Automated Immunohistochemistry Scoring: A Review of Advanced Computational Methods in Histopathology

Dominika Petríková, Ivan Cimrák
University of Žilina
Žilina, Slovakia
dominika.petrikova, ivan.cimrak@fri.uniza.sk

***Abstract*—Machine learning methods, including both deep learning and traditional approaches, have achieved significant advancements in medical image analysis, particularly in histopathology. Histopathological images play a vital role in disease diagnosis and prognosis, driving the growing interest in automating histological image analysis using machine learning. This paper provides a comprehensive overview of recent developments in machine learning-based techniques for analyzing immunohistochemistry images, with a focus on studies published in recent years. The review categorizes the research from two perspectives: 1) the training approach and model design, including supervised deep learning, unsupervised, and hybrid methods, and 2) the application type, such as tissue grading and positive cell quantification. The study aims to guide future research by offering a structured analysis of current methodologies and highlighting the potential of machine learning, including both deep learning and traditional algorithms, to enhance the efficiency and accuracy of histopathological analysis**

I. Introduction

Histological staining techniques are indispensable in clinical practice for the visualization and differentiation of cellular and tissue structures, enabling accurate diagnosis and prognosis. While a variety of specialized stains are employed to detect specific components or abnormalities within tissues, immunohistochemistry (IHC) stands as a powerful technique using antibodies to detect specific antigens. This approach is widely utilized in clinical pathology to provide crucial information for diagnosing cancers, infectious diseases, and autoimmune disorders. IHC can also be used to assess the expression of tumor markers such as human epidermal growth factor receptor 2 (HER2) in breast cancer [1]. The interpretation of IHC results, however, is inherently subjective and prone to significant inter-pathologist variability. This can lead to inconsistent diagnostic and prognostic results, especially given that manual scoring generally lacks the precision required for accurate quantification of antigen expression levels [2]. This subjectivity represents a significant limitation in the diagnostic workflow, highlighting a clear and pressing need for objective, automated solutions. Significant advances in machine learning (ML), particularly in deep learning (DL), coupled with the increasing volume of digitized whole-slide images (WSIs), have created a unique opportunity to address these challenges. Several studies have explored the application of computer vision and ML to support and automate histological analysis [3]–[11].

The objective of automating IHC scoring is not only to reduce the workload of pathologists but also to improve consistency, provide precise quantification, and enable large-scale analysis that is currently not feasible. Understanding the state-of-the-art methods for IHC quantification is particularly important as these techniques can also be leveraged to tackle related research problems, such as the prediction of IHC biomarker values from hematoxylin-eosin (H&E) stained images. A major obstacle in this area is the lack of high-quality, expertly annotated datasets. This forces researchers to develop their own annotation solutions. Therefore, a comprehensive overview of current automated methods, which can be adapted for dataset generation, is essential. Automating the evaluation of IHC images has the potential to substantially enhance the efficiency and reliability of research in this field. In this paper, we present a comprehensive overview of the most recent and relevant ML and DL approaches developed for IHC score estimation. We systematically reviewed articles from leading databases, with the methodology detailed in the following section. This survey distinguishes itself by focusing specifically on the period from 2020 up to June 2025, providing a current and highly relevant snapshot of the field. The rest of this paper is organized to provide a structured analysis: Section 2 details a systematic approach to conducting the literature review. In Section 3, a basic overview of ML approaches in the context of computational histopathology is presented. Section 4 discusses in detail methods and approaches used in digital pathology for IHC assessment. In Section 5, we discuss the histopathological point of view by classifying the methods according to their area of application. In Section 6, we conclude the paper.

II. Methodology

The literature review was carried out by systematically searching two key scientific databases, PubMed and arXiv, for articles published between January 2020 and June 2025. The search strategy combined terms related to the subject matter and the analytical methods. The following search terms were used to identify relevant papers:

- **Subject Keywords:** ”IHC”, ”immunohistochemistry”

- **Methodology Keywords:** ”automated scoring”, ”automated labeling”, ”automated annotating”, ”automatic quantification”, ”automatic calculation”

The search queries were constructed by combining keywords to ensure a targeted and comprehensive retrieval of relevant literature (e.g., (”IHC” OR ”immunohistochemistry”) AND (”automated scoring” OR ”automated labeling” OR ”automatic calculation”)). Papers were initially screened by title and abstract to assess their relevance to the review's scope. For final inclusion, a stringent set of criteria was applied. We selected only articles that addressed the problem of IHC quantification using a novel research method, specifically excluding studies that relied solely on existing medical software solutions. Furthermore, the review was limited to studies focused on human tissue sections and those that utilized patch-level annotations, as opposed to weak annotations applied at the whole-slide image (WSI) level. We also excluded articles that worked with multiple or mixture antibody stainings, focusing only on studies with a single antibody stain. This structured approach ensures the review is both thorough and reproducible.

To ensure the transparency and reproducibility of our systematic review, we have included a PRISMA flow diagram in Fig. 1 . This diagram visually summarizes the entire search and selection process, from the initial identification of records to the final set of studies included in the review. It details the number of articles found, screened, and ultimately deemed eligible, along with the specific reasons for exclusion at each stage. The PRISMA flow diagram was generated using [12].

III. Overview of learning schemas

This section provides an overview of different learning approaches within the context of ML as applied to computational pathology. The problem of evaluating IHC images can be approached in both ways by supervised learning (SL) as well as unsupervised learning, or by a combination of both approaches.

