

Digital Systems & Technology

## Applying Deep Learning to Transform Breast Cancer Diagnosis

Deep convolutional neural networks can assist pathologists in breast cancer diagnosis by automatically filtering benign tissue biopsies, identifying malignant regions and labeling important cellular features like nuclei for further analysis. Automatic detection of diagnostically relevant regions-of-interest and nuclei segmentation reduces the pathologist's workload, while ensuring that no critical region is overlooked, rendering breast cancer diagnosis more reliable, efficient and cost-effective.

# Executive Summary

According to the World Health Organization, breast cancer is the most commonly diagnosed cancer and is the leading cause of cancer deaths among women worldwide. On average, a woman is diagnosed with breast cancer every two minutes and one woman dies of it every 13 minutes worldwide.<sup>1</sup> In 2019, an estimated 268,600 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S. alone, along with 62,960 new cases of noninvasive (in situ) breast cancer,<sup>2,3,4</sup> with about 41,760 mortalities.

Since 1989, early detection and diagnosis have increased treatment success and survival rates. Screening or detection is generally conducted by self-examination or clinical breast palpation followed by mammography or ultrasound imaging. This typically identifies the presence of lesions/lumps that could be cancerous. Finally, conclusive breast tissue biopsy and histopathological analysis is done to ascertain the presence, type, grade and malignancy of cancer.

Pathologists typically use a light microscope to manually identify various cellular markers, such as nuclear atypia, tubule formation and mitotic cells in stained biopsy tissues. Such visual diagnosis, however, is very tedious and subjective, with average diagnostic concordance between pathologists being relatively low (approximately 75%).<sup>5,6</sup>

The recent introduction of slide scanners that digitize the biopsy into multi-resolution images, along with advances in deep learning methods, has ushered in new possibilities for computer-aided diagnosis of breast cancer. This will potentially automate all intermediate steps from localization of the tissue and focus plane selection to image enhancement, segmentation, annotation and quantitative diagnosis of histopathology slides using advanced forms of artificial intelligence (AI) such as machine learning (ML) and computer vision. These innovations help make diagnosis accurate, reliable, efficient and cost-effective.

Our deep-learning-based breast cancer grading approach uses convolutional neural networks (ConvNet), which automatically classifies the biopsy slide as benign or malignant and then identifies tumor regions with saliency maps for further analysis by the pathologist. Automatic detection of diagnostically relevant regions-of-interest (Rols) reduces the pathologist's workload while also assuring that no critical region is overlooked. Secondly, ConvNet segmentation models label cells and nuclei that the pathologist uses to identify atypical (misshaped, abnormal) nuclei, tubules and mitotic (dividing) nuclei.

Our approach is aimed at automating the tedious and error-prone pre-diagnostic steps pathologists perform manually, and to enable them to make faster and more accurate diagnoses.

The classification and segmentation models were trained on publicly available data sets. Since these data sets are small, special data augmentation and transfer learning techniques were used to offset the sparsity of training data.

The adoption of AI/ML in medical imaging will transform disease diagnosis, though regulatory and privacy concerns and the need for high reliability, security and safety need to be addressed.

This white paper highlights the advances in breast cancer histopathological diagnosis, our approach and the potential opportunity for AI/ML in computational pathology.

Automatic detection of diagnostically relevant regions-of-interest reduces the pathologist's workload while also assuring that no critical region is overlooked.

## Breast cancer 101

One in eight women are diagnosed with breast cancer in their lifetime.<sup>7</sup> According to the American Cancer Society, the average mortality rate for women diagnosed with breast cancer between 2008 and 2014 is 10%,<sup>8</sup> meaning that about one in 10 women diagnosed die from this disease. The impact of timely diagnosis can be seen from the fact that while the five-year breast cancer survival rate for localized stage (when cancer cells do not grow beyond the organ where they began) is 99%, it is 85% for regionalized stage (when cancer cells have spread beyond the organ where they began, but the spread is limited) and 27% for distant stage (when cancer cells have spread to other parts of the body).<sup>9</sup>

Awareness of symptoms and risk factors, regular self-examination and clinical screening are necessary to detect breast cancer early, thereby improving survival chances. Symptoms include a lump or swelling in the breast, an imbalance in the shapes or sizes of the two breasts, a difference in color of the breasts, nipple discharge, etc.<sup>10</sup> Common risk factors include age (over 40), family history, hormonal therapy, obesity, etc.

Clinical screening tests are intended to detect the breast cancer early, either before it causes symptoms or when in its early stages (i.e., when a lump is felt). Numerous screening tests for this purpose can be employed, including clinical and self-breast exams, mammography, genetic screening, ultrasound and magnetic resonance imaging (MRI).

Typical screening methods include:

- I Breast exam:** A clinical or self-breast exam involves feeling the breast for lumps or other abnormalities.
- I Mammography:** Screening mammography is a low-dose x-ray examination of the breast. Since it is relatively fast and widely available, it is the most common screening method for women with moderate risk after 40.
- I Magnetic resonance imaging:** MRI can be used to screen women who have a high risk of breast cancer. An MRI is more likely than mammography to find a breast mass that is not cancer.

## Breast cancer diagnosis process

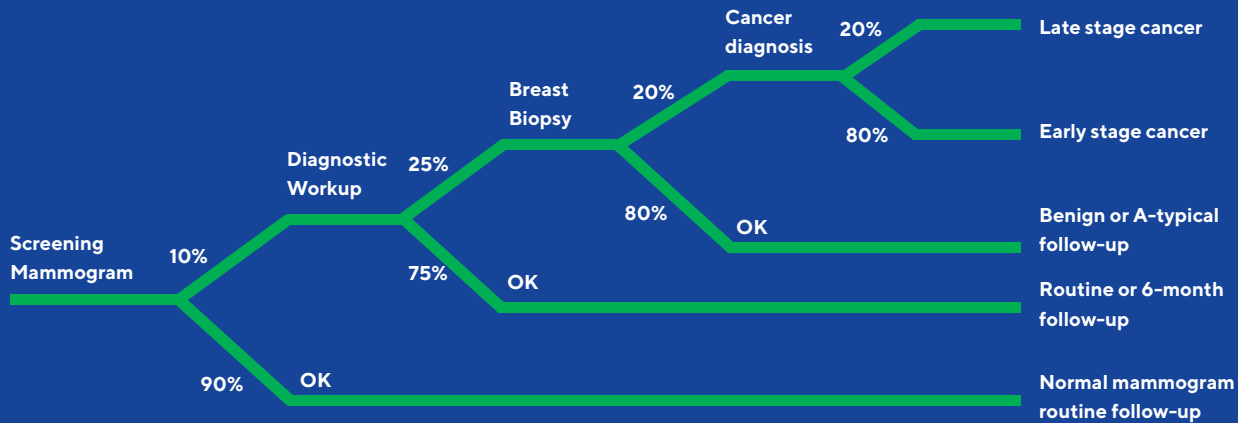


Figure 1

Source: <http://www.imagins.com/breast-cancer-diagnosis/overview-of-breast-cancer-diagnosis-decision-process>

Conclusive diagnosis after screening can only be done using breast tissue biopsy and histopathological analysis.

