An empirical analysis of machine learning frameworks for digital pathology in medical science

S.K.B.Sangeetha ¹, R.Dhaya², Dhruv T Shah³, Dharanidharan R⁴ and K.Praneeth Sai Reddy⁵

¹Assistant Professor(Sr.G), Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai-600026, India
²Associate Professor, Department of Computer Science and Engineering, King Khalid University, Kingdom of Saudi Arabia.
³, ⁴, ⁵UG Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai-600026, India

skbsangeetha@gmail.com

Abstract. Digital pathology is a technology that allows pathological information created from a digital slide to be accessed, handled, and interpreted. Using optical pathology scanners, glass slides are collected and transformed to digitized glass slides that can be viewed on your computer monitor. Relevant support for education and the practice of human anatomy is offered by digital pathology. With the recent developments in digital pathology led to computer-aided diagnosis using machine learning approaches. So, machine learning frameworks assist physicians in diagnosing critical cases such as cancer, tumors, etc and improve patient management. With an ever growing number of choices, it can be hard to pick a better machine learning method for pathological data. Big potential attempts are made in this paper to research the full context of digital pathology with the specifics of how artificial intelligence has contributed to digital pathology. This review also analyzes various machine learning frameworks by providing as much information as possible and quantifying what the tradeoffs will be. This paper ultimately provides the improvements in the frameworks available that will be required in the near future applications.

1. Introduction

Pathology is the bridge between science and medicine, which supports every aspect of patient management such as diagnostic testing, treatment advice, preventing disease, etc. In general, pathology is the study of the causes and effects of the disease. However, the pathology has roots with all other medical specialties when looking back at the history of pathology [29, 36, 44, 52] and excellent surveys of algorithms [1, 3, 18, 23, 24, 28, 48, 49, 52] in related pathology have been published for many years which is beyond the scope of this article. Pathologists are experts in illness and disease. They use their expertise to treat patients with life-threatening conditions. Usually, pathologists will be involved in new treatments to fight viruses, infections, and cancer, etc. And this is how a pathologist could save you or could kill you. When technological innovation in health care across all specialties is growing at an increasingly fast pace, digital pathology plays a major role in the diagnosis and treatment of disease [9, 30]. Digital methods could seem financially daunting and provide better patient management at lower costs. Digitizing pathology allows pathologists to derive deeper insights that include the patient’s history and unique risk factors. By enabling this digital transformation also ensures patients are diagnosed accurately. The modern digital pathology workflow requires the diagnosis of diseases from the capture and management of whole slide images. The
traditional glass slide is digitized via a slide scanner and brought about the ability for pathologists to easily analyze [4].

The workflow of digital pathology starts the procedure from the patient either through a biopsy or a resection. The pathology division is associated with appropriate clinical information to register the samples in the laboratory information system. Whole slide imaging technology has evolved in which digital slide scanners produce high-resolution digital images from the report for a computer environment [16]. The subsystems of the digital pathology system and complete workflow are depicted in Figure 1. Digital methods in pathology offer speed, accuracy, efficiency, and provide a reasonable application platform in medical diagnosis. With the availability of a high volume of quality digitized data, finding the informative, discriminative, and independent set of features requires interpretations based on machine learning. Using machine learning techniques can solve the variety of image analysis problems and its fast adoption rate can be attributed to the improved results [55].

On the technical side, machine learning-based interpretation should have clear decision support algorithms to work on digital pathology images [5]. The availability of dense planar data requires more intelligent and efficient approaches for analyzing digitized slide images. Apart from depending only on pathologists in decision making, the use of machine learning frameworks could enable better diagnosis for improved disease risk characterization [31, 38]. The purpose of this review is to provide the reader with available machine frameworks being tailored for digital pathology images and make assistance for pathologists to compare the frameworks. The main contribution of this work is

1. Originated with a complete groundwork of digital pathology.
2. Provided a clear road map for the requirement of machine learning in digital pathology.
3. Described the list and complete trade-offs of available machine learning frameworks.

The above-said points clearly distinguish this survey from other recent surveys. It gives the detail as broad as earlier works. The paper is organized as follows: Section II reviews the complete study of digital pathology origin to till. Section III discusses the influence of machine learning in digital pathology. Section IV lists and briefly describes the machine learning frameworks for digital pathology followed by a conclusion in section V.

2. Study of digital pathology

2.1 Overview
The first digitized pathology slides were found in 1968 and were sent to Massachusetts General Hospital from Boston’s Logan Airport even with insufficient computing resources and computer technology. The widespread invention of data transmission systems during the 1990s made digital
pathology a flourishing field. The first digital microscope device costs about $300,000 for scanning a single slide for a whole day. Digital microscopy describes the processing, management, and storage of digitized microscopic images and capable of the histology and cytology slides being digitized fully. Also known as Whole-Slide Imaging (WSI) [55]. Today whole-slide scanners create images that replicate glass slides in high resolution. The entire histology glass slide is scanned with an image scanner and then communicated with a pathologist. Also, for a second opinion, this whole-slide scanned image is remotely consulted. Whole slide imaging is a relatively modality of digital imaging. It saves time, costs, and the physical transportation of slides. Departments of pathology use digital imaging from simple image archiving to complex image analysis. Hence, digital pathology is a mixture of pathology and computers to replace the traditional diagnosis based on a microscope. This has evolved over the past 5 years with the emergence of new cost-effective scanner technology and cloud storage innovations in the medical sector.

Applications in the health sciences typically include treatment and clinical study. This digital pathology makes the sharing and annotation of slides much simpler and offers new opportunities for e-learning in health science applications. Digital pathology is regulated by the Food and Drug Administration (FDA) for improving diagnostic accuracy.

