IDC OPINION

Current global efforts to improve the clinical and financial outcomes in healthcare require innovation to be successful. Technology, and in the case of this research artificial intelligence (AI) and deep learning, can play a key role in that innovation – from improving the timeliness and accuracy of diagnoses to informing the treatment of diseases to helping address issues of physician shortages. AI and deep learning can have significant impact on healthcare organizations and the ecosystem at large as they are incorporated into ongoing digital transformation initiatives.

IDC defines digital transformation as the continuous process by which organizations adapt to or drive disruptive changes in their external ecosystem, leveraging digital competencies to improve efficiencies and performance. By 2019, 40% of digital transformation initiatives will be supported by cognitive/AI capabilities, providing critical on-time insights.

While deep learning is at the early stage in the adoption cycle, healthcare organizations must recognize that the maturity of the technology is not the gating factor to adoption; rather, the skepticism of physicians, the regulatory environment, reimbursement policy, and other issues must be addressed to move the application of deep learning from bench to bedside.

This white paper explores the application of deep learning in two key areas of healthcare: the augmentation or assistance to physicians and predicting the onset of an illness or adverse health event. In all cases, there are initiatives across the broad adoption spectrum, from early stage to production. You will note that those applications of deep learning that are closest to the delivery of care are in earlier stages of adoption than other applications as they show promise in improving clinical and financial outcomes.

WHAT IS DEEP LEARNING?

Deep learning is a type of machine learning where the learning happens in successive layers — each layer of the neural network adding to the knowledge of the previous layer. Each layer teaches the next successive layer. Deep learning trains the machine to do what the human brain does naturally. Feeding unlimited data to an algorithm is successful only when feature extraction is present to guide the machine about the specific learning objects. In traditional machine learning, the programmer had to intervene and guide the machine on which features it should be looking for in the learning process. The burden on the programmer is significant, but more significant is the potential for human error. In deep learning, programmer intervention is not required, resulting in increased accuracy.
Deep learning is a type of machine learning that leverages representation learning methodologies to allow computers to process natural data in its raw form using neural network algorithms. It takes advantage of both supervised and unsupervised machine learning:

- **Supervised machine learning** begins with examples of training data paired with identifying labels (e.g., right or wrong and positive or negative) selected from the categories to be learned. Using these pairs of example data and labels (training data), the system learns parameters of statistical models that it can then generalize to unlabeled examples of data items that were not seen in the training data (test data). In most cases, the learned models improve over time via a feedback loop that adjusts the model parameters to better reflect additional sets of training or production data. The performance of a learned model can be measured by simple prediction accuracy or by the business metric the learned model is designed to support. Performance depends on the degree to which the training data matches the real world, the choice of algorithm, the algorithm’s parameters, and the quantity of data.

- **Unsupervised machine learning** is another variation of machine learning where algorithms detect and discern attributes and features without the benefit of labeled training data. Some algorithms cluster data into meaningful groups by finding centers of data density. Other unsupervised algorithms use dimensionality reduction techniques (such as singular-value decomposition) to uncover the essential attributes of the data without requiring a human to define those attributes in advance. This is particularly useful for "unstructured" data, such as images or text, where an underlying structure can be automatically inferred, enabling other algorithms to leverage the data.

Chief among the benefits of deep learning is lack of programmer intervention, which improves the accuracy of results. Healthcare data is diverse, comprising both structured and unstructured data, both of which can be ingested by deep learning. This ability is driving the growing adoption of deep learning in healthcare, particularly in the radiology and pathology domains where there is voluminous data in the form of images, clinician notes, reports, and tissue slides.

**APPLICATION OF DEEP LEARNING IN HEALTHCARE**

Deep learning in healthcare is augmenting clinical decision making in areas ranging from analyzing medical research findings and best practices to prioritize and recommend treatment options to detecting abnormalities in radiology images and pathology slides to identifying genomic markers in tissue samples. In predictive analytics, deep learning is being applied to the early detection of disease, the identification of clinical risk and its drivers, and the prediction of future hospitalization. While there are opportunities for the application of deep learning in other aspects of healthcare, this white paper focuses on clinician augmentation and prediction of the risk of disease and/or adverse clinical events.

