

Literature Review

This chapter delves into a comprehensive exploration of cutting-edge techniques in image analysis aimed at enhancing image quality, particularly focusing on the challenge of addressing specular reflections within smart colposcopy images. Specular reflections, often appearing as troublesome noise, can significantly impact image clarity and information extraction. To mitigate this issue, researchers have turned to a combination of traditional image processing methods and advanced deep learning techniques, which have demonstrated promising results in enhancing the quality of smart colposcopy images. In the realm of traditional image processing methods, a prominent strategy involves the identification and reduction of specular reflections on smart colposcopy images. This is thoroughly discussed in Section 2.2, where a literature review showcases various approaches adopted to tackle reflection problem on digital images. By employing techniques rooted in traditional image processing, investigators have developed algorithms to detect and remove specular reflections, thereby restoring image fidelity and improving visual interpretation. The Section 2.3 discussed further by delving into the realm of deep learning methods. Here, the focus shifts to the utilization of cutting-edge deep learning learning techniques for segmenting glare regions within smart colposcopy images. Deep learning, with its ability to automatically learn intricate features from data, has been employed to create models that accurately identify and isolate areas affected by glare. This segmentation process plays a pivotal role in subsequent image enhancement steps. A critical aspect of image restoration lies in addressing missing regions, which may result from glare removal or other factors. Section 2.3 highlights the use of deep learning inpainting methods to fill in these gaps seamlessly. Inpainting involves predicting and generating plausible content for the missing areas, ensuring the visual coherence of the enhanced image.

Finally, Section 2.4 delves into the application of deep learning in image classification for the grading of smart colposcopy images. Leveraging the discriminative capabilities of deep learning models, the classification systems that assign grades to images based on the severity of detected anomalies. These systems aid healthcare professionals in making informed decisions and recommendations, thereby contributing to improved patient care. In conclusion, this chapter provides a thorough review of state-of-the-art approaches to

enhance image quality in smart colposcopy images by addressing specular reflections and related challenges. It encompasses the utilization of traditional image processing techniques, deep learning-based glare region segmentation, advanced inpainting methods, and deep learning-driven image classification. By combining these methodologies, paved the way for more accurate grading and informative, and clinically relevant smart colposcopy image analysis, ultimately benefiting both medical practitioners and patients. As technology continues to evolve, the integration of these diverse methods holds the promise of further revolutionizing image quality improvement in the field of medical imaging.

2.1. Searching Strategies

The comprehensive literature review paper incorporates articles that address three primary areas. Firstly, it includes papers discussing the application of color-based threshold methods for the detection of specular reflections in medical images. Secondly, it integrates research articles that concentrate on the utilization of deep learning inpainting techniques for the segmentation of digital images. Furthermore, within the realm of medical images, it delves into inpainting methods designed to enhance image quality and provides insights into papers covering classification models for grading purposes. This literature review encompasses a wide array of research articles, with a specific search string, as detailed in Table 1, used to identify the most pertinent contributions in these fields.

Table 2.1 Search String used for the Searching Strategies

Topic	Search String
Color Space Threshold Method for the Detection of Glare Region	“RGB color for medical image glare detection” OR “HSV color space for glare detection ” OR “XYZ color for medical image glare detection” OR “YUVcolor space for glare detection ” OR “Lab color for medical image glare detection” OR “SUVcolor space for glare detection ” OR “glare detection on endoscopy” OR “glare detection on colposcopy” OR “glare detection on thorascopy images” OR “glare detection on digital heart images” OR “glare detection on colonoscopy images “OR “Specular Reflection detection on medical images”
Deep Learning segmentation model	Deep learning segmentation using masked images “Deep learning segmentation on medical images” OR “Deep segmentation model” OR “Deep learning segmentation on colposcopy images” OR”Reflection segmentation using deep learning models”
Deep Learning inpainting model	“Supervised inpainting models”, OR “Deep learning supervised inpainting method” OR “CNN inpainting Method” OR “CNN inpainting on digital images” OR “CNN inpainting on medical images” OR “Inpainting model filters” “Filter for inpainting models”
Deep Learning for Grading	“Classification of deep learning model” OR “Grading of medical image” OR “Deep learning classification model for medical images” OR

In this review article, all non-English review papers have been excluded. To determine which papers to be included, selection criteria and a filtering process is represented in the PRISMA diagram Figure 2.1 The research papers are excluded which are not pertinent to our research, those published in languages other than English, abstract-only publications, papers lacking full-text availability, and articles not indexed in specified databases such as IEEE, Web of Science (WOS), PubMed, and Scopus.

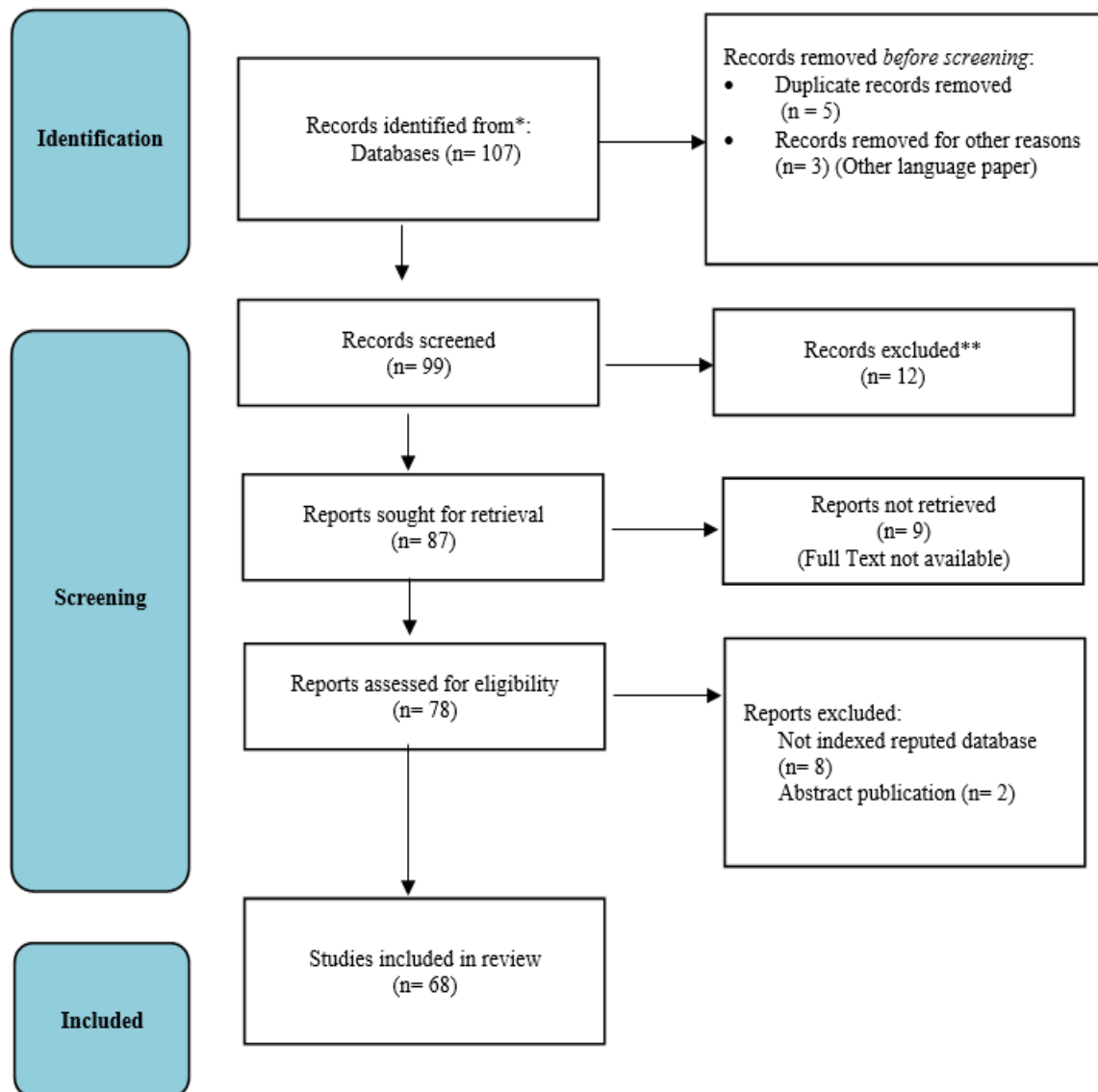


Figure.2.1 Prisma representation of the study selection

2.2. Specular Reflection Detection using Color Space Method

Specular reflection refers to the mirror-like reflection of light off a smooth and shiny surface. Unlike diffuse reflection, which scatters light in different directions, specular reflection occurs at specific angles and preserves the color and intensity of the incident light. Specular reflections are common in images taken in the presence of glossy or reflective surfaces like glass, water, or polished metals. The goal is to distinguish these reflections from other elements in the image. The specular reflection detection relies on color space transformations or analysis. Color spaces are mathematical models that represent colors in various ways, such as RGB (Red-Green-Blue), HSV (Hue-Saturation-Value), or LAB color

spaces. Analyzing color information can be valuable in identifying specular reflections because these reflections often exhibit distinct color characteristics.

