the internal microstructures in both materials and biological systems. This technique utilizes the measurement of back-scattered or back-reflected light to generate detailed visualizations of the examined samples. By harnessing the power of light, OCT offers remarkable insights into the internal structures, facilitating precise analysis and evaluation in diverse scientific and medical applications [23], [24]. Fig. 1 and Table 1, present details about the datasets used in this study. All the images are standardized to the size of 224 × 224 pixels. Subfigure (a) exhibits the examples of H&E-stained images belonging to the two distinct classes: the normal tissues labeled as NORMAL, and the tumorous tissues labeled as TUM. Subfigure (b) displays the examples of chest X-ray images representing two classes: the normal lung images labeled as NORMAL and the lung images with pneumonia labeled as PNE. Subfigure (c) presents the examples of OCT images depicting two different classes: the normal retina images without any retinal fluid/edema labeled as NORMAL and the images exhibiting signs of choroidal neovascularization labeled as CNV. Additionally, subfigures (a), (b), and (c) provide a comprehensive overview of the datasets under investigation, across the training and test sets and highlighting the distribution strategy for the model development and evaluation.

III. METHODS

In the initial phase of the study, various adjustments and filtering techniques are applied to enhance the quality of the raw images. Additionally, different normalization ranges are chosen to standardize the pixel values across the dataset, allowing fair comparison and reliable analysis. The extraction of key features from the pre-processed images is essential for capturing the distinctive characteristics of the data. Here the VGG16, a pre-trained network along with its pre-trained parameters, is used to extract meaningful features from the pre-processed images. The output of VGG16 for a given image comprises high-level feature maps or representations with spatial dimensions (7, 7, 512). In this tensor, each channel corresponds to a specific feature. The resulting tensor maintains a spatial resolution of 7 × 7 and consists of 512 channels. Consequently, every image is transformed into an array of features with 25088 elements. Both the average and max pooling modes are employed for feature extraction in this step. Extracting features for each dataset across all the 248, and 296 pre-processing combinations for the sequence 1 and sequence 2 took approximately seven days.

Next, the focus shifted toward designing robust binary classification models. The PCA-LDA model, a widely used technique for dimensionality reduction and classification, is chosen as the classifier. We conducted a comprehensive examination of PCA components, ranging from 1 to 100. By iteratively training and testing with each PCA component, we identified the optimal number that resulted in the highest mean sensitivity for the test set. This maximal mean sensitivity value is subsequently used to compare the preprocessing techniques. Combining the power of the Principal Component Analysis (PCA) for the feature dimensionality reduction and the Linear Discriminant Analysis (LDA) for the classification, the model aimed to distinguish between the defined classes accurately. The estimated time for model training was approximately two hours per dataset, accounting for all the incorporated pre-processing combinations. Following this training phase, generating predictions for the test set with each model was accomplished in less than one minute.

The main The main concern of this study is to examine the impact of various pre-processing combinations on the performance of classification models rather than focusing on model optimization. As a result, instead of implementing deep learning models and controlling different parameters during training, which can significantly influence the model performance, we opted for the PCA-LDA as a basic Machine Learning (ML) approach. This choice allowed us to place greater emphasis on the effect of pre-processing combinations in our study while minimizing any interference from other influential factors originating from the classification models themselves.

To provide a deeper insight into the overall process, the subsequent sections cover the adopted processing techniques, the feature extraction methodology and the design and implementation of the binary classification models using the PCA-LDA approach. We are not just investigating the effects of various pre-processing techniques and their respective

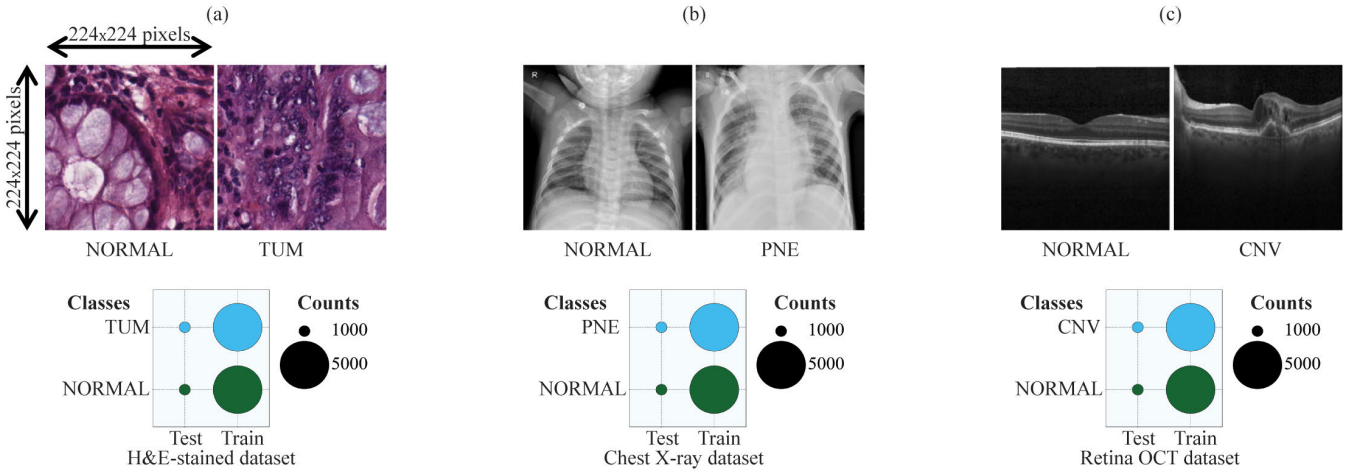

FIGURE 1. Characteristics of the studied images within each dataset. We provide an overview of the key features of the analyzed images within each dataset. All the images have been uniformly resized to the dimensions of 224 × 224 pixels for consistency. Subfigure (a) illustrates the examples of H&E-stained images belonging to two distinct classes: normal tissues labeled as NORMAL and tumorous tissues labeled as TUM. Subfigure (b) displays the examples of chest X-ray images representing two classes: normal lung images labeled as NORMAL and lung images with pneumonia labeled as PNE. Subfigure (c) presents the examples of OCT images showing two different classes: normal retina images without any retinal fluid/edema labeled as NORMAL and images exhibiting signs of choroidal neovascularization labeled as CNV. Additionally, subfigures (a), (b), and (c) provide the overview of the datasets under investigation, visually displaying the classes and the respective number of images in each class, as well as the division between the training and test sets for all the three datasets.

TABLE 1. Overview of the studied datasets.