**Augment Clinicians**

Among the many drivers to augment clinicians’ abilities are the overwhelming volume and rapid pace of new treatment options, the high volume of screenings, the urgency to get results when disease is suspected and, in some places, physician shortages. These can be addressed with the application of deep learning. Improving the accuracy of evaluating radiologic images and pathology slides and the ability to organize fragmented patient data and deliver it in the context of a physician’s inquiry are all benefits that can be realized.
Reduce Cognitive Burden

One of the key advantages of deep learning is its ability to digest large volumes of data and create models that can determine the impact on clinical decisions. It is virtually impossible for physicians to stay abreast of advances in medical care in real time because of an explosion of medical information. For example, currently, approximately 50,000 oncology research papers are published annually,¹ and by 2020, medical information is projected to double every 73 days,² outpacing the ability of humans to keep up with the proliferation of medical knowledge. Applying deep learning to voluminous data enables the model to synthesize millions of data points and draw out correlations that can guide clinical decisions. This reduces the cognitive burden on physicians, enabling them to keep pace with current research and treatment protocols.

Improve Accuracy and Speed to Diagnoses and Treatment

Another area where deep learning has shown promising results is in enabling physicians to improve the accuracy and speed of diagnosis. Medical images, pathology samples, and genomics information are among the most important tools doctors use in diagnosing conditions. However, analyzing this data can often be a difficult and time-consuming process. Deep learning can assist by automating some of the analysis, screening for abnormalities or areas of interest within the medical data rapidly, and enabling physicians to better focus their expertise. These combined approaches not only can achieve improved accuracy and speed but could also increase the objectivity of diagnosis as well as concordance rate.

Researchers at Mayo Clinic concluded that using deep learning could lead to an earlier and more accurate way to diagnose and treat brain tumors. It is also a way to track progress or response to treatment without surgery. Using tissue from the tumor and a medical resonance imaging (MRI), the team identified a type of chromosomal damage that is important for predicting how well patients with a low-grade type of tumor will respond to chemotherapy and radiation. The research was conducted on 155 patients with glioblastoma multiforme, which is the most common primary brain tumor, accounting for 45% of all malignant primary central nervous system tumors, and with a median survival of around 14 months. The work, called radiogenomics, reflects the thought that genomic properties of tumors can be determined in the appearance of images (in this case, medical resonance imaging).³

Another area where artificial intelligence is being applied is cardiology, where deep learning is being used with MRIs to measure the volume of blood transported with each pump of the heart's four chambers. Cardiologists typically need 30-60 minutes to calculate the volume; AI comes up with the answer within seconds.

Pathology is another area that benefits from the application of deep learning to improve the accuracy of diagnoses. A pathologist's report is often the gold standard in the diagnosis of many diseases. Review of pathology slides is a very complex task, requiring years of training to gain the required expertise and experience. Even with this extensive training, there can be substantial variability in the diagnoses given by different pathologists for the same patient. For example, agreement in diagnosis for some forms of breast cancer can be as low as 48% and similarly low for prostate cancer. The lack of agreement is not surprising given the massive amount of information that must be reviewed to make an accurate diagnosis. Pathologists are responsible for reviewing all the biological tissues visible on a slide. However, there can be many slides per patient, the equivalent of a thousand 10MP photos where every pixel is important. Most biopsies are negative (e.g., 70% of breast biopsies are negative) and would warrant consideration of the use of technology to evaluate images and tissue samples.
Patients would be better served if radiologists and pathologists turned their attention to questionable and more complex cases.

Combining the knowledge of clinicians with deep learning improves the accuracy of diagnosis for breast cancer. For example, MIT and Harvard researchers teamed up with clinicians and Beth Israel Deaconess and Massachusetts General Hospital to evaluate the effectiveness of automated evaluation of lymph node tissue in patients with breast cancer versus results from pathologists. The automated diagnostic method was accurate 92% of the time; pathologists alone were accurate 96% of the time. By combining the automated approach with pathologists’ result, an accuracy rate of 99.5% was achieved. This represents a significant reduction in errors. Clinicians will not be replaced by the application of deep learning, but rather we are seeing increased evidence of the benefit deep learning provides to assist clinicians. This is important in high-volume conditions where death rates are significant such as breast cancer. Breast cancer is the number 1 cause of death in women worldwide; in 2017, it is estimated that there will be over 40,000 deaths attributed to breast cancer in the United States.

