International Journal for Quality Research 16(1) 77–92
ISSN 1800-6450

Abdulrhman Samman Al-Asmari[1]

Article info:
Received 19.03.2021.
Accepted 15.07.2021.

UDC – 37.014.6
DOI – 10.24874/IJQR16.01-05

APPLICATIONS OF DEEP LEARNING TO IMPROVE THE QUALITY OF HEALTHCARE OUTCOMES

Abstract: *The aim of this paper is only to provide a systematic review of important research undertaken thus far in Deep Learning (DL) applications in healthcare and biomedicine. A total of 47 papers were shortlisted for this review. The review revealed that DL can be and has been applied in a number of healthcare contexts to improve outcomes. The two most important ones among them, indicated by the number of available papers, are cancer and medical imaging. A good number of papers on drugs and their development were also available. However, it must be highlighted that such categorisations may be somewhat arbitrary. It may be possible to categorise one paper in more than one category. The primary implication of this research for the academia is that there is a large deficit of papers on many of the chronic and lifestyle related illnesses such as diabetes as well as some diseases caused by immunodeficiency. There is also a complete deficit of literature on most of the acute health problems. This may be indicative of the need for more intensified research in the deficit areas. The primary implication of this research for health practitioners is that there is a plethora of substantial research that is currently available and accessible regarding the applications of DL to cancer and medical imaging which may be utilised for their practice.*

Keywords: *Deep Learning; Healthcare; Biomedicine; Outcomes; Quality.*

1. Introduction

Deep learning and its application are new areas of research that fall within the ambit of Machine Learning (ML). Deep learning or DL and its application has the potential to move ML closer to one of its original goals: Artificial Intelligence (Gulcehre, 2015). Investopedia (Hargrave, 2019) describes DL as a subcategory of ML in artificial intelligence (AI) which also has networks that are capable of learning unsupervised from data that is unstructured or unlabelled. This is also known as deep neural learning or a deep neural network. The evolution of big data makes it easier to have access to a range of details regarding any subject. However, extracting relevant information from such a voluminous data by human beings is practically impossible. There is great potential that can be realised by analysing this wealth of information for which AI systems can be adapted to provide automated support. The most common AI application used for big data analysis is ML. This algorithm improves itself the more it is used and the more data is added to it.

[1] Corresponding author: Abdulrhman Samman Al-Asmari
Email: aabdulrhman985@gmail.com

DL, which is also a subcategory of ML, adopts a hierarchical level of artificial neural networks in order to undertake the processes associated with ML. The artificial neural networks bear resemblance with the human brain, with neuron nodes that are connected in a manner that is similar to a web. Whereas earlier and more traditional programs develop their analysis by incorporating data in a linear way, the hierarchical function of DL systems allows machines to compute data by utilising a nonlinear approach. Deep learning mimics the working system of the human brain to process data for decision making. Deep learning can learn from both unstructured and unlabelled data.

Deep learning has numerous applications in various fields as the potential indicates from the above descriptions. One of them is in biomedicine. This paper reviews the progress of deep learning in biomedicine.

2. Methodology

It should be stated at the outset that this paper does not aim to provide an exhaustive review and it is almost impossible due to the recent explosion of research and papers on various aspects of DL applications in biomedicine. The intention here is to provide major pointers to the direction towards which the whole concept of DL is moving in the realm of biomedicine. Google Scholar was searched to select papers of various types dealing with DL applications in biomedicine. By this type of search, 47 papers were selected. The results from this will be elaborated on in the sections below.

3. Result

3.1 Scope of DL in biomedical applications

Volumes of data have been made available in biological and medical fields. These data may be found in a multitude of forms. Medical images, electroencephalography, genomic and protein sequences, electronic patient records are only a few of them. These data can be used for learning about human health and diseases. Some of these data can provide clues for interventions leading to cure or improve the management of the disease. A review of DL techniques and a few of its state-of-the-art applications in the biomedical field was provided by Cao et al. (2018). DL tries to derive generalised models from a large volume of data using multi-layered deep neural networks (DNNs), enabling it to derive meanings out of a variety of data forms like images, sounds, and texts. DL has two properties in general. Multiple layers of nonlinear processing units is one of them. The other is the supervised or unsupervised learning of feature presentations layer by layer. Automatic speech recognition, image recognition, natural language processing, drug discovery, and bioinformatics are only some areas of medical fields in which DL has been used. The timeline of DL's development along with the most commonly used ML algorithms was provided by the authors reproduced in Fig 1.

In Fig 1, the development of DL and neural networks is shown in the top panel, and several commonly-used ML algorithms are shown in the bottom panel. Abbreviations: NN, neural network; BP, backpropagation; DBN, deep belief network; SVM, support vector machine; AE: auto-encoder; VAE: variational AE; GAN: generative adversarial network; WGAN: Wasserstein GAN.

Many functional logistics are possible in DL, which have been discussed with mathematical treatment. A long list of applications of DL in biomedical fields is tabulated by Cao et al. (2018). In medical image classification and segmentation, distinct features for medical image interpretation are manually designed for classification as in the case of detection of lesions or abnormalities and in marking regions of interest such as tissues or organs in different medical applications. This requires special expertise of the physician. However, the complexity of machine learning applications in this area have limited its

applicability as has the lack of adequate expertise in medical image interpretation and the large amounts of annotated data required for analysis. However, high levels of success has been achieved by using DL have attained success in several computer vision tasks like object recognition, localization, and segmentation in natural images. Thus, an active field of ML in medical image analysis has evolved. The authors discussed many applications in which DL has been used. Considering the merits of DL, its low uptake in biomedicine was also pointed out by Mamoshina et al. (2016).