**I Biopsy:** Although some abnormalities may be typical characteristics of malignancy, screening tests like mammograms or MRIs alone cannot provide proof of a tumor. If mammography raises a significant suspicion of cancer, a biopsy is done. A breast biopsy involves removing samples of tissue for examination under a microscope. This is the only way breast cancer can be diagnosed definitively. Approximately 65–80% of breast biopsies reveal benign (non-cancerous) conditions.

Identifying the exact type and severity of the disease is essential for determining treatment options and disease prognosis. Figure 2 shows the various types of breast cancer.

## Breast cancer types

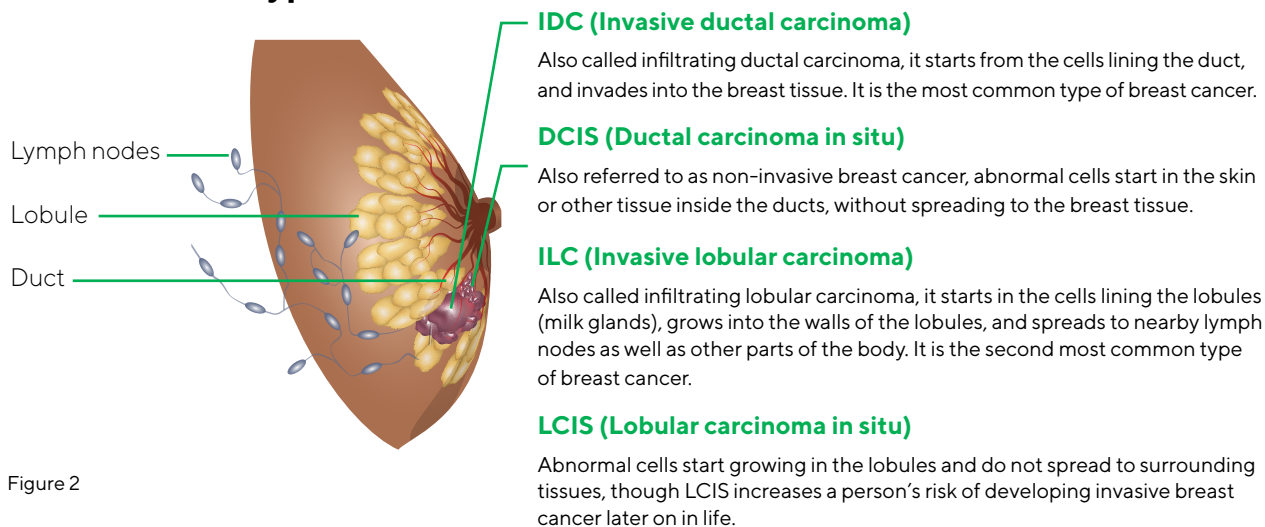


Figure 2



Severity is determined by the grade and stage of breast cancer. Grade refers to the “aggressive potential” of the tumor, while stage refers to the size and spread of the tumor to other parts of the body.

A tumor’s grade and stage determine the possible treatment options and prognosis. Treatment options include chemotherapy, radiation, partial or full mastectomy (surgical removal of breast tissue), hormonal therapy and molecularly targeted therapy.

## Biopsy, slide preparation & digitization

Breast tumor biopsies are performed in an operating room using fine needle aspiration (FNA), core needles or surgical incision – after which the collected tissue is sent for analysis to a pathology lab. The first step of the tissue preparation process is formalin fixation and embedding in paraffin. From the paraffin blocks, sections with a thickness of 3–5  $\mu\text{m}$  are cut using a microtome (a high precision cutting instrument) and mounted on glass slides. The structures-of-interest in the tissue, mainly the nuclei and cytoplasm, are not readily visible on the mounted sections. They therefore need to be dyed with stains that highlight them (see Figure 3).

### The anatomy of a biopsy

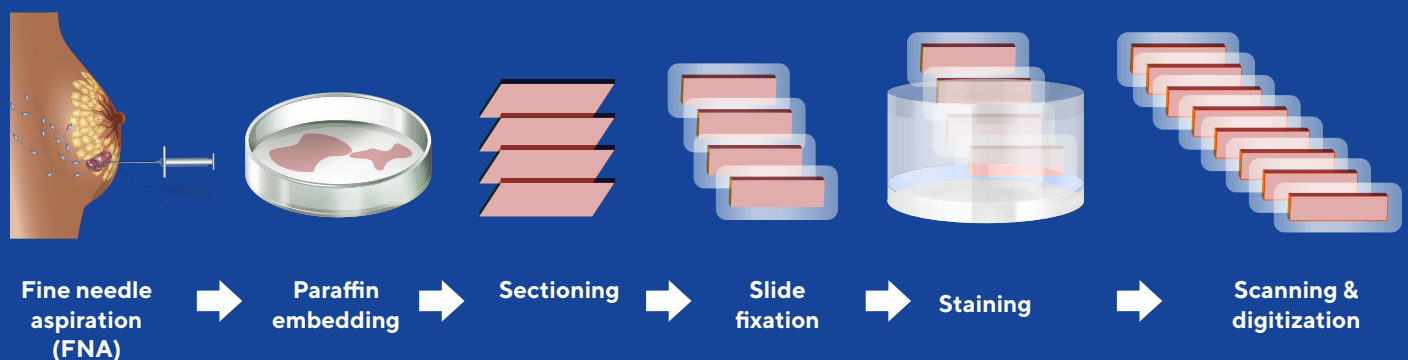
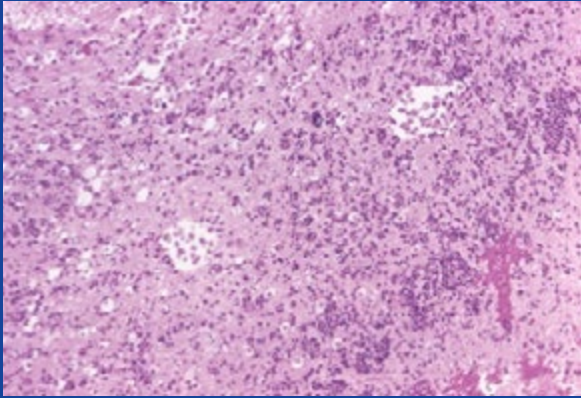