Deep learning algorithms are ingesting current medical literature and best practices and patient-specific data to provide treatment recommendations to physicians. Results from deep learning are showing a high degree of confluence with recommendations of hospital tumor boards. S. P. Somashekhar, MD, an oncologist and chairman of the Manipal Comprehensive Cancer Center, Manipal Hospitals, in Bengaluru, India, recently released the results of a study to assess the “accuracy” of deep learning recommendations compared with those of Manipal’s multidisciplinary tumor board — a group of 12-15 oncologists who meet weekly to review cases at the hospital. Somashekhar and colleagues studied the cases of 638 breast cancer patients who had been treated at Manipal Hospitals. Ninety percent of standard treatment recommendations were in accord with those of the tumor board. Concordance rates varied depending on the type of breast cancer and in cases where new medical literature may have been unknown to tumor board members.

**Address Limited Access to Physicians**

Within the Unites States, the Association of American Medical Colleges estimates that the country will need 91,500 additional doctors by 2020, and this figure rises to 130,600 by 2025. The issue of physician shortage is not exclusive to the United States. The World Health Organization has estimated that the global needs-based shortage of healthcare workers is projected to be over 14 million in 2030, with poorer countries disproportionately impacted.

Although a more significant problem in rural, underserved geographies, there exists a shortage of radiologists and pathologists in some countries. In China, there are 1.4 billion lung scans performed per year and only 80,000 radiologists, while Japan, a country with the most CT and MRI scanners per unit population in the world, has the lowest number of radiologists per unit population (approximately 36 per million population) of any of the 26 OECD countries.

The aforementioned examples of the Manipal Comprehensive Cancer Center and other hospitals using deep learning to improve treatment decisions are good examples of how this technology can be used to address limited access to physicians. The greatest adoption of these technologies is in geographies that have large rural areas or where the lack of specialists is prevalent.

Skin cancer, among the most prevalent cancers, is found most often in sunny rural areas where access to dermatologists may be limited and transportation a challenge. It is estimated that over 5 million cases of skin cancer are diagnosed in the United States. Skin cancer is diagnosed visually,
beginning with an initial clinical screening and followed potentially by dermoscopic analysis, a biopsy, and histopathological examination. Automated classification of skin lesions using images is a challenging task owing to the fine-grained variability in the appearance of skin lesions. Advances in the diagnoses of skin cancer include the work of researchers at Stanford University. Using deep learning algorithms, the department of dermatology and researchers at the Artificial Intelligence Labs found that photos of skin lesions can be classified as benign or malignant with the same accuracy as a dermatologist. 

Predict the Future Risk of a Disease or Negative Health Event

In the past, the ability to predict the future risk of an individual developing an illness has been a blunt instrument at best. This is true in part because the available data was limited to structured clinical data and healthcare claims. This limited view typically represents a point in time and risks treatment of symptoms rather than a holistic view of the patient. Waiting until a patient has symptoms of an illness reduces the ability to slow or reverse the progression of a disease. Examples of early detection of illness using deep learning are emerging and are showing promise of early treatment that might well change the course of the illness, for example, Alzheimer's disease (AD), which is widespread.

An individual in the world develops dementia every three seconds. There were an estimated 46.8 million people worldwide living with dementia in 2015, and this number is believed to be close to 50 million people in 2017. This number will almost double every 20 years, reaching 75 million in 2030 and 131.5 million in 2050. Much of the increase will be in developing countries. Already 58% of people with dementia live in low- and middle-income countries, but by 2050, this will rise to 68%. The fastest growth in the elderly population is taking place in China and India and their south Asian and western Pacific neighbors. In addition, Alzheimer's is difficult to accurately and objectively diagnose in the early pathogeny because of its mild symptoms. To date, AD is generally detected at a late stage at which treatment can only slow the progression of cognitive decline. Hence early detection of AD is important to improve preventive and disease-modifying therapies. The functional change process of AD is thought to begin many years before the diagnosis of AD dementia. This long preclinical phase of AD would provide a critical opportunity for therapeutic intervention. It is especially important in individuals with mild cognitive impairment (MCI), who have high risk to develop AD.