A. Supervised learning

SL is a fundamental ML paradigm in which an algorithm is trained on labeled data, where the input data is paired with the correct output. The goal is for the model to learn the mapping between inputs and outputs, allowing it to make predictions or classifications on unseen data [13]. Among the supervised learning techniques, we identify three major canonical deep learning models based on the nature of tasks that are solved in digital histopathology: classification, detection and segmentation based models.

Key models for segmentation include U-Net, which uses an encoder-decoder architecture to capture both local and global features, and Mask R-CNN, which extends Faster R-CNN by adding a segmentation mask prediction to the object detection framework. For object detection, popular models include Faster R-CNN, which uses a Region Proposal Network (RPN) to generate proposals for bounding boxes, and YOLO (You Only Look Once), which performs detection in a single step and is known for its speed [14].

B. Unsupervised learning

Unsupervised learning is a ML paradigm where models are trained on data without labeled outputs. The goal of unsupervised learning is to identify hidden patterns or intrinsic structures in the input data, such as grouping similar data points together or reducing the dimensionality of the data. Clustering techniques, such as K-means or hierarchical clustering, group pixels or image regions into clusters based on similarity in color, texture, or intensity. These methods are useful in segmenting different tissue structures or identifying specific features in a tissue sample without requiring labeled data [13]. Thresholding, on the other hand, involves segmenting the image based on pixel intensity values, typically applying a fixed or adaptive threshold to separate foreground objects from the background. In histopathology, DAB (3,3'-Diaminobenzidine) color deconvolution is frequently used to separate overlapping color channels in IHC images. This technique allows for the isolation of individual stain intensities (e.g., DAB for detecting specific antigens) from complex color mixtures, enabling clearer analysis of cellular structures [15].

IV. IHC estimation

This section provides a general overview of recent publications using DL, ML and image analysis methods for the problem of automated IHC quantification or scoring. The focus of this work is solely on studies that address the estimation of biomarkers from IHC images using a custom approach without the use of available histological image processing software. We also included papers that did not explicitly aim at IHC estimation, but solved this problem as part of some other problem, e.g. prediction of IHC values directly from another type of staining. This section is divided according to the type of algorithms used into three subsections: supervised DL, unsupervised methods and combined (hybrid) approach.

A. Supervised Deep Learning

One of the primary advantages of DL and SL as well is that it provides high accuracy when sufficient labeled data is available. However, it can be limited by the need for large, annotated datasets, which needs to be created by experts [16]. Papers using this approach are summarized in Table I. For an explanation of staining proteins or result accuracy metrics, please see the original article. Researchers of [3] combined segmentation and detection task to evaluate proliferation index of brain tumor. First they segmented cells from background using U-Net. Then modified YOLOv3 was employed to detect and classify cells. For YOLOv3 model they replaced original backbone DarkNet-53 network with the SqueezeNet to reduce computational requirements. In [17], authors addressed the problem of detection three IHC expression categories: nuclear, cytoplasmic and membranous. Separate Mask-R-CNN and YOLOv5 models were developed for each category scoring tumor tissue of several organs. Detection networks were utilized also in [18], where authors compared dedicated model PathoNet with generic models such as Faster R-CNN and others. In [19] scholars utilized Faster

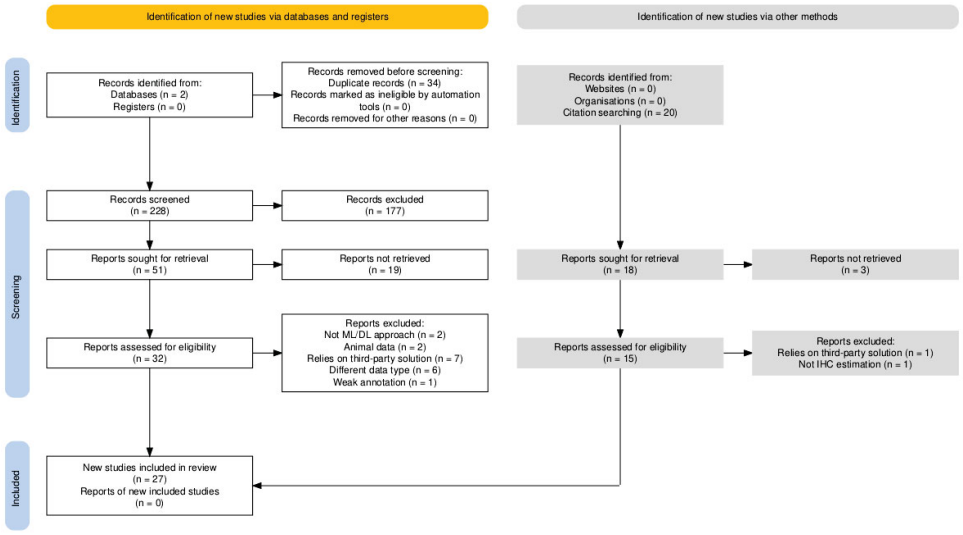

Fig. 1. PRISMA flow diagram of our review, generated via [12]

R-CNN and HoLy-Net for segmentation and detection of IHC positive and negative cells on three different biomarkers in IHC images and classification of tumor and non-tumor cells in HE images.