2.2 Background study
Digital pathology technology goes back more than 10 decades when specialized equipment was first used to capture images from a microscope on photographic plates. The idea of telepathology that conveys microscope images among remote areas has been there for about 50 years. In the last decade, however, pathology has begun to undertake an effective digital transition, shifting from the analog to the modern way [39]. Prior to this, scientists could not create images of the entire slide sample, forcing them to use their sampling to assess regions of interest. The 1990s also saw the technological expansion of computer technology in general, making sure that the hardware and software were finally available to perform the data processing required by digital pathology [41]. A broad range of software systems was developed over the ensuing framework to maintain multi-resolution two-dimensional dataset traversals [55]. When digital cameras were widely used in the late 1990s as well as in the early 2000s, pathology slide photography with microscopy was extremely popular. It was during this period as well that WSI was first introduced.

2.3 Advances in digital pathology

Digital pathology is not yet widespread but developments in digital pathology can eventually change that [19]. The rapid convergence of Artificial Intelligence (AI) with digital pathology is now one of the main medicine revolutions over the next decade, and histopathology is at the very center of this transition. In addition to developing AI-based applications that could provide tumor grading and prognostic insights that are currently inaccessible via traditional methodologies. Figure 2 depicts the desired component for the conduct of AI and other automated laboratory pathology instruments. The
series of occurrences is provided as a guide only, and in some situations capacity will be needed [8]. In particular, we have seen a growing use of Machine Learning (ML), Deep Learning (DL), and AI methods finding their way into healthcare, and also a procedure for diagnostic pathology [37, 42]. With today’s ease of cloud computing accessibility, powerful processors, and robust infrastructure, a pixel-pipeline-based workflow can be built that allows for the creation of AI-based predictive or diagnostic algorithms. In recent years several kinds of research have shown the tremendous value of digital pathology and AI solutions [55].

2.4 The advent of whole slide imaging

In pathology, digital imaging has endured an incremental period of growth and development, facilitated by improvements in hardware imagery and benefits in computational processing. Digitization of whole glass slides at near optical light resolution limits can take place today in the 60s. Whole slides may be imaged in fluorescence or using multi-spectral imaging [22]. The advent of whole-slide imaging in digital pathology has contributed to the development of computer-aided tissue exams through digital image processing. Whole slide imaging is the ability to scan glass slides for digital images, which has advanced slide imagery through the use of simple cameras. It demands a two-step process: First, a scanner is used to digitize the glass slide, producing a broad, descriptive digital slide. Then specialized software is required to display the digital image. The program is often referred to as a virtual slide viewer [4]. Before whole slide imaging, microscope-mounted cameras were only able to capture different slide areas and thus had controlled empirical use. The early virtual microscope used to microscopy a robotic computer device and was able to scan the glass slide to create a collection of mosaic image tiles that could be assembled to shape the full slide image. This application has been restricted by a lengthy scan time.

Modern whole slide imaging was created by the use of automated, high-speed image capture systems. Glass slides can now be scanned in less than one minute and digital images can be produced with high resolution [10]. In addition, continuous processing will automate entire slide imaging technology. This means that while another is being scanned a slide can be submitted. Slide labeling is also easily achieved, as the scanners can read one- and two-dimensional barcodes that can be incorporated into glass slides. Some modern whole slide imaging systems can digitize slides on different vertical focal planes so that there is no lack of precision compared to a traditional microscope’s fine focus control. In addition, DL with WSIs has accomplished remarkable performance in digital pathology image classification. However, in realistic cases, high processing costs of WSIs complicate applications, and most pathologists do use microscopy images in their workflow [59]. Recent developments in digital pathology were used to develop diagnostic tests based on histological images of diseases such as cancer. Because differences in the molecular expression of a disease can be claimed in tissue architecture and improvements in nuclear morphology, the development of automated tissue classification and diagnostic testing was based on whole slide imaging technology.

2.5 Advantages and challenges in Digital pathology

Digital pathology provides greater efficiency in terms of diagnostics, with a lowered turnaround time for documenting clinical cases. There is also the chance to expand the quality of the output by decreasing error rates. Telepathology transmits data images of pathology among areas related to research, diagnosis, and skills training. Telepathology has the logistical benefit in terms of laboratory workflows by not involving physical shipment and storage by glass slides between locations. The efficiency and quality of image analysis can also be enhanced through the reduction of manual errors via an automated system. Learners can understand from a broader variety of case studies, as digital pathology possesses the opportunity to scan a single sample of tissue that can be used numerous times in classes [5, 6]. Digital pathology allows more accurate images to be easily transmitted by data; however, obstacles remain to be addressed before digital pathology is widely introduced. Rapid scanning must be achieved in order to integrate the digital pathology into a clinical setting. There also needs to be a global standard for digital archiving. It involves reworking laboratory procedures and
designing new guidelines for scanning to try and standardize digital imaging practices. Real-time reporting requires fast data transfer.

For medical applications, strong internet connectivity, laboratory systems engineering, and electronic medical records are needed. In addition, because of the large file dimensions, high-resolution images involve additional storage systems. Considering the collection of a large amount of medical data, high data protection standards are expected [36]. In digital pathology, treatment has to be enhanced with graphics cards and high-resolution displays. And, real-time data reporting often requires fast data transfer. But, the introduction of digital pathology technology is a bigger problem in low-resource countries. There is a shortage of experienced pathologists in rural parts and in these areas, the only telepathology can enhance the capacity of image analysis, but in low-resource countries, the restriction of high bandwidth requirements will become a significant obstacle. Digital slides could be used in local networks, but with complex systems, dial-up connections contribute to repeated timeouts and images of poor quality [20, 21, and 32]. However, obstacles such as the high cost of machinery and restricted connectivity are present. While we have many benefits in adopting digital pathology techniques, there are still many problems that need to be addressed in countries with high and low capital. Whatever a universally growing digital pathology market may be in the future.

3. Influence of machine learning in Digital pathology

Digital pathology plays an important role in contemporary medical care and is progressively a technological necessity in the scientific laboratory environment. The introduction of whole-slide imaging, quicker network access, and affordable data storage has made life simpler for pathologists to handle and exchange digital slide images for clinical research. In addition, exponential developments in machine learning have allowed the convergence of artificial intelligence and digital pathology, providing possibilities for image-based assessment once restricted only to radiology and cardiology. The integration of digital slides into the workflow of pathology, complex technologies, and computer-aided diagnostic methodologies extends the boundaries of the pathologist’s point of view further than a microscopic slide and allows for the true use and application of information beyond human boundaries and clear possibilities for artificial intelligence developments in the pathology setting. Given the recent emergence of AI in image processing applications, many scientists and physicists expect AI to be ready to aid in various digital pathological tasks. Although prospects are both evident and quantifiable, in computational pathology there are obviously many barriers that need to be overcome in order to utilize the AI opportunities.