Researchers at Harvard University, Harvard Medical School, and Massachusetts General Hospital published the results of a study they conducted to determine the ability of deep learning algorithms to identify Alzheimer's using MRI images. Deep learning has advantages in the field of image processing as it can extract and analyze the features. It can find links between different parts of the image and produce the overall cognition of the image, which has great significance in the brain networks. Start-ups such as Winterlight and Canary Speech are using AI natural language processing to detect patterns in speech that are preliminary indicators of reduced cognitive ability.

As a data source, electronic health records (EHRs) contain a wealth of medical information; however, EHR data is challenging to represent and model because of its high dimensionality, noise, heterogeneity, sparseness, incompleteness, random errors, and systematic biases. Applying deep learning has shown positive results in deriving a general-purpose patient representation from EHR data that facilitates clinical predictive modeling to inform better clinical decision making and treatment protocols. Sutter Health and Mount Sinai Hospital researchers tested the accuracy of deep learning algorithms to predict the onset of disease using a large data set (700,000 records) from EHRs. A common approach with EHRs is to have a domain expert to designate the patterns to look for (i.e., the learning task and the targets) and specify clinical variables in an ad hoc manner. Researchers found
supervised definition of the feature space scales poorly, does not generalize well, and missed opportunities to discover novel patterns and features. Unsupervised feature learning attempts to overcome limitations of supervised feature space definition by automatically identifying patterns and dependencies in the data to learn a compact and general representation that makes it easier to automatically extract useful information when building classifiers or other predictors. Researchers showed that unsupervised deep feature learning applied to preprocessed patient-level aggregated EHR data results in representations that are better understood by the machine and significantly improve predictive clinical models for a diverse array of approximately 80 clinical conditions.12

In a study, the results of which were published in late 2016 by the Journal of the American Medical Association (JAMA), researchers compared the results of automated deep learning algorithms with manual grading by ophthalmologists to identify diabetic retinopathy in retinal fundus photographs. Deep learning algorithms had high sensitivity and specificity for detecting diabetic retinopathy and macular edema in retinal fundus photographs. A convolutional neural network was trained using a retrospective development data set of 128,175 retinal images, which were graded three to seven times for diabetic retinopathy, diabetic macular edema, and image gradability by a panel of 54 U.S. licensed ophthalmologists and ophthalmology senior residents from May to December 2015. The resultant algorithm was validated in January and February 2016 using two separate data sets, both graded by at least seven U.S. board-certified ophthalmologists with high intragrader consistency. Automated grading of diabetic retinopathy has potential benefits such as increasing efficiency, reproducibility, and coverage of screening programs; reducing barriers to access; and improving patient outcomes by providing early detection and treatment. To maximize the clinical utility of automated grading, an algorithm to detect referable diabetic retinopathy is needed.13

In an ongoing effort with Boston area hospitals, including the Boston Medical Center and the Brigham and Women's Hospital, researchers found that they could predict hospitalizations because of diabetes and heart diseases about a year in advance with an accuracy rate of as much as 82%. Results of this study were published in the Harvard Business Review.

Based on a study of a year's worth of hospital admissions, the U.S. Agency for Healthcare Research and Quality (AHRQ) estimated that 4.4 million of those admissions in the United States, totaling $30.8 billion in costs, could have been prevented. Of that $30.8 billion, $9 billion was for patients with heart diseases and $5.8 billion for patients with complications from diabetes. That's half of all unnecessary hospitalizations. Just 5% of Medicaid's 70 million beneficiaries account for 54% of Medicaid annual expenditures of more than $500 billion and 1% account for 25% of the total. Of this 1%, 83% have at least three chronic conditions.14

ADOPTION OF DEEP LEARNING IN HEALTHCARE

The journey map for deep learning in healthcare as applied to clinician augmentation and assistance is most mature as it identifies the optimal treatment plan to physicians, particularly for complex illnesses such as cancer. In geographies that are underserved, access to personalized treatment plans that are in part based on the analysis of medical journals and expert advice is well served by the application of deep learning and the insights and recommendations.