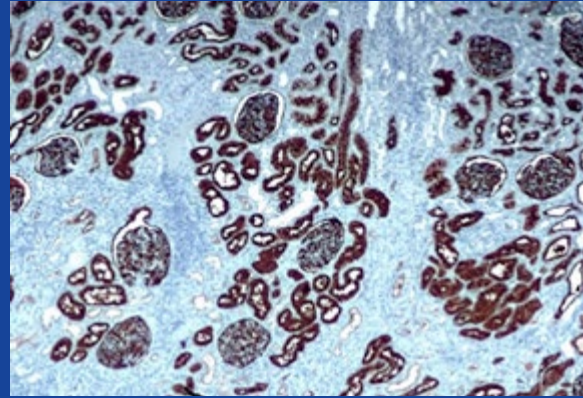
Figure 3

## H&E (hematoxylin and eosin) stained whole slide imaging (WSI)



Source: [http://www.publicdomainfiles.com/show\\_file.php?id=13546030819304](http://www.publicdomainfiles.com/show_file.php?id=13546030819304)

## IHC (immunohistochemical) stained whole slide imaging (WSI)



Source: [https://commons.wikimedia.org/wiki/File:Kidney\\_cd10\\_ihc.jpg](https://commons.wikimedia.org/wiki/File:Kidney_cd10_ihc.jpg)

Figure 4

Another common staining technique – immunohistochemistry (IHC) – is a more advanced staining technique which makes use of antibodies to highlight specific antigens in the tissue. In breast cancer, IHC is commonly used to highlight the presence of estrogen (ER), progesterone (PR) and human epidermal growth factor 2 (HER2) receptors, as well as to assess the proliferation of the tumor.

Pathology labs are currently transforming to a fully digital workflow that includes the digitization of histopathology slides and the use of digital monitors for viewing them. The ultimate aim of this approach, however, is to replace the optical microscope, which has been the rudimentary tool used by pathologists till now. State-of-the-art whole-slide imaging (WSI) scanners are tabletop devices that take glass slides as input and produce whole-slide digital images (sometimes referred to as digital or virtual slides) as output.

## Histopathology analysis & diagnosis

Examination of hematoxylin and eosin (H&E)-stained tissue under a microscope or using digital monitors is still the mainstay of pathology. The popularity of H&E is due to its low cost and ability to reveal tissue structure and nuclear morphology that is sufficient for primary diagnosis of breast cancer. Doing so, it eliminates 80% of cases as benign or atypical conditions.<sup>12,13</sup>

First, histological grading of the biopsy tissue is done. Histological grade<sup>14</sup> refers to the microscopic similarity of breast cancer cells to normal breast tissue. Low-grade tumor cells have an appearance similar to normal cells, while high-grade tumor cells appear distinctly abnormal. The Nottingham modification of the Bloom-Richardson grading system grades breast carcinomas by semi-quantitative assessment of nuclear pleomorphism (abnormal nuclei), tubule formation and mitotic (dividing cells) activity.<sup>15</sup>