Dataset	H&E-stained images		Chest X-ray images		Retina OCT images	
Classes	TUM	NORMAL	PNE	NORMAL	CNV	NORMAL
Train set	5000	5000	5000	5000	5000	5000
Test set	1000	1000	1000	1000	1000	1000

combinations; we are also studying the influence of the sequencing of pre-processing steps. To analyze this aspect, two experiments have been designed to explore the effect of applying pre-processing steps in different sequences. The experimental design and the specific sequences of the pre-processing steps are outlined below (See Fig. 2):

- Sequence 1:**Adjustment** + **Filtering** + Normalization + Feature extraction + Classification
- Sequence 2:**Filtering** + **Adjustment** + Normalization + Feature extraction + Classification

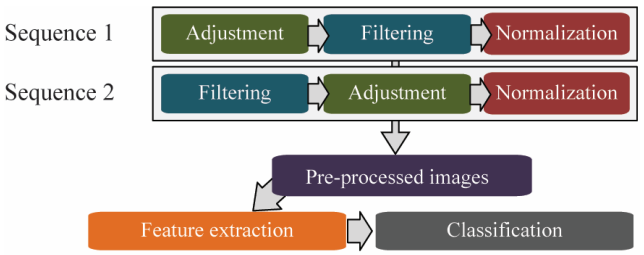

FIGURE 2. Illustration of the different pre-processing sequences. This figure illustrates the two distinct pre-processing sequences, denoted as "Sequence 1" and "Sequence 2," demonstrating the application of pre-processing steps.

A. PRE-PROCESSING

To generate pre-processed images, well-established and widely used methods in the image pre-processing have been considered to modify and enhance the image quality. They enable users to manipulate image features while emphasizing specific and important aspects or eliminating undesired elements. Various operations, such as smoothing, sharpening, and edge enhancement, can be achieved through image pre-processing. These methods can be categorized into two subgroups: frequency domain methods (FDMs) and spatial domain methods (SDMs). In this study we mainly focus on the second group, SDMs. In SDMs, the transformations are applied directly to the pixel values of the images, manipulating them to achieve the desired enhancement. On the other hand, FDMs involve multiple steps. It starts with computing the image's Fourier Transform (FT), followed by applying

the enhancement operations to the FT, and finally, using the inverse Fourier transform to obtain the enhanced image [25]. Our study focuses on enhancing medical image quality using the most common and well-established spatial domain filtering techniques to address noise reduction and visual quality enhancement.

1) ADJUSTMENT

For the adjustment step, we perform brightness adjustment, contrast adjustment, and histogram equalization. Brightness adjustment is a commonly employed and straightforward method for image manipulation. It involves modifying the image's brightness by adding or subtracting a constant value to each pixel, resulting in a brighter or darker image. This technique alters the overall luminance or brightness uniformly across all the channels. It serves various purposes, such as enhancing visibility, addressing exposure issues, and preparing images for analysis. Contrast adjustment is another technique that falls within the SDMs group. Contrast refers to the difference in brightness between the light and dark regions of an image. Contrast adjustment aims to manipulate and redistribute the pixel values of an image to improve the separation of obscured structural variations in the pixel intensity into a more visually differentiable structural distribution. The goal is to achieve a more visually noticeable distribution of these structural differences [26]. Additionally, histogram equalization enhances image quality by evenly distributing the most common intensity values.

Essentially, histograms serve as the statistical depictions of an image, where intensity values typically occupy a narrow range. By employing histogram equalization, the image is modified to achieve an equal distribution of intensity values throughout the enhanced image [23].

2) FILTERING

For the filtering phase, widely recognized filters used in image processing, namely median, Gaussian, and mean filtering, have been employed. Median filtering effectively reduces noise in an image or signal by replacing each pixel value with the median value of its neighboring pixels. By sliding a window (kernel) across the image, the median value is calculated from the pixel values within the window (kernel) [27]. In mean filtering, the same approach as in the median filtering is generally followed, with one key difference: instead of replacing the pixel value with the median value of the surrounding pixels, it is replaced by the mean (average) value of the surrounding pixels [28]. The Gaussian filter convolves the image with a Gaussian kernel, a matrix of values derived from the Gaussian function. Each pixel in the image is replaced with a weighted average of its neighboring pixels, with the weights determined by the Gaussian kernel [29], [30]. All these applied filtering techniques are associated with several advantages and disadvantages. For example, median filtering stands out for its reliability in preserving sharp edges and handling outliers, while effectively eliminating noise. However, it is essential to consider that median filtering may lead to a potential loss of fine details, particularly in high-frequency regions, due to its smoothing effect. Moreover, the computational complexity of median filtering increases with larger neighborhood sizes, making it less suitable for real-time applications or processing large images.

Considering mean filtering, it offers simplicity and efficiency in noise reduction, making it well-suited for real-time scenarios and fast processing. Nevertheless, a notable drawback is the potential blurring or loss of fine details, as the uniform averaging process smooths sharp edges, textures, and significant image details, especially noticeable in high-frequency areas. It may also introduce unrealistic values near sharp transitions, affecting important image features. Gaussian filtering applies a weighted average based on neighboring pixels using a Gaussian kernel, effectively reducing noise while preserving essential image details. It is worth mentioning that Gaussian filtering can introduce blurring on sharp edges and fine details as neighboring pixels contribute to the smoothing process, reducing overall image sharpness. These pros and cons encouraged us to study these filtering techniques in a very systematic procedure along with different kernel sizes to monitor the power of each method for suppressing noise and check their potential effect on the outcomes of the classification models. Fig.3 presents the visual representation of the pre-processed images for the three sets of studied data. Each transformation depicted in Fig. 3, is applied to the raw images as a part of the pre-processing pipeline. The set of the adjustments featured in Fig.3 are brightness adjustment, contrast adjustment, and histogram equalization. By manipulating the brightness, contrast levels, and histograms, the images become more visually enhanced while the important features for image analysis are highlighted. Mean, median and Gaussian filtering techniques have been applied with different kernel and sigma sizes. Fig.3 clearly shows that as the kernel/sigma sizes increase, the images become smoother and more blurred. However, it is important to note that excessively large sizes may remove the key features necessary for the image classification tasks. Therefore, finding the right balance between noise reduction and feature preservation is critical.