3.1 Challenges of AI in Digital pathology

Most AI algorithms involve a wide collection of images of high-quality training images. Optimally these training images need to be labeled. This essentially means that all images require a pathologist to manually distinguish the region of interest. When dealing with low-resolution images, a thorough annotation of large numbers of images is difficult for pathologists [57]. Excessive polymorphism makes tissue recognition extremely difficult by image algorithms [23]. Pathologists often use careful terminology for challenging situations. Such terminology has implications for future control and care. Thus, in simple situations, binary terminology can only be beneficial. For a clinical experience that is rarely so [17]. Patching is a possible alternative to the general methods of computer vision. But, even for updates, to be able to support them into a deep network, one absolutely requires to down sample them. Down sampling such patches can contribute to the loss of important details [27]. Deep AI applications rely heavily on the use of Graphical Processing Units (GPUs), highly technical digital images, and graphics electronic circuits. Training and using deep approaches with Central Processing Units is extremely expensive on ordinary computers. Nevertheless, pathology laboratories are still under enormous financial strain to embrace WSI technology, and the processing and storage of gigapixel histopathology scans present a daunting challenge to digital pathology adoption [58]. Targeted exploitation of a quite small pixel density within an image, which is called an adversarial attack, could even mislead a heavy-duty deep network. The path toward a reliable diagnosis has to be straightforward and fully understandable. This is also critical if there is a need to repair a deep learning algorithm as well as gain regulatory authorization for its utilization in clinical practice. Findings of AI
failures in medical services are not specifically connected to failed technology but instead to challenges in the practical deployment of AI frameworks.

3.2 End to end usage of machine learning in Digital pathology

Digital pathology has started to emerge with the digitization of patient clinical specimens and the use of WSIs, in particular. These may be widely dispersed for the development of diagnostics, teaching, and research. As the adoption of digital pathology expands, automated tissue morphology image analysis further establishes itself in pathology and reduces pathologists’ workload with standardized clinical practices. Successful application of machine learning to WSIs has the opportunity to introduce new clinical frameworks that exceed recent treatment approaches in terms of effectiveness, robustness, and rationality while offering new insight into various pathologies as well [53]. Using dimensions cataloged by a pathologist, images are cropped that slide into patches of equal size. Each image can produce thousands of patches and is classified as either tumor or regular. Most of the machine learning models on WSIs do not use the entire image as input and use only a tiny proportion of patches. In patch level annotations, all the extracted training patches have class labels and patch-based annotations are usually derived from pixel-level annotations that necessitate specialists to annotate all the pixels [7]. Transfer learning uses a pre-trained model to extract features from image patches and then use Apache Spark to train a binary classifier to predict tumor vs. normal patches. The wide range of digital pathology image analysis tasks influencing performance and tracking (e.g., mitotic events), optimization (e.g., nuclei), and categorization of tissues (e.g., cancerous versus non-cancerous).

Modern methods that manually define domain-specific signs and evolve into task-specific “handcrafted” applications that require comprehensive tuning to manage these variances. Nonetheless, machine learning takes a more agnostic approach to the domain that incorporates both exploration and implementation of features to mitigate discrimination between groups of interest. Although over the last few years there have been a variety of papers in the field of machine learning in digital pathology for object detection and quantification applications, two key drawbacks tend to existent strategies. First, designing task-specific solutions appears to require long periods of research and development. For example, to establish a method for nuclei segmentation, one must first recognize all the probable differences in appearances of morphology, texture, and color. Consequently, it is vital to building an algorithmic strategy that can account for as many of these differences as probable while not being too broad to result in false-positive results or too specific in false-negative errors. This relates to the second disadvantage with current approaches; the tacit awareness of how optimal parameters can be modified always remains exclusively with the algorithm developers and is thus not intuitively understood by externals [1]. In general, diagnosis of invasive ductal carcinomas, diagnosis of mitosis [13], segmentation of the neurons [15], segmentation of the colon gland [14], segmentation and diagnosis of nuclei [43], recognition of brain tumors [56], segmentation of the epithelium, and rating of gliomas [40] are previously addressed through machine learning techniques.

In addition, we have witnessed a tremendous use of ML frameworks finding their way into health care on digital pathology in the last 5 years. Major machine learning methodologies for medical image analysis with over 300 accomplishments in different medical fields as well as in the field of digital pathology; accomplishments were well classified as per their inherent type of image analysis: classification, detection, segmentation, registration, etc. The digitization of tissue glass slides clearly opens up exciting issues and constraints for computational imaging scientists worldwide. It is evident that while computational imaging can obviously play a role in enhancing the characterization of disease and precision medicine, there is still a range of significant technological and practical challenges that need to be addressed before computer-assisted digital pathology image analysis can become part of the standard clinical diagnostic workflow. Advances in computer imaging for digital pathology will soon proceed to develop pathology more objective, a discipline that has so far lagged far behind radiology in this regard. This transition from qualitative to quantitative pathology is in the not too distant future, by all evidence.
4. Machine learning frameworks for Digital pathology

In both experimental and diagnostic settings, digital pathology has heavily featured, with technical advancements in imaging and the development of new, powerful computational frameworks. The core of this transition was whole slide imaging, enabling us to digitize pathology slides effectively into high-resolution images. In seeking to make slides easily available and interpretable, WSI has also improved reliability and encouraged remote pathology professionals. Currently, the digitization of whole slides at very maximum resolution will take place efficiently in less than a few minutes. In reality, many more healthcare providers and research fields have been having big digitized slide catalogs. Using machine learning, these large datasets are being used to create automated diagnostics which can classify slides as describing certain pathology, or acquire observable biomarkers specifically from slides. The digital slides can be viewed by machine learning in just a short amount of time. It offers enormous opportunities for departments of pathology, doctors, and scientists to diagnose and treat cancer and infectious diseases in order to improve productivity and effectiveness.