Modified U-Net model was employed in [20] for nerve detection in thyroid tissue samples. Model output may contain predicted positive instances that are too small, or a cluster of predicted positive instances that are separated from each other. Therefore, post-processing method such as binary morphology was applied to combine prediction results for nerve quantification. U-Net model was also employed in [21] for rheumatoid arthritis tissue images. Trained model is available for use or further fine-tuning. Instance based segmentation model was used in [22]. Authors trained SOLOv2 segmentation model using transfer learning to quantify and grade four type of biomarkers used in breast cancer treatment process. They also compared their model performance to Mask-R-CNN on all IHC biomarkers.

In [4] authors introduced LYSTO, the Lymphocyte Assessment Hackathon, where participants had to assess the number of lymphocytes in histopathological images of colon, breast, and prostate cancer stained with CD3 (cluster of differentiation) and CD8 immunohistochemistry. They summarized methods used by all five teams attending the hackathon. All teams used model based on DL neural networks. Problem of melanoma classification and grading was addressed in [23]. They used weak annotations approach to train ResNet classifier. Weak annotations approach means all patches from the same slide has the same label of the slide. Predictions for all patches, were then averaged into final prediction for each slide. For classification into 4 grades, authors in [24] devised a pyramid sampling strategy to capture the multiscale nature of tissue morphology and HER2 expression patterns. This approach involved systematically extracting small patches from original high-resolution tissue images. These patches were then fed into DenseNet. The study [25] follows a two-stage approach for automating the analysis of PD-L1 in histopathology images. First, the framework segments tumor areas into ”positive” and ”negative” regions using a modified version of the UNet or DeepLabV3+ neural network. In the second stage, the method uses a specialized neural network StarDist to detect individual cell nuclei within the previously segmented regions. Based on the nuclei’s location (within a positive or negative region), the cells are classified, and a final score is calculated.

B. Unsupervised Machine Learning

Acquiring high-quality datasets for training deep learning models is often a labor-intensive and time-consuming process. In many cases, the challenges associated with data acquisition are further exacerbated by the specific nature of the problem being addressed, particularly when the data is scarce or difficult to obtain. This scarcity of labeled data can significantly hinder the development and effectiveness of machine learning models, as large, diverse, and well-annotated datasets are typically required for robust model training. In these situations, it is necessary to use unsupervised learning methods that do not require annotated data. Table II summarizes studies analyzing IHC images with unsupervised methods. Authors of [29] developed method to predict the image scores on a 5-point scale. The proposed method first converts the RGB (red, green, blue) images into optical density using the Beer–Lambert law. After this, the following three stages are carried out: stain separation, feature extraction, and prediction of the scores. The images are scored using a k-means clustering algorithm equipped with beta divergences with each centroid representing one score. In [7], researchers presented an unsupervised four-stage pipeline for detecting protein markers of human epidermal keratinocyte differentiation on IHC images. The pipeline consisted of these steps: color normalization, color deconvolution to acquire color channels of the stains used,

TABLE I. SUMMARY OF PAPERS USING SUPERVISED DL

Ref	Method	Cancer type	Staining	Dataset	Result	Application
[17]	Mask-R-CNN, YOLOv5	Colon, breast, prostate	Ki67, PMS2, PTEN	in house	Acc 0,91	grading
[3]	U-Net, YOLOv3	Brain	Ki67	in house	mAP 0,87	counting
[18]	multiple detection	Breast	Ki67	SHIDC-B-Ki-67, LSOC-Ki-67	x	counting
[24]	DenseNet	Breast	HER2	in house	Acc 0.84	grading
[19]	Faster-R-CNN, HoLy-Net	B Lymphocytes	Ki67, CD3, ERG	LyNSeC	F1 0,84	counting
[20]	U-Net	Thyroid	PGP9.5	ref in article	Precision 0.75	counting
[22]	SOLOv2	Breast	ER, PR, HER2, Ki67	in house	mAP 0,77	counting
[4]	multiple DL	Breast, colon, prostate	CD3, CD8	available at zenodo	x	counting
[21]	U-Net	Rheumatoid arthritis	CD20, CD68, CD138	in house	Dice score 0.863	segmentation
[23]	ResNet	Melanoma	MART1	in house	AUROC 0.92	grading
[26]	AlexNet	Breast	Ki67	AIDPATH	F1 0,47	grading
[27]	cycle consistent GAN, U-Net	Pancreas	Ki67	in house	F1 0,813	grading
[28]	ResNet101	Prostate	PTEN	in house	AUC 0,964	counting
[25]	UNet, StarDist	Lung	PD-L1	in house	F1 0.93	counting

morphological operations and k-means clustering using DAB stain intensity.

To solve the problem of predicting mismatch repair (MMR) status from HE images in [5], researchers needed to extract MMR score from IHC images and create annotation for HE patches. To address the problem of IHC estimation they implemented unsupervised approach consisting of color deconvolution into H and DAB channels with following thresholding. Similarly also in [6] the researchers needed to create a label based on the quantification of the IHC images. Estimated labels were then used to train deep neural networks for melanocytic cell segmentation from HE, obtaining labels for HE patches from adjacent IHC tissue sections. To quantify IHC patches, they applied color deconvolution, histograms thresholding and binary morphology.