3) NORMALIZATION

Normalization, scaling, and mapping techniques can be helpful to improve the accuracy and efficiency of data analysis tasks and model training [31], [32]. Normalizing the pixel values transforms the image data so that they have similar statistical properties, regardless of differences in lighting conditions, exposure levels, or color distributions. This helps to make the image data more suitable for various machine learning algorithms, as it reduces the influence of variations in pixel values unrelated to the underlying structure or content of the image. In this study, three different normalization ranges have been considered. The first range that has been explored

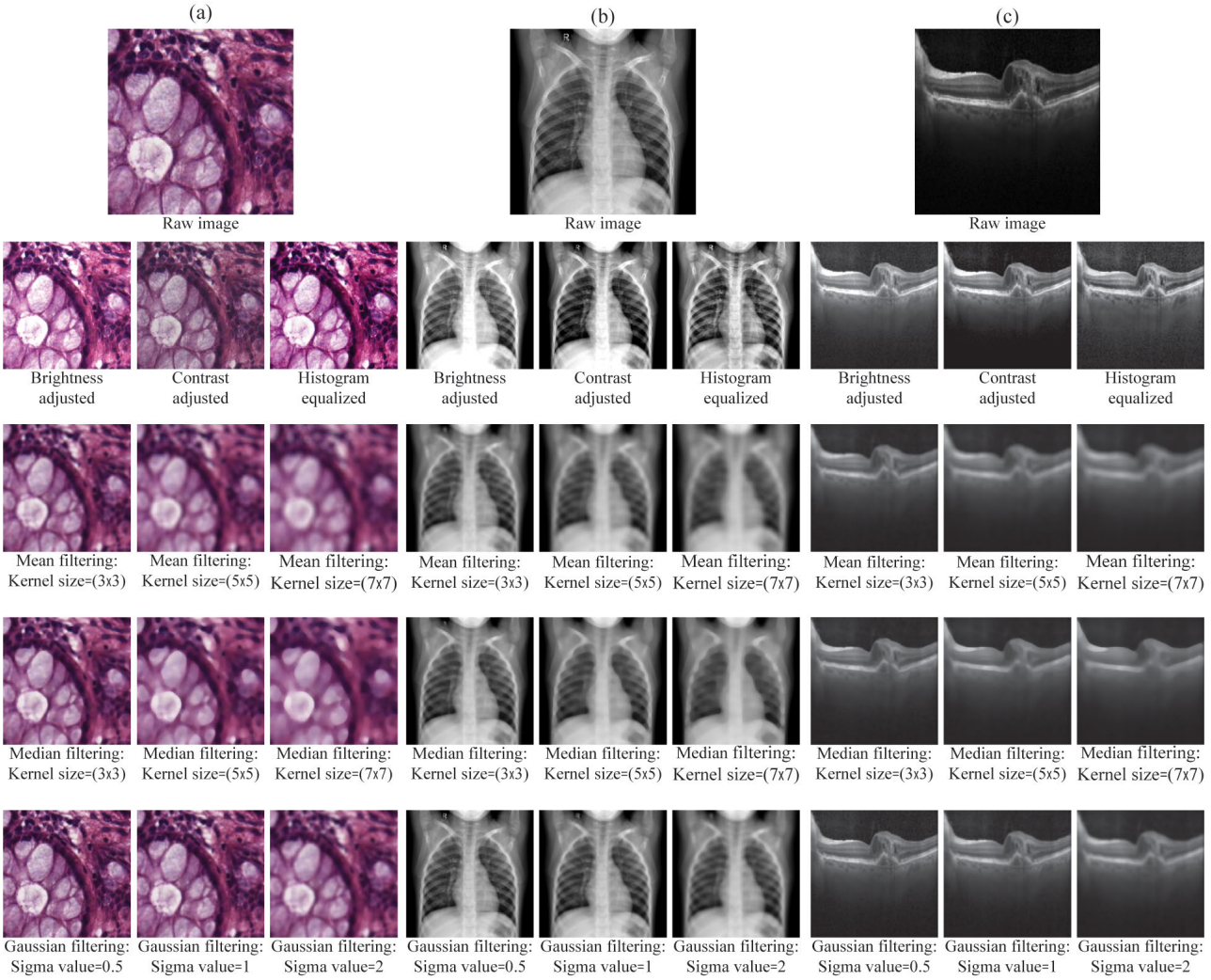

FIGURE 3. **Illustration of the image transformation with various adjustments and filtering techniques. This figure presents the transformation of the raw image using various adjustment methods and filtering techniques, employing different kernel and sigma sizes for (a) the H&E images, (b) the X-ray images, (c) the OCT images.**

is the commonly used (min, max) range, which scales the pixel values to the minimum and maximum values within the image.

The second range that has been examined is $(-1, 1)$. Additionally, we have also investigated the effect of normalizing the pixel values by dividing them by their maximum value (norm to max). These normalization techniques have been applied to standardize the pixel values across all the images, enabling fair comparisons.

Subsequently we can also assess the influence of these normalization ranges on the classification outcomes and explore whether any patterns emerge between the selected normalization ranges and the range initially employed to train the model used as the feature extractor in this study. In the following section, we explore our model and feature extraction approach. In Table 2, an overview of the adjustment methods, filtering techniques, and normalization ranges is shown.

4) FEATURE EXTRACTION

Transfer learning is a powerful approach that involves taking a model initially trained on one task and customizing it for a different, often related task. For instance, the Luo et al. demonstrated the effectiveness of the transfer learning as their primary classification method when establishing their analysis [33]. In the context of our present study, we have employed transfer learning as the central technique for feature extraction. This process transforms the raw data into the numerical features that include key information from the dataset. This approach allows us to benefit from the prior knowledge learned by the pre-trained model and adapt it to the specific requirements of the current task. These extracted features capture the most informative aspects of the data, enabling machine learning models to achieve higher accuracy and efficiency. To extract features from the pre-processed images, a pre-trained network called VGG16 [15] with two pooling modes, the average and max pooling have been

TABLE 2. Overview of the pre-processing methods applied on the raw images of three datasets.

Filtering technique Kernel/Sigma size		Adjustment method	Normalization range
1. Median filtering	(3×3)	1. Brightness adjustment	1. (min, max)
2. Mean filtering (Three kernel sizes were employed)	(5×5) (7×7)	2. Contrast adjustment	2. (-1, 1)
3. Gaussian filtering (Three sigma values were employed)	0.5 1 2	3. Histogram equalization	3. max
4. No filtering		4. No adjustment	4. No normalization

utilized. Developed by the Visual Geometry Group at the University of Oxford, VGG16 has been trained on the large dataset, ImageNet [34], without scaling for large-scale image classification tasks. As mentioned before, pre-trained models like VGG16 can be used as they are or with customized layers, making them adaptable to new tasks. With this critical information for the image analysis is driven while, saving time and resources. This approach allows for efficient feature extraction and making predictions.

B. PCA-LDA AS THE CLASSIFIER

There are many possible techniques for data classification. The Principal Component Analysis (PCA) and the Linear Discriminant Analysis (LDA) are the two commonly used techniques for dimensionality reduction and data classification respectively. In this study combination of PCA with LDA has been implemented to develop the binary classification models.