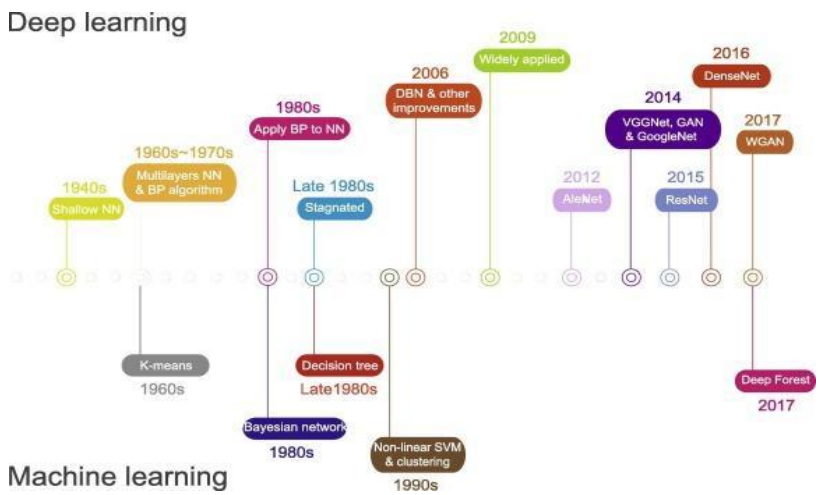

Figure 1. Timeline of deep learning development (Cao et al., 2018)

3.2 Reviews

In a review, Faust et al. (2018) pointed out that DL was suitable for large and varied datasets. However, its full potential has not yet been realised. This is because its use and application in healthcare has been quite limited. This is especially the case with the area of physiological signal analysis which leaves room for further examination. In another review on big data, their sources and methods of processing and using them for various purposes, including elements of DL, were discussed by Costa (2014). The types, nature and sources of big data, their various limitations, models and methods of using them for solving many biomedical problems were reviewed in detail by Kocheturov, Pardalos, and Karakitsiou (2019). Three broad areas of machine learning applications in biomedicine (clinical diagnostics, precision treatments, and health monitoring) were discussed by Goecks et al. (2020). Using Google Scholar search, a review of machine learning and DL was done by Alanazi and Alanazi (2020). The topics covered in this review were Precision Medicine, Machine Learning (ML) Empowered Biometric Methods, Visible Machine Learning for Biomedicine and Deep Learning and Biomedicine. The review highlighted that the healthcare sector is in the process of being reorganised owing to concurrent developments in electronics, communication and computers are leading biomedicine to a stage of data revolution. Applications of DL and ML in biomedicine include predictive and inferential analysis using medical records, imaging data, sequencing data, genotypes and sensor data.

In a similar review of DL applications, the topics related to structured biomedical data, image processing, DL models and potential trends for the future were discussed by Bacciu

et al. (2018). Traditional mining of complex, high-dimensional, heterogeneous, poorly annotated and largely unstructured biomedical data and statistical learning approaches require feature engineering first to transform them into effective and robust data. Only then, clustering and prediction models can built on them. Many challenges exist in this respect. DL has provided answers to many of these problems. How DL enables to overcome these challenges was the subject of the review by (Miotto et al., 2018).

3.3 Specific applications

More specifically, the ability of DL to predict more or less successfully about the cellular processes involved in pathogenesis due to genetic variations, modulation of activities of therapeutically relevant proteins and small molecules, radiographic images indicative of disease were highlighted by Wainberg et al. (2018). On the other hand, some challenges also exist due to the very flexibility of DL. There is no guarantee about the performance of deployed systems. There are problems of trust with stakeholders, clinicians and regulators, as they will need a rationale for the use of DL for these purposes of decision making. This problem can be overcome by using the same flexibility itself for training the models to provide rationale for their predictions also.

The main goal of using AI in biomedicine is to translate the patient data into successful therapies. However, problems of handling the extreme heterogeneity of data and paucity of mechanistic insight into predictions exist. Michael et al. (2018) proposed "visible" approaches to guide the model structure using experimental biology. In this model, ML will not act as a substitute for experimental cell and tissue biology, but will be highly enabled by such knowledge if the precise visible intelligence infrastructure is provided. The visible model proposed by the authors was claimed to resemble the Visible V8 engines in terms of similarities of hierarchical nature of components as given in Fig 2. Visible learning has close relationship with model interpretation paradigm itself in the form of model internal logic after it has been trained. Models for various biomedical purposes have been researched.

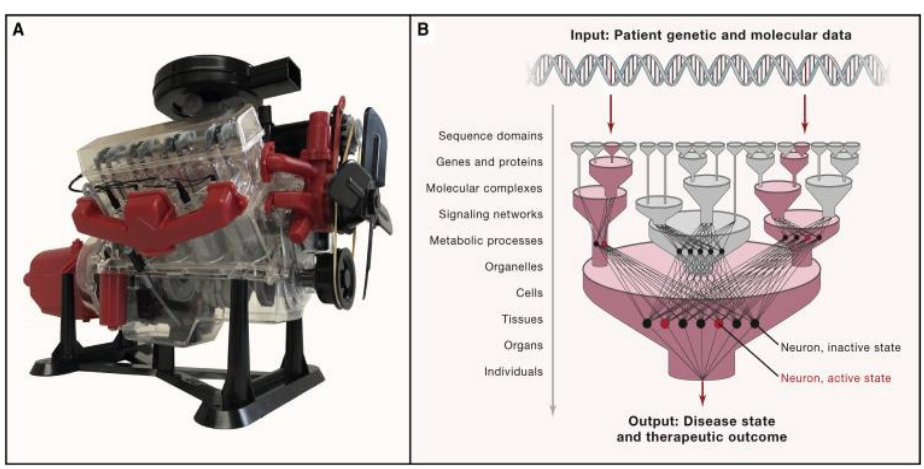

Figure 2. Visible healthcare model resembles Visible V8 engine model (Michael et al., 2018)

In the ML part of DL, relevant data like DNA or RNA sequences, gene sets or pathways, gene interaction or co-expression networks, ontologies, and phylogenetic trees have a rich structure not easily encodable as predictors of real value. Some recent machine learning

models used data structure to constrain model architecture or to combine structured data into model training. In biomedicine, while the availability of data is low due to small sample size, model interpretability is highly critical. This means that prior understanding and experience with structured data has the potential to be incorporated as a solution to this type of problem (Crawford & Greene, 2020).