Akbari et al. (2018) proposed a method for the prediction of specular reflection in smart colonoscopy images using the HSV color space. The analysis suggests that the S (saturation) and V (value) channels exhibit higher saturation properties, making them suitable for detecting specular reflections in colonoscopy video frames. By extracting and aggregating the HSV channels, the generate a cost function that aids in detecting reflection regions using a threshold method. To automatically identify reflection regions in the video frames extracted from colonoscopy images, the authors employ a Support Vector Machine (SVM) model, trained on the extracted features. It allows the system to predict glare regions on the images with an impressive accuracy of 99.68%. However, despite its high accuracy; the proposed method encounters limitations when dealing with large intensity values. To evaluate their method's performance, the authors utilized the CVC-ColonDB database, adapting the approach to identify reflection regions based on intensity values. While the results are promising and demonstrate the effectiveness of their approach for most cases, the challenges with large intensity values warrant further investigation and refinement to enhance the method's robustness in such scenarios [67].

Nie et al. (2023) proposed a method for detecting reflections in endoscopy images using RGB and HSV color space-based threshold techniques. To ensure consistency in the analysis, the contrast of the images was initially equalized, as the dataset contained images with varying contrast levels. The authors employed an adaptive thresholding approach on the Green (G) and Blue (B) color channels to calculate the threshold values and extract high-intensity regions in the images. Subsequently, these threshold values were utilized to identify glare regions, which are indicative of specular reflections. Based on their analysis, images with high brightness were assigned a threshold value of 0.3, images with medium brightness had a threshold value of 0.8, and images with low brightness were given a threshold value of 1.1 for the detection of specular reflections in endoscopy images. After successfully identifying the reflections, the authors applied an exemplar-based inpainting method to remove the reflections, thereby improving the overall image quality. This reflection removal process aimed to make the endoscopy images more suitable for various medical applications, such as analysis and diagnosis [68].

Yu et al. (2023) focuses on identifying large intensity-based reflection regions in endoscopy images. The RGB color distribution and specifically the ratio of the red channel to both the green and blue channels to characterize the features of the reflection on the images. The threshold method is applied to distinguish the highlighted pixels based on the calculated red-to-green and red-to-blue ratios. This thresholding process allows them to identify regions with intense highlights in the images. To identify small intensity values, the mean distance of the green and blue channels is computed. This step aids in predicting glare regions, further enhancing the accuracy of the model. The adaptive thresholding technique proposed in the method takes advantage of the difference between the red channel and the other channels. By integrating this technique with overlapped window are able to effectively handle large highlight regions with high intensity. As a result of their approach, the model achieves a high accuracy of 99.60% in predicting glare regions in endoscopy images. This indicates that the method is robust and efficient in dealing with large intensity values and effectively identifying specular reflections on the images [69].

Wang et al. (2019) The colposcopy plays an important role in the early stage of cervical lesion identification. Due to the physiological mucus in the tissue body the glare region appears as the bright pixel on the images captured through the colposcopy device. The representation of the highlighted area in the images has high certain similarity with the abnormal metaplasia epithelium (i.e., acetowhite or lesion region). During the medical image processing, like segmentation of lesion region and grading process it is difficult to differentiate because of its similarity properties. So, in this paper threshold-based segmentation is applied on the color images by adopting the nonlinear filtering method to normalize and predict the specular reflection accurately. The color balancing ratio and window size is along with the median filter is computed to detect some minute specular pixel points from the colposcopy images where the specular region is too weak [70].

Asif et al. (2021) Endoscopy is also the digital images which are captured to observe the internal structure of the human body. Specular reflection is mostly a result of the bright areas and intense light that appear on endoscopic pictures, which has an impact on the success of minimally invasive surgery. So the author proposed a method for the automatic identification of specular reflection using the intrinsic image layer separation (IILS) method. The method involves the process like image normalization followed by the extraction of high gradient area which helps in the separation of the specular reflection using the color

model. After predict the patch-based approach is used for the construction of reflection region to improve the quality of the endoscopy images. The proposed method acts as preprocessing method for the further analysis of the endoscopic images. Although the proposed model predicts the specular reflection with the accuracy of 99%, but it fails to predict the when the specular reflection is larger region on the digital images [71].

Kudava et al (2017) proposed the algorithm for the prediction of glare region using the standard deviation filters. It acts as the preprocessing method after identification the curvilinear structure enhancement is applied on the cervical image for the segmentation of the cervix regions. To identify the glare region, the color channel which has the property of high intensity value of the color space model is extracted. These color channels are computed and then the standard deviation filter is applied on the computed channels to predict the glare region on the colposcopy images. The algorithm is applied on each stage of the cervix image i.e., on the graded the glare region predicted are manually calculated and also the algorithm is tested at different illuminations using the GIMP 2 software. Based on the comparison analysis of the proposed method, the glare region is predicted with the sensitivity rate of 0.809. The preprocessed image gives the better segmentation image than the unprocessed images [72].

Yue et al., (2021) presented a new computer-aided diagnosis algorithm in this paper, which addresses the challenges faced by existing algorithms in segmenting acetowhite lesions in colposcopy images. The current methods struggle due to issues such as specular reflections, limited training data, and the inability to focus on semantically relevant lesion regions. The proposed algorithm introduces a novel approach to automatically segment acetowhite lesions in cervigrams. To overcome the negative impact of specular reflections on segmentation performance, incorporate a specular reflection removal mechanism. This mechanism effectively identifies and inpaints these areas with high precision, reducing their interference in the segmentation process. To evaluate the performance of the proposed approaches, conducted comprehensive experiments on a dataset of 3,045 clinical cervigrams. The results demonstrate that this method surpasses state-of-the-art techniques in acetowhite lesion segmentation. Specifically, this algorithm achieves superior Dice similarity coefficient and Hausdorff Distance values, indicating its improved accuracy and precision in identifying and delineating acetowhite lesions [73].

Gao et al (2017) proposed a method for removing the specular reflections on endoscopic images to improve diagnostic accuracy. The method separates the highlights based on supervised learning and restores the images by propagating the regions with similar structure features to specular regions. The multiscale dynamic image expansion and fusion based method is applied to restore the highlighted region on the endoscopic imaging. The specular reflection separation method is based on the statistical threshold method along with the SVM learning algorithm to automate the reflection separation process. To refill the empty region the candidate block are selected to fill the empty region with the similar region. The model is trained with the 55 images which is having the specular spots which are collected from the ColonDB which have the annotated video sequence of the colonoscopy video. The proposed method measures the predicted inpainted image using the COV value of 0.3168 [74].

Li et al. (2015) proposed an Adaptive-RPCA approach aimed at effectively removing highlights by optimizing the distance between the identified highlight image and the sparse result image. It consists of three key components: an adaptive threshold algorithm for detecting specular reflections, a robust similarity measurement algorithm based on grid feature extraction from binary images, and an improved RPCA-based low-rank + sparse decomposition method. This approach allows for adaptive optimization of the sparse decomposition parameters, making it applicable to various endoscopic minimally invasive surgery (MIS) procedures. By employing a low-rank representation to describe highlight regions as a stable feature, the method achieves superior highlight removal results compared to existing techniques. The distance metric algorithm employed in this method relies on grid feature extraction from binary images. Both the detected highlight image and the sparse result image are divided into rectangular grids, and the count of highlight pixels is computed for each grid. Subsequently, the similarity between the two feature vectors is measured using the Euclidean distance. A total distance criterion is designed based on the similarity ratio between the two images, with a smaller distance indicating a better highlight removal outcome [75].