While many sectors of life and health science recognize the potential influence of applying machine learning to whole slide images, managing an automated slide analytics pipeline persists challenging. An effective WSI pipeline must be able to consistently handle a high digitizer slide output at a low cost. Three main challenges that prevent organizations from implementing advanced digital pathology workflows with support for machine learning:

1. WSI images are typically very wide (approximately 2 GB / slide), and may require significant file pre-processing.
2. Training a machine learning algorithm may require days or even weeks along with a reasonably sized dataset of hundreds of WSIs on one single node. There is rapidly developing research on large datasets. While the separation of machine learning tasks across distributed devices can reduce latency, this is an innovative technique that is outside the reach of a conventional biological data scientist.
3. This is very helpful to be able to reproduce findings when it comes to novel methods based on patient data. Current solutions are often ad-hoc and do not require appropriate methods to track experiments and data variations used throughout the implementation of machine learning models.

Conventionally, pathologists use their peers and other frameworks, such as manuals, to look at the issue at hand. Current technology helps computers to scan through thousands of images to get details nearest to the query [50].

4.1 Similar Medical Images Like Yours (SMILY)

For various reasons, a doctor can refer to the same document and it differs from case to case. The developers optimized the method to make SMILY more engaging by allowing end-users to communicate what similitude means on-the-fly:

1. Refine-by-region helps pathologists to image an area of focus, restricting the search to that region.
2. Refine-by-example allows users to select a subset of the search queries and receive more outcomes like those.
3. Refine-by-definition sliders can be used to indicate the existence of more or less of a clinical term in the search queries. The implementation of the protocol tasks and the reporting phase involve several difficulties:
   - Monitoring and tissue slide analysis are 2 distinct tasks carried out in various systems, and tracking in various tasks is fragmented.
   - Images are not a component of a report regarding pathology.
   - Due to the complexity of the tissue features and details, pathologists’ visual memory can be overloaded.
4. No diagnostic decision-making information is included in the study that hindered a second pathologist’s reproducibility, coordination, and evaluation.

Figure 3. Schematic of the steps in constructing the SMILY database and the mechanism by which patches of the input image are used to scan similar images [25].

Figure 3 shows the steps in constructing the SMILY database and the mechanism by which patches of the input image are used to scan similar images. PathoVA can allow the pathologist consultant to examine the observations without carrying out a tissue regeneration slide analysis but only beginning from the areas selected by the first pathologist. Digital pathology promotes interdisciplinary gatherings where a histo-pathological review is taken into account when agreeing on a treatment plan. The framework works in these cases, allowing oncologists and pathologists to interactively evaluate a case, and guide for the right care. Finally, this application will address an educational environment in which students will study an experienced pathologist’s workflow to concentrate on various parts of one tissue slide; which areas the pathologists focused at, which conclusions, and other related details [16].

The above principles don’t need to be implemented into a model of machine learning. Instead, this approach allows end-users to construct new post-hoc concepts, customizing the search algorithm to concepts that are appropriate for each particular application scenario [5]. A first step in creating SMILY was to implement a deep learning model, trained using 5 billion normal, non-pathological images (e.g., dogs, trees, man-made objects, etc.). Throughout the testing, the network learned to discriminate among similar and dissimilar objects by analyzing the embedding. Embedding is a description of the mathematical vector which is a compilation of all the images fed in during training into the deep learning techniques. SMILY helps a user to pick a region of interest, and to get matches that are visually identical. Picking a tiny area in a slide allows SMILY to provide quality in a few seconds for a database of trillions of cropped images. Since pathology images can be presented at various magnifications (zoom levels), SMILY scans the image enhancement at the same intensity as the image input. While these machine learning frameworks for specific image retrieval dramatically shorten the number taken by conventional methods, they lack understanding of the user’s purpose. The framework allows users to pick a region of interest, and to get matches that are visually identical. SMILY’s ability to obtain images along a pre-specified similarity axis (e.g. histological feature or tumor grade) is assessed using images of breast, colon, and prostate tissue. SMILY showed positive results despite not being explicitly educated on pathology images or using any identified examples of histological characteristics or grades of tumors. Finally, pathologists will be able to use methods such as SMILY to detect all instances of a function in the tissue sample of the same patient to further understand the nature of the disease and make decisions on cancer therapy [26].
4.2 Concentriq Dx
The Concentriq Dx digital pathology platform has been made accessible for use in the study and reporting of digital pathology slides in the United States by Proscia, a technology provider of AI-enabled digital pathology devices. Concentriq Dx necessitates that pathologists have an Internet browser to interpret occurrences from remote areas, such as home offices, securely. The software can be mounted conveniently on network infrastructure and functions with any laboratories and machine [33]. The Concentriq app combines powerful image and data management technology with an open user interface to enable a comprehensive digital pathology environment. The networks of Visiopharm and Proscia establish a connected digital pathology ecosystem, and the power of the image workflow and management of Procia to provide the connected digital pathology ecosystem is shown in figure 4.

1. Data management-Image and data storage accessible in the cloud, with features that include importing and cataloging.
2. Whole slide viewing-Full-slide image viewer and versatile data fields enable robust imaging, scanning, and annotation.
3. Sharing and collaboration-In real-time, it is possible to share study, to conduct remote seminars, and to confer from any place.
4. Workflow management-Laboratory-defined workflows rigidly regulate and enhance images to improve performance and efficiency across the laboratory.
5. Reporting analytics-Detailed documentation of output assessment gains useful insights into the efficacy and quality of laboratory services.