According to Bzdok and Ioannidis (2019) when research data are used as prior available data to enhance the sample size in DL, it should be remembered that null hypothesis testing for significance should be useful for modelling goals. The progressive use of methods is: exploration for a first cursory glance to indicate potentially interesting data available now; inference aiming at isolating variables which may be individually important beyond the level of mere chance indicated by P values and prediction to identify sets of variable which may help to guess outcomes precisely based on other or future data. It may be also noteworthy that P values are not meant to measure the predictive accuracy of a model. Variables found significantly important by null hypothesis may be dissimilar from variables which may maximize predictive performance in new individuals or settings.

3.4 Biochemical processes

Alternative splicing (AS) is a genetically and epigenetically regulated pre-mRNA processing method to create greater diversities of transcriptome and proteome. It is possible to obtain deeper insights into many biological contexts involving AS, for example, development and diseases. For this purpose, comprehensive decoding of these regulatory mechanisms is done. Xu et al. (2017) assembled splicing (epi)genetic code, called DeepCode (a DL method) for differentiation of human embryonic stem cell (hESC) by meshing together the heterogeneous features of genomic sequences along with a multi-label deep neural network. The epigenetic features enabled Deepcode to predict the splicing patterns and their changes during hESC differentiation. Thus the importance of incorporating the epigenetic properties while assembling a comprehensive splicing code was demonstrated. Deepcode also exhibited robust prediction qualities for cell lineages and datasets. A novel candidate mechanism linking histone modifications to hESC fate decision was, thus, demonstrated. When applied and adapted to varied contexts, DeepCode could be expanded to many other biological and biomedical fields. A schematic view of DeepCode and its applications given in the paper is reproduced in Fig 3.

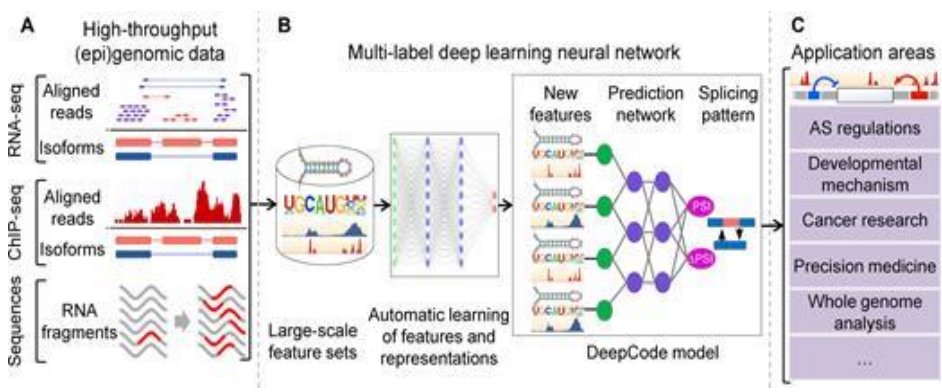

Figure 3. A schematic view of DeepCode and its applications (Xu et al., 2017)

A platform, Galaxy-ML was developed by Gu, et al. (2020) to increase the accessibility of supervised machine earning to biomedical researchers. The platform allows using web browser and can be used to perform end-to-end reproducible machine learning analyses at large scale. The architectural structure, features and various applications of Galaxy-ML with detailed methodology have been described in the paper.

3.5 Cancer

A specific case of cancer prediction using DL was reported by Xiao et al. (2018). Gene expression can be the basis of efficient cancer prediction leading to precise and effective treatment decisions. It is now possible to differentiate between cancer patients from healthy persons using machine learning methods. In this DL-based multi-model ensemble, informative gene data were selected by differential gene expression analysis leading to five different classification models. DL is used over the output of the selected five classifications. The model was tested on three types of cancer affecting lungs, stomach and breast to find increased accuracy of predictions of the three cancers. Another DL-based cancer prediction model, a stacked sparse auto-encoder (SSAE) based model, was proposed by Xiao et al. (2018). The model consisted of sparsity and training. The model used the three public RNA-seq datasets for prediction of lung, stomach and breast cancer.

The model predicted the three cancers better than other methods. An attempt was made for building a DL-based model that could predict the disease's progression in head and neck squamous cell carcinoma (HNSCC) patients by Zhao et al. (2020). The method used was integration of multi-omics data. RNA sequencing, miRNA sequencing and methylation data from The Cancer Genome Atlas (TCGA) were used as the input for the autoencoder which is a DL approach. An autoencoder-based prognosis model was built by SVM algorithm, using which three confirmation tests with three data sets were done. The model was then contrasted against two other approaches in order to test its predictive performance. The differential expression analysis for mRNAs, microRNAs (miRNA) and methylation was also done. Functional annotation of differentially expressed genes (DEGs) was obtained using function enrichment analysis. The results showed good fitness index for identified prognosis models for two subgroups of patients with highly differing progress free survival (PFS). The two subgroups were validated in the three confirmation tests. Compared to principal component analysis (PCA) or individual Cox-PH-based models, DL-based model was more accurate and efficient. A number of pathways and gene targets were uncovered, which had possible implications in cancer progression. The model has the potential to allow for and enable the development of individualized therapy for HNSCC patients with improved prognosis.