Wang et al (2021) introduced a technique aimed at enhancing the accuracy of cervical cancer diagnosis through colposcopic images. Their approach focused on identifying and removing reflection regions present in the images, which was crucial for both diagnostic improvement and image quality enhancement. However, the initial method encountered

challenges related to visual visibility and the preservation of fine image details. To address these challenges, the researchers proposed a novel approach. Initially, they tackled the issue of identifying glare or reflection regions. This was achieved by converting the RGB image into the HSV color space. From the enhanced images, extracted the chromacity value, which helped pinpoint the regions affected by glare. The comparison between pixel luminance and color luminance played a vital role in detecting glare regions, where pixel luminance values exceeding a certain threshold indicated the presence of reflection areas. Once the glare regions were successfully detected, applied an exemplar-based inpainting method to eliminate the glare regions from the colposcopic images. This technique involved utilizing known information from the surrounding areas to fill in the glare-affected regions, effectively improving the overall quality of the images [76].

Oak et al (2020) focus on identification of reflection region on the medical image analysis, which can negatively impact diagnostic accuracy. It compared the detection techniques and proposes the combination of the ‘Alsaleh’ detection method with the Kittler auto selection techniques as the most appropriate for the identification of glare region on the cervix image. This paper introduced the two main steps: Scrambling the plain image using Arnolds transformation and encrypting the selected scramble images using a keystream generated by the word oriented snow steam cipher. The Kittler method exhibits a maximum deviation of 0.09 and 0.14 for threshold values in the I and S images respectively. The paper suggests a combination of Kittler's automatic threshold selection technique and the Alsaleh method for detecting specular reflections (SR). This combined approach yields a high sensitivity of approximately 0.9, indicating its effectiveness in identifying SRs within the medical images [77].

Shen et al. (2020) introduced a technique for suppressing and removing reflections from endoscopic images, aiming to mitigate their adverse effects. The strategy involves a three-step process encompassing reflection detection, reflection region classification, and reflection concealment. To identify regions affected by glare in endoscopic images, the authors employed a combination of the threshold method and morphological dilation operations. The algorithm for classifying reflection-affected regions was developed based on local image content, allowing for parameter adjustments within the proposed image inpainting algorithms. Experimental outcomes demonstrated that the proposed scheme

achieved higher Structural Similarity Index (SSIM) values compared to existing methods. Notably, the reflection suppression scheme outperformed established methodologies [78].

Das et al. (2017) introduces a comprehensive approach to detecting and addressing reflection region. This is achieved through a sequence of methods, including top-hat filtering, global thresholding, diverse morphological operations, and a novel filling algorithm, all aimed at identifying and filling SR areas. A key focus is to discern the areas of interest (AW) which contain relevant information for detecting various cervix tissues. The non-region of interest (ROI) portions, which hold unnecessary information that can hinder tissue detection, are omitted. The ROI extraction involves the utilization of hue (H) and value (V) parameters from the HSV color space. This enables the isolation of the cervix boundary within the colposcopy image. The authors' algorithm for processing the ROI is then evaluated against expert markings, and statistical metrics are computed for comparison. This innovative approach represents a promising step forward in humanitarian technology, particularly in the context of detecting cervical cancer. By effectively addressing Specular Reflections and enhancing the accuracy of cervix image analysis, this ongoing research holds the potential to contribute significantly to early cervical cancer detection and diagnosis [79].

Bidokh et al. (2017) delve into the domain of Wireless Capsule Endoscopy (WCE), a non-invasive medical imaging tool employed by gastroenterologists to explore gastrointestinal tract disorders. Within the realm of WCE images, a common challenge arises in the form of specular reflections (SRs). These reflections stem from intense light and luminous areas within WCE images, which can significantly impact both abnormality detection methods and the analysis conducted by physicians. The core objective of the present research is to introduce an automated technique capable of segmenting and eliminating SRs in WCE images, thereby enhancing the precision of abnormality detection methods and aiding physicians in accurate diagnoses. The study introduces a novel approach for SR elimination while preserving image texture using a homomorphic filter. The proposed method involves three key steps. Initially, a robust segmentation technique leveraging the U-Net model is applied to semantically segment SRs. Following this, the homomorphic filter is employed to disentangle the illumination and reflection components present in the image. Since SRs affect these components differently, the proposed method enhances them individually, rather than uniformly enhancing the entire image. Finally, the Navier-Stokes inpainting

technique based on fluid dynamics is utilized to reconstruct SR regions. Experimental evaluations were conducted across various WCE datasets, encompassing three distinct types [80].

2.3. Deep Learning Segmentation Model using the Binary Masking

Deep learning is a subset of machine learning that focuses on using artificial neural networks to model and solve complex tasks. It has been particularly successful in various computer vision tasks, including image segmentation. A segmentation model is a type of neural network architecture designed to partition an input image into multiple segments or regions based on certain characteristics. These segments are typically associated with objects or areas of interest within the image. Binary masking is a technique used in image processing and computer vision. It involves creating a binary mask, which is essentially a binary image where pixels are labeled as either foreground (usually representing the object of interest) or background (everything else). In the context of image segmentation, binary masking can be used to separate objects from their backgrounds. This could mean that the model learns to generate binary masks that highlight the regions of interest within input images, a common approach in semantic segmentation tasks.

Jha et al (2020) employed pixel-wise labelling for precise segmentation in their study. The U-Net model is the major choice for medical image segmentation, known for its encoder-decoder architecture. Introducing the Double U-Net, the innovative approach involves stacking two U-Net structures. The initial U-Net utilizes a pre-trained VGG-19 as an encoder, gaining insights from ImageNet's features. Subsequently, another U-Net captures semantic details effectively, enhancing segmentation. This methodology was applied across diverse medical image datasets. Through comparative analysis, it became evident that the Double U-Net model surpassed the performance of baseline models. The promising outcomes achieved across a range of medical image segmentation datasets underscore the potential of the Double U-Net as a robust foundation. Its efficacy extends to serving as a solid baseline for both medical image segmentation tasks and the assessment of Deep Learning (DL) models' generalizability through cross-dataset evaluations [81].

Zhou et al. (2018) introduced a robust framework named UNet++, designed specifically for advanced medical image segmentation. This architecture encompasses a deeply supervised encoder-decoder network, intricately linking the encoder and decoder sub-networks through a sequence of nested, dense skip pathways. The objective of these skip connections is to

bridge the semantic gap between the feature maps of the encoder and decoder sub-networks. When the feature maps from the decoder and encoder networks exhibit semantic similarity, an optimizer is employed for the learning process. UNet++ was subjected to a comparative analysis against the UNet and Wide UNet architectures in diverse medical image segmentation tasks. These tasks encompassed nodule segmentation in low-dose CT scans of the chest, nuclei segmentation in microscopy images, liver segmentation in abdominal CT scans, and polyp segmentation in colonoscopy videos. The experimental outcomes were compelling, showcasing that the UNet++ model achieved an average Intersection over Union (IoU) gain of 3.9 compared to UNet, and 3.4 compared to Wider UNet. This emphasizes UNet++'s superiority in medical image segmentation applications [82].

Cao et al (2022) proposed Swin-Unet model, aiming to enhance medical image segmentation. This approach incorporates a pure Transformer framework, resembling the structure of the Unet architecture. The process involves dividing the image into tokenized patches, which are then processed through a U-shaped Encoder-Decoder configuration, featuring skip-connections for holistic local-global semantic understanding. The encoder leverages a hierarchical Swin Transformer with shifted windows to capture contextual information effectively. On the other hand, the decoder employs a symmetric Swin Transformer-based design with an added patch expanding layer, facilitating up-sampling and the restoration of spatial detail in the resultant feature maps. The input and output are directly down-sampled and up-sampled by a factor of 4x4, respectively. Through comprehensive experiments focused on multi-organ and cardiac segmentation tasks, the Transformer-based U-shaped Encoder-Decoder network of Swin-Unet showcases superior performance compared to methods reliant on full-convolution techniques or a hybrid approach of transformers and convolutions [83].