C. Hybrid approach

Even in situations where annotations are available, it may be advantageous to use a combination of DL and unsupervised methods for pre- or post-processing the image. These papers are organized in Table III. In [30], authors have developed a hybrid model that effectively calculates and grades the proliferation index on Ki-67 images of neuroendocrine tumors. The proposed system first performs preprocessing using Gaussian function. Then segmentation is performed using the U-Net architecture to separate nuclei from background. The identified nuclei are then evaluated as Ki67 positive or negative employing mathematical morphology and color and shape information extracted from RGB or HSV (hue, saturation, value) image. Researches in [31] used a two-step approach based on segmentation and classification. Unlike the previous study, they used simple methods such as watershed and highpass for segmentation and SVM (support vector machine), KNN (K-nearest neighbor) and random forest (RF) as the posterior classifier. A DL-based decision support system for IHC scoring of invasive ductal carcinoma was presented in [32]. Authors modified U-Net model to segment regions of interest (ROIs) of four different IHC biomarkers. On segmented ROIs, Mask-R-CNN model was employed to detect nuclei and CMYK color space with combination of value thresholding to classify cells.

V. DISCUSSION

From a histological standpoint, the IHC assessment methods described in the literature can be broadly categorized into two primary types: grading or scoring (semi-quantification) and counting (fully quantification).

The studies reviewed in the previous chapter highlight several effective approaches for applying automated image processing techniques and ML to the task of IHC estimation in histology, across various cancer types. DL has been most frequently applied to breast cancer, which remains a leading cause of cancer-related mortality in women worldwide [39]. From a histological standpoint, the IHC assessment methods described in the literature can be broadly categorized into two primary types: grading or scoring (semi-quantification) and counting (fully quantification).

1) Fully Quantitative Methods: By a fully quantitative approach, we refer to methods that provide precise, numerical values to measure the extent and intensity of antigen expression in tissue samples. These approaches utilize advanced image analysis and computational methods to objectively measure specific parameters, such as staining intensity or the percentage of positively stained cells. One of the biomarkers for which fully quantified estimation is often applied is the Ki67 protein, a proliferation marker commonly used in breast cancer assessment. The estimation of the Ki67 positive cells ratio has been addressed in many studies including [22], [33], [37].

2) Semi-Quantitative Methods: A semi-quantitative technique in IHC estimation refers to methods that provide an

TABLE II. SUMMARY OF PAPER USING UNSUPERVISED METHODS

Ref	Method	Cancer type	Staining	Dataset	Result	Application
[29]	K-means	Colon	TEM/MET	in house	Acc 0,87	grading
[7]	K-means	Epidermis	Ki67, fillagrin, keratin10, HSPA2	in house	Acc 0,87	counting
[5]	Thresholding	Colorectal	MMR	COMET	x	grading
[6]	Thresholding	Melanoma	MART1	in house	x	grading

TABLE III. SUMMARY OF PAPERS COMBINING UNSUPERVISED METHODS WITH SUPERVISED DL

Ref	Method	Cancer type	Staining	Dataset	Result	Application
[30]	U-Net	Neuroendocrine	Ki67	in house	F1 0,86	counting
[31]	SVM, KNN, RF	Breast	ER, PR	in house	Acc 0,9	grading
[32]	U-Net, Mask-R-CNN	Breast	ER, PR, HER2, Ki67	in house	Acc 0,9	counting
[33]	UV-Net	Breast	Ki67	Deepslides, in house	F1 0,833	counting
[34]	MLP	Breast	Ki67	in house	F1 0,767	grading
[35]	Seg-Net	Melanoma	MART1, Ki67	in house	Acc 0,9	counting
[36]	Inception V1, RF	Breast	ki67	in house	Acc 0.9	counting
[37]	CNN	Melanoma	Ki67	in house	MAE ¡ 0.04	counting
[38]	SVM	Breast	HER2	in house	Acc 0,88	grading

approximate measure of antigen expression in tissue samples based on visual assessment, typically using a subjective scale. These techniques are not fully quantitative but instead categorize the intensity of staining (e.g., weak, moderate, or strong) [40]. This method is predominantly utilized for estimating HER2, ER (estrogen receptor), or PR (progesterone receptor) in breast cancer image analysis [24], [31], [38]. Nevertheless, it is also applicable to other biomarkers, including Ki67 [26], [34].

Although instances of both primary tasks: counting and grading are present across all method categories (Supervised, Unsupervised, and Hybrid), a critical analysis necessitates differentiating their inherent difficulties to understand methodological choices.

A. Task Complexity: Counting Versus Grading

It is crucial to emphasize that Counting (precise enumeration of individual cells) represents a significantly more challenging task than Grading (overall tissue classification based on IHC properties). This complexity manifests in two critical areas:

Data Annotation: Counting demands exceptionally intensive and detailed dataset annotation, often requiring the precise segmentation or explicit marking of every single cell. Conversely, Grading often relies on simpler annotations applied at the patch or WSI level.

Model Architecture: While the Grading task can often be successfully addressed using a simpler ML classifier, such as a CNN, the Counting task mandates more complex architectures, including image segmentation and object detection models.

B. Limitations in Comparative Evaluation

A critical limitation we identified when summarizing the existing body of work is the inconsistency in evaluation metrics used across studies. Given that various papers utilize and report different final metrics (e.g., Accuracy, AUC, F1-score), direct, quantitative comparison of the performance of individual approaches becomes ad hoc impossible. This widespread lack of benchmark standardization presents a significant hurdle for accurately determining the most effective methods and severely complicates the interpretation of results within the broader research context.