1) PRINCIPAL COMPONENT ANALYSIS (PCA)

The Principal Component Analysis (PCA) [35] is a powerful technique that aids in simplifying the complexity of high-dimensional data while preserving essential trends and patterns. Its primary function is to transform the data into a reduced the number of dimensions, which serve as informative summaries of the original features. This is particularly valuable in the field of biology, where high-dimensional data is commonly encountered. As an unsupervised learning method, the operation of PCA is not dependent on any pre-existing knowledge regarding the relationships or differences between samples. It seeks to identify underlying patterns only based on the data itself. By geometrically projecting the data onto lower-dimensional spaces called principal components (PCs), PCA effectively reduces the data dimension while aiming to capture the most significant information using a limited number of PCs. This process enables the creation of a condensed representation that retains the core essence of the original data [36].

2) LINEAR DISCRIMINANT ANALYSIS (LDA)

The Linear Discriminant Analysis (LDA) [37] is one of the most popular dimensionality reduction techniques used for supervised classification problems in the machine learning. While LDA shares similarities with PCA, one crucial distinction exists between them. Unlike PCA, which seeks to identify new axes or dimensions that maximize the data's overall variation, LDA specifically aims to maximize the distinguishability or separability among predefined categories or classes.

This is the reason that we can use LDA not only for the dimension reduction but most importantly for our study as the classifier. In this study, the LDA models were trained using the images with unique pre-processing combinations as the input and subsequently tested on the independent test sets. By adopting this approach, we were able to compare the results obtained from the different models and determine the best-performing pre-processing combination based on the highest achieved mean sensitivity

3) SOFTWARE AND COMPUTATIONAL ANALYSIS

The computations were conducted on a commercially available PC system featuring an AMD Ryzen Threadripper 3960X 24-Core Processor (48 CPUs, 3.79 GHz, 128 GB RAM). Additionally, an NVIDIA GeForce RTX 3090 GPU with 10496 CUDA cores, a base clock speed of 1440 MHz, and a boost clock speed of 1710 MHz was utilized. The GPU is equipped with 24576 MB of GDDR6 memory, boasting a 936.10 GB/s memory bandwidth, 384-bit memory interface, and a memory data rate of 19.50 Gbps. All the analyses were carried out based on in-house written functions in the programming language Python version 3.10.9. In the supplementary materials a simplified version of the Pseudocode

for our implemented workflow within this study has been provided.

IV. RESULTS

This study undertakes a comprehensive analysis to thoroughly investigate the effects of different combinations of the pre-processing techniques on the outcomes of classification models. By examining the influence of various pre-processing procedures, this research addresses critical questions, such as ''Which filtering techniques and normalization ranges contribute to the most effective models, resulting in the highest mean sensitivities?'' or ''Does the order of pre-processing application significantly impact the final results?''. To address these questions, two distinct sequences, namely sequence 1 and sequence 2 as outlined in the methods section, are carefully explored. These sequences are employed to train the classifiers using specific pre-processing combinations. Subsequently, the trained models are then evaluated on the independent test sets allowing for comparison and analysis of the results. With the aim of determining the pre-processing combination that exhibits the highest performance, this approach enables informative decision-making regarding the optimal pre-processing techniques that enhance the predictive capabilities of the classification models.

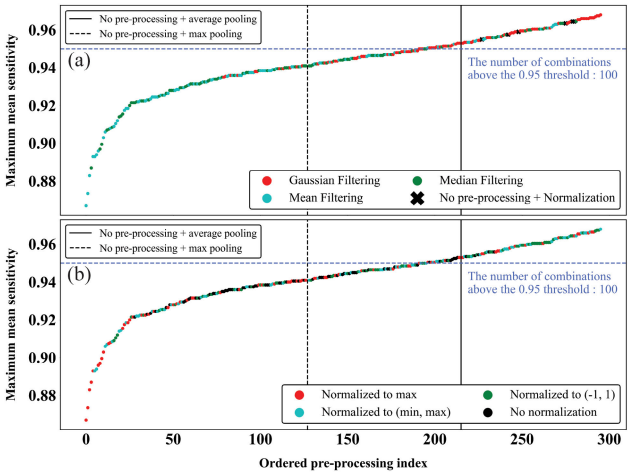

FIGURE 4. Optimizing the classification models for the H&E images with different pre-processing combinations. (a) Analysis of the filtering techniques on the classification outcomes. Each filtering technique is represented by a distinct color, while no pre-processing is indicated by vertical solid and dashed line. (b) Investigation of the normalization range impact on the classification model performance. Each normalization is shown with a specific color, while no pre-processing is presented with two solid and dashed vertical lines. Among the various filtering techniques, the Gaussian filtering stands out as a prominent enhancer of the classification model performance, while we could not see such a prominent color for the normalization ranges. The dashed blue lines in (a) and (b) represent the established threshold (around 0.95) and the 100 classification models exceeding that threshold.

1) EFFECT OF THE FILTERING AND NORMALIZATION

In this section, the performance of binary classification models has been compared, considering the filtering techniques and normalization ranges. The objective is to identify the best performing filtering method and normalization range based on the highest mean sensitivity. To distinguish the key factors that drive the best performance of the classification models, firstly our emphasis lies in exploring the impact of filtering techniques and normalization ranges when utilizing the H&E images as the training and test sets, while following the sequence 2pipeline for generating the pre-processed images. Fig. 4 illustrates the mean sensitivity of the classification task versus the ordered pre-processing combination index. The analysis presented in Fig. 4a clearly indicates that Gaussian filtering significantly improves the performance of the classification, leading to the higher mean sensitivities. Conversely, mean filtering yields the lowest mean sensitivities among the examined filtering techniques. In Fig. 4b, it can be observed that selecting an appropriate normalization range can indeed enhance the performance of the classification task. However, the absence of a dominant color indicates that there is no definitive ''best'' or the ''worst'' choice for the normalization. Furthermore, it's important to note that VGG16 was originally trained using non-normalized data. As a result, establishing any potential correlations between the normalization ranges we have chosen and the original training data which VGG16 was initially trained on is not feasible. To expand our investigation, we established the mean sensitivity threshold of 0.95. The models surpassing this standard were recognized as the well-suited for diagnostic tasks. Remarkably, over a hundred of the classification models demonstrated mean sensitivities exceeding 0.95, revealing their precise class identification capabilities and underlining the model's enhanced reliability through carefully tuned pre- processing combinations.

Investigation of the filtering techniques and the normalization for the X-ray and OCT images is presented in Fig. S1 and Fig.S2, respectively. Fig.S1 highlights the findings for the X-ray images, revealing that Gaussian filtering emerged as one of the most effective techniques, resulting in higher mean sensitivities. On the other hand, mean and median filtering exhibited lower mean sensitivities. Similar to the observations made for the H&E images, no noticeable color code was observed for the normalization effect in the X-ray image analysis. Turning to Fig. S2, the results for the OCT dataset indicate that there is no clear trend for identifying the best performing filtering technique. Figure S3 illustrates two instances of confusion tables related to the H&E-stained dataset within the test set. However, a notable trend emerges when examining the impact of normalization. It becomes evident that images without any form of normalization were associated with the lower mean sensitivities.