Zhang et al (2021) addressed the crucial role of the thyroid gland in producing essential hormones that facilitate the proper functioning of all body cells. To facilitate early-stage diagnosis and treatment of thyroid cancer, the focus was on accurately characterizing thyroid nodules. For this purpose, introduced two models called cascade UNet and CH-UNet, designed for segmenting thyroid nodules and distinguishing between benign and malignant nodules. The cascade UNet approach comprises two components: UNet-I and UNet-II. UNet-I operates at a consistent resolution, while UNet-II operates at the original image resolution, both serving the purpose of segmenting nodules. CH-UNet, on the other

hand, capitalizes on segmentation as an auxiliary task to enhance the accuracy of nodule classification. The effectiveness of the proposed method was validated using the TNSCUI 2020 Challenge's test dataset. Remarkably, their approach achieved an Intersection over Union (IoU) of 81.73% in the segmentation task and an F1 score of 0.8551 in the classification task. This achievement secured the top position in the classification track and fell only 0.81% short of the leading score in the segmentation track [84].

Zhou et al (2021) tackled the challenge posed by the computationally intensive nature of 3D convolutional neural networks (CNNs) in the context of medical image segmentation. While the encoder-decoder structure had displayed promise, its resource demands remained a concern. To address this, the researchers introduced the dimension fusion UNet (DUNet) model, a novel approach that ingeniously merges both 2D and 3D convolutions within the encoding phase. Through this innovation, the DUNet model achieves superior segmentation performance compared to solely 2D networks, all the while exhibiting reduced computation time when compared to full-fledged 3D networks. The method involves the incorporation of a weighted focal coefficient and the combination of two traditional loss functions, enhancing its functionality. The efficacy of this proposed approach was evaluated using the ATLAS dataset and benchmarked against three state-of-the-art methods. The results prominently highlight the success of the method, as it achieved the highest quality performance based on metrics such as Dice Similarity Coefficient (DSC) with a gain of $0.5349 + 0.2763$ and precision with an increase of $0.6331 + 0.295$, as compared to the alternatives [85]

Park et al (2022) emphasized the significance of precise cervix region segmentation for augmenting the effectiveness of computer-aided diagnostics through colposcopy, particularly within the realm of cervical intraepithelial neoplasia (CIN) classification. This work introduced innovative enhancements, including a novel loss function called CT loss, which combines cross-entropy and total-variation loss. It also devised a creative scheduling strategy tailored to the baseline W-Net architecture, collectively contributing to the realization of fully unsupervised cervix region segmentation in colposcopy images. The effectiveness of their approach was validated using the Kaggle dataset. The experimental outcomes underscore the method's capability to consistently derive segmentation masks for the region of interest (ROI) in colposcopy, even when working with a limited training dataset. By incorporating the proposed CT loss and EW (early and weight) learning

techniques, the segmentation performance exhibited a substantial improvement, with the Dice coefficient increasing from 0.6120 to 0.7100, when compared against the baseline graphcut W-Net and CNN-based methods. Moreover, the optimized model training time was substantially reduced from 10.6 hours to 3.4 hours [86].

Huang et al (2020) addressed the limitations of the UNet, a prevalent deep learning network with an encoder-decoder architecture, in the domain of medical image segmentation. While the UNet model is commonly employed, it falls short in harnessing comprehensive information across different scales, leaving substantial room for enhancement. To tackle this, the authors introduced a novel variant known as UNet 3+. The UNet 3+ capitalizes on the power of full-scale skip connections and deep supervisions. These full-scale skip connections seamlessly combine detailed low-level elements and high-level semantic features from diverse scale feature maps. Furthermore, integrated a classification-guided module and a hybrid loss function into the framework. An added advantage of UNet 3+ is its ability to streamline network parameters, leading to improved computational efficiency. Rigorous experimentation using liver and spleen datasets underscored the remarkable superiority of UNet 3+ over all prior state-of-the-art methodologies. The method not only enhances the delineation of organs but also yields coherent boundaries, further solidifying its effectiveness in the field of medical image segmentation [87].

Niu et al (2022) introduced an innovative approach featuring a nested Unet model that incorporates an EfficientNet backbone. This model strategically incorporates multiple skip pathway connections across various layers to mitigate the semantic gap between networks. Pneumothorax is a condition where air accumulates between the lung and chest wall, can arise from factors like chest injuries, underlying lung disease, or even Covid-19 infection. In this context, utilizing CT scans to screen high-risk individuals represents a crucial task undertaken by medical professionals and institutions. Leveraging the advancements in machine learning, computer-aided diagnosis has become increasingly prevalent in pneumothorax detection. Based on the experimentation analysis, the proposed model demonstrated superior performance when compared to both the basic Unet model and other competing models. The results showcased the model's effectiveness, highlighted by its ability to achieve enhanced performance metrics and notably lower Dice loss. This underscores the potential of their nested Unet model with an EfficientNet backbone in

elevating pneumothorax detection and aiding medical practitioners in providing accurate diagnoses [88].

Shinohara et al. (2023) proposed the U-Net deep learning network to effectively segment lesions. The key innovation lies in processing colposcopic images captured before applying the acetic acid solution alongside those taken after its application. By using these combined image sets, particularly those taken before acetic acid solution application, the model gains the ability to more accurately segment CIN lesions. The methodology's effectiveness was validated through experiments involving 30 real colposcopic images showcasing acetowhite epithelium, a representative CIN type. The results highlighted notable improvements in accuracy, precision, and F1 scores. Specifically, when employing both pre- and post-acetic acid solution application images, the accuracy, precision, and F1 scores reached 0.894, 0.837, and 0.834, respectively. This outperformed the results achieved using only post-acetic acid solution images, which were 0.882, 0.823, and 0.823, respectively [89].

Yu et al (2022) introduced an enhanced version of the Faster Region-Convolutional Neural Network (R-CNN) to accurately extract the cervical region from colposcopic images, eliminating interference from surrounding tissues and instruments. Subsequently, proposed the CLS-Net, a deep convolutional neural network that leveraged EfficientNet-B3 to extract features specific to the cervical region. The CLS-Net also employed a redesigned atrous spatial pyramid pooling (ASPP) module, tailored to the lesion region's size and feature map after subsampling, enabling the capture of multiscale features. Cross-layer feature fusion was employed to achieve precise segmentation of the lesion region, and the segmentation outcome was mapped back to the original image. Through rigorous experimentation involving 5455 colposcopic post-acetic-acid images related to low-grade squamous intraepithelial lesions (LSIL+), which encompass cervical intraepithelial neoplasia and cervical cancer, the proposed model exhibited impressive performance metrics. Specifically, the accuracy, specificity, sensitivity, and Dice coefficient of the model were 93.04%, 96.00%, 74.78%, and 73.71%, respectively. The CLS-Model introduced in this study excelled in segmenting cervical lesions within colposcopic post-acetic-acid images, making it a valuable tool for assisting colposcopists in enhancing their diagnostic capabilities [90].

2.4. Deep Learning inpainting model for missing region

In the realm of computer vision and image processing, the development of advanced techniques for image restoration and enhancement is of paramount importance. One critical facet of this endeavor is the task of "inpainting," which involves the reconstruction of missing or corrupted regions within images. With the exponential growth of deep learning, particularly the advent of neural networks with multiple layers, a new era of inpainting capabilities has emerged. This progress is poised to revolutionize various fields, from art restoration to medical imaging. In this context, this paper explores the concept of a "Deep Learning Inpainting Model for Missing Regions," delving into the mechanisms and applications of this innovative approach. Through the integration of neural networks and advanced algorithms, such models have the potential to significantly enhance image quality, filling voids with contextually accurate information and opening doors to a myriad of practical and creative possibilities.

Liu et al., (2018) introduced a novel approach called the partial convolutional inpainting algorithm (P.Conv) to address the challenge of generating accurate masks for missing regions. The innovation behind P.Conv lies in its ability to yield higher-quality images in a single forward pass, incorporating skip links to predict and fill in the absent pixels. The proposed method seamlessly combines the concepts of partial convolutions and binary mask updates to effectively perform inpainting on digital images. This operation entails applying a convolutional filter W to binary-masked M images (images with missing pixels), utilizing feature value x along with corresponding bias b values. The element-wise multiplication comes into play to "unmask" the input images, iteratively refining the images by updating the binary mask after each partial operation. The objective is to convert all pixels to valid ones in the digital images. For training, the partial convolutional inpainting process is fine-tuned with a learning rate of 0.00005, employing techniques such as batch normalization and the Adam optimizer. This approach not only contributes to more accurate mask generation for missing regions but also offers a streamlined and efficient process for enhancing image quality [91].