There are differences between the 5-years disease-free survival (DFS) rates of non-invasive adenocarcinoma (non-IA) and invasive adenocarcinoma (IA) stage-I lung adenocarcinoma. The aim of the study by Xia et al. (2020) was to utilise AI schemes for the development of CT images in order to facilitate categorisation between non-IA and IA nodules and incorporate deep learning (DL) and radiomics features to enhance the categorisation process. Surgical pathological confirmed ground-glass nodules (GGNs) were collected from patients in two centres consisting of both non-IA and IA samples. First, a recurrent residual CNN based on U-Net was used for segmentation of the GGNs. Next, two schemes (DL scheme and radiomics scheme) were designed to distinguish between non-IA and IA. In order to enhance the performance that may be gained from such a distinction, the prediction scores of two schemes were fused by applying an information fusion method. In the final stage, an observation study was done to contrast the execution of this scheme with

those of two radiologists by testing on an independent dataset. The new fusion scheme performed better than the DL scheme and radiomics scheme with respect to the risk classification performance. Also, higher accuracy was obtained for the fusion scheme than with two radiologists, although agreement between radiologists was high. Therefore, applying the AI method has been proven to be a useful manner in which the effectiveness of risk prediction in the performance of GGNs may be enhanced. In another study on classification of brain CT images, Gao, Hui, and Tian (2017) obtained high levels of precision and accuracy for fused CNN architecture, 2D CNN for three classes of AD, lesion and normal, 2D SIFT and 2D KAZE and 3D SIFT and 3D KAZE. The latter two were significant contributions, especially to fulfil the highly felt needs of Alzheimer's disease for DL methods of categorisation of brain images.

The aim of the study by Arnaldo et al. (2020) was to study an MRI radiomics-powered ML model's effectiveness and execution while determining the deep myometrial invasion (DMI) in endometrial cancer (EC) patients while also attempting to explore its clinical applicability. Pre-operative MRI scans of EC patients were used. The Random Forest wrapper identified the three most informative variables, which were used for ML training. The classifier reached an accuracy of 86% and 91%. Performance of radiologists also improved to 100% when they used ML. Thus, the feasibility of a radiomics-powered ML model for DMI detection on MR T2-w images to improve the performance of radiologists was demonstrated.

The progress achieved so far on ML with respect to cancer detection and therapy were evaluated by El Naqa and Murphy (2015). Over half of the patients with cancer receive radiotherapy as part of their interventions. This has been found to be a significant treatment and is the only possible and viable treatment at advanced stages of localised cancer. Radiotherapy consists of a set of several processes starting from initial consultation to beyond treatment to ensure that the patients have received the prescribed radiation dose and are responding as expected. These processes have varying degrees of complexities and numbers involve several stages of refined human-machine interactions and decision making. The characteristics of these processes demand the use of ML algorithms for optimizing and automating them. These needs include radiation physics quality assurance, contouring and treatment planning, image-guided radiotherapy, respiratory motion management, treatment response modelling, outcomes prediction and many other related aspects. As ML algorithms have the ability to learn from current context and generalize them into unseen tasks, improvements are possible both in the safety and efficacy of radiotherapy practices for better patient outcomes.

A novel Computer-Aided Diagnosis (CAD) system for the identification and categorisation of breast masses in mammograms was proposed by Al-Masni et al. (2018) based on one of the regional DL techniques, a ROI-based CNN which was termed You Only Look Once (YOLO). The advantage of this model over the previous ones was that YOLO-based CAD systems have the potential to manage the identification and categorisation simultaneously in the same framework. Test results showed high precision in detecting mass locations of cells. It also differentiated malignant and benign lesions. Other similar models were proposed by Ribli et al. (2018) and Dhungel, Carneiro, and Bradley (2017) with more or less similar conclusions.

It is essential to track the individual-cell/object over a period of time in order to fully understand the drug treatment and the effects it has on cancer cells and video surveillance. One of the key issues of such an individual-cell/object tracking is to be able to simultaneously address the cell/object appearance variations which may be as a result of either intrinsic or extrinsic factors. Based on DL architecture, a robust feature

learning method has been put forward by Zhong et al. (2016) in order to construct discriminative appearance models without much pretraining. The DL component occurs after an unsupervised method is used first for learning the abstract feature using PCA analysis. The developed model validated the utilisation of two standard individual cell/object tracking benchmarks.

3.6 Hypertension

A new deep feature selection method based on deep architecture was proposed by Nezhad et al. (2016). The method used stacked auto-encoders for feature representation in higher-level abstraction. A novel feature learning approach was developed and applied to a specific problem of precision medicine problem. This problem was related to assessment and prioritisation of high risk factors for hypertension (HTN) in an especially vulnerable demographic subgroup (African-American). Some of the key risk factors impacting the left ventricular mass indexed to body surface area (LVMI) was found to be a useful indicator of heart damage risk and gave better results than other methods.

3.7 Diabetes

With regard to diabetic retinopathy, red lesions belonging to the group of microaneurysms and haemorrhages were found to be a prominent early sign. The traditional method of manual detection of fundus photographs has been found to be both time consuming and tedious due to their small size and lack of contrast. DL has the potential to address this and help here. This is because consistency, accuracy and feedback are possible with DL. An ensemble method for adapting to assess the red lesions in fundus images was proposed by Orlando et al. (2018). Both deep learning coupled with subject matter expertise were used here. CNN was used for learning features. These were enhanced by combining hand crafted features. Such ensemble vector of descriptors was used to identify true lesion candidates later with a Random Forest classifier. The integration of manual and DL methods trained by lesion-level annotated data was very useful in improving precision and accuracy of detection and evaluation of red lesions in diabetic retinopathy.

3.8 Imaging

Currently available DL platforms are flexible. But they do not serve the specific purpose of medical image analysis. Considerable effort is required for adapting them for application in medical image analysis. TensorFlow APIs was used by Gibson et al. (2018) in an open source platform (NiftyNet) for deep learning in the medical imaging domain. The aim of NiftyNet was to speed up and simplify the development of solutions for these problems and to also offer a standard method for the dissemination of the research outputs that may be utilised, adapted and built upon. The steps involved in its implementation in typical medical imaging machine learning are: starting with established pre-trained networks, adapting the existing neural network architectures to new problems and rapid prototyping of new solutions. TensorBoard visualization of both 2D and 3D images are possible by default. Concrete examples of three deep-learning applications, including segmentation (of multiple abdominal organs by CT), regression (image regression for prediction of CT attenuation maps from brain MR images), image generation (simulation of ultrasound images of some anatomical poses) and representation learning have been presented for illustration of the important aspects of the platform. The NiftyNet architecture diagrammatised in the paper is reproduced in Fig 4.