C. Data Requirements and Method Selection

The choice of estimation method is critically dependent on the availability and quality of ground truth annotations.

From our analysis of the literature, it is evident that Supervised DL methods are primarily applicable where large volumes of densely annotated data are available. This typically translates to hundreds of WSIs, which, when divided into patches, yield tens of thousands of data points. This scale reflects the high data demands inherent to Supervised DL.

In contrast, Unsupervised methods are typically deployed in situations facing severe data limitations. Our review indicated that datasets used for these methods consisted of only around 10 data samples on average. While this assessment is challenging to generalize given the lower volume of articles in this category, it highlights their reliance on minimal data.

The Hybrid approach serves as a practical compromise. By leveraging human intervention—for example, in expert-driven feature extraction—the complexity of the task for the subsequent Supervised algorithm is reduced. This allows for the construction of a viable pipeline even with a smaller data sample, as the model is not required to discover domain features from scratch.

Our systematic review of literature published since 2020 revealed a notable absence of Transformer-based models and Foundation Models being successfully applied to complex

IHC quantification tasks. While Vision Transformers (ViT) are currently revolutionizing general computer vision, their specialized application in histopathology is still in its nascent stages compared to established CNN models [41], [42]. ViT and Foundation Models, pre-trained on massive datasets, represent the next frontier, promising superior generalization capabilities [43]. The observed deficit suggests that researchers are still navigating the unique technical challenges of WSI analysis—namely, the gigapixel-scale size of WSIs (often exceeding 150,000 x 150,000 pixels), the inherent high dimensionality of the data, and the critical need for accurate micro-level spatial localization across the entire slide [42], [44], [45]. Successfully adapting these powerful, data-hungry architectures to process WSI efficiently, without compromising detailed cellular information, remains a significant computational hurdle that must be overcome before they can be adopted as standard methodologies for practical IHC estimation.

D. Allocation of Domain Expertise

A final critical factor distinguishing these methodologies is the allocation of domain expertise required from the research team.

Supervised DL: Domain knowledge is almost entirely delegated to the algorithm. The model autonomously learns and extracts relevant features from the image to solve the specific task (counting/grading). This approach reduces the burden on the scientist to explicitly specify color intensity thresholds or geometric criteria.

Unsupervised / Traditional Methods: The scientist must actively take on the domain expertise. They are required to clearly specify the explicit criteria and features upon which data classification will be based—such as defining precise thresholds for color intensity or specific cellular morphology. This requirement for deep medical or biological knowledge can be extremely challenging for technical researchers lacking clinical experience and poses a significant risk of method failure if the defined criteria are imprecise or incomplete.

E. Addressing Data Scarcity and Privacy

The dependency on large, richly annotated datasets, particularly for supervised deep learning methods, presents the single largest bottleneck to the clinical adoption of AI in histopathology. To overcome the scarcity of data and, more critically, the privacy barriers associated with sharing patient records across institutions, researchers are increasingly employing sophisticated decentralized and augmentation strategies:

1) Federated Learning (FL): FL offers a privacy-preserving solution that directly tackles the data silo problem. This paradigm allows multiple institutions to collaboratively train a single global model by only exchanging model parameters (weights), rather than transferring sensitive raw WSI data [41], [46]. Studies in computational pathology have demonstrated that FL can achieve model quality comparable to centralized training, thus improving model generalizability while adhering to strict privacy regulations like HIPAA and GDPR [47], [48].

2) Self-Supervised Learning (SSL): When data is plentiful but labels are scarce, SSL is utilized to leverage the vast amount of unlabeled histopathology imagery. By generating "pretext tasks" (e.g., predicting rotated image patches or reconstructing masked sections), SSL models learn robust, high-quality visual representations from the data itself. These representations can then be effectively transferred and fine-tuned using minimal labeled data for specific IHC counting or grading tasks [45], [48].

3) Synthetic Data Generation: A third strategy involves utilizing Generative Adversarial Networks or Latent Diffusion Models to create new, synthetic image patches that possess realistic tissue and staining variations [49]. This approach allows researchers to effectively augment limited training datasets, balance class imbalances, and expose models to a wider range of domain shifts and visual variability, significantly enhancing the robustness and generalization capabilities of DL classifiers [50], [51].

VI. Conclusion

This article presents a comprehensive systematic review of the latest automated methods based on deep learning and machine learning for the analysis and quantification of IHC biomarkers in histological images. The analysis of papers published in the last five years demonstrates that automated IHC processing has been applied to a wide range of tumor types to inform treatment plans and advance research in computational pathology. Our survey reveals several key conclusions:

- **Application Areas:** Automated IHC estimation has been successfully applied to a variety of cancer types, including breast, colon, brain, and melanoma, as well as autoimmune diseases. The reviewed works have primarily focused on two main tasks: tissue grading (semi-quantification) and the precise counting (quantification) of biomarker-positive cells.
- **Deep Learning Dominance:** The rapid growth of deep learning and the development of advanced neural network models have led to their widespread adoption in histology. It is evident that supervised methods are currently a major focus in research, despite their high demand for annotated data. This approach has become the most prevalent method for tackling complex visual tasks, owing to its ability to efficiently process and analyze complex image data. In this context, many researchers use available medical software for IHC estimation, such as QuPath and ImageJ, not only for analysis but also for generating the datasets necessary for training these supervised models.
- **Hybrid Approaches:** Despite the significant success of supervised deep learning, unsupervised methods remain highly valuable. When used in conjunction with deep learning techniques for tasks like preprocessing and post-processing, these methods can substantially enhance the overall performance and robustness of DL models.