Some applications of deep learning require government approval by organizations such as the Federal Drug Administration (FDA) in the United States. While government approval is often a lengthy process, the FDA recently authorized the use of deep learning for ventricle segmentation by MRI. This is the first instance of such an approval and likely to encourage others to move toward approval. Progress is
being made in the application of deep learning to the fields of radiology and pathology. Deep learning is particularly well suited to the application of large volumes of data such as those available through routine screenings. The ability to increase the efficiency and accuracy of diagnoses is critical to the early diagnoses and treatment of disease. This will be of great use in geographies where physician shortages result in long delays in evaluating images and slides. Finally, research is ongoing in the application of deep learning for enhancing both diagnoses and treatment. Findings released by Mayo Clinic demonstrate the power of deep learning in identifying genomic properties in brain cancer tissues to determine the most effective treatment. Figure 1 identifies the journey for deep learning in the augmentation of clinicians for the diagnoses and treatment of disease.

FIGURE 1
From Bench to Bedside — The Journey of Deep Learning to Assist Clinicians Diagnose and Treat

Stanford remote skin cancer diagnoses
Augment diagnoses in radiology and pathology
Treatment plan recommendations for cancer

Mayo brain tumor tissue texture analysis
China lung cancer detection
Ventricle segmentation via MRI
FDA, CE Mark

MIT, Harvard, Beth Israel, Mass. General — breast lymph node evaluation
Dx osteoporosis via existing CT scans CE Mark

Regulatory approval
Manipal Comprehensive Cancer Center, 37 other hospitals globally

Source: IDC, 2017

Deep learning is also being applied to predictive analytics. There appear to be two emerging use cases: the first is the prediction of the onset of a disease like diabetic retinopathy or an adverse event like hospital admission and the second is the prediction of clinical and financial risk for a patient and the identification of the drivers of risk. The ability to predict either an illness before a patient is symptomatic or an adverse health event provides huge opportunity in the improvement of clinical and financial outcomes. The ability to intervene before an individual is diagnosed with a chronic illness has the potential to prevent the development of the disease.
Work is being done in several hospitals to detect the onset of sepsis, a high-impact adverse event, in hospitalized patients. Early diagnosis and treatment of sepsis can mean the difference between life and death, not to mention extended hospitalization and increased costs. While not widely commercialized, hospitals are taking the results of their research and deploying the results within their institutions. New processes and protocols are being developed to alert physicians and initiate treatment in a timely manner.

A growing number of organizations, both suppliers and users of technology, are using deep learning, as well as other forms of machine learning, to predict hospitalization of patients with chronic illness and predict readmissions (CMS levies penalties, amounting to tens of millions of dollars, for readmission to hospitals for selected conditions within 30 days).

Less of a direct clinical impact is the use of deep learning to predict patients at high clinical and financial risk and provide insight into the clinical, behavioral, societal, and economic drivers of risk. With physicians and hospitals taking on increased financial risk, they are looking for more precise identification and insight into next best action for their high-risk patients. It appears that the further from direct care delivery a deep learning application gets, the greater the market adoption. Figure 2 shows a depiction of the journey of deep learning in the prediction of a disease or an adverse event.

FIGURE 2

The Journey of Deep Learning in Prediction in Healthcare

Pre-onset of disease
Mount Sinai, Sutter Health
Preventable hospitalizations
Boston University, Boston Medical Center
Early diagnoses and treatment of sepsis
Recent regulatory approval
Predict cardiovascular events
Early stage United States-based vendors
Clinical risk identification

Source: IDC, 2017

ADOPTION CHALLENGES OF DEEP LEARNING IN HEALTHCARE

Among the most significant challenges facing the adoption of deep learning is physician acceptance. One of the concerns expressed by physicians is their skepticism of the validity or repeatability of results delivered by deep learning. The closer deep learning applications come to the delivery of clinical processes, the greater the level of concern. Physicians believe that the results of deep learning have not been clinically validated across a broad population of patients. The stigma of "black box"
technology continues to plague deep learning. Broader acceptance and positive assurance from colleagues will defuse physician skepticism.