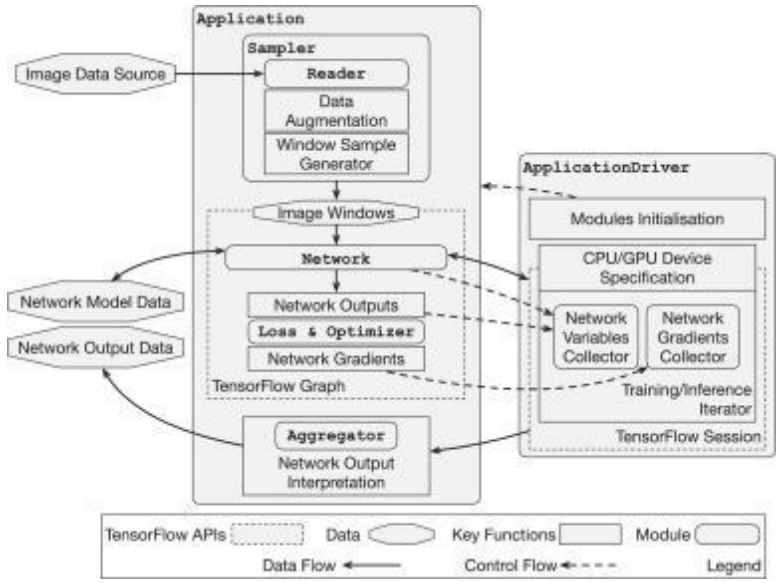

Figure 4. NiftyNet architecture showing its components (Gibson et al., 2018)

Some methods to overcome the need for a large volume of training data for segmentation of histopathological data required for DL learning were used by Van Eycke, Foucart, and Decaestecker (2019). The strategies for a significant section of annotated images required to automate the segmentation of histological images utilising deep learning were found through a review. The strategies consisted of the use of immunohistochemical markers as labels and realistic data augmentation, Generative Adversarial Networks (GAN) and transfer learning. Use of imperfect annotations and artificially generated data were also tested. However, the addition of real data coupled with high-quality annotations in the training set, even if fewer in numbers, was identified as being the safest way to enhance the functioning and effectiveness of a well configured deep neural network.

Sudre et al. (2017) in their earlier studies, found that of late, DL has become a powerful and useful method that may be adopted in image analysis while also gaining popularity for segmentation of both 2D and 3D medical images. While the choice of network architecture is important, choice of loss function also appears to be equally important. When rare observations are targeted in the segmentation process, there is a likelihood of class imbalance of severe nature between candidate labels. This has the potential to result in sub-optimal performance. Strategies such as weighted cross-entropy function, the sensitivity function or the Dice loss function have been proposed as possible solutions. This paper evaluated the behaviour of these loss functions and their sensitivity to learning rate when tuned in the presence of different rates of label imbalance across 2D and 3D segmentation tasks. Class re-balancing properties of the Generalized Dice overlap was also evaluated.

In addition to this, further research pertaining to the use of segmentation using DL, in order to tackle the concern regarding the lack of image-specific adaptation and the inadequate generalizability to previously unseen object classes related to CNN, Wang et al. (2018) suggested the adoption of a novel DL-based interactive segmentation framework by incorporating CNNs into a bounding box and scribble based segmentation pipeline. Image-

specific fine tuning to make a CNN model adaptive to a specific test image was also used for both unsupervised (without additional user interactions) and supervised (with additional scribbles). Also, a weighted loss function was tested considering network and interaction-based uncertainty for the fine tuning. This network was applied to 2-D segmentation of multiple organs. In one case, foetal magnetic resonance (MR) slices with only two types of these organs annotated for training and 3-D segmentation of brain tumour core with the exclusion of oedema and whole brain tumour which included oedema from different MR sequences. In another case, only the tumour core in one MR sequence was annotated for training. The image-specific fine tuning with the proposed weighted loss function significantly improved segmentation accuracy. Increased robustness and precise results with fewer user interactions and lesser time were obtained compared to currently used methods. In respect of multiscale modelling, Sloot and Hoekstra (2010) observed that only models existed, but no methodology for modelling. In the paper, a dynamic systems approach was used for transmission of resistance in human immunodeficiency virus spreading and in-stent restenosis in coronary artery disease.

In a review, Shen, Wu, and Suk (2017) noted that recent advances in ML and DL have the potential to assist in categorising, locating as well as quantifying patterns in medical images. Ability to exploit hierarchical feature representations learned solely from data is exploited here. This new tool replaces hand-designed features based on the domain-specific knowledge. Thus DL is the state of the art developing very fast. There is enhanced performance in various medical applications, especially those related to medical imaging. A similar survey of DL in image analysis was also done by Litjens et al. (2017). The survey led to identification of the most successful DL method for medical imaging. CNN and their derivatives were rated as the top applications in most medical image analyses. However, the exact architecture, augmentation and pre-processing, multi-scale network, network components and model hyperparameters were also key in determining the solutions of the problems. Similar conclusions were drawn in an editorial by Greenspan, Van Ginneken, and Summers (2016) in the IEEE Transactions in Medical Imaging journal.

3.9 Drug categorisation and development

The purpose of the paper by Aliper et al. (2016) was to show how the DNN can be adapted and augmented on large transcriptional response data sets in order to categorise various drugs into therapeutic categories being contingent upon their transcriptional profiles. DNN was trained using both gene level transcriptomic data and transcriptomic data processed using a pathway activation scoring algorithm for a pooled data set of samples perturbed with various concentrations of the drug from 6 to 24 hours. In both pathway and gene level classification, DNN achieved high classification accuracy. It also significantly outperformed the support vector machine (SVM) model on all multiclass classification problems. Models based on pathway levels were better than gene level. Thus, a deep learning neural net trained on transcriptomic data was noted as being helpful in recognising the pharmacological properties of multiple drugs across different biological systems and conditions. The next stage is the utilisation of deep neural net confusion matrices for drug repositioning. This will be a proof of principle for the application of deep learning to drug discovery and development.