## Grading breast cancer's severity

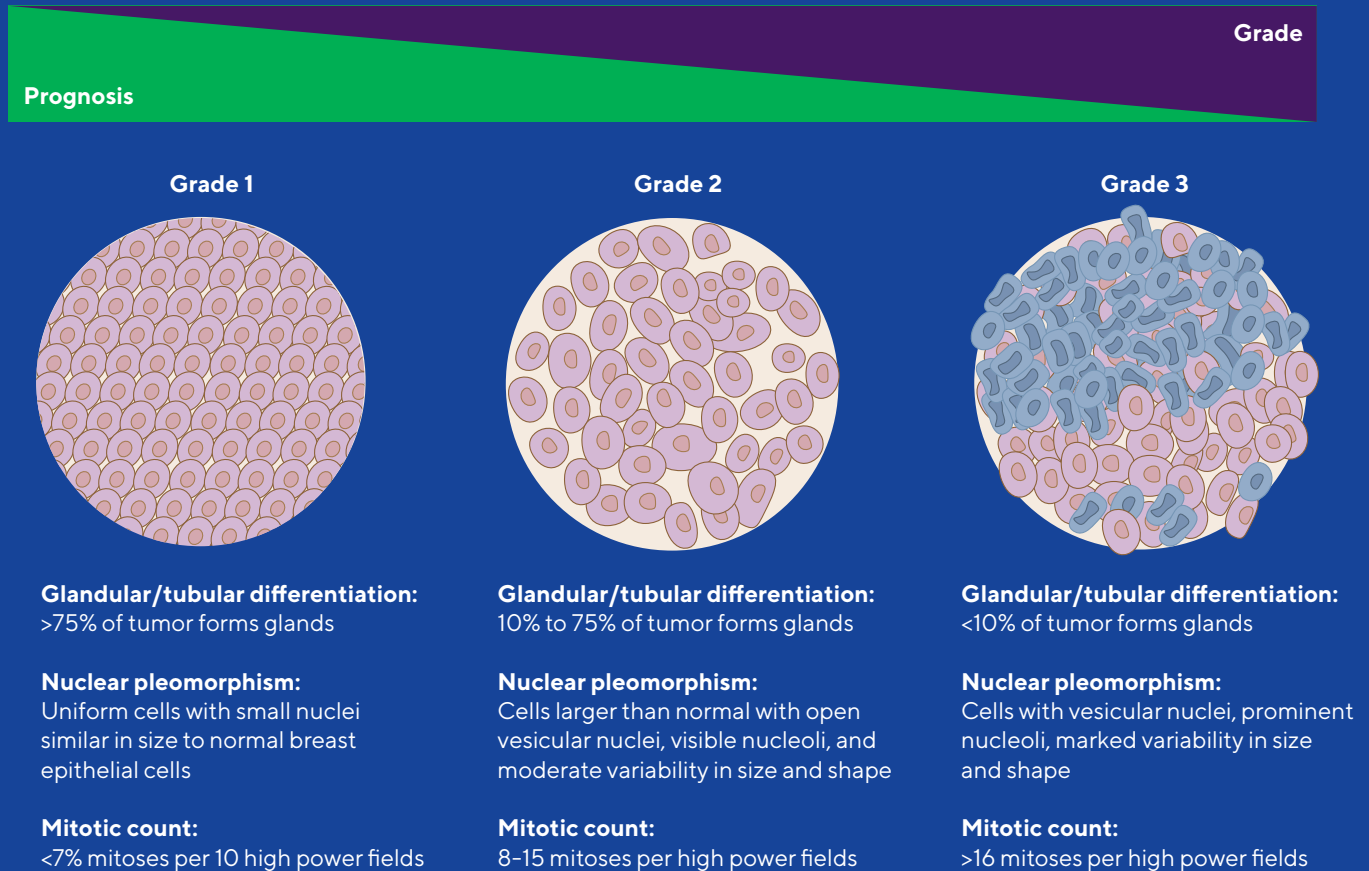


Figure 5




Source: <https://pathology.jhu.edu/breast/my-results/staging-grade>

The next step, clinical staging,<sup>16</sup> is the process to determine how much the cancer has spread and where it is located. It typically requires histopathological and multi-modal radiological imaging. It requires more detailed analysis than grading and involves multiple tests including hormone receptor status with IHC stains, biopsies of lymph, lung, liver and body fluids and other advanced tests including x-ray, MRI and CT scans. Staging the breast cancer is essential to make appropriate treatment choices. The staging system most often used for breast cancer is the American Joint Committee on Cancer (AJCC) TNM<sup>17</sup> classification system, which is based on the following:

- **T describes the size of the primary tumor** and whether it has invaded nearby tissue.
- **N describes nearby lymph nodes** that are involved.
- **M describes distant metastasis** (spread of cancer from one part of the body to another).
- **Estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) are also often used in staging.**



## The breast cancer grading methodology

Parameter	Score		
	1	2	3
 <b>Tubule formation</b>	>75%	10-75%	<10%
 <b>Nuclear pleomorphism</b>	Absent*	Moderate	Marked
 <b>Mitotic count**</b>	<9	9-17	>17

Final score	3	4	5	6	7	8	9
	Grade I			Grade II		Grade III	

Figure 6

Source: <http://ar.iiarjournals.org/content/35/7/4219/F1large.jpg>

## Challenges in breast cancer diagnosis

Breast cancer and other similar (though benign) conditions like cysts, calcifications, etc. are quite complex in nature and require multifactorial analysis that includes patient health, age, family history, race/ethnicity and finally histological grading and staging.

### Whole slide images at various resolutions

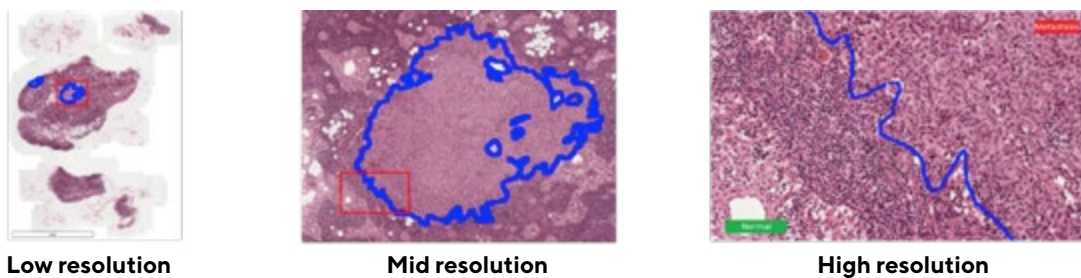


Figure 7

Source: <https://camelyon17.grand-challenge.org/Background/>

## A magnified view

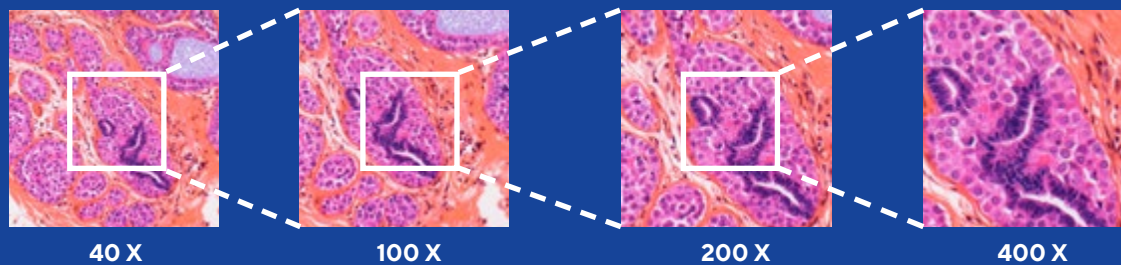


Figure 8

Source: <https://web.inf.ufr.br/vr/databases/breast-cancer-histopathological-database-breakhis/>

The diagnostic grading process using biopsy WSI is quite tedious as the images are so large; in fact, they are three to five times bigger than radiology images. A typical histopathology slide contains a tissue area of approximately 15x15 mm, but these high resolution images are of up to several gigapixels. Pathologists must cover all the regions of the slide at high magnification to identify and quantify cellular and nuclear markers according to the AJCC TNM classification system to grade and stage the tumor. This is very tedious and error-prone, with high observer variability,<sup>18</sup> and usually is referred to another pathologist for a second opinion.

Another challenge is the variations in tissue appearance partly due to different capturing, fixation and staining methods. Furthermore, an inherent heterogeneity in the appearance of epithelial cancerous nuclei, along with overlapping, clustered or clumped nuclei makes them difficult to analyze.

Assessing tumor grade in pathology labs depends mainly on mitosis counting. The task of identifying and counting mitotic nuclei is notoriously time-consuming and difficult, due to the fact that many other objects such as apoptotic and necrotic nuclei may have similar appearance, which renders it difficult even for trained experts to spot.

Another important indicator for advanced malignancy is the lack of tubule formation. Tubules are generally round or oval structures formed by a layer of epithelial cells and lumen. The major challenge in tubule segmentation is the similarity to other structures, such as adipose tissue, or tears, formed during the tissue preparation process.

Primarily because of the subjective nature of evaluating the three components that constitute the grading system – nuclear pleomorphism, tubule formation and mitotic count – histologic grading of breast cancer with average diagnostic concordance between pathologists is generally low (approximately 75%).

## AI approach for computer-assisted diagnosis

Our approach is aimed at automating tedious, error-prone pre-diagnostic grading steps that a pathologist performs manually. Key steps involved in breast cancer grading from FNA biopsy on are illustrated in Figure 9. We applied Deep ConvNets for localization and segmentation tasks, which the pathologist can use to perform further quantitative analysis and grading of the biopsy tissue.

Deep networks require large training data sets to train, whereas available public breast cancer datasets are small. This necessitates special methods to be viable. Data augmentation and transfer learning techniques are used to offset the training data sparsity.