2) EFFECT OF THE PRE_PROCESSING ORDER

Another, often neglected, parameter is the order in which pre-processing techniques are applied. Hence, a thorough investigation has been conducted to address this aspect as well. Two distinct sequences in the study were employed to comprehensively assess the impact of pre-processing order:

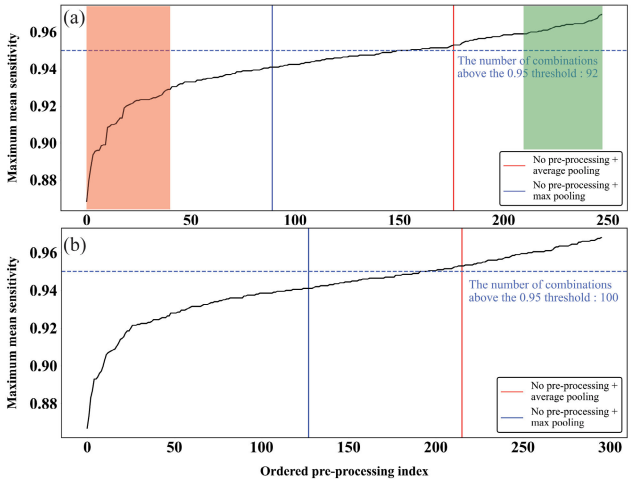

FIGURE 5. Exploring the mean sensitivity for the classification models (The H&E dataset) by comparing the two pre-processing orders with different sequences. (a) Pre-processing sequence 1, and (b) Pre-processing sequence 2. It can be clearly seen that the sequence does not have a significant effect on the results. The highlighted areas in red and green represent two distinct sets of the models, each comprising 40 These models are classified into the two groups: the best models, indicated by the green area, and the worst models, represented by the red area. The dashed blue lines indicate the established threshold at round 0.95, and the models exceeding that threshold.

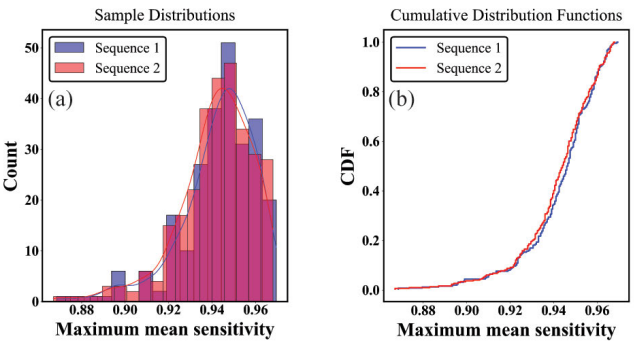

FIGURE 6. Comparison of the classification outcomes and the probability distributions for the two pre-processing sequences: sequence 1 and sequence 2 and the H&E dataset. (a) Visualizes the sample distributions of the binary classification results coming from two different pre-processing orders. Not only the sample distributions are not following a normal pattern, but also the shape of the distributions illustrates a high level of similarity. (b) Shows Cumulative Distribution Functions (CDFs) for two distinct orders of pre-processing. Through the comparison of their CDFs, it becomes apparent that there is no significant distinction in terms of their probability distribution and spread.

- Sequence 1: **Adjustment** + **Filtering** + Normalization + Feature extraction + Classification
- Sequence 2: **Filtering** + **Adjustment** + Normalization + Feature extraction + Classification

Fig. 5, displays the mean sensitivity versus the ordered pre-processing index for the H&E images by applying the pre-processing methods while sequence 1 (Fig. 5a), and sequence 1 (Fig. 5b) are followed. Interestingly, the analysis reveals that the order in which the pre-processing techniques are applied does not significantly affect the results.

To further support our findings, we conducted a test aimed at comparing the classification outcomes derived from employing the two distinct sequences during the application of the pre-processing steps. The Kolmogorov-Smirnov test [38] was chosen as our analytical tool, which relies on the use of Cumulative Distribution Functions (CDFs) in its methodology. This non-parametric method evaluates whether a significant difference exists between the outcomes originating from the sequence 1 and sequence 2.

Fig. 6a, displays the distribution of the results obtained through the application of both sequence 1 and sequence 2 for generating the pre-processed images. This visual presentation reveals two significant insights. First, the way the data is spread out doesn't match a normal pattern, which highlights why we are using the Kolmogorov-Smirnov test for the comparison. Furthermore, it becomes clear that both distributions exhibit the same shapes, indicating their similarity. In Fig.6b, we designed a thorough examination of the probability distributions within outcomes derived from the two distinct pre-processing sequences, utilizing CDFs. The CDF, as a fundamental statistical concept, offers essential insights into the probability distribution of a specific variable.

Running the Kolmogorov-Smirnov test, the maximum vertical distance between two CDFs, which is known as the KS statistic was 0.06 with the p-value $=$ 0.5 (larger than the significance level, $\alpha = 0.05$). Considering the obtained KS statistic, and the p-value we can strongly claim that the data distributions in the sequence 1 and sequence 2 are similar and there is no significant evidence to suggest a difference between these two distributions.

This observation suggests that the performance and effectiveness of the classification models are independent of the specific sequence in which the pre-processing procedures are applied.

Fig. S4 and Fig. S5, showing the classification results obtained for the X-ray and OCT datasets, respectively. Similar to the findings for the H&E images, no significant differences are observed when comparing the outcomes obtained using the pre-processing sequence 1 and sequence 2 for both the X-ray and OCT images. Generally, considering all the three sets of data, it can be concluded that, regardless of the sequence in which the pre-processing techniques were applied, the performance and effectiveness of the classification models remained consistent and reliable. To provide a clear understanding of the findings, Fig. 7 offers an informative overview of the results obtained for the three studied datasets. The box plot highlights the similarity in the performance, reinforcing that the specific order of pre-processing steps did not significantly influence the classification models' overall outcomes.

3) THE BEST-PERFORMING AND WORST-PERFORMING COMBINATIONS

Next, the similarities between the best-performing and worst-performing pre-processing combinations have been

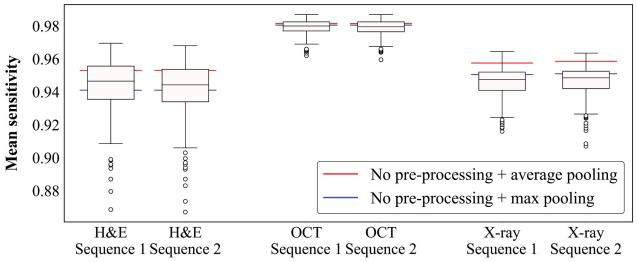

FIGURE 7. Comparative analysis of the results across the H&E, X-ray, and OCT datasets with the two different pre-processing sequences. This figure provides an overview of the results obtained from the H&E, X-ray, and OCT images using two distinct pre-processing sequences, denoted as Sequence 1 and Sequence 2. This comparison reveals that the choice of pre-processing sequence has a negligible impact on the overall performance of the classification models.