![Figure 4. Connected digital pathology ecosystem [45].](image)

Concentriq Dx provides pathology research labs with the capacity to acquire manage and optimize the crucial data they generate, thus operating as a launch platform for analytical pathology suite and also third-party image analytics solutions. Concentriq Dx helps pathologists to provide a clinical diagnosis of diseases such as cancer from digitized images of tissue biopsies of patients, helping physicians have more accurate and higher-quality treatments. Concentriq Dx focuses on automating time-consuming and error-prone tasks and streamlines access to specialized expertise by focusing on the strategy of pathology around images rather than physical glass slides [47].

4.3 Databricks
Databricks Unified Data Analytics Platform can be used to handle a wide variety of pre-processing images, minimize latency by parallelizing deep learning workloads across multiple nodes, and replicate results. It deploys the WSI image data with an end-to-end automated deep learning workflow. Pathologists may use the very high resolution to navigate WSI photos and annotate the slide to identify
clinically important locations. Since there can be several more patches in operation, classification using deep neural networks will greatly increase the accuracy.

Training an image classifier to identify regions in a diaper containing cancer metastases requires the following steps:

1. **Patch generation**—Slide images are cut into similarly sized patches using coordinates. The image can produce thousands of patches and is classified as either tumor or regular.
2. **Deep learning**—Transfer learning uses a pre-trained model for extracting features from patches of images and then uses Apache Spark to train a binary algorithm to classify tumors versus regular patches.
3. **Scoring**—The trained model that is logged to a given slide using MLflow to generate a heat-map of probability.

![Figure 5. Horovod framework [46].](image)

Distributed training methods can be used in such cases to scale out the training process. The HorovodRunner toolkit on the Databricks platform distributes the workload assignment across a wide cluster with only slight modifications to the machine learning code [34]. Because of the massive quantity of WSI images the size of a modest WSI dataset can easily become difficult to handle on a single node. Training jobs must be run on a cluster of GPUs to improve efficiency. The Horovod system is used to spread training across several GPUs in a cluster that is connected across machines. The framework of horovod system is shown in figure 5. To minimize latency this method scales out training. Although Horovod involves complex setup such as GPU drivers, and Message Passing Interface (MPI) setup and configuration, the Databricks platform makes it simple to set up a technique such as HorovodRunner. HorovodRunner is enabled and initialized by default when building a cluster with the ML Runtime inside the cluster development tab [46].

### 4.4 HistoQC

HistoQC is fast quality control digital pathology framework not only to recognize and define objects but also to detect outliers at the cohort level. Even as digital pathology is making rapid progress globally with more physicians viewing tissue images on smart machines to monitor the disease, there are no consistent requirements for the tissue slides itself being processed and digitized. That means low-quality slides are mixed in with simple and precise slides, potentially confounding or manipulating a computer program trying to understand, for instance, what tumor cells look like. The framework uses a set of metrics and classification methods to help flag corrupted images for users and help preserve those that will assist technicians and doctors in their diagnoses [51]. The framework uses image metrics (e.g., color histograms, brightness, contrast), features (e.g., edge scanners), and supervised classification (e.g., pen detectors) to distinguish artifact-free regions on digitized slides. Currently, most quality control processes are conducted manually for clinical and testing purposes, rendering the procedure subjective, tedious, and vulnerable to error. HistoQC framework used to facilitate slide quality evaluation within a public slide archive comprising artifacts. In HistoQC, the
user provides a configuration file specifying the quality control pipeline specifications, such as the components to run, and in what sequence [2].

4.5 PathoVA
PathoVA is a method of visual processing for the diagnosis of computer-assisted pathology. The diagnostic work includes characterizing cells and making histological portions of tissues visible, which are illuminated by procedures that highlight unique elements. The system consists of a viewer with a tracking feature that logs all communication events with the WSI, and the subsequent magnification level and provenance of visualization). It also offers a range of techniques for image processing to assist in particular diagnostic tasks. A descriptive trace is created from the reported incidents, which the pathologist may use for reporting on the pathology. The analysis and trace are related so that observations in the slide image can be traced back to the source area. The principal tasks originate directly from the protocol:

1. Select the required measurements from among those taken and use them in the analysis.
2. Nottingham score includes ranking the tubular histology, measuring cell morphometry, and counting mitosis.
3. Quantify areas and margins of tumors, and note any specific observable.

![Figure 6](image.png)

Figure 6. A schematic view of the distinction between both the standard workflow and PathoVA workflow [12].

The implementation of the protocol tasks and the reporting phase involve several difficulties: Monitoring and tissue slide analysis are 2 distinct tasks carried out in various systems, and tracking in various tasks is fragmented. Images are not a component of a report regarding pathology. Due to the complexity of the tissue features and details, pathologists’ visual memory can be overloaded. No diagnostic decision-making information is included in the study that hindered a second pathologist’s reproducibility, coordination, and evaluation. PathoVA can allow the pathologist consultant to examine the observations without carrying out a tissue regeneration slide analysis but only beginning from the areas selected by the first pathologist. Figure 6 illustrates the distinction between the standard workflow and PathoVA workflow. Digital pathology promotes interdisciplinary gatherings where a histopathological review is taken into account when agreeing on a treatment plan. The framework works in these cases, allowing oncologists and pathologists to interactively evaluate a case, and guide for the right care. Finally, this application will address an educational environment in which students will study an experienced pathologist’s workflow to concentrate on various parts of one tissue slide; which areas the pathologists focused at, which conclusions, and other related details [11].

