Classification of skin lesions via neural networks: How deep learning can improve medical diagnosis and raise health data literacy

Leopold Ringmayr
Occidental College

December 13, 2019

Abstract

With the rapid surge in medical data volume and velocity, computational techniques to assist in the analysis and evaluation of this data become increasingly necessary and critical. Image processing and machine learning offer novel and useful opportunities for medical diagnosis and research–for the medical professional as well as for the patient. Based on prior research on image processing, convolutional neural networks are the most prominent machine learning algorithm for image-based classification problems. This work employs convolutional neural networks to classify skin lesion images from the HAM10000 dataset. Three neural networks are evaluated and compared using the project dataset: a custom neural network, an AlexNet model, and a ResNet18 model. Overall, these networks delivered an average classification accuracy of approximately 78 percent, with none of the three models having significantly different accuracy than the other. Additionally, this project delves into data exploration and visualization as a means of communication. Visualizations convey findings and information about skin lesion predictions, characteristics, and occurences to the user. Using d3.js, possible visualizations of the HAM10000 dataset and the neural network results are explored, using Bubble Charts, spider charts, among others. The objective of this work is to establish the feasibility of using neural networks for medical image classification, and to display this medical data using D3 visualizations.

1 Introduction

This work examines how machine learning and data analytics can improve medical diagnosis, patient care, and health data literacy among patients. Research on existing machine learning approaches to disease diagnosis and prediction shed light on how these computational techniques can be leveraged to communicate medical data–both to the patient and to the medical professional. One application of machine learning to medicine is the classification and diagnosis of skin

lesions. Machine learning algorithms could provide more accurate diagnoses than a dermatologist, at a quicker rate. This machine learning-based skin lesion classification should distinguish between different classes of lesions, and should mainly establish whether a lesion is malignant or benign. Skin lesion diagnosis serves as one example that illustrates how machine learning can be applied to the medical field. This project not only entails extensive research on machine learning and data analysis, but it also delves into fields such as medicine, public health, or bioethics. The main emphasis of this work lies on image processing via convolutional neural networks. One aspect that makes this project challenging is the analysis and evaluation of the dataset containing dermatoscopic images of skin lesions and training a machine learning model that is capable of correctly classifying a given image, with a sufficient accuracy. The second component of this work deals with data visualizations of the skin lesion findings, to show relationships between classes and potentially useful findings for the user. The goal with this visual component of the project is to represent this skin-lesion data visually, in a fashion that may be useful to medical professionals in their workflow. Patients may benefit from these visualizations as well, and may gain a broader understanding of their own health through this data-driven approach to health. Ultimately, machine learning should serve as a means to solve problems in other disciplines–medicine, in this case–and the data visualizations are a form of communication of these highly technical findings and algorithms to the user. Visualizations help to increase health data literacy among patients and practitioners, and to promote a new self-sufficiency and self-empowerment through medical data. Fully leveraging the potential of medical data and making this information widely accessible, in turn, could have a profound impact on how patients view and treat their own health.

The term "big data", widely used in both academic and popular literature, generally encompasses the collection, analysis, evaluation, and presentation of large volumes of heterogeneous data. Oftentimes, the analysis of data relies on machine learning techniques, which are becoming increasingly prevalent in data-driven industries such as finance, commerce, scientific research, and healthcare. With the rapid increase in health data collection, health professionals are turning to machine learning and data science to organize and analyze the data. However, the age of big data poses several main challenges: Sharma et al. (2018) characterizes these challenges as the three attributes of big data, or "the 3 V's": volume, velocity, and variety of data (Sharma et al. 