Patel et al., (2021) harnessed the effectiveness of a robust partial convolutional deep learning algorithm to address the challenge of completing the empty portions within digital images. To facilitate training, they introduced a modified U-Net model that featured seven encoders and a decoder. In this proposed architecture, ReLu activation functions were

utilized in each encoding layer, while Leaky ReLU was adopted for the decoding process except for the initial layer. The image dimensions were determined by the i -th encoder, with automatic adjustments made to the image size in conjunction with the kernel size. The effectiveness of their approach was evaluated across two distinct applications: automated segmentation for object removal and manual mask generation, as well as the inpainting of the removed pixels. This method focused keenly on structural features such as colors for image inpainting, emphasizing that pixel-per-information inpainting yields superior efficiency in image reconstruction when compared to semantic inpainting [92].

Mohite et al., (2020) presented a method that combines contextual attention with partial convolutional techniques to address the task of filling in missing image portions. By leveraging contextual attention, the approach focuses on pixel-wise inpainting, effectively utilizing the surrounding content within digital images. In this method, the conventional P.conv algorithm is substituted with stochastic gradient descent, which aids in iteratively updating the network's weights based on the training data. During the training process, masked images are generated, and the model is trained using original images. A notable advantage of this proposed technique is that it obviates the need for post-processing steps to eliminate noise in image inpainting. The contextual attention mechanism allows for a deep understanding of available pixel information, albeit at the cost of increased training time [93].

Kaur et al., (2020) introduced a deep learning methodology for inpainting manuscript images, employing a combination of partial convolutional and U-Net architecture. The training of these models follows two distinct phases. Initially, the models are trained with batch normalization, utilizing a learning rate of 0.0002 over a span of 120 epochs. Subsequently, batch normalization during the encoding stage is deactivated, and the models are trained with a learning rate of 0.00005 for a total of 80 epochs. To evaluate the quality of inpainted images, the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) were employed. The results indicated that the optimal quality was achieved with a masking size of 16x16 for PSNR and 4x512 for SSIM [94].

Chen et al., (2019) introduced a novel approach for inpainting Dunhuang mural images, incorporating partial convolutional techniques alongside a sliding window strategy. The sliding window operates as an augmentation process, effectively reducing the image patch size. This augmentation not only aids in maintaining a manageable computation time for the

model but also assists in improving its efficiency. Through the proposed method, missing pixels are effectively restored with commendable accuracy, while also contributing to a reduction in processing time. However, it's important to note that in some instances, certain missing regions might exhibit blurriness along with visible traces due to the inpainting process [95].

Yan et al., (2022) introduced the Partial Convolution Attention Mechanism (PCNet) as a solution for digital image inpainting. The encoder layer is constructed with a combination of convolutional and down-sampling layers. Generators and discriminators are then applied to these encoder stages. The decoders, meanwhile, employ deconvolutional and up-sampling layers to extract features from the digital images. A key aspect of PCNet is the attention mechanism, which is employed to capture long-range pixel information, making it particularly effective for addressing large holes in digital images. This mechanism replaces batch normalization, serving to normalize pixel features across the digital images. The convolutional kernel accelerates the learning of feature distribution, contributing to the effective inpainting of pixels in large holes within images with extensive scenery or broad areas [96].

Kang et al., (2020) introduced a novel technique called weighted convolution (WConv) aimed at addressing the instability imbalances among pixels and normalization challenges encountered in relation to invalid pixels. This innovative approach involves a weighting scheme that serves to normalize layers and balance the influence of invalid pixels, which may arise due to holes and zero-padding. To evaluate the weighted convolutional inpainting ratio of the mask, the convolution operation incorporates equation (4) to iteratively update the binary mask. Furthermore, the updated mask undergoes a normalization process, effectively enhancing the stability of dynamic holes within the images while simultaneously reducing complexity. This approach is particularly useful within the context of a partial convolutional neural network designed for the purpose of inpainting digital images. Kang et al.'s method not only contributes to improving the stability of pixel imbalances but also enhances the effectiveness of the inpainting process for images, making it a valuable advancement in the field [97].

Cai et al., (2017) introduced a blind inpainting method named Blind Inpainting Convolutional Neural Network (BICNN), designed to effectively remove invalid pixels from images. This approach utilizes a fully convolutional neural network and incorporates

stochastic gradient descent with standard backpropagation for training. The BICNN is trained to predict missing pixels, ultimately generating a clean output through a single forward pass. The BICNN architecture comprises multiple convolutional layers and intentionally avoids pooling or fully connected layers. This choice contributes to the network's ability to perform inpainting efficiently without unnecessary complexity. This method excels in generating higher-quality inpainting results for small, scattered regions within images. However, when the size of missing pixels surpasses that of the sub-images, the inpainted pixels tend to become blurred, accompanied by a notable increase in computation cost [98].

Liu et al., (2019) introduced an efficient blind inpainting approach employing deep convolutional neural networks, featuring encoder and decoder architecture to extract essential features from digital images. This method utilizes image gradient calculations to restore missing information within the images. A robust loss function is incorporated to accurately fill the outline of the missing details, as outlined in the corresponding equation. The proposed network consists of an impressive forty layers, delivering enhanced depth for capturing intricate image features. This architecture underwent rigorous testing across various datasets, including SET5, Set14, Urban, and BSDS500, achieving the desired outcomes. Notably, this method boasts an improved average runtime, with computations taking just 0.03 seconds. However, it's worth mentioning that the method faces limitations in scenarios where crucial structures or intricate details are concerned. For instance, when addressing human facial images, important features like the nose and mouth can become distorted or corrupted during the inpainting process [99].

Wang et al., (2020) introduced a novel approach for blind image inpainting, aimed at predicting patterns in unknown missing regions. Their method, called the Visual Consistency Network (VCN), operates in two distinct stages. The first stage involves estimating the region to be filled through the Mask Prediction Network, while the second stage focuses on determining what to fill into the missing region using the Robust Inpainting Network. This method showcases better performance compared to existing techniques and is particularly well-suited for addressing small holes. It offers the added benefit of automating the identification of masking and refilling processes. However, it exhibits limitations when dealing with larger holes. In cases where the dominant part within the missing region cannot be determined effectively, the method's performance significantly

deteriorates. This shortcoming is particularly evident when faced with common occlusions in images [100].

Schmalfluss et al., (2023) introduced an innovative enhancement to the blind inpainting approach by incorporating sparse filters. Their method not only improved the performance of the conventional CNN model but also expedited the network design process. This improvement was particularly notable in terms of network coverage. To achieve this, they employed the Structured Receptive Field Network (SRFN) filter in each CNN filter during the training phase. This strategic choice efficiently captured salient features such as edges and corners within digital images, enhancing the inpainting process. The model was trained on a dataset comprising 225,100 images, each sized 256x256. Remarkably, the proposed method achieved higher-quality inpainting results even with a reduced amount of training data. Furthermore, it significantly cut down computation time to just 2.453 milliseconds. It's important to note that while the proposed method yielded higher-quality inpainting results, there was a decrease in the Peak Signal-to-Noise Ratio (PSNR) value as the pixel loss in the digital image ranged from 20% to 25%. This implies that as pixel loss increases, the quality of the image declines [101].

Yu et al., (2019) introduced an innovative image inpainting method that harnesses gated convolution (G.Conv) to learn images without the need for explicit labeling. Unlike vanilla convolutional methods, which treat all empty pixels as valid pixels, this proposed approach is tailored for free-form masked images and effectively learns the soft mask automatically. The sigmoid activation function and another activation function are represented. The convolution filters W_g and W_f are distinct and serve to inpaint the missing portions. This method dynamically learns features for each channel of input images. Gated convolutional techniques are adept at highlighting the masked regions and capturing the sketched information across each channel, which contributes to generating realistic inpainting outcomes. This novel approach excels when compared to existing methods, generating results that exhibit seamless boundary transitions and a high level of realism [102].

Chai et al., (2020) introduced the Spatial Temporal-based Gated Convolution Network (STGCN) for cloud removal utilizing a deep learning approach. This method aims to restore invalid pixels, which correspond to removed portions, by leveraging auxiliary information from multi-temporal images. The process involves gated convolution layers and sub-pixel convolutions to effectively differentiate between cloud and non-cloud areas within images.