4.6 Pathviewer
An interactive visualization, evaluation, and annotation framework designed specifically for digital pathology, and the most efficient, easiest-to-use WSI data visualization and annotation tool accessible. Pathviewer manages the image data and helps the viewer to transfer and overlay a certain number of channels with ease. It greatly improves the reach and understanding of the probes and pathways researchers can use multiplexed imagery. It continues to promote standard workflows and different
data for digital pathology attempting to make the visualization, annotation, communicating, and strategic planning of digital pathology images straightforward [54]. Future application tissue sample evaluation is vital in many broad varieties of medical uses. In the segment for analysis and predictive scoring, the tissue samples are stained with solvents to visualize a biomarker of interest. Recent advances in computing capacity and machine learning algorithms have facilitated the creation of effective computer-assisted analytical frameworks for biomedical data with the rapid emergence of whole-slide digital scanners. There are several frameworks available for digital pathology which are open-source or publicly accessible for scholarly, study, and clinical purposes, in some cases. A detailed review article can be found covering various aspects of digital pathology, including that of the recently established methods for image analysis as well as particular issues related to histopathology. It also addresses algorithms for the selection of functions, reduction of dimensional space, and machine learning [23]. A survey paper discusses research into digital pathology image processing algorithms for the identification, segmentation, and classification of nuclei [28]. A recent review paper includes the identification of biomarkers through digital pathology and image analysis. A number of methods for image analysis are studied including the study of nuclear morphometry and tissue architecture but with a focus on immune histochemistry and tissue biomarker fluorescence study [24]. Along with these surveys, several computing resources are available with the emergence of cloud computing and high-end devices like none before. The environment is essential to a novel approach to digital pathology difficulties of image analysis, known as deep learning, a learning system with multilayered architectures of the neural network [35]. With traditional machine learning approaches, there are many limitations to handcrafted features. Downsides are being highlighted as the possibilities of applying deep learning in digital pathology are mind-boggling which is left for future analysis.

5. Conclusion
The revolutionary step of digitization of tissue glass slides has started new endless opportunities in the digital pathology world. Over the years, we have seen gradual evolution aiming to reduce the manual involvement and automate the process flow of digital pathology. With rapid innovations in digital pathology, the algorithm designers, pathologists across the world have a large proportion of quality digitized data available. In the early stages of digital pathology, for tissue detection, segmentation, morphometry, etc., conventional computer vision conceptual frameworks more equipped for radiology have been used. The biggest hurdles to get perceived quality were fluctuation in slide staining, slide preparation and various scanners offer access on the market. The constrained pixel intensity space data was inadequate to deliver stable and accurate results. In addition, the domain experts give their perspectives based on years of experience required to be obtained in the frameworks for image processing. The diversity and difficulties with the image analysis required the use of machine learning models to address these problems. In this survey, maximum effort is put forth to study about digital pathology and various machine learning frameworks available for digital pathology.

References
[1] Janowczyk A and Madabhushi A. 2016. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of Pathology Informatics 7 (July 2016), 7–29. https://doi.org/10.4103/2153-3539.186902.
[2] Janowczyk A, Zuo R, Gilmore H, Feldman M, and Madabhushi A. 2019. HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. JCO Clinical Cancer Informatics 3 (April 2019), 1–7. https://doi.org/10.1200/CCI.18.00157
[3] Madabhushi A and Lee G. 2016. Image analysis and machine learning in digital pathology: Challenges and opportunities. Medical Image Analysis 33 (October 2016), 170–175. https://doi.org/10.1016/j.media.2016.06.037
[4] Famke Aeffner, Mark Zarella, Nathan Buchbinder, MarilynM Bui, MatthewR Goodman, DouglasJ Hartman, GiovanniM Lujan, MariamA Molani, Anil Parwani, Kate Lillard, OliverC Turner, VenkataN Vemuri, AnaG Yuil-Valdes, and Douglas Bowman. 2019. Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology
Association. Journal of Pathology Informatics 10, 1 (March 2019), 9. https://doi.org/10.4103/jpi.jpi_82_18

[5] Madabhushi Anant. 2009. Digital Pathology Image Analysis: Opportunities and Challenges. Imaging in Medicine 1, 1 (August 2009), 7–10. https://doi.org/10.2217/IIM.09.9

[6] Brachtel, Elena, Yagi, and Yukako. 2011. Digital imaging in pathology – Current applications and challenges. Journal of biophotonics 5, 327-335 (December 2011). https://doi.org/jbio.201100103Citations

[7] Hye Yoon Chang, Chan Kwon Jung, Junwoo Isaac Woo, Sanghun Lee, Joonyong Cho, Sun Woo Kim, and Tae-Yeong Kwak. 2019. Artificial intelligence in pathology. Journal of Pathology and Translational Medicine 53 (December 2019), 1–12. https://doi.org/10.4132/jptm.2018.12.16

[8] Richard Colling, Helen Pitman, Karin Oien, Nasir Rajpoot, Philip Macklin, David Snead, Tony Sackville, and Clare Verrill. 2019. Artificial intelligence in digital pathology: A roadmap to routine use in clinical practice. The Journal of Pathology 249, 2 (October 2019), 143–150.

[9] Lee A. D. Cooper, Alexis B. Carter, Alton B. Farris, Fusheng Wang, Jun Kong, David A. Gutman, Patrick Widener, Tony C. Pan, Sharath R. Cholleti, Ashish Sharma, Tahsin M. Kurc, Daniel J. Brat, and Joel H. Saltz. 2012. Digital Pathology: Data-Intensive Frontier in Medical Imaging. In Proceedings of the IEEE, Vol. 100. 991–1003. https://doi.org/10.1109/JPROC.2011.2182074

[10] Toby Cornish, Ryan Swapp, and Keith Kaplan. 2012. Whole-slide imaging: routine pathologic diagnosis. Advances in Anatomic Pathology 19, 3 (May 2012), 152–9. https://doi.org/10.1097/PAP.0b013e318253459e

[11] Alberto Corvo. 2019. Visual Analytics in Digital Computational Pathology. Ph.D. Dissertation. Eindhoven University of Technology, Eindhoven. 978-90-386-4931-3.