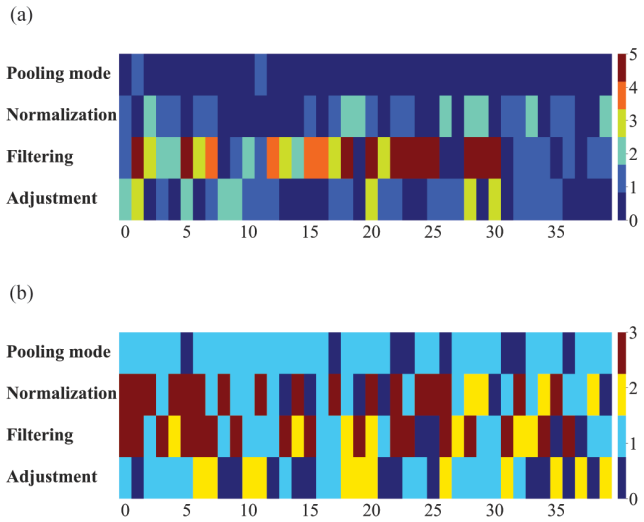

FIGURE 8. Exploring the best-performing and the worst-performing solutions using the H&E dataset. This figure provides an overview of the best-performing (a) and the worst-performing (b) classification models, along with the pre-processing combinations they utilize. These combinations contain various pre-processing steps including adjustments, filtering techniques with different kernel and sigma sizes, normalization ranges, and pooling modes, all clearly color-coded for better interpretation. It can be seen that in the both shown cases, except for the pooling mode, the rest (Adjustment, Filtering, Normalization) do not exhibit any similarities.

investigated. To explore this aspect, two distinct sets comprising 40 models each (see the highlighted areas in Fig. 5a) were analyzed. The first set consists of the best-performing models that have been selected based on their higher mean sensitivities, and the other set contains the worst-performing models that have been chosen for their lower mean sensitivities. Analyzing the pre-processing combinations in these sets allows us to uncover the patterns influencing their performances. In Fig. 8, the best and worst pre-processing combinations for the H&E images are presented. In this false color plot, each step (Adjustment, Filtering, Normalization), along with pooling mode, is displayed using its own distinct color, enabling a clear and visually informative representation. This visualization offers valuable insights into the pre-processing pipeline and makes it easier to distinguish the order in which each process occurred. In Fig. 8a, 40 of the best-performing combinations are shown, while Fig.8b illustrates the worst-performing pre-processing combinations. In Fig. 8a, it is evident that, except for the pooling mode, a uniform color pattern is not observed. However, when considering the remaining steps, the wide variety of color combinations indicate the absence of any specific trend or pattern. The same observation is apparent in Fig. 8b, where the pre-processing combinations for the worst-performing models are shown. Except for the pooling mode, this finding highlights the absence of a one-size-fits-all approach, emphasizing the need for a case-oriented pre-processing strategy.

Fig. S6 and Fig. S7 focus on the exploration of common features and patterns between the 40 of the best-performing and worst-performing models, determined by their higher and lower mean sensitivities, for the X-ray and OCT datasets. Consistent with the observations made for the H&E images, the results demonstrate minimal shared features among the models, except for the pooling mode, where partial similarities can be observed due to the uniform color code. Based on our findings, the selection of appropriate pooling mode for the feature extraction appears as the most influential factor affecting the overall performance of the classification models. One possible explanation for the observed behavior is the implementation of highly standardized protocols during the generation of the H&E [39], X-ray [40], and OCT [41] images. These standardized protocols ensure consistent image acquisition and processing, which may reduce the sensitivity of these imaging modalities to the additional image pre-processing steps.

Figs S8, S9, and S10 present detailed representations of false color plots depicting the performance of both the best and worst-performing models. Annotations are included to provide insights into the underlying factors represented by the colors in the plots.

V. PRINCIPLE FINDINGS

Our findings within this comprehensive study can be shortly outlined as:

- Our study pioneered a systematic examination of image pre-processing's impact on the model performance, addressing key questions to enhance understanding of its influence on the analysis results.
- The wise selection of the pre-processing methods concerning the data and the objective of the analysis can improve the analysis outcomes.
- The impact of pre-processing on the performance of binary classification models, in terms of sensitivity scores, observed within the ranges of 0.87-0.97 for H&E, 0.96-0.99 for Retina OCT, and 0.92-0.96 for the chest X-ray dataset.
- Across all the incorporated image modalities, a significant number of models surpassed the threshold of 0.95 for the mean sensitivity.

- Almost half of the worst-performing pre-processing combinations led to lower mean sensitivities compared to cases without pre-processing, highlighting the substantial negative influence on the overall model outcomes.
- Exploration of the pre-processing sequence impact on the model performance using two distinct sequences (sequence 1 and sequence 2), revealing consistent model performance irrespective of the specific order of pre-processing steps, as indicated by the Kolmogorov-Smirnov test.
- Minimal similarities in pre-processing steps for the best-performing and worst-performing models were observed, except for the pooling mode.
- No relationship was found between the original normalization range for VGG16 and the chosen ranges in the current study.

VI. CONCLUSION

In this systematic investigation, our goal was to undertake a comprehensive and thorough examination of how pre-processing influences the outcomes of binary classification models. Our focus was to determine whether pre-processing consistently enhances model performance or if this assumption doesn't universally hold true. Furthermore, we aimed to uncover any potential best-performing or worst-performing pre-processing combinations that might exist and also to explore whether the sequence of applying pre-processing steps can significantly impact the results of the developed models.

To establish the analysis, three different datasets: the H&E, the X-ray, and the OCT images within two distinct approaches for combining the pre-processing methods were explored. The first approach involved the application of pre-processing steps in the following sequence: First, the adjustment procedure was applied, followed by filtering and the normalization. In the second approach, the sequence was altered, starting with the filtering, followed by the adjustment and finally the normalization. All the possible combinations of these three methods were systematically explored in our analysis.

The findings of this study highlighted the significant impact of pre-processing procedures on the performance of classification models. By appropriately selecting and optimizing the pre-processing steps, it was possible to enhance the models' performance. Specifically, for the H&E dataset, careful selection and optimization of the pre-processing strategy led to a remarkable increase in the mean sensitivity from 0.86 to 0.97. Similarly, for the X-ray dataset, the mean sensitivity improved from 0.91 to 0.96, and for the OCT images, it increased from 0.96 to 0.99. These results demonstrated the effectiveness of correctly applying the pre-processing techniques in improving the accuracy of classification models.