### Histopathological analysis for breast cancer grading

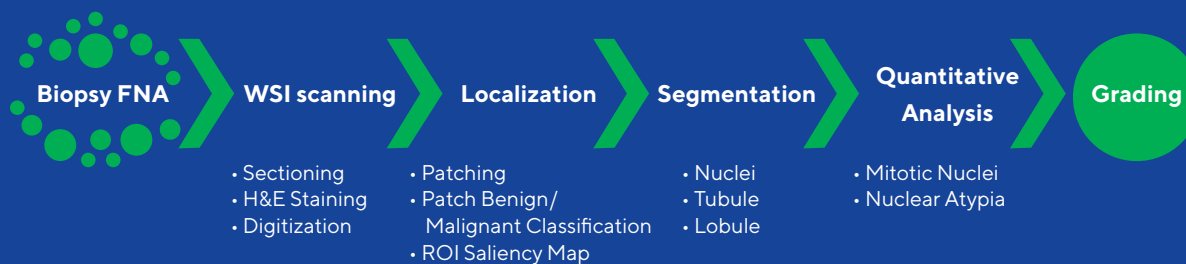
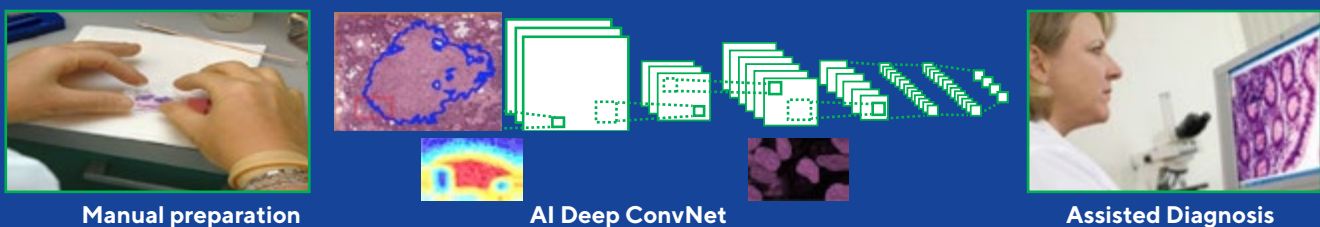


Figure 9

## Tumor localization

As mentioned earlier, a typical histopathology slide contains a tissue area of approximately 15x15 mm. At a resolution at which digital slides are captured, this results in image sizes up to several gigapixels. Processing of these very large images is computationally expensive, so common practice is to identify the regions of the slides that are of interest prior to performing more detailed image analysis. For histological grading, only the tumor tissue is of interest and non-tumor regions need to be excluded from the analysis.

Localization refers to identifying these RoIs, particularly areas of potential malignant tissue that require further analysis. For pathologists, this is a very tedious and time-consuming undertaking, since they have to methodically cover all parts of the large slide at different magnification factors and mark these areas for further quantitative analysis. Automatic detection of diagnostically relevant RoIs focuses pathologists' attention to those areas, thereby reducing their workloads while ensuring that no critical region is overlooked.

To train the ConvNets, the training data first is augmented by standard affine transformations like flipping, rotations, translation, etc. Then, the WSI images are normalized for brightness and staining variations, using mean image subtraction and local contrast normalization.

Next, we use transfer learning for classifying WSI and patch images as malignant or benign. Transfer learning is a deep learning technique where a model that is developed for a different task and trained on large data sets such as ImageNet<sup>19</sup> (1.2 million images) is repurposed for a different task that has limited data to train on – in our case, breast cancer grading.

A fine-tuned ConvNet reuses pretrained layers that were trained on a large data set like ImageNet, followed by a custom classification (softmax) layer that is trained on the smaller breast cancer datasets (see Figure 10). This technique is used both for WSI and patch level benign/malignant classification.