[12] Alberto Corvo, Marc Driel, and Michel Westenberg. 2017. PathoVA: A visual analytics tool for pathology diagnosis and reporting. In Proceedings of the IEEE Workshop on Visual Analytics in Healthcare (VAHC), Vol. 9041. https://doi.org/10.1109/VAHC.2017.8387544

[13] Angel Cruz-Roa, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi. 2014. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In Proceedings of the SPIE Medical Imaging, Vol. 9041. 904103. https://doi.org/10.1117/12.2043872

[14] Ciresan D, Giusti A, Gambardella LM, and Schmidhuber J. 2012. Deep neural networks segment neuronal membranes in electron microscopy images. In Proceedings of Advances in Neural Information Processing Systems, Vol. 2. 2843–2851. https://doi.org/10.5555/2999325.2999452

[15] Ciresan DC, Giusti A, Gambardella LM, and Schmidhuber J. 2013. Mitosis detection in breast cancer histology images with deep neural networks. In Proceedings of Medical Image Computing and Computer-Assisted Intervention, Vol. 8150. 411–418. https://doi.org/10.1007/978-3-642-40763-5_51

[16] Sandeep Dwivedi, Madhu Swamy, Amita Dubey, and Yamini Verma. 2019. The advent of digital pathology: A depth review. Journal of Entomology and Zoology Studies 7, 2 (2019), 43–49.

[17] Joann G Elmore, Gary M Longton, Patricia A Carney, Berta M Geller, Tracy Onega, Anna N A Tosteson, Heidi D Nelson, Margaret S Pepe, Kimberly H Allison, Stuart J Schnitt, Frances P O’Malley, and Donald L Weaver. 2015. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA Network 313, 11 (March 2015), 1122–1132. https://doi.org/10.1001/jama.2015.1405

[18] Xing F and Yang L. 2016. Robust nucleus/Cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. IEEE Reviews in Biomedical Engineering 9 (January 2016), 234–263. https://doi.org/10.1109/RBME.2016.2515127

[19] Shelley Farrar. 2020. News Medical Life Sciences. Retrieved March 24, 2020 from https://www.news-medical.net/life-sciences/Advances-in-Digital-Pathology.aspx
[20] Shelley Farrar. 2020. News Medical Life Sciences. Retrieved April 21, 2020 from https://www.news-medical.net/life-sciences/Digital-Pathology-Challenges.aspx

[21] Paul Fontelo, John Faustorilla, Alex Gavino, and Alvin Marcelo. 2012. Digital Pathology – Implementation Challenges in Low-Resource Countries. Analytical cellular pathology 35, 1 (August 2012), 31–6.

[22] Farzad Ghaznavi, Andrew Evans, Madabhushi Anant, and Michael Feldman. 2013. Digital Imaging in Pathology: Whole-Slide Imaging and Beyond. Annual review of pathology:Mechanisms of Disease 8 (January 2013), 331–359. https://doi.org/10.1146/annurev-pathol-011811-120902

[23] M N Gurcan, Boucheron L, Can A, Madabhushi A, Rajpoot N, and Yener B. 2009. Histopathological Image Analysis: A Review. IEEE Reviews in Biomedical Engineering 2 (October 2009), 147–171. https://doi.org/10.1109/RBME.2009.2034865

[24] Peter W. Hamilton, Peter Bankhead, Yinhai Wang, Ryan Hutchinson, Declan Kieran, Darragh G. McArt, Jacqueline James, and Manuel Salto-Tellez. 2014. Digital pathology and image analysis in tissue biomarker research. Methods 70, 1 (November 2014), 59–73.

[25] Narayan Hegde. 2020. Google AI Blog. Retrieved May 16, 2020 from https://ai.googleblog.com/2019/07/building-smily-human-centric-similar.html

[26] Narayan Hegde, Jason Hipp, Yun Liu, Michael Emmert-Buck, Emily Reif, Daniel Smilkov, Michael Terry, Carrie Cai, Muhammad Amin, Craig Mermel, Phil Nelson, Lily Peng, Greg Corrado, and Martin Stumpe. 2019. Similar image search for histopathology: SMILY. Nature Partner Journal Digital Medicine (2019).

[27] Le Hou, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel H Saltz. 2016. Patch-based convolutional neural network for whole slide tissue image classification. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2424–2433.https://doi.org/10.1109/CVPR.2016.266

[28] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu. 2013. Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential. IEEE Reviews in Biomedical Engineering 7 (December 2013), 97–114. https://doi.org/10.1109/RBME.2013.2295804

[29] jg Van Den Tweel and Clive Taylor. 2010. A brief history of pathology: Preface to a forthcoming series that highlights milestones in the evolution of pathology as a discipline. Virchows Archiv : an international journal of pathology 457 (May 2010), 3–10. https://doi.org/10.1007/s00428-010-0934-4

[30] Nirag Jhala. 2017. Digital pathology: Advancing frontiers. IEEE Signal Processing in Medicine and Biology Symposium (SPMB) (December 2017), 1–1.

[31] Lewis JSJ, Ali S, Luo J, Thorstad WL, and Madabhushi A. 2014. A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. The American Journal of Surgical Pathology 38, 1 (January 2014), 128–137. https://doi.org/10.1097/PAS.0000000000000086

[32] Upadhyaya Kafle and Santosh. 2018. Scope and Challenges of Digital Pathology Practice (DPP) in a Low Developing Country-Nepal. American Journal of Clinical Pathology 150 (September 2018), S118. https://doi.org/10.1093/ajcp/aqy099.283

[33] Dr. Keith J. Kapalan. 2020. Proscia. Retrieved June 14, 2020 from https://tissuepathology.com/2020/04/27/proscia-offers-concentriq-dx-for-remoteuse-in-clinical-practice-during-covid-19

[34] Amir Kermany and Frank Austin Nothaft. 2020. Databricks. Retrieved July 4, 2020 from https://databricks.com/blog/2020/01/31/automating-digitalpathology-image-analysis-with-machine-learning-on-databricks.html

[35] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A W M van der Laak, Bram van Ginneken, and Clara I Sánchez. 2017. A survey on deep learning in medical image analysis. Medical Image Analysis 42 (December 2017), 60–88. https://doi.org/10.1016/j.media.2017.07.005
[36] Yingci Liu and Liron Pantanowitz. 2019. Digital Pathology: Review of Current Opportunities and Challenges for Oral Pathologists. Journal of Oral Pathology Medicine 48, 4 (April 2019), 263–269. https://doi.org/10.1111/jop.12825

[37] Marit Lucas, Ilaria Jansen, C Dilara Savci-Heijink, Sybren L Meijer, Onno J de Boer, Ton G van Leeuwen, Daniel M de Bruin, and Henk A Marquering. 2019. Deep learning for automatic Gleason pattern classification for grade group determination of prostate biopsies. Virchows Arch 475 (May 2019), 77–83.