Why ML becomes a transformative tool for drug discovery was explained by Duran-Frigola et al. (2019). Over the years, customary chemo-informatics methods were adopted in order to study the vector descriptors of compound structures as the input of their prediction tasks in a standardised manner. If a common vector format is available to represent biology and chemistry together, it can push biological

information into many current phases involved in the drug discovery pipeline. This enhances the accuracy of predictions and uncovering interlinkages between small molecules and other biological entities such as targets or diseases for more possible candidates of new drugs. This process has been diagrammatised in Fig. 5 reproduced from the paper.

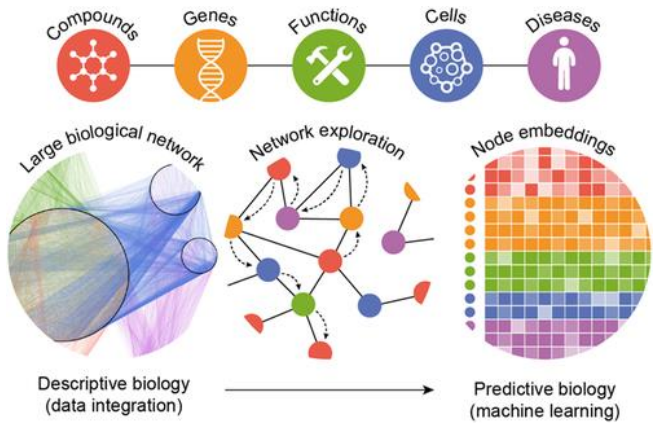

Figure 5. How DL helps drug discovery (Duran-Frigola et al., 2019)

In-silico toxicology is important in the context of efforts to reduce the use of animal experiments. In-silico toxicology can be useful for identifying hazards of compounds before synthesis, that is, at very early stages of drug development. It is helpful as a method to fill the gaps in the knowledge so that risks can be minimised through suitable strategies of product development. Thus, it is useful for both for chemical industries and regulatory agencies. ML and DL can be used in silico toxicology, even if data is scarce. The concept of adverse outcome pathways sets all techniques into a broader context. This can help to elucidate predictions by perfunctory insights (Hemmerich & Ecker, 2020). While making similar comments on use of DL in chemical toxicity issues, Gini et al. (2019) also mentioned about (Q)SAR model development, which leads to discovery of new products with less toxicity by learning the structural-chemical-biological relationships and their associated properties. The process of developing (Q)SAR model has been explained in the paper.

3.10 Proteases

Protease enzymes cleave by hydrolysing peptide bonds between specific amino acids in the target substrate proteins. Many of the substrates and the cleavage sites of the functional proteolysis have not been found yet. If accurate predictors of the substrates and cleavage sites are available, it would facilitate comprehension of the features and physiological roles of proteases. DL offers a promising approach for the solution of this problem. Li, et al. (2020) proposed DeepCleave as the first such deep DL-based predictor. The input for DeepCleave was the protein substrate sequence data and CNN was used with transfer learning to train accurate predictive models. High quality cleavage site features extracted from the substrate sequences, the implementation of transfer learning, multiple kernels and attention layer in the conception of the deep network makes the predictor highly efficient. Empirical tests against many conventional methods proved superiority of DeepCleave in predicting the substrate-cleavage sites of caspase and matrix metalloprotease.

3.11 Human emotions

DL methods are being increasingly attractive to analyse multimodal physiological signals to recognize human emotions. Conventional models of deep emotion classifiers are not effective enough. Hence, Yin et al. (2017) proposed a multiple-fusion-layer based ensemble classifier of stacked autoencoder (MESAE). It recognised emotions by identifying their deep structure based on a physiological data-driven approach. The performance of this model was validated by comparing it against the existing best emotions classifier and the superior nature of the proposed model in terms of classification rate was established.

3.12 Micro- and nano robotics

A possible extension of DL may be application of micro and nano- robots for various biomedical needs. A review of such tools based on their use types have been discussed with diagrammatic examples by Li, Esteban-Fernández de Ávila, Gao, Zhang, and Wang (2017). Many developments and many research works have been reported on this topic, which need to be treated as a separate topic.

3.13 Text mining

According to Spasic et al. (2005) for sophisticated text mining in biomedicine, lexical, syntactic and semantic layers of text annotation are required. Rule-based ML are necessary for such tasks. Ontologies can be used to represent specific domains of knowledge. Thus, both ontologies and terminological lexicons can be combined to obtain good results with biomedical text mining. Other variables of differentiation can be used at finer levels of analysis. Many recent developments in the area of text mining can be tried and adapted to biomedicine also.

In another work on the same topic, Jiang, Li, Huang, and Jin (2015) proposed a model in which biomedical domain-specific word embedding incorporating stem, chunk and entity to train word embeddings. Two DL architectures were used for two biomedical text mining tasks. Using them, word embeddings were comparatively evaluated with other models. These two models outperformed other general purpose word embeddings used for biomedical text mining tasks.

4. Discussions and Conclusions

The This overview of DL and its potential to contribute to the field of biomedicine has revealed that DL can be applied to almost the entire range of healthcare. However, the solutions must be context specific and should be tailored to all healthcare requirements and challenges. For each aspect of healthcare system, appropriate DL needs to be developed. Many times, it may involve a combination of DL with non-DL or one DL with another DL method. These also consist of certain limitations and obstacles. The solutions for them also depend on the context and the way the DL system has been used.

In this review, papers were discussed in various sections. Such categorisations are largely arbitrary. Many papers can be categorised in more than one way. Such problems are inherent in any review.

From the number of papers reviewed in different sections, it seems more research was done on cancer and medical imaging aspects. A notable miss in the papers was that no paper was available on DL applications for acute health problems and even in the case of chronic problems, not all diseases have been addressed. There was only one paper each on hypertension, diabetic retinopathy and psychological problems of emotions. Probably these are areas for future intensive research. It is also possible that papers on these aspects were missed due to the nature of literature search as explained in the limitations of this work below.