## Fine-tuned ConvNet for WSI and patch classification

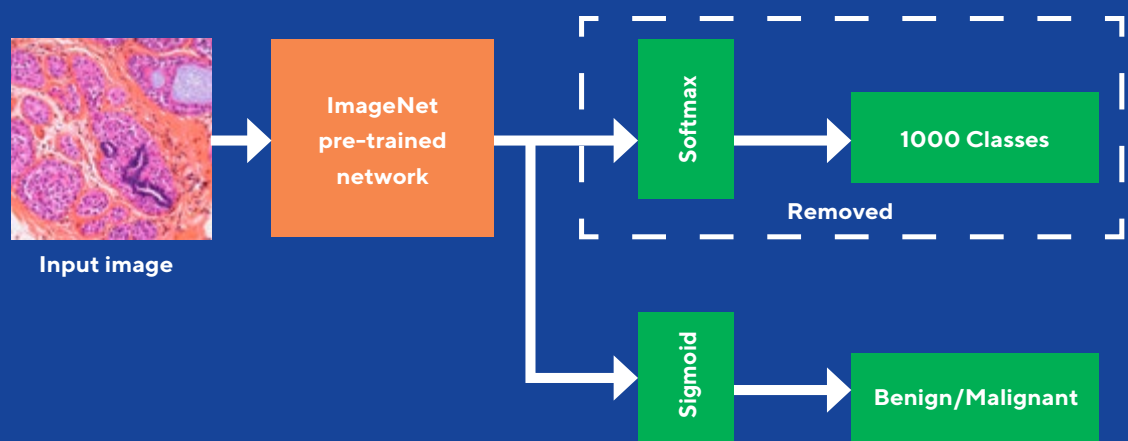


Figure 10

## Breast cancer saliency map to segment malignant regions

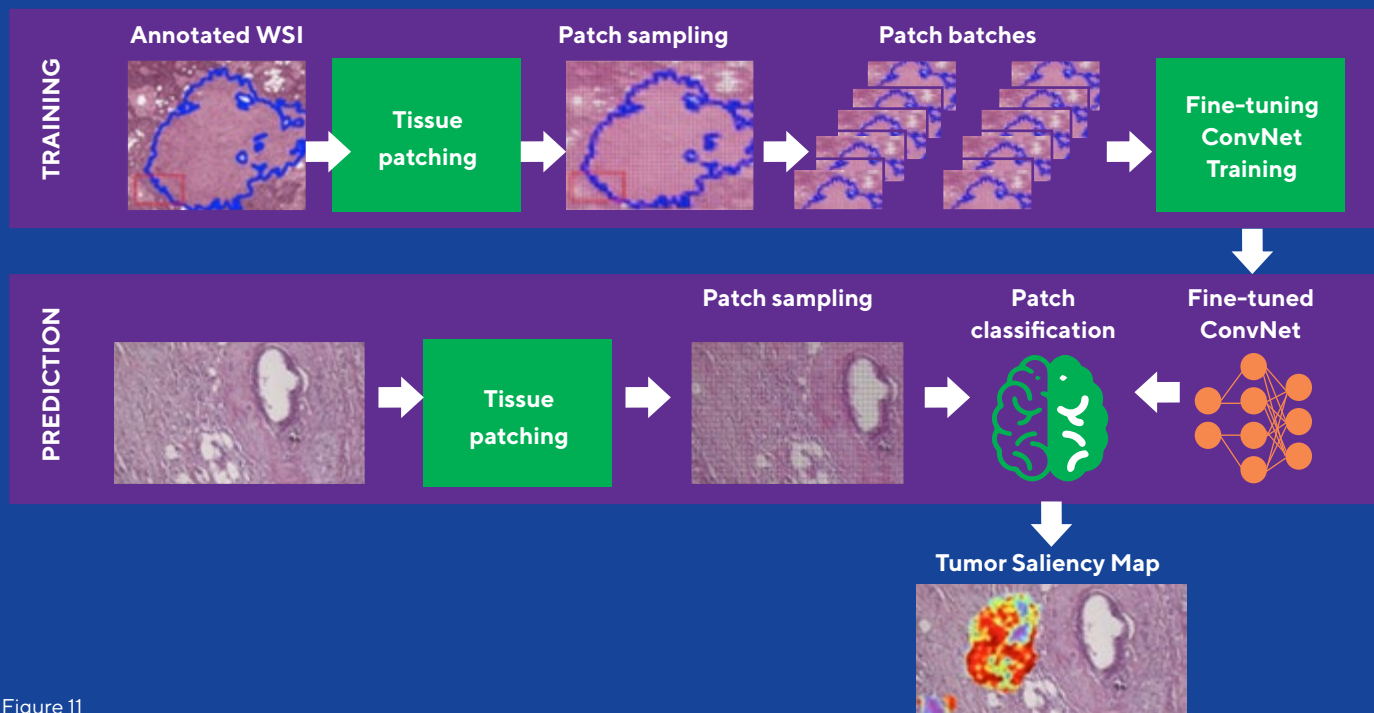


Figure 11

Our RoI localization approach is a three-step process:

- 1. WSI benign/malignant classification:** Classify WSIs as benign or malignant using a fine-tuned deep ConvNet trained on the BreakHis<sup>20</sup> dataset. The model is optimized for sensitivity or true positive rate, a metric that measures how correctly actual malignant slides are labeled as malignant by the classifier. The fine-tuned ConvNet achieved state-of-the-art sensitivity of 99.04% and accuracy outperforming all other previous models with 98.5±1.25%. This classifier was used for identifying whole slides that are malignant, while slides marked benign with high confidence are filtered out.
- 2. Patch benign/malignant classification:** A single WSI often contains multiple areas with different levels of diagnostic importance. The pathologist's final assessment can be based on even a small region that manually is easy to miss. Patch level classification models, though not as accurate as WSI classification models, help the pathologist decipher the cancerous tissue at a granular level, thereby reducing false positives. For this, patches of 256 x 256 pixels were created from the WSI from the Camelyon<sup>21</sup> data set. Equal samples of malignant and other patches (benign, empty, stroma, necrosis, etc.) were used to train another fine-tuned ConvNet benign/malignant patch classifier.

Since the number of patches per WSI is quite large, an intelligent sampling mechanism is used for selected patches using lower magnification to eliminate empty areas and stroma. Next, a seed set of patches selected with large strides are classified into benign or malignant with moderate magnification. Subsequently, at high magnification, patches close to malignant seed patches are selected for classification, thereby creating aggregate regions of malignant areas.



- 3. Malignant region segmentation:** Finally, at high magnification, a fine-tuned classifier – again optimized for sensitivity – produces a saliency map based on classifier confidence to segment the malignant region for further analysis (see Figure 11).

## Nuclei segmentation

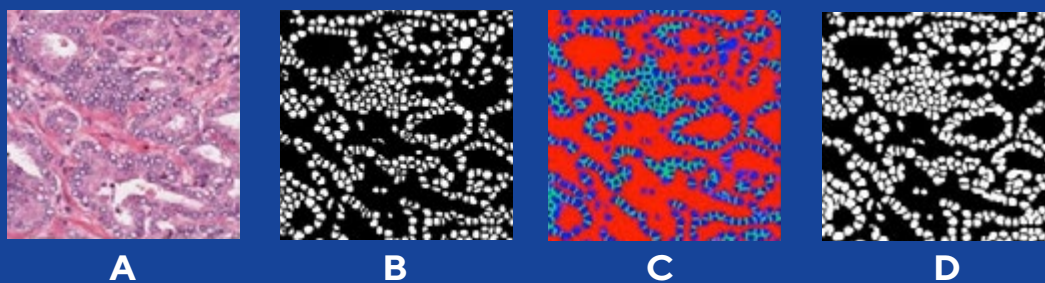
The segmentation of nuclei in breast cancer histopathology images is a basic step for extraction of relevant nuclei morphometric features (size, shape, chromatin texture). Such extraction is necessary for identifying nuclear pleomorphism, counting mitotic nuclei and assessing the extent of tubule formation.

Nuclei segmentation remains a very challenging problem, particularly for routinely stained H&E sections, due to the variability of tissue appearance due to imperfections in the staining process. Additionally, nuclei may be overlapping, clustered or tightly clumped, which makes them difficult to distinguish.

MoNuSeg<sup>22</sup> is a multi-organ nuclei segmentation data set of H&E stained tissue images captured at 40x magnification. Due to the diversity of nuclei appearances across multiple organs and patients, and the richness of staining protocols adopted at multiple hospitals, the training data set enables the development of robust and generalizable nuclei segmentation models.

In our nuclei segmentation approach, images are preprocessed by color normalization and training data is augmented using random cropping, flipping, rotation, scaling, elastic deformation and adding Gaussian noise. We leveraged U-Net<sup>23</sup> network architecture for nuclei segmentation trained on the MoNuSeg grand challenge data set. U-Net is a convolutional neural network that is developed for biomedical image segmentation. The network is based on the fully convolutional network and its architecture is modified and extended to work with fewer training images and to yield more precise segmentations (see Figure 12).

## Nuclei segmentation with U-Net



**A)** Tissue image, **B)** corresponding training mask with segmented nuclei, **C)** predicted output with segmented nuclei with overlap boundaries, **D)** predicted output with segmented nuclei with overlaps removed.

Figure 12

Tubule detection and segmentation is even more challenging, primarily due to a lack of publicly available annotated training data sets and its diverse manifestations. Currently, this must be analyzed and assessed by pathologists, and represents an opportunity in the future for automated tubule detection and segmentation using AI/ML and computer vision.