In particular, considering the established threshold of 0.95 for the mean sensitivity, our investigation has revealed that across all the incorporated image modalities (the H&E, Chest X-ray, and OCT retinal images), significant number of the pre-processing combinations have surpassed this threshold when employed as the inputs for the classification models.

An even more interesting finding emerged when considering the impact of utilizing the worst-performing pre-processing combinations on the model's performance. By analyzing classification models that incorporated all the three image modalities, it was uncovered that almost half of these combinations resulted in lower mean sensitivities compared to the cases when no pre-processing applied to the images. This finding served as a reminder of the substantial negative influence that the worst-performing pre-processing techniques can have on the overall outcome of the developed models.

The impact of the pre-processing sequence on the model's performance, by applying two distinct sequences (sequence 1 and sequence 2), was also explored. Utilizing the Kolmogorov-Smirnov test as an analytical instrument allowed for assessing the similarity of the results arising from two distinct pre-processing sequences. The results of the test highlighted insignificant differences in the data distribution for both sequence 1 and sequence 2. This finding suggested that the performance of the models remained consistent regardless of the specific order in which pre-processing steps were applied.

Lastly, when considering the best-performing and worst-performing models, minimal similarities were observed in the pre-processing steps across almost all the three image modalities, except for the pooling mode. We extended our investigation to unveil any potential correlation between the original normalization range employed for training VGG16 initially and the chosen normalization ranges in our current study. Considering that VGG16 underwent its initial training without data scaling, no noticeable relationship or pattern can be identified in this scenario.

We integrated a diverse range of methods and datasets, allowing us to draw conclusions from a rich subset of data and techniques. Using open repositories, our study utilized image data from three major medical domains: Radiology, Pathology, and Ophthalmology. This systematic examination was conducted by exploring all the possible combinations of the 15 pre-processing methods, comprising three adjustment techniques, three normalization ranges, and nine filtering techniques with VGG16 as the feature extractor. We believe that while the numerical outcomes may vary for a given task and modality, they are generalizable and consistently fall within a comparable and similar range. Testing all potential combinations with hyperparameter optimization is not feasible. For differing tasks, other modalities and data combinations, the results have to be re-verified using the approach outlined in this research. Continuous improvement is a fundamental aspect of any work, and further studies will include the interplay between the DL network, pre-processing, and modality.

The study holds critical significance in advancing diagnostics and patient care. With the implemented approach

and accomplished extensive study, the improvement of optimizing the pre-processing steps can be judged. Through the optimization of the pre-processing strategies, the robustness and reliability of AI-based decisions, particularly when dealing with the challenges posed by inherent noise and visual variations in medical images, will be enhanced. In industrial applications, the improved performance of classification models not only ensures more accurate disease detection but also positively influences treatment planning, ultimately impacting patients with improved outcomes.

ACKNOWLEDGMENT

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of European Union or European Research Council. Neither European Union nor the granting authority can be held responsible for them.

REFERENCES

[1] A. Babayan et al., "A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults," *Sci. Data*, vol. 6, no. 1, pp. 1–21, 2019.

[2] E. Hatef, K. M. Searle, Z. Predmore, E. C. Lasser, H. Kharrazi, K. Nelson, P. Sylling, I. Curtis, S. D. Fihn, and J. P. Weiner, "The impact of social determinants of health on hospitalization in the veterans health administration," *Amer. J. Preventive Med.*, vol. 56, no. 6, pp. 811–818, Jun. 2019.

[3] J. Gerretzen, E. Szymańska, J. J. Jansen, J. Bart, H.-J. van Manen, E. R. van den Heuvel, and L. M. C. Buydens, "Simple and effective way for data preprocessing selection based on design of experiments," *Anal. Chem.*, vol. 87, no. 24, pp. 12096–12103, Dec. 2015.

[4] J. Engel, J. Gerretzen, E. Szymańska, J. J. Jansen, G. Downey, L. Blanchet, and L. M. C. Buydens, "Breaking with trends in pre-processing?" *TrAC Trends Anal. Chem.*, vol. 50, pp. 96–106, Oct. 2013.

[5] R. Arun, M. S. Nair, R. Vrinthavani, and R. Tatavarti, "An alpha rooting based hybrid technique for image enhancement," *Image*, vol. 9, no. 10, pp. 1–10, 2011.

[6] N. Ali, E. Quansah, K. Köhler, T. Meyer, M. Schmitt, J. Popp, A. Niendorf, and T. Bocklitz, "Automatic label-free detection of breast cancer using nonlinear multimodal imaging and the convolutional neural network ResNet50," *Translational Biophotonics*, vol. 1, nos. 1–2, Dec. 2019, Art. no. e201900003.

[7] P. Pradhan, T. Meyer, M. Vieth, A. Stallmach, M. Waldner, M. Schmitt, J. Popp, and T. Bocklitz, "Computational tissue staining of non-linear multimodal imaging using supervised and unsupervised deep learning," *Biomed. Opt. Exp.*, vol. 12, no. 4, p. 2280, 2021.

[8] A. Giełczyk, A. Marciniak, M. Tarczewska, and Z. Lutowski, "Pre-processing methods in chest X-ray image classification," *PLoS ONE*, vol. 17, no. 4, Apr. 2022, Art. no. e0265949.

[9] A. K. Pradhan, K. Das, D. Mishra, and P. Chithaluru, "Optimizing CNN-LSTM hybrid classifier using HCA for biomedical image classification," *Expert Syst.*, vol. 40, no. 5, Jun. 2023, Art. no. e13235.

[10] A. M. Q. Farhan and S. Yang, "Automatic lung disease classification from the chest X-ray images using hybrid deep learning algorithm," *Multimedia Tools Appl.*, vol. 82, no. 25, pp. 38561–38587, Oct. 2023.

[11] A. Hameed, M. Umer, U. Hafeez, H. Mustafa, A. Sohaib, M. A. Siddique, and H. A. Madni, "Skin lesion classification in dermoscopic images using stacked convolutional neural network," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 4, pp. 3551–3565, Apr. 2023.

[12] H. Shrestha, S. C. B. Jaganathan, C. Dhasarathan, and K. Suriyan, "Detection and classification of dermatoscopic images using segmentation and transfer learning," *Multimedia Tools Appl.*, vol. 82, no. 15, pp. 23817–23831, Jun. 2023.

[13] S. Al-Fahdawi, A. S. Al-Waisy, D. Q. Zeebaree, R. Qahwaji, H. Natiq, M. A. Mohammed, J. Nedoma, R. Martinek, and M. Deveci, "Fundus-DeepNet: Multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus images," *Inf. Fusion*, vol. 102, Feb. 2024, Art. no. 102059.