This differentiation is used to create masks for inpainting and accurately replenishing pixels, ultimately producing non-cloud images. The ASPP (Atrous Spatial Pyramid Pooling) module, with varying dilation rates, is a crucial component that aids in capturing diverse features from receptive fields. These features are then combined through concatenation to generate the missing regions within the images. During training, the model employs a learning rate of 10^{-4} and iterates over 500 epochs. Notably, this approach outperforms existing methods in restoring irregular cloud masks. However, it's essential to acknowledge that the proposed method involves complexity and demands increased computational resources due to its advanced techniques [103].

Chang et al., (2019) introduced a novel approach involving a 3D gated convolutional neural network for inpainting free-form masks. They incorporated a temporal patch GAN loss to enhance temporal consistency in the generated outputs. One notable advantage of this method is its faster computation speed, achieved by reducing the number of parameters, leading to superior performance when compared to existing techniques. However, it's important to highlight that this method faces limitations in generating natural results, particularly when dealing with thick masks. Despite its strengths, this approach might struggle to produce satisfactory outcomes under certain challenging conditions [104].

Wei et al., (2022) introduced an innovative network utilizing four channels in conjunction with binary masked images, where the value '1' denotes pixels to be repaired within the digital images. The receptive field of the convolutional kernel within the upper layer, denoted as 'R', and 'k', represents the size of the convolutional kernel, while 's' signifies the convolutional step in the layers. In these equations, 'k' denotes the convolutional kernel size, along with the dilation rate 'd'. The method's performance was assessed using both human perception and objective data. Image inpainting was conducted across three groups, employing four different inpainting algorithms to evaluate the method's effectiveness. Particularly, the proposed technique was applied to restore images of the human face's five sensory organs using available image data. Notably, the proposed approach exhibited improved detail along the edges of inpainted regions, matching the attributes of the entire image. However, it's important to mention that this method is not applicable to images with free-form masking, which could potentially limit its versatility [105].

Yan et al., (2018) introduced the Shift_Net deep learning approach to address the issue of blurry images and enhance the generation of plausible results. The core innovation lies in

the integration of shift connections within the U-Net architecture. This allows for the completion of missing regions while preserving the sharp structures and intricate texture details of digital images. In this method, the encoder extracts feature from known regions of the pixels, and the decoder employs these features to estimate the missing regions within the images. The dataset employed for training consists of 5000 samples, with 900 for testing and 100 for validation purposes. During training, the model is optimized over 30 epochs with a learning rate of 2×10^{-4} . The proposed approach demonstrates the ability to restore fine details within the missing regions while ensuring speedy computational processing. However, it's essential to note that the method's performance in dealing with large holes within digital images is not explicitly stated, potentially indicating a limitation in handling such scenarios [106].

Zhou et al., (2021) presented an advanced variation of the Shift-Net architecture known as Bishift-Net, designed to enhance image inpainting. This architecture modifies the encoder and decoder components and strategically reorganizes the generated features to achieve sharp textures in digital images. The Bishift-Net model excels in generating fine details along the edges of the images. The model's training utilized a dataset of 14,900 images, each with a size of 256x256 pixels. Masking regions covering 20-40% of the center of the images were used. The training process involved 30 epochs, utilizing the Adam optimizer with a learning rate of 2×0.0001 and a beta value of 1. The model was trained using both center and random masks. It achieved PSNR values of 28.14 dB for center masks and 28.657 dB for random masks. Additionally, the Structural Similarity Index (SSI) for center masks was 0.937, and for random masks, it was 0.919. This method surpasses the original Shift-Net model in its ability to effectively inpaint digital images, particularly in generating sharp textures and details. However, it's important to note that the performance of this approach might not be explicitly stated for addressing large holes within digital images [107].

Hong et al., (2019) introduced a fusion block approach that employs a flexible alpha composition map to combine both known and unknown regions within digital images, ultimately generating comprehensive structural and textural information. This fusion block operates by initially deriving a raw completion image through feature maps. It then predicts the alpha composition map to combine these components and achieve the final outcome. The experiment involved applying masks to images, where different ranges of pixel removal were tested, such as [0-10%), [20-30%), [30-40%), and [40-50%). During model training, a

batch size of 6 was used for each GPU processor, and the entire process took three days to complete. The learning rate was gradually reduced from $2e^{-3}$ to $2e^{-6}$ over the course of 20 epochs. Remarkably, the proposed method exhibited superior performance compared to existing techniques across various masking ratios, ranging from 0% to 50%. The quality of image completion at boundary regions was notably higher in the digital images processed by this method. However, it's important to highlight that this model might not be suitable for addressing larger holes, as the information transmission to the inner regions of images might be limited [108].

Zeng et al., (2019) introduced the Pyramid-context Encoder Network (PEN-Net) as an approach for digital image inpainting. This network is constructed upon the U-Net architecture, effectively harnessing its ability to encode input image resolution and learn semantic features during the decoding phase. To facilitate model training and testing, images were resized to a dimension of 256x256. During the training process, digital images underwent masking, where square areas of sizes 32x32, 64x64, and 128x128 were masked. The model's performance was evaluated, and it demonstrated superior results when compared to existing methods, all while requiring minimal computational time. However, it's important to note that in some instances, the predicted images lacked the fine structural details of facial features. This suggests that there might be limitations in capturing intricate facial details using this method [109].

Salem et al., (2019) employed deep learning techniques using convolutional filters to analyze images, and mean values were employed to replace missing regions. The utilization of dilated convolutional layers and multiscale context aimed to effectively fill random missing regions in digital images without compromising their resolution. Traditionally, inpainting processes tend to be time-consuming. To address this, the feature parameter dilated method was employed to expedite the process of filling in missing portions. The model was trained using a dataset of 149,000 images selected from the Paris Street Review dataset. Out of these, 36,501 were earmarked for validation, and 328,501 were allocated for testing. The Adam optimizer was chosen for its efficiency, and the training process was carried out using NVIDIA GPU processing to reduce computational time and achieve faster results. While this method significantly reduced computational time and improved the quality of the filled regions by minimizing blurriness, there were instances where the

refilled regions were marked by fine lines. Additionally, the resolution of the filled pixels was observed to be lower compared to the original pixels [110].

Prabhu et al., (2019) employ a U-Shape architecture characterized by a feature encoder and a feature decoder. The feature encoder incorporates four convolutional layers, while the feature decoder aids in fusing pixel information. Enhancements in synthesis images are achieved through the utilization of multiscale dilated convolution, up-sampling, and down-sampling modules. The model architecture involves four convolutional layers with respective kernel sizes of 3x3, 1x1, 3x3, and 1x1, and channel counts of 128, 32, 32, and 128. This approach surpasses existing methods, improving performance while reducing computation time. However, some larger fingerprints exhibit slight blurring, and background regions appear smudged [111].

2.4.1 Deep Learning Inpainting Model for Medical images

Deep learning model utilizes multi-layered networks to discern and extract patterns from data, has gained considerable prominence across diverse domains, notably in the realm of medical image analysis. In this context, 'inpainting' refers to the automated process of reinstating or rectifying sections within medical images that are either incomplete, missing, or have suffered corruption. This task holds immense significance within the field of medical imaging, where the quality and fidelity of images are paramount for precise diagnosis and analysis. Medical images encompass a wide array of imaging modalities, ranging from X-rays, CT scans, and MRI scans to ultrasound images and digital colposcopy. Image quality is often compromised by various factors, and inpainting plays a pivotal role in eliminating noise and replacing missing regions with high-quality pixel data. This session delves into relevant research papers that explore the application of deep learning techniques for inpainting tasks in the context of medical images.

Pimkin et al., (2020) introduced a model designed specifically for the removal of distorted metal artifacts in sinograms. They adapted the partial convolutional inpainting method to enhance the quality of the reconstructed images. When compared to the original partial convolutional inpainting approach, this modified version yielded superior image quality improvements. The sinogram inpainting U-net, integrated with partial convolutions, was trained through 500 epochs while incorporating a learning rate multiplication factor of 0.5. Additionally, the learning rate underwent successive multiplications of 100, 200, and 300 at

epoch values 400, 450, 475, and 490 respectively. The results demonstrated the effectiveness of their proposed approach, surpassing existing methods with a remarkable 64% reduction in MAE. While the proposed method is designed to be applicable across multiple domains, the analysis and evaluation have thus far been confined exclusively to brain CT images [112].