## Quantitative analysis & grading

Pathology is still mostly a subjective, semi-quantitative scientific discipline that is performed by expert human pathologists. Grading nuclei pleomorphism and mitotic staging (especially S-Phase<sup>24</sup> detection and tubule formation) are specialized tasks still performed by expert pathologists. This may change with the availability of annotated data sets, evolving trust and adoption with further AI advances.

Also, automated image analysis of H&E-stained tissue sections have failed to provide general reproducible results because of the variation in the staining intensity of the slide. Variability in the staining intensity and other steps of tissue collection and preparation significantly influence the outcomes of automated image analysis, whereas pathologists are able to effectively integrate these artifacts and differences between slides in their decisional process.

Our approach reduces the pathologist's workload by automatic detection of diagnostically relevant Rols (localization) and nuclei segmentation, with further scope for AI-based automation of other breast cancer grading and staging tasks.

### Benefits

Our approach (in limited proofs of concepts and pilots with hospitals and other clinical research partners) enables pathologists to be more efficient, consistent and accurate by using AI to automate error-prone steps of the breast cancer grading process. More specifically, it achieves the following:

- I High reliability and accuracy**, leading to fewer misdiagnoses (false positives) or missed diagnoses (false negatives).
- I Reduced diagnostic variability**, leading to informed treatment and decision-making, hence facilitating better outcomes and prognoses.
- I Automatic Rol detection and nuclei segmentation** decreases the pathologist's workload, making them more efficient and enabling them to devote their time to more complex tasks.
- I AI-based aids/tools for computational pathology** can be applied to similar types of cancer with minimal changes.
- I Lower cost of diagnosis** due to higher efficiency and reliability, hence decreasing the overall cost of diagnosis.

## Looking ahead

AI- and ML-based technologies have the potential to transform healthcare by delivering new and important insights from the vast amount of clinical data. High-value applications include faster disease detection, more accurate diagnosis, and the development of personalized diagnostics and therapeutics. The availability of digitized wellness, clinical and medical imaging data along with the widespread success of deep learning methods is accelerating this transformation. From radiology and drug discovery to disease risk prediction and patient care, deep learning is transforming healthcare from every angle.

Diagnostic services currently are strained by growing and aging populations, with physicians and technicians exposed to long hours and high-stress work environments.<sup>25</sup> Another gap is the shortage of highly trained and experienced physicians and technicians. Traditional manual methods of diagnosis over the wide range of life-threatening human ailments are time-consuming, error-prone and expensive.







AI will transform the diagnostic medical imaging industry too, in terms of improved diagnostic accuracy, better treatment planning and clinical outcomes, and increased productivity. AI will play a key role in enabling diagnostic departments to cope with the ever-increasing volume of diagnostic imaging procedures, despite the chronic shortage of trained technicians. AI-based software-as-a-medical-device (SaMD)<sup>26</sup> diagnostics will leverage digitized clinical data – patient medical records and medical images – to identify patterns for faster, cheaper and more accurate diagnosis and treatment. Accelerated adoption of AI/ML in healthcare and life sciences is therefore critical to building digitized disease diagnosis solutions to help enable early diagnosis and effective patient care.<sup>27,28</sup>

AI-based pathology and oncology diagnostics, though, are still in a nascent state and represent an opportunity. Pathology is an excellent candidate for AI, especially for labor-intensive tasks such as histologic cell counts, structure and morphological estimation, etc. Computational pathology leverages the power of AI, ML, image analytics and clinical big data integration to enhance the diagnostic precision of pathologists. Future directions in this area include enhanced histopathological imaging, clinical decision support and even pathology diagnosis at home. Similarly, clinical decision support leverages statistical and ML models that can assimilate information from various sources such as pathology, cytology, radiology, etc. to facilitate accurate and timely decision-making.

In the recent past, we have developed AI/ML-based solutions for skin cancer and diabetic retinopathy detection using retinal fundus photographs.<sup>29</sup> AI/ML is beginning to transform the disease diagnosis and treatment processes. Challenges such as the reliability of diagnosis, patient safety, privacy and security must still be addressed to realize the technology's full potential and mass adoption.

## Endnotes

- <sup>1</sup> [www.nationalbreastcancer.org/breast-cancer-facts](http://www.nationalbreastcancer.org/breast-cancer-facts).
- <sup>2</sup> [www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf](http://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf).
- <sup>3</sup> U.S. Breast Cancer statistics, [www.breastcancer.org/symptoms/understand\\_bc/statistics](http://www.breastcancer.org/symptoms/understand_bc/statistics).
- <sup>4</sup> American Cancer Society, <https://cancerstatisticscenter.cancer.org/>.
- <sup>5</sup> Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson ANA, et al. Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. 2015 Mar;pp,1122-1132.
- <sup>6</sup> Variability in Pathologists' interpretations of individual breast biopsy slides: A population perspective. 2016 Joann Elmore et al
- <sup>7</sup> [www.nationalbreastcancer.org/breast-cancer-facts](http://www.nationalbreastcancer.org/breast-cancer-facts).
- <sup>8</sup> [www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html](http://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html).
- <sup>9</sup> <https://ww5.komen.org/BreastCancer/ChancesForSurvivalBasedOnCancerStage.html>.
- <sup>10</sup> [www.cancercenter.com/cancer-types/breast-cancer/symptoms](http://www.cancercenter.com/cancer-types/breast-cancer/symptoms).
- <sup>11</sup> Stathonikos, et al., 2013, [www.ncbi.nlm.nih.gov/pmc/articles/PMC3709427/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3709427/).
- <sup>12</sup> [www.imaginis.com/breast-cancer-diagnosis/overview-of-breast-cancer-diagnosis-decision-process](http://www.imaginis.com/breast-cancer-diagnosis/overview-of-breast-cancer-diagnosis-decision-process).
- <sup>13</sup> [www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf](http://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf).
- <sup>14</sup> <https://pathology.jhu.edu/breast/my-results/staging-grade>
- <sup>15</sup> Elston et al. (1991); Rakha et al. (2010), <https://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr2607>.
- <sup>16</sup> <https://pathology.jhu.edu/breast/my-results/staging-grade>.
- <sup>17</sup> [www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/stages-of-breast-cancer.html](http://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/stages-of-breast-cancer.html).
- <sup>18</sup> Meyer et al. (2005); Perez et al. (2006), [www.sciencedirect.com/science/article/pii/S0344033811802633](http://www.sciencedirect.com/science/article/pii/S0344033811802633).
- <sup>19</sup> The ImageNet dataset is a large image database designed for training visual object recognition models, with over 14 million images that have been hand-classified into 20,000+ categories. Large datasets like ImageNet have enabled deep learning models to be adapted to other tasks using transfer learning techniques where annotated data is difficult to obtain.
- <sup>20</sup> F. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Transactions of Biomedical Engineering, Volume: 63, Issue: 7, July 2016, pp. 1455 – 1462.
- <sup>21</sup> <https://camelyon17.grand-challenge.org/Data/>.
- <sup>22</sup> <https://monuseg.grand-challenge.org/Data/>.
- <sup>23</sup> <https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/>.
- <sup>24</sup> [https://en.wikipedia.org/wiki/S\\_phase](https://en.wikipedia.org/wiki/S_phase).
- <sup>25</sup> [www.prnewswire.com/news-releases/strained-healthcare-system-hurts-doctors-and-affects-patient-care-300522622.html](http://www.prnewswire.com/news-releases/strained-healthcare-system-hurts-doctors-and-affects-patient-care-300522622.html).
- <sup>26</sup> The term software as a medical device is defined by the International Medical Device Regulators Forum (IMDRF) as “software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.”, [www.fda.gov/medical-devices/digital-health/software-medical-device-samd](http://www.fda.gov/medical-devices/digital-health/software-medical-device-samd).
- <sup>27</sup> [www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device](http://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device).
- <sup>28</sup> <https://medium.com/syncedreview/ai-powered-fda-approved-medical-health-projects-a19aba7c681>.
- <sup>29</sup> [www.cognizant.com/whitepapers/how-ai-enhances-and-accelerates-diabetic-retinopathy-detection-codex2972.pdf](http://www.cognizant.com/whitepapers/how-ai-enhances-and-accelerates-diabetic-retinopathy-detection-codex2972.pdf).