The primary implication of this research for the academia is that there is a large deficit of

papers on many chronic diseases like cardiology, diabetes and immunodeficient diseases and a total deficit of acute health problems. This may be indicative of the need for more intensified research in the deficit areas. The primary implication of this research for health practitioners is that there is a lot of substantial research available on the applications of DL to cancer and medical imaging which may be adapted to be utilised not only in theory but also in practice.

References:

Alanazi, A. M., & Alanazi, F. O. (2020). Machine Learning and Biomedicine. *Journal of Contemporary Scientific Research (ISSN (Online) 2209-0142), 4*(1), 7 pp. Retrieved July 8, 2020, from http://www.jcsronline.com/wp-content/uploads/2020/06/Volume4Issue1 Paper1.pdf

Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., & Zhavoronkov, A. (2016). Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Molecular pharmaceutics, 13*(7), 2524-2530. doi:10.1021/acs.molpharmaceut.6b00248

Al-Masni, M. A., Al-Antari, M. A., Park, J.-M., Gi, G., Kim, T.-Y., Rivera, P., . . . Kim, T.-S. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Computer methods and programs in biomedicine, 157*(April), 85-94. doi:10.1016/j.cmpb.2018.01.017

Arnaldo, S., Cuocolo, R., Anna, N., Valeria, R., Antonio, T., Antonio, R., & Giuseppe, B. E. (2020). Deep Myometrial Infiltration of Endometrial Cancer on MRI: A Radiomics-Powered Machine Learning Pilot Study. *Academic Radiology*, In press. doi:10.1016/j.acra.2020.02.028

Bacciu, D., Lisboa, P. J., Martín, J. D., Stoean, R., & Vellido, A. (2018). Bioinformatics and medicine in the era of deep learning. *arXiv, 1802* (Volume 1) (February), 09791. Retrieved July 8, 2020, from https://arxiv.org/pdf/1802.09791.pdf

Bzdok, D., & Ioannidis, J. P. (2019). Exploration, inference, and prediction in neuroscience and biomedicine. *Trends in neurosciences, 42*(4), 251-262. doi:10.1016/j.tins.2019.02.001

Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., . . . Xie, Z. (2018). Deep learning and its applications in biomedicine. *Genomics, proteomics & bioinformatics, 16*(1), 17-32. doi:10.1016/j.gpb.2017.07.003

Costa, F. F. (2014). Big data in biomedicine. *Drug discovery today, 19*(4), 433-440. doi:10.1016/j.drudis.2013.10.012

Crawford, J., & Greene, C. S. (2020). Incorporating biological structure into machine learning models in biomedicine. *Current Opinion in Biotechnology, 63*(June), 126-134. doi:10.1016/j.copbio.2019.12.021

Dhungel, N., Carneiro, G., & Bradley, A. P. (2017). A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical image analysis, 37*(April), 114-128. doi:10.1016/j.media.2017.01.009

Duran-Frigola, M., Fernández-Torras, A., Bertoni, M., & Aloy, P. (2019). Formatting biological big data for modern machine learning in drug discovery. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 9*(6), e1408. doi:10.1002/wcms.1408

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In I. El Naqa, R. Li, & M. Murphy (Eds.), *Machine Learning in Radiation Oncology* (pp. 3-11). Springer, Cham. doi:10.1007/978-3-319-18305-3_1

Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer methods and programs in biomedicine, 161*(July), 1-13. doi:10.1016/j.cmpb.2018.04.005

Gao, X. W., Hui, R., & Tian, Z. (2017). Classification of CT brain images based on deep learning networks. *Computer methods and programs in biomedicine, 138*(January), 49-56. doi:10.1016/j.cmpb.2016.10.007

Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D. I., Wang, G., & Eaton-Rosen, Z. e. (2018). NiftyNet: a deep-learning platform for medical imaging. *Computer methods and programs in biomedicine, 158*(May), 113-122. doi:10.1016/j.cmpb.2018.01.025

Gini, G., Zanoli, F., Gamba, A., Raitano, G., & Benfenati, E. (2019). Could deep learning in neural networks improve the QSAR models? *SAR and QSAR in Environmental Research, 30*(9), 617-642. doi:10.1080/1062936X.2019.1650827

Goecks, J., Jalili, V., Heiser, L. M., & Gray, J. W. (2020). How machine learning will transform biomedicine. *Cell, 181*(1), 92-101. doi:10.1016/j.cell.2020.03.022

Greenspan, H., Van Ginneken, B., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging, 35*(5), 1153-1159. doi:10.1109/TMI.2016.2553401

Gu, Q., Kumar, A., Bray, S., Creason, A., Khanteymoori, A., Jalili, V., . . . Goecks, J. (2020). Accessible, Reproducible, and Scalable Machine Learning for Biomedicine. *bioRxiv*, 15 pp. doi:10.1101/2020.06.25.172445

Gulcehre, C. (2015, December 1). *Deep learning.* Retrieved July 7, 2020, from Deep learning.net: http://deeplearning.net/

Hargrave, M. (2019, April 30). *Deep Learning.* Retrieved July 7, 2020, from Investopedia: https://www.investopedia.com/terms/d/deep-learning.asp#:~:text=Deep%20learning%20is%20a%20subset,learning%20or%20deep%20neural%20network.