[38] Veta M, Kornegoor R, Huisman A, Verschuer-Maes AHJ, Viergever MA, Pluim JPW, and van Diest. 2012. Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer. Modern Pathology 25, 12 (December 2012), 1559–1565. https://doi.org/10.1038/modpathol.2012.126

[39] Editorial members of Digital Pathology. 2020. Leica Bio Systems. Retrieved February 28, 2020 from https://www.leicabiosystems.com/knowledgepathway/digital-pathology

[40] Ertosun MG and Rubin DL. 2015. Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks. In Proceedings of AMIA Annual Symposium. 1899–1908.

[41] Saarah Moore. 2020. News Medical Life Sciences. Retrieved March 2, 2020 from https://www.news-medical.net/life-sciences/History-of-Digital-Pathology.aspx

[42] Kainz P, Pfeiffer M, and Urschler. 2017. Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization. PeerJ 5 (October 2017). https://doi.org/10.7717/peerj.3874

[43] Kothari S, Phan JH, Stokes TH, and Wang MD. 2013. Pathology imaging informatics for quantitative analysis of whole-slide images. Journal of the American Medical Informatics Association 20, 6 (November 2013), 1099–1108. https://doi.org/10.1136/amiajnl-2012-001540

[44] Al-Janabi S, Huisman A, and Van Diest PJ. 2012. Digital pathology: Current status and future perspectives. Histopathology 61 (July 2012), 1–9. https://doi.org/10.1111/j.1365-2559.2011.03814.x

[45] Proscia. 2020. Visiopharm and proscia. Retrieved June 28, 2020 from https://www.prweb.com/releases/proscias_concentriq_dx_receives_ce_mark_for_primary_diagnosis/prweb16701974.htm

[46] Gretchen Roede. 2020. CISION PRweb. Retrieved June 19, 2020 from https://www.prweb.com/releases/proscias_concentriq_dx_receives_ce_mark_for_primary_diagnosis/prweb16701974.htm

[47] Proscia. 2020. Visiopharm and proscia. Retrieved June 28, 2020 from https://www.prweb.com/releases/proscias_concentriq_dx_receives_ce_mark_for_primary_diagnosis/prweb16701974.htm

[48] Al-Janabi S, Huisman A, and Van Diest PJ. 2012. Digital pathology: Current status and future perspectives. Histopathology 61 (July 2012), 1–9. https://doi.org/10.1111/j.1365-2559.2011.03814.x

[49] Kothari S, Phan JH, Stokes TH, and Wang MD. 2013. Pathology imaging informatics for quantitative analysis of whole-slide images. Journal of the American Medical Informatics Association 20, 6 (November 2013), 1099–1108. https://doi.org/10.1136/amiajnl-2012-001540

[50] Ram sagar. 2020. Analytics Magazine India. Retrieved May 28, 2020 from https://analyticsindiamag.com/google-introduces-new-machine-learning-tool-to-boost-digital-pathology
[51] Dr. Tim Sandle. 2020. Pharmaceutical Microbiology. Retrieved July 17, 2020 from https://www.pharmamicroresources.com/2019/06/new-technology-to-improve-tissue-imaging.html?m=0

[52] Luiz Silva, Paulo Saldava, and Venancio Alves. 2016. History and prospects of Pathology in Medicine. Revista de Medicina 95 (July-August 2016), 68–72. https://doi.org/10.11606/issn.1679-9836.v95ispe2p68-72

[53] David R J Snead, Yee-Wah Tsang, Aisha Meskiri, Peter K Kimani, Richard Crossman, Nasir M Rajpoot, Elaine Blessing, Klaus Chen, Kishore Gopalakrishnan, Paul Matthews, Navid Monttahan, Sarah Read-Jones, Shatrughan Sah, Emma Simmons, Bidis Sinha, Sari Suortamo, Yen Yeo, Hesham El Daly, and Ian A Cree. 2016. Validation of digital pathology imaging for primary histopathological diagnosis. Histopathology 68 (September 2016), 1063–1072. https://doi.org/10.1111/his.12879

[54] Jason Swedlow. 2020. CISION PRweb. Retrieved July 26, 2020 from https://www.prnewswire.com/news-releases/glencoe-softwares-pathviewer-digital-pathology-solution-easily-handles-multiplexed-imaging-data-300499758.html

[55] Hamid Tizhoosh and Liron Pantanowitz. 2018. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. Journal of Pathology Informatics 9 (November 2018), 9–38. https://doi.org/10.4103/jpi.jpi_53_18

[56] Jun Xu, Lei Xiang, Qingshan Liu, Hannah Gilmore, Jianzhong Wu, Jinghai Tang, and Anant Madabhushi. 2016. Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Transactions on Medical Imaging 35, 1 (January 2016), 119–130. https://doi.org/10.1109/TMI.2015.2458702

[57] Yan Xu, Zhipeng Jia, Liang-Bo Wang, Yuqing Ai, Fang Zhang, Maode Lai, and Eric I-Chao Chang. 2017. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. BMC Bioinformatics 18 (May 2017), 281. https://doi.org/10.1186/s12859-017-1685-x

[58] Chen XW and Lin X. 2014. Big data deep learning: Challenges and perspectives. IEEE Access 2 (2014), 514–525. https://doi.org/10.1109/ACCESS.2014.2325029

[59] Yifan Zhang, Hanbo Chen, Ying Wei, Peilin Zhao, Jiezhang Cao, Hailing Liu Jinlong Hou Xiao Han Jianhua Yao Qingyao Wu Xinjuan Fan, Xiaoying Lou, Mingkui Tan, and Junzhou Huang. 2019. From Whole Slide Imaging to Microscopy: Deep Microscopy Adaptation Network for Histopathology Cancer Image Classification. In Proceedings of Medical Image Computing and Computer-Assisted Intervention, Vol. 11764. 360–368. https://doi.org/10