## About the authors



### Harpreet Kanwar

#### CTO, Cognizant Life Sciences

Harpreet Kanwar is Chief Technology Officer within Cognizant's Life Sciences Practice. He is responsible for driving innovation, emerging technology adoption, architecture enablement and governance. Harpreet currently focuses on helping life sciences companies adopt automation, AI/ML/NLP/deep learning and blockchain technologies. He has over 25 years of IT industry experience covering consulting, pre-sales and delivery. Harpreet can be reached at [Harpreet.Kanwar@cognizant.com](mailto:Harpreet.Kanwar@cognizant.com) | [linkedin.com/in/harpreetkanwar](https://www.linkedin.com/in/harpreetkanwar).



### Praveen Koduganty

#### Principal Architect, Cognizant Global Technology Office

Praveen is a senior member of Cognizant's Global Technology Office, where he leads teams in developing innovative and cutting-edge solutions using emerging technologies. He is an avid technologist with over 17 years of product and solution development roles, focusing on enterprise architecture and apps, AI, NLP, enterprise search and computer vision. Praveen can be reached at [Praveen.Koduganty@cognizant.com](mailto:Praveen.Koduganty@cognizant.com) | [linkedin.com/in/pkoduganty](https://www.linkedin.com/in/pkoduganty).



### Saravanan Radhakrishnan

#### Architect, Cognizant Global Technology Office

Saravanan R is an Architect within Cognizant's Global Technology Office. He has over 13 years of experience in innovation, research and product development in computer vision, AI and data science. Saravanan obtained a master's degree in computer applications from Madras University. He can be reached at [Saravanan.Radhakrishnan@cognizant.com](mailto:Saravanan.Radhakrishnan@cognizant.com) | [linkedin.com/in/saravanan-radhakrishnan-4308579](https://www.linkedin.com/in/saravanan-radhakrishnan-4308579).



### Adya Jha

#### Senior Business Analyst, Global Technology Office

Adya Jha is a Senior Business Analyst within Cognizant's Global Technology Office. She obtained a bachelor's degree in technology in computer sciences and engineering from the National Institute of Technology, Arunachal Pradesh, and a post-graduation degree in business management from the Institute of Management Technology, Ghaziabad. She can be reached at [Adya.Jha@cognizant.com](mailto:Adya.Jha@cognizant.com) | [linkedin.com/in/adya-jha-6989bb5a](https://www.linkedin.com/in/adya-jha-6989bb5a).

## About Cognizant's Global Technology Office

Cognizant's Global Technology Office (GTO) is a research and development incubator that explores how emerging technologies can be applied to solve our clients' business challenges – today and tomorrow. We partner with leading academic and research institutions worldwide, as well as our business units and practices, to research and experiment with a cornucopia of technologies – from advanced forms of artificial intelligence and swarm robotics, through quantum technologies and mixed reality – via pilots and proofs of concept that determine their real-world application and potential business benefit. To learn more, contact us at [inquiry@cognizant.com](mailto:inquiry@cognizant.com).

## About Cognizant Life Sciences

Cognizant's Life Sciences business unit is committed to helping change millions of lives for the better by partnering with clients to build solutions to healthcare challenges, continually improve the way they do business, set the pace in clinical development, strengthen their regulatory infrastructure and increase competitiveness. Cognizant serves the top 30 global pharmaceutical companies, 9 of the top 10 biotech companies and 12 of the top 15 medical device companies. With a large team of dedicated professionals including doctors, pharmacologists, physicians, biomedical engineers, pharmacists, biostatisticians, medical writers and GxP consultants, the practice provides domain-aligned consulting, IT, business process, software and as-a-service solutions globally. Visit us at <http://cognizant.com/life-sciences>.

## About Cognizant

Cognizant (Nasdaq-100: CTSH) is one of the world's leading professional services companies, transforming clients' business, operating and technology models for the digital era. Our unique industry-based, consultative approach helps clients envision, build and run more innovative and efficient businesses. Headquartered in the U.S., Cognizant is ranked 195 on the Fortune 500 and is consistently listed among the most admired companies in the world. Learn how Cognizant helps clients lead with digital at [www.cognizant.com](http://www.cognizant.com) or follow us @Cognizant.

# Cognizant

### World Headquarters

500 Frank W. Burr Blvd.  
Teaneck, NJ 07666 USA  
Phone: +1 201 801 0233  
Fax: +1 201 801 0243  
Toll Free: +1 888 937 3277

### European Headquarters

1 Kingdom Street  
Paddington Central  
London W2 6BD England  
Phone: +44 (0) 20 7297 7600  
Fax: +44 (0) 20 7121 0102

### India Operations Headquarters

#5/535 Old Mahabalipuram Road  
Okkiyam Pettai, Thoraipakkam  
Chennai, 600 096 India  
Phone: +91 (0) 44 4209 6000  
Fax: +91 (0) 44 4209 6060

© Copyright 2019, Cognizant. All rights reserved. No part of this document may be reproduced, stored in a retrieval system, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the express written permission from Cognizant. The information contained herein is subject to change without notice. All other trademarks mentioned herein are the property of their respective owners.