Hemmerich, J., & Ecker, G. F. (2020). In silico toxicology: From structure–activity relationships towards deep learning and adverse outcome pathways. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, e1475. doi:10.1002/wcms.1475

Jiang, Z., Li, L., Huang, D., & Jin, L. (2015). Training word embeddings for deep learning in biomedical text mining tasks. *IEEE international conference on bioinformatics and biomedicine, 9-12 Nov. 2015, Washington, DC, USA* (pp. 625-628). IEEE. doi:10.1109/BIBM.2015.7359756

Kocheturov, A., Pardalos, P. M., & Karakitsiou, A. (2019). Massive datasets and machine learning for computational biomedicine: trends and challenges. *Annals of Operations Research, 276*(1-2), 5-34. doi:10.1007/s10479-018-2891-2

Li, F., Chen, J., Leier, A., Marquez-Lago, T., Liu, Q., Wang, Y., & Revote, J. e. (2020). DeepCleave: a deep learning predictor for caspase and matrix metalloprotease substrates and cleavage sites. *Bioinformatics, 36*(4), 1057-1065. doi:10.1093/bioinformatics/btz721

Li, J., Esteban-Fernández de Ávila, B., Gao, W., Zhang, L., & Wang, J. (2017). Micro/nanorobots for biomedicine: Delivery, surgery, sensing, and detoxification. *Science Robotics, 2*(4), eaam6431. doi:10.1126/scirobotics.aam6431

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., . . . Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis, 42*(December), 60-88. doi:10.1016/j.media.2017.07.005

Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Molecular pharmaceutics, 13*(5), 1445-1454. doi:10.1021/acs.molpharmaceut.5b00982

Michael, K. Y., Ma, J., Fisher, J., Kreisberg, J. F., Raphael, B. J., & Ideker, T. (2018). Visible machine learning for biomedicine. *Cell, 173*(7), 1562-1565. doi:10.1016/j.cell.2018.05.056

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics, 19*(6), 1236-1246. doi:10.1093/bib/bbx044

Nezhad, M. Z., Zhu, D., Li, X., Yang, K., & Levy, P. (2016). Safs: A deep feature selection approach for precision medicine. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 15-18 Dec. 2016, Shenzhen, China* (pp. 501-506). IEEE. doi:10.1109/BIBM.2016.7822569

Orlando, J. I., Prokofyeva, E., del Fresno, M., & Blaschko, M. B. (2018). An ensemble deep learning based approach for red lesion detection in fundus images. *Computer methods and programs in biomedicine, 153*(January), 115-127. doi:10.1016/j.cmpb.2017.10.017

Ribli, D., Horváth, A., Unger, Z., Pollner, P., & Csabai, I. (2018). Detecting and classifying lesions in mammograms with deep learning. *Scientific reports, 8*(1), 1-7. doi:10.1038/s41598-018-22437-z

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering, 19*(June), 221-248. doi:10.1146/annurev-bioeng-071516-044442

Sloot, P. M., & Hoekstra, A. G. (2010). Multi-scale modelling in computational biomedicine. *Briefings in bioinformatics, 11*(1), 142-152. doi:10.1093/bib/bbp038

Spasic, I., Ananiadou, S., McNaught, J., & Kumar, A. (2005). Text mining and ontologies in biomedicine: making sense of raw text. *Briefings in bioinformatics, 6*(3), 239-251. doi:10.1093/bib/6.3.239

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In M. e. Cardoso (Ed.), *Deep learning in medical image analysis and multimodal learning for clinical decision support, DLMIA 2017, ML-CDS 2017.* (Vols. Lecture Notes in Computer Science, vol 10553, pp. 240-248). Springer, Cham. doi:10.1007/978-3-319-67558-9_28

Van Eycke, Y.-R., Foucart, A., & Decaestecker, C. (2019). Strategies to reduce the expert supervision required for deep learning-based segmentation of histopathological images. *Frontiers in Medicine, 6*(October), 222. doi:10.3389/fmed.2019.00222

Wainberg, M., Merico, D., Delong, A., & Frey, B. J. (2018). Deep learning in biomedicine. *Nature biotechnology, 36*(9), 829-838. doi:10.1038/nbt.4233

Wang, G., Li, W., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., & Doel, T. e. (2018). Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE transactions on medical imaging, 37*(7), 1562-1573. doi:10.1109/TMI.2018.2791721

Xia, X., Gong, J., Hao, W., Yang, T., Lin, Y., Wang, S., & Peng, W. (2020). Comparison and Fusion of Deep Learning and Radiomics Features of Ground-Glass Nodules to Predict the Invasiveness Risk of Stage-I Lung Adenocarcinomas in CT Scan. *Frontiers in Oncology, 10*, 418. doi:10.3389/fonc.2020.00418

Xiao, Y., Wu, J., Lin, Z., & Zhao, X. (2018). A deep learning-based multi-model ensemble method for cancer prediction. *Computer methods and programs in biomedicine, 153*(January), 1-9. doi:10.1016/j.cmpb.2017.09.005

Xiao, Y., Wu, J., Lin, Z., & Zhao, X. (2018). A semi-supervised deep learning method based on stacked sparse auto-encoder for cancer prediction using RNA-seq data. *Computer methods and programs in biomedicine, 166*(November), 99-105. doi:10.1016/j.cmpb.2018.10.004

Xu, Y., Wang, Y., Luo, J., Zhao, W., & Zhou, X. (2017). Deep learning of the splicing (epi) genetic code reveals a novel candidate mechanism linking histone modifications to ESC fate decision. *Nucleic acids research, 45*(21), 12100-12112. doi:10.1093/nar/gkx870

Yin, Z., Zhao, M., Wang, Y., Yang, J., & Zhang, J. (2017). Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer methods and programs in biomedicine, 140*(May), 93-110. doi:10.1016/j.cmpb.2016.12.005

Zhao, Z., Li, Y., Wu, Y., & Chen, R. (2020). Deep learning-based model for predicting progression in patients with head and neck squamous cell carcinoma. *Cancer Biomarkers, 27*(1), 19-28. doi:10.3233/CBM-190380

Zhong, B., Pan, S., Wang, C., Wang, T., Du, J., Chen, D., & Cao, a. L. (2016). Robust individual-cell/object tracking via PCANet deep network in biomedicine and computer vision. *BioMed research international* (August), Article ID 8182416. doi:10.1155/2016/8182416

Abdulrhman Samman Al-Asmari
Master of Cardiovascular Nursing,
Ministry of Health,
Academic Affairs and Training Department
Madinah, KSA
aabdulrhman985@gmail.com
ORCID 0000-0002-8796-7171