Manjón et al., (2020) introduced a pioneering 3D blind inpainting approach aimed at cleansing the lesion-affected regions to create a healthier version of MRI brain images. This marks the first instance of a 3D blind inpainting technique employed within medical image inpainting. The method involves inpainting the lesion-affected images to generate a healthy variant of MRI images. This is achieved by masking the lesion regions within the multiple sclerosis (MS) images and then applying the inpainting procedure. The 3D-UNet architecture is harnessed for this purpose, leveraging the ReLU activation function and Batch Normalization during the model's training process. In terms of computational efficiency, the proposed method exhibits a processing time of approximately 0.5 seconds on a GPU and about 8 seconds on a CPU. However, it's important to note that there exists a high correlation loss for the tested images, suggesting potential challenges or limitations in the inpainting performance [113].

2.5. Image Classification for using Deep learning Classification Models

Classification models are algorithms or neural network architectures designed for the purpose of categorizing data into different classes. In this context, these models are specifically tailored for image classification tasks. So, in this cover the application of advanced deep learning techniques, particularly deep neural networks, for the task of automatically and accurately categorizing medical images into predefined classes.

Zhou et al., (2022) employed deep learning models to distinguish between five subtypes of intracranial hemorrhage and normal images in non-contrast-enhanced CT scans. Their study comprised 351 patients who underwent such scans, resulting in a total of 2768 images. The researchers chose ResNet-18 and DenseNet-121 deep learning models and divided the dataset into 80% for training, 10% for validation, and 10% for final evaluation. Two experienced radiologists' diagnostic reports guided the process. The evaluation metrics encompassed accuracy, sensitivity, specificity, and AUC values. Notably, the ResNet-18 model achieved an overall accuracy of 89.64%, whereas the DenseNet-121 model reached

82.5%. Both models exhibited sensitivity and specificity exceeding 0.80 for distinguishing between the five subtypes and normal images. However, when identifying specific subtypes like intraventricular hemorrhage and cerebral parenchymal hemorrhage, the sensitivity of the DenseNet-121 model fell below 0.80, specifically at 0.73 and 0.76, respectively. Both models demonstrated AUC values above 0.9 [114].

Kumar et al., (2022) brain tumor detection achieved enhanced accuracy while reducing error rates and computational complexity through the utilization of a Deep Convolutional Neural Network (DCNN) combined with a Transfer Learning model based on the Nature-inspired ResNet 152 architecture. The research proposed a strategy that involved CNN and Transfer Learning techniques for the purpose of detecting and categorizing brain images. To optimize the image quality, preprocessing techniques were applied to eliminate noise and enhance image quality using the Otsu binarization method. This hybrid model replaced the batch normalization layer of the Deep CNN with a ResNet 152 layer. This model was designed to classify images into normal, benign, and malignant categories. The weight parameters of the Hyb-DCNN-ResNet 152 TL model were fine-tuned using the Covid-19 optimization algorithm (CoV-19 OA). The proposed method demonstrated superior performance metrics, including accuracy rates of 99.57%, 97.28%, 94.31%, 95.48%, 96.38%, 98.41%, and 96.34%, as well as reduced error rates of 0.012(s), 0.014(s), 0.011(s), 1.052(s), 0.013(s), 0.016(s), and 0.015(s), when compared to existing methods such as BTC-Deep CNN-Dolphin-SCA, BTC-Deep CNN-WHHO, BTC-AFDNN-FLA, BTC-MLPNN-IWOA, BTC-ANN-PSO, BTC-RF-WSO, and BTC-WRF-ACO [115].

Huang et al. (2020) harnessed supervised convolutional neural networks (CNNs), particularly DenseNet, as a powerful and adaptable approach for medical image classification tasks. However, this approach demands substantial labeled data and involves intricate and time-intensive training procedures. Conversely, unsupervised CNNs, exemplified by the principal component analysis network (PCANet), sidestep the need for labeled data during training but compromise on achieving satisfactory classification accuracy. Addressing the challenge of precise medical image classification with limited training data, the researchers introduced a lightweight hybrid neural network. This hybrid model combines a modified PCANet with a simplified DenseNet. The modified PCANet incorporates two stages, generating effective feature maps through convolutions with diverse learned kernels. These feature maps serve as inputs to the subsequent simplified

DenseNet, which incorporates a reduced number of weights. The DenseNet employs dense shortcut connections to facilitate accurate medical image classification, leveraging the feature maps produced by the PCANet. Experimental findings underscore the viability of the proposed hybrid neural network. It demonstrates ease of training and surpasses the performance of prominent CNN models such as PCANet, ResNet, and DenseNet in terms of classification accuracy, sensitivity, and specificity [116].

Sarwinda et al., (2021) conducted a study focusing on employing deep learning techniques, specifically utilizing the ResNet architecture, for the classification of colorectal cancer in medical images. The remarkable success achieved by deep learning classification methods has encouraged researchers to apply them to the realm of medical image analysis. This investigation concentrated on training ResNet-18 and ResNet-50 models using images of colon glands. The primary objective was to train these models to effectively differentiate between benign and malignant forms of colorectal cancer. To validate the performance of their models, the researchers subjected them to three different testing scenarios involving subsets of the entire dataset (20%, 25%, and 40%). The empirical findings consistently indicated that the implementation of ResNet-50 yielded the most dependable outcomes in terms of accuracy, sensitivity, and specificity compared to ResNet-18 across all three testing scenarios. Across these three distinct testing scenarios, the study highlighted the most robust performance on the 20% and 25% test sets. In these cases, the models achieved a classification accuracy exceeding 80%, a sensitivity surpassing 87%, and a specificity greater than 83% [117].

Pattanaik et al., (2020) addressed the significance of detecting breast cancer at an early stage, given its association with abnormal cell changes in milk ducts that can lead to fatalities. To simplify and expedite the classification process, the researchers introduced an innovative approach—a hybrid model that combines DenseNet121 with an Extreme Learning Machine (ELM). This model aimed to classify breast cancer from mammogram images. The workflow involved preprocessing and data augmentation of mammogram images, followed by feature collection after the pooling and flatten layer in the initial classification stage. The gathered features were then input into the fully connected layer of the proposed DenseNet121-ELM model, wherein the traditional fully connected layer was replaced by the ELM model. The study utilized mammogram images from the Digital Database for Screening Mammography (DDSM) dataset. The proposed DenseNet121+ELM

model demonstrated promising results, achieving training and testing accuracies of 99.47% and 99.14% respectively, for a batch size of 128. Additionally, the model exhibited specificity and sensitivity values of 99.37% and 99.94%, while requiring 159.7731 minutes for computational time [118].

Chandran et al. (2021) highlight the limitations of traditional cervical cancer type classification, which heavily relies on the experience of pathologists and is associated with lower accuracy. The integration of medical image processing with convolutional neural network (CNN) models, a subset of deep learning, offers promising prospects for improving cervical cancer type classification. This study introduces two distinct CNN architectures for automated cervical cancer detection using colposcopy images: the VGG19 (TL) model and CYENET. The VGG19 (TL) model adopts transfer learning, specifically utilizing the VGG19 architecture for the task. Additionally, a novel Colposcopy Ensemble Network (CYENET) is devised to enable automatic classification of cervical cancers from colposcopy images. The performance of the models is evaluated in terms of accuracy, specificity, and sensitivity. The VGG19 model achieves a classification accuracy of 73.3%, with moderately satisfactory results. The kappa score of the VGG19 model indicates a moderate classification level. In contrast, the proposed CYENET demonstrates noteworthy outcomes, boasting high sensitivity (92.4%), specificity (96.2%), and a substantial kappa score (88%). The CYENET model exhibits a significantly improved classification accuracy of 92.3%, surpassing the VGG19 (TL) model by 19% [119].

Nayak et al., (2022) discuss the prevalence of brain tumors, which commonly affect both children and the elderly. This type of cancer stems from uncontrolled growth of brain cells within the skull and poses a significant health concern. The classification of tumor cells is particularly challenging due to their heterogeneous nature. Convolutional neural networks (CNNs) have emerged as the prominent machine learning algorithm for visual learning and the recognition of brain tumors. In their study, the researchers proposed a novel approach employing a CNN-based dense EfficientNet model alongside min-max normalization. The objective was to classify 3260 T1-weighted contrast-enhanced brain magnetic resonance images into four distinct categories: glioma, meningioma, pituitary, and no tumor. The developed network, a variant of EfficientNet, incorporated additional dense and drop-out layers. To further enhance the contrast of tumor cells, the authors combined data augmentation with min-max normalization. The key advantage of the dense CNN model lies

in its ability to effectively categorize a relatively small image database. Consequently, the proposed approach demonstrated exceptional overall performance. Experimental findings showcased a remarkable training accuracy of 99.97% and a testing accuracy of 98.78%. Notably, the model's high accuracy was accompanied by a favorable F1 score, indicating its potential as a valuable decision-making tool for brain tumor diagnostic assessments [120].