In summary, the advancements in automated IHC image analysis are poised to significantly improve the efficiency and consistency of diagnostic workflows. The continued development

of hybrid DL/ML models and the creation of standardized, high-quality datasets will be crucial for the further maturation of this field. Our review provides a foundational overview for researchers and practitioners, paving the way for more objective and data-driven pathological assessment.

ACKNOWLEDGMENT

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00705.

REFERENCES

[1] J. Duraiyan, R. Govindarajan, K. Kaliyappan, and M. Palanisamy, "Applications of immunohistochemistry," *J. Pharm. Bioallied Sci.*, vol. 4, no. Suppl 2, pp. S307–9, Aug. 2012.

[2] M. G. Kohale, A. V. Dhobale, N. J. Bankar, O. Noman, K. Hatgaonkar, and V. Mishra, "Immunohistochemistry in pathology: A review," *J. Cell. Biotechnol.*, vol. 9, no. 2, pp. 131–138, Dec. 2023.

[3] F. A. Dzulkifli, M. Y. Mashor, R. A. A. Raof, and H. Jaafar, "A proposed framework for improving the detection and classification of ki67 expression in astrocytoma histopathological images," in *2023 International Workshop on Artificial Intelligence and Image Processing (IWAIIP)*, 2023, pp. 315–320.

[4] Y. Jiao, J. van der Laak, S. Albarqouni, Z. Li, T. Tan, A. Bhalerao, J. Ma, J. Sun, J. Pocock, J. P. W. Pluim, N. A. Koohbanani, R. M. S. Bashir, S. E. A. Raza, S. Liu, S. Graham, S. Wetstein, S. A. Khurram, T. Watson, N. Rajpoot, M. Veta, and F. Ciompi, "Lysto: The lymphocyte assessment hackathon and benchmark dataset," 2023.

[5] R. Awan, M. Nimir, S. E. A. Raza, M. Bilal, J. Lotz, D. Snead, A. Robinson, and N. Rajpoot, "Deep learning based prediction of msi using mmr markers in colorectal cancer," 2022.

[6] M. Tada, U. E. Lang, I. Yeh, E. S. Keiser, M. L. Wei, and M. J. Keiser, "Learning melanocytic cell masks from adjacent stained tissue," 2024.

[7] D. Zamojski, A. Gogler, D. Scieglinska, and M. Marczyk, "Epidermaquant: Unsupervised detection and quantification of epidermal differentiation markers on h-dab-stained images of reconstructed human epidermis," 2024.

[8] D. Pilutti, V. Della Mea, E. Pegolo, F. La Marra, F. Antoniazzi, and C. Di Loreto, "An adaptive positivity thresholding method for automated ki67 hotspot detection (AKHoD) in breast cancer biopsies," *Comput. Med. Imaging Graph.*, vol. 61, pp. 28–34, Nov. 2017.

[9] R. Paulik, T. Micsik, G. Kiszler, P. Kaszál, J. Székely, N. Paulik, E. Várhalmi, V. Prémusz, T. Krenács, and B. Molnár, "An optimized image analysis algorithm for detecting nuclear signals in digital whole slides for histopathology," *Cytometry A*, vol. 91, no. 6, pp. 595–608, Jun. 2017.

[10] R. S. Gomolka, A. Korzynska, K. Siemion, K. Gabor-Siatkowska, and W. Klonowski, "Automatic method for assessment of proliferation index in digital images of dlbcl tissue sectionmungle2017149," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 30–37, 2019.

[11] R. S. Geread, P. Morreale, R. D. Dony, E. Brouwer, G. A. Wood, D. Androutsos, and A. Khademi, "Ihc color histograms for unsupervised ki67 proliferation index calculation," *Frontiers in Bioengineering and Biotechnology*, vol. 7, 2019.

[12] N. R. Haddaway, M. J. Page, C. C. Pritchard, and L. A. McGuinness, "PRISMA2020: An R package and shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis," *Campbell Syst. Rev.*, vol. 18, no. 2, p. e1230, Jun. 2022.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[14] T. Hoeser and C. Kuenzer, "Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends," *Remote Sensing*, vol. 12, no. 10, 2020.

[15] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analytical Quantitative Cytology and Histology*, vol. 23, no. 4, pp. 291–299, 2001.

[16] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2009.

[17] W. Kildal, K. Cyll, J. Kalsnes, R. Islam, F. M. Julbø, M. Pradhan, E. Ersvær, N. Shepherd, L. Vlatkovic, OSBREAC, X. Tekpli, Ø. Garred, G. B. Kristensen, H. A. Askautrud, T. S. Hveem, H. E. Danielsen, and OSBREAC - Oslo Breast Cancer Consortium, "Deep learning for automated scoring of immunohistochemically stained tumour tissue sections - validation across tumour types based on patient outcomes," *Heliyon*, vol. 10, no. 13, p. e32529, Jul. 2024.