[14] R. Houhou and T. Bocklitz, "Trends in artificial intelligence, machine learning, and chemometrics applied to chemical data," *Anal. Sci. Adv.*, vol. 2, nos. 3–4, pp. 128–141, Apr. 2021.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[16] J. N. Kather, N. Halama, and A. Marx, "100,000 histological images of human colorectal cancer and healthy tissue," *Zenod*, vol. 10, p. 5281, Aug. 2018.

[17] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (OCT) and chest X-ray images for classification," *Mendeley Data*, vol. 2, no. 2, p. 651, 2018.

[18] D. Wittekind, "Traditional staining for routine diagnostic pathology including the role of tannic acid. 1. value and limitations of the hematoxylin-eosin stain," *Biotechnic Histochemistry*, vol. 78, no. 5, pp. 261–270, Oct. 2003.

[19] M. Titford, "Progress in the development of microscopical techniques for diagnostic pathology," *J. Histotechnology*, vol. 32, no. 1, pp. 9–19, Mar. 2009.

[20] H. He, S. Yan, D. Lyu, M. Xu, R. Ye, P. Zheng, X. Lu, L. Wang, and B. Ren, "Deep learning for biospectroscopy and biospectral imaging: State-of-the-art and perspectives," *Anal. Chem.*, vol. 93, no. 8, pp. 3653–3665, Mar. 2021.

[21] J. D. Bancroft and C. Layton, "The hematoxylins and eosin," in *Bancroft's Theory and Practice of Histological Techniques*, 7th ed. Amsterdam, The Netherlands: Elsevier, 2012, pp. 173–186.

[22] A. Stanton, "Wilhelm conrad rontgen on a new kind of rays: Translation of a paper read before the Wurzburg physical and medical society, 1895," *Nature*, vol. 53, no. 1369, pp. 274–276, 1895.

[23] A. Maier, S. Steidl, V. Christlein, and J. Hornegger, *Medical Imaging Systems: An Introductory Guide*. New York, NY, USA: Springer, 2018.

[24] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[25] D. L. Galata, L. A. Mészáros, N. Kállai-Szabó, E. Szabó, H. Pataki, G. Marosi, and Z. K. Nagy, "Applications of machine vision in pharmaceutical technology: A review," *Eur. J. Pharmaceutical Sci.*, vol. 159, Apr. 2021, Art. no. 105717.

[26] R. D. Putra, T. W. Purboyo, and A. L. Prasasti, "A review of image enhancement methods," *Int. J. Appl. Eng. Res.*, vol. 12, no. 23, pp. 13596–13603, 2017.

[27] Y. Zhu and C. Huang, "An improved median filtering algorithm for image noise reduction," *Phys. Proc.*, vol. 25, pp. 609–616, Aug. 2012.

[28] K. Gupta and S. Gupta, "Image denoising techniques-a review paper," *IJITEE*, vol. 2, pp. 6–9, Apr. 2013.

[29] M. Mafi, H. Martin, M. Cabrerizo, J. Andrian, A. Barreto, and M. Adjouadi, "A comprehensive survey on impulse and Gaussian denoising filters for digital images," *Signal Process.*, vol. 157, pp. 236–260, Apr. 2019.

[30] Y.-L. You and M. Kaveh, "Fourth-order partial differential equations for noise removal," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1723–1730, May 2000.

[31] S. G. K. Patro and K. K. Sahu, "Normalization: A preprocessing stage," 2015, *arXiv:1503.06462*.

[32] L. A. Shalabi, Z. Shaaban, and B. Kasasbeh, "Data mining: A preprocessing engine," *J. Comput. Sci.*, vol. 2, no. 9, pp. 735–739, Sep. 2006.

[33] R. Luo and T. Bocklitz, "A systematic study of transfer learning for colorectal cancer detection," *Informat. Med. Unlocked*, vol. 40, Jun. 2023, Art. no. 101292.

[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conference Computer Vision Pattern Recognition*, Jun. 2009, pp. 248–255.

[35] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, Jul. 2010.

[36] J. Lever, M. Krzywinski, and N. Altman, "Points of significance: Principal component analysis," *Nature methods*, vol. 14, no. 7, pp. 641–643, 2017.

[37] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis—A brief tutorial," *Inst. Signal Inf. Process.*, vol. 18, pp. 1–8, Jul. 1998.

[38] K. An, "Sulla determinazione empirica di una legge didistribuzione," *Giorn Dell'inst Ital Degli Att*, vol. 4, pp. 89–91, Jul. 1933.

[39] W. B. Yoon, H. Kim, K. G. Kim, Y. Choi, H. J. Chang, and D. K. Sohn, "Methods of hematoxylin and erosin image information acquisition and optimization in confocal microscopy," *Healthcare Informat. Res.*, vol. 22, no. 3, p. 238, 2016.
[40] P. Kirkpatrick and A. V. Baez, "Formation of optical images by X-Rays," *J. Opt. Soc. Amer.*, vol. 38, no. 9, p. 766, 1948.
[41] D. M. Sampson, A. M. Dubis, F. K. Chen, R. J. Zawadzki, and D. D. Sampson, "Towards standardizing retinal optical coherence tomography angiography: A review," *Light, Sci. Appl.*, vol. 11, no. 1, p. 63, Mar. 2022.

PEGAH DEHBOZORGI received the B.Sc. degree in nuclear physics and photonics from the University of Mazandaran, in 2011, and the M.Sc. degree in nuclear physics and photonics from the Institute for Advanced Studies in Basic Sciences, Zanjan, in 2015. She is currently pursuing the Ph.D. degree with the Research Group of Photonic Data Science, University of Jena, supervised by Dr. Thomas Bocklitz. Her research interests include the intersection of image analysis and machine learning methods and driving innovation in fields, such as computer vision and medical diagnostics.

OLEG RYABCHYKOV received the B.Sc. degree in applied physics from V. I. Vernadsky Taurida National University, the M.Sc. degree in high technology from Taras Shevchenko National University of Kyiv, and the Ph.D. degree from Friedrich Schiller University Jena, in 2019. Currently, he is a Postdoctoral Researcher, a Data Scientist, and a Software Developer. His research interests include advancing spectral data processing automation, exploring the realms of data fusion, and employing machine learning methods for photonics data to automate diagnostics in the medical and biological fields.

THOMAS BOCKLITZ received the Diploma degree in theoretical physics, in 2007, the Ph.D. degree in chemometrics, in 2011, and the degree in physics and mathematics from Friedrich Schiller University Jena. He is currently a Full Professor with the University of Bayreuth and leading the AI in Spectroscopy and Microscopy Group. He is fascinated by the fact that problems can be understood and solved using mathematical methods. His research interests include machine learning for photonic image data, chemometrics/machine learning for spectral data, correlation of different measurement methods, and data fusion.