Second, some physicians fear that they will be replaced by deep learning. On the contrary, deep learning should be viewed as assisting physicians by reducing cognitive burden; automating routine or cumbersome tasks, allowing physicians to apply their expertise to more critical areas; raising insights to inform the physician's diagnosis and treatment plan; and meeting unmet demands in underserved areas. As deep learning becomes more widely used as an assist to physicians, often doing tasks they neither have time for nor the interest in taking on, the physician concerns may well subside. In addition, as users of the technology quantify and publicize the benefits they are receiving from deep learning, physicians may become more comfortable with the technology.

The regulatory hurdles in the United States and other regions play a part in the slow adoption of deep learning. In areas where deep learning drives a medical device's ability to autonomously read images or evaluate tissue samples, the need for regulatory approval exists. The closer the application is to potentially creating "harm" to a patient, the more rigorous the review process. Those applications in which minimal harm could occur require FDA registration but not a long evaluation process. While a daunting process, FDA approval in the United States and CE approval in the EU have been acquired by a small handful of companies.

PARTING THOUGHTS

AI is increasingly being adopted in clinical research and is gaining acceptance as regulatory agencies in the United States and other regions are bringing advances to the larger healthcare industry. These use cases are resulting in increased evidence of the benefits of applying deep learning to the delivery and management of healthcare. Drivers include the rapid development and availability of more powerful technologies to enable massive compute power, sophisticated neural network algorithms modeled on the human brain, and access to the explosion of data.

The key to deep learning technologies is that the algorithms self-program based on the data provided and significant compute resources, typically based on GPUs today. Using these technologies, organizations can develop solutions that tune themselves automatically at a speed that human programmers cannot hope to emulate. In addition, since deep learning can be based strictly on data, the quality of solutions is improved as more data is provided. The inclusion of deep learning capabilities into applications and solutions will make it possible for healthcare organizations to become more responsive to changes in the ecosystem as well as shifting standards of quality care.

While adoption challenges exist, especially for those applications closest to care delivery, healthcare organizations and technology suppliers should continue to push forward on those initiatives that deliver on their digital transformation objectives.

If healthcare as an industry is going to reach its goals of improved quality and managed costs, innovations such as those being made through the application of deep learning must continue and, as in some cases, must move beyond bench research to have a sustainable impact. Challenges ahead are less technological and more financial and societal and as such must be addressed to realize the full potential of deep learning in healthcare.
APPENDIX — STUDY CITATIONS

1 PubMed (www.ncbi.nlm.nih.gov/pubmed/)

2 Densen, Peter, Challenges and Opportunities Facing Medical Education, 2011. Transactions of the American Clinical and Climatological Association (www.ncbi.nlm.nih.gov/pmc/articles/PMC3116346/)

3 Panagiotis Korfiatis, Timothy L. Kline, Lucie Coufalova, Daniel H. Lachance, Ian F. Parney, Rickey E. Carter, Jan C. Buckner, Bradley J. Erickson, MRI Texture Features as Biomarkers to Predict MGMT Methylation Status in Glioblastomas, Med. Phys. 43(6), June 2016


5 Centers for Disease Control and Prevention (www.cdc.org)


7 Physician Shortages to Worsen Without Increases in Residency Training, AAMC Center for Workforce Studies (www.aamc.org/download/150584/data/physician_shortages_factsheet.pdf)


Masatoshi Matsumoto, Soichi Koike, Saori Kashima, Kazuo Awai. Geographic Distribution of Radiologists and Utilization of Teleradiology in Japan: A Longitudinal Analysis Based on National Census Data, September 2015 (journals.plos.org/plosone/article?id=10.1371/journal.pone.0139723)


10 Dementia statistics, Alzheimer’s Disease International (www.alz.co.uk/research/statistics)


About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data – Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2017 IDC. Reproduction without written permission is completely forbidden.