[18] M. Karol, M. Tabakov, U. Markowska-Kaczmar, and L. Fulawka, "Deep learning for cancer cell detection: do we need dedicated models?" *Artif. Intell. Rev.*, vol. 57, no. 3, Feb. 2024.

[19] T. Kataria, S. Rajamani, A. B. Ayubi, M. Bronner, J. Jedrzkiewicz, B. S. Knudsen, and S. Y. Elhabian, "Automating ground truth annotations for gland segmentation through immunohistochemistry," *Mod. Pathol.*, vol. 36, no. 12, p. 100331, Dec. 2023.

[20] I. P. Astono, J. S. Welsh, C. W. Rowe, and P. Jobling, "Objective quantification of nerves in immunohistochemistry specimens of thyroid cancer utilising deep learning," *PLoS Comput. Biol.*, vol. 18, no. 2, p. e1009912, Feb. 2022.

[21] A. Gallagher-Syed, A. Khan, F. Rivellese, C. Pitzalis, M. J. Lewis, G. Slabaugh, and M. R. Barnes, "Automated segmentation of rheumatoid arthritis immunohistochemistry stained synovial tissue," 2023.

[22] B. M. Priego-Torres, B. Lobato-Delgado, L. Atienza-Cuevas, and D. Sanchez-Morillo, "Deep learning-based instance segmentation for the precise automated quantification of digital breast cancer immunohistochemistry images," *Expert Systems with Applications*, vol. 193, p. 116471, May 2022.

[23] C. Wies, L. Schneider, S. Haggenmüller, T.-C. Bucher, S. Hobelsberger, M. V. Heppt, G. Ferrara, E. I. Krieghoff-Henning, and T. J. Brinker, "Evaluating deep learning-based melanoma classification using immunohistochemistry and routine histology: A three center study," *PLOS ONE*, vol. 19, no. 1, p. e0297146, Jan. 2024.

[24] S. Y. Selcuk, X. Yang, B. Bai, Y. Zhang, Y. Li, M. Aydin, A. F. Unal, A. Gomatam, Z. Guo, D. M. Angus, G. Kolodney, K. Atlan, T. K. Haran, N. Pillar, and A. Ozcan, "Automated HER2 scoring in breast cancer images using deep learning and pyramid sampling," *BME Front.*, vol. 5, p. 0048, Jul. 2024.

[25] S. Kabir, M. E. H. Chowdhury, R. Sarmun, S. Vranić, R. M. Al Saady, I. Rose, and Z. Gatalica, "A novel deep learning framework for automatic scoring of PD-L1 expression in non-small cell lung cancer," *Biomol. Biomed.*, Mar. 2025.

[26] Z. Swiderska-Chadaj, J. Gallego, L. Gonzalez-Lopez, and G. Bueno, "Detection of ki67 hot-spots of invasive breast cancer based on convolutional neural networks applied to mutual information of h&e and ki67 whole slide images," *Applied Sciences*, vol. 10, no. 21, 2020.

[27] X. Zhang, T. C. Cornish, L. Yang, T. D. Bennett, D. Ghosh, and F. Xing, "Generative adversarial domain adaptation for nucleus quantification in images of tissue immunohistochemically stained for ki-67," *JCO Clinical Cancer Informatics*, no. 4, pp. 666–679, 2020.

[28] S. A. Harmon, P. G. Patel, T. H. Sanford, I. Caven, R. Iseman, T. Vidotto, C. Picanço, J. A. Squire, S. Masoudi, S. Mehralivand, P. L. Choyke, D. M. Berman, B. Turkbey, and T. Jamaspishvili, "High throughput assessment of biomarkers in tissue microarrays using artificial intelligence: PTEN loss as a proof-of-principle in multi-center prostate cancer cohorts," *Mod. Pathol.*, vol. 34, no. 2, pp. 478–489, Feb. 2021.

[29] A. Sarmiento, I. Durán-Díaz, I. Fondón, M. Tomé, C. Bodineau, and R. V. Durán, "A method for unsupervised semi-quantification of inmunohistochemical staining with beta divergences," *Entropy*, vol. 24, no. 4, 2022.

[30] Z. Yücel, F. Akal, and P. Oltulu, "Automated AI-based grading of neuroendocrine tumors using ki-67 proliferation index: comparative evaluation and performance analysis," *Med. Biol. Eng. Comput.*, vol. 62, no. 6, pp. 1899–1909, Jun. 2024.

[31] H. H. Razzaq, R. Ghazali, and L. E. George, "Analytical use of ihc dataset by using segmentation and classification techniques," in *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*, 2022, pp. 343–347.

[32] S. K. Jha, P. Mishra, S. Mathur, G. Singh, R. Kumar, K. Aatre, and S. Rengarajan, "Development and validation of fully automatic deep learning-based algorithms for immunohistochemistry reporting of invasive breast ductal carcinoma," 2024.

[33] S. Mirjahanmardi, M. Dawe, A. Fyles, W. Shi, F.-F. Liu, S. Done, and A. Khademi, "Preserving dense features for ki67 nuclei detection," in *Medical Imaging 2022: Digital and Computational Pathology*. SPIE, Apr. 2022, p. 35.

[34] L. Roszkowiak, A. Korzynska, K. Siemion, J. Zak, D. Pijanowska, R. Bosch, M. Lejeune, and C. Lopez, "System for quantitative evaluation of DAB&H-stained breast cancer biopsy digital images (CHISEL)," *Sci. Rep.*, vol. 11, no. 1, p. 9291, Apr. 2021.