Phine et al., (2023) shed light on the critical nature of pneumonia, a potentially fatal ailment that predominantly affects the elderly population. Swift pneumonia diagnosis significantly contributes to saving lives. Leveraging the capabilities of deep learning in medical imaging, the paper endeavors to advance the detection and classification of pneumonia cases based on chest X-ray images. The study introduces a deep learning-based approach for the detection and classification of pneumonia patients using a carefully designed model. Specifically, the VGG-19 deep learning architecture is harnessed to address the challenge of pneumonia classification. To train the model, the researchers employ a dataset of chest X-ray images from Kaggle, which focuses on pneumonia cases. The implementation of the proposed system spans both balanced and unbalanced datasets of chest X-ray images. Remarkably, the developed system attains an accuracy of 86% when confronted with the unbalanced dataset, and this performance improves to 94% when dealing with the balanced dataset [121].

Shah et al., (2022) employed a deep convolutional neural network (CNN) based on the EfficientNet-B0 architecture, fine-tuned with custom layers to effectively classify and detect brain tumor images. The study integrated image enhancement techniques involving diverse filters to augment image quality. Data augmentation methods were also deployed to expand the dataset, thereby improving model training. The findings demonstrate that the proposed fine-tuned EfficientNet-B0 model outperforms alternative CNN models, achieving superior classification accuracy, precision, recall, and area under the curve values compared to state-of-the-art models. Notably, the proposed model achieved an overall classification and detection accuracy of 98.87%. The study also conducted a comparative analysis, evaluating other deep learning algorithms including VGG16, InceptionV3, Xception, ResNet50, and InceptionResNetV2 [122].

Munien et al. (2021) delved into the gravity of breast cancer as a fatal ailment, serving as a leading cause of female mortality worldwide. The conventional diagnostic process based on biopsy tissue analysis is intricate, time-intensive, and susceptible to human error. Moreover,

conflicts in final diagnoses can emerge due to discrepancies in observer interpretations. To address these challenges, computer-aided diagnosis systems have been developed and implemented. These systems play a pivotal role in enhancing diagnostic efficiency, accuracy, and cost-effectiveness. However, there remains room for improvement to bolster the reliability of their determinations. This study explored the suitability of the EfficientNet architecture for classifying hematoxylin and eosin-stained breast cancer histology images obtained from the ICIAR2018 dataset. Specifically, seven variants of the EfficientNet model underwent fine-tuning and assessment. Their performance was evaluated in categorizing images into four classes: normal, benign, in situ carcinoma, and invasive carcinoma. Additionally, the study examined the influence of two standard stain normalization techniques—Reinhard and Macenko—on model performance. The results unveiled that the EfficientNet-B2 model, fine-tuned with the Reinhard stain normalization method, achieved an impressive accuracy and sensitivity of 98.33% on training images. Meanwhile, utilizing the Macenko stain normalization method led to an accuracy and sensitivity of 96.67%. These favorable outcomes underscore the feasibility of transferring generic features from natural images to medical images through fine-tuning EfficientNet models, ultimately attaining satisfactory diagnostic results [123].

Zhaputri et al. (2021) shed light on the burgeoning interest in deep learning, a prominent machine learning approach that has garnered substantial attention from researchers in recent times. This methodology has proven to be particularly successful in addressing intricate challenges within the realm of medical image analysis. For instance, brain tumor afflictions are grave conditions associated with limited life expectancy post-diagnosis. Expert diagnosis is both time-consuming and heavily reliant on the expertise of the practitioner. This underscores the necessity for an automated classification system that can aid experts in decision-making processes. This has resulted in the development of numerous cutting-edge CNN models. Among these, the EfficientNet architecture stands out for its capacity to deliver high accuracy while demanding fewer computational resources. In line with this, the research proposes the adoption of the EfficientNet architecture for categorizing different types of brain tumors, including glioma, meningioma, and pituitary tumors. The EfficientNet architecture encompasses eight tiers, ranging from EfficientNet-B0 to EfficientNet-B7. The study reveals that the most favorable outcomes were attained with EfficientNet-B1 and EfficientNet-B2, achieving a commendable accuracy rate of 96% [124].

Alyafeai et al (2020) highlight the global significance of cervical cancer, ranking as the fourth most prevalent cancer among females worldwide, with approximately 528,000 new cases annually. To address these challenges, presented a fully automated pipeline aimed at detecting the cervix and classifying cervical cancer from cervigram images. The pipeline comprises two pre-trained deep learning models: one for automatic cervix detection and another for cervical tumor classification. The first model, focused on cervix detection, offers a detection speed 1000 times faster than existing data-driven models, achieving an intersection of union (IoU) measure accuracy of 0.68. The second model employs self-extracted features for cervix tumor classification. These features are acquired through lightweight models based on convolutional neural networks (CNN). The proposed deep learning classifier not only surpasses existing models in terms of classification accuracy and speed but also maintains a lightweight architecture. With an area under the curve (AUC) score of 0.82, this classifier classifies each cervix region 20 times faster [63].

Yoon et al., (2022) address a significant challenge in the realm of medical imaging, which involves the occurrence of light reflection during the process of lesion diagnosis. This phenomenon, where light reflections form, obscures the diagnostic view of the lesion and hampers accurate observation for diagnosis. Current methods attempt to mitigate this issue by adjusting the camera's angle to suppress light reflection, resulting in inconvenient diagnostic procedures. To overcome these limitations, the researchers present a novel approach that involves the rotation of a linear polarization filter to eliminate light reflection in a diagnostic imaging camera. By controlling vertical and horizontal polarization through the rotation of the filter, the polarization is optimized for horizontal polarization. Notably, the filter's rotation angle for horizontal polarization control is set at 90° , creating a 90° phase difference between the vertical and horizontal polarization waves. The study effectively demonstrates that the proposed method efficiently removes light reflection during the imaging process. This removal of light reflection serves to enhance the visibility of the lesion's field of view. Ultimately, this elimination of light reflection aids medical practitioners in achieving more accurate diagnoses. The results of this research hold potential for robust reliability and future commercialization, with direct applications in the domain of diagnostic medicine [125].

2.6. Summary

In this chapter, explores the various methods and techniques employed in the realm of image analysis, specifically targeting the enhancement of smart colposcopy images using color space method. The chapter commences by discussing a range of color space-based threshold methods detailed in the existing literature. These methods serve as the foundational step in identifying the regions affected by glare in smart colposcopy images. By analyzing the pixel values in different color spaces, researchers have developed techniques to pinpoint glare areas. Building upon the initial identification of glare regions, the chapter shifts its focus to the cutting-edge domain of deep learning. In this chapter explored the utilization of deep learning techniques to segment and isolate the glare-affected regions within smart colposcopy images. Deep learning models, renowned for their ability to autonomously learn complex patterns; have proven highly effective in this regard. With glare regions successfully identified and segmented the chapter progresses to address the issue of missing or damaged areas in the images. This may result from the removal of glare or other factors. To seamlessly restore visual coherence, researchers have harnessed deep learning inpainting methods, which predict and generate high-quality pixel content for these gaps. A critical aspect of this exploration involves the application of deep learning in image classification. By leveraging the discriminative power of deep learning models, researchers have developed systems capable of grading smart colposcopy images based on the severity of detected anomalies. These systems empower healthcare professionals to make informed decisions and recommendations, Throughout the chapter, an in-depth analysis of the salient features and characteristics of each method is conducted. This analysis highlights the strengths and weaknesses of individual models, providing insights into their suitability for specific applications. Additionally, the potential for combining multiple algorithms to achieve optimal results and enhance overall efficiency in document organization is explored. In summary, this chapter represents a comprehensive journey through the world of image analysis, focusing on the enhancement of smart colposcopy images by addressing challenges related to glare and image quality. It underscores the importance of leveraging both traditional color space-based threshold methods and cutting-edge deep learning techniques in a synergistic manner. By analyzing the advantages and disadvantages of each approach and considering their adaptability to different applications, this chapter lays the foundation for more precise and efficient image analysis in the context of medical imaging.