[35] S. Alheejawi, M. Mandal, H. Xu, C. Lu, R. Berendt, and N. Jha, "10 - deep learning-based histopathological image analysis for automated detection and staging of melanoma," in *Deep Learning Techniques for Biomedical and Health Informatics*, B. Agarwal, V. E. Balas, L. C. Jain, R. C. Poonia, and Manisha, Eds. Academic Press, 2020, pp. 237–265.

[36] M. Feng, Y. Deng, L. Yang, Q. Jing, Z. Zhang, L. Xu, X. Wei, Y. Zhou, D. Wu, F. Xiang, Y. Wang, J. Bao, and H. Bu, "Automated quantitative analysis of ki-67 staining and HE images recognition and registration based on whole tissue sections in breast carcinoma," *Diagn. Pathol.*, vol. 15, no. 1, p. 65, May 2020.

[37] S. Alheejawi, R. Berendt, N. Jha, S. P. Maity, and M. Mandal, "Automated proliferation index calculation for skin melanoma biopsy images using machine learning," *Computerized Medical Imaging and Graphics*, vol. 89, p. 101893, 2021.

[38] S. Tewary, I. Arun, R. Ahmed, S. Chatterjee, and S. Mukhopadhyay, "AutoIHC-Analyzer: computer-assisted microscopy for automated membrane extraction/scoring in HER2 molecular markers," *J. Microsc.*, vol. 281, no. 1, pp. 87–96, Jan. 2021.

[39] Z. Hameed, B. Garcia-Zapirain, J. J. Aguirre, and M. A. Isaza-Ruget, "Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network," *Sci. Rep.*, vol. 12, no. 1, p. 15600, Sep. 2022.

[40] J. Bencze, M. Szarka, B. Kóti, W. Seo, T. G. Hortobágyi, V. Bencs, L. V. Módis, and T. Hortobágyi, "Comparison of semi-quantitative scoring and artificial intelligence aided digital image analysis of chromogenic immunohistochemistry," *Biomolecules*, vol. 12, no. 1, p. 19, Dec. 2021.

[41] L. A. Schoenpflug, Y. Nie, F. Sheikhzadeh, and V. H. Koelzer, "A review on federated learning in computational pathology," *Comput. Struct. Biotechnol. J.*, vol. 23, pp. 3938–3945, Dec. 2024.

[42] J. Rahnfeld, M. Naouar, G. Kalweit, J. Boedecker, E. Dubruc, and M. Kalweit, "A comparative study of explainability methods for whole slide classification of lymph node metastases using vision transformers," *PLOS Digit. Health*, vol. 4, no. 4, p. e0000792, Apr. 2025.

[43] M. Bilal, Aadam, M. Raza, Y. Altherwy, A. Alsuhaibani, A. Abduljabbar, F. Almarshad, P. Golding, and N. Rajpoot, "Foundation models in computational pathology: A review of challenges, opportunities, and impact," 2025. [Online]. Available: https://arxiv.org/abs/2502.08333

[44] R. J. Chen, C. Chen, Y. Li, T. Y. Chen, A. D. Trister, R. G. Krishnan, and F. Mahmood, "Scaling vision transformers to gigapixel images via hierarchical self-supervised learning," 2022. [Online]. Available: https://arxiv.org/abs/2206.02647

[45] X. Jin, T. Huang, K. Wen, M. Chi, and H. An, "HistoSSL: Self-supervised representation learning for classifying histopathology images," *Mathematics*, vol. 11, no. 1, p. 110, Dec. 2022.

[46] N. Talebi and M. Azimi, "Federated learning for multi-institutional AI in healthcare via digital pathology," Sep. 2025.

[47] J. Shi, Y. Zhang, Z. Li, X. Han, S. Ding, J. Wang, and S. Ying, "Pseudo-data based self-supervised federated learning for classification of histopathological images," 2023. [Online]. Available: https://arxiv.org/abs/2205.15530

[48] R. Yan, L. Qu, Q. Wei, S.-C. Huang, L. Shen, D. L. Rubin, L. Xing, and Y. Zhou, "Label-efficient self-supervised federated learning for tackling data heterogeneity in medical imaging," *IEEE Transactions on Medical Imaging*, vol. 42, no. 7, p. 1932–1943, Jul. 2023. [Online]. Available: http://dx.doi.org/10.1109/TMI.2022.3233574

[49] J. L. Ruiz-Casado, M. A. Molina-Cabello, and R. M. Luque-Baena, "Enhancing histopathological image classification performance through synthetic data generation with generative adversarial networks," *Sensors (Basel)*, vol. 24, no. 12, p. 3777, Jun. 2024.

[50] S. Doerrich, F. D. Salvo, and C. Ledig, "Self-supervised vision transformer are scalable generative models for domain generalization," 2024. [Online]. Available: https://arxiv.org/abs/2407.02900

[51] Y. Xue, J. Ye, Q. Zhou, L. R. Long, S. Antani, Z. Xue, C. Cornwell, R. Zaino, K. C. Cheng, and X. Huang, "Selective synthetic augmentation with histogan for improved histopathology image classification," *Medical Image Analysis*, vol. 67, p. 101816, Jan. 2021. [Online]. Available: http://dx.doi.org/10.1016/j.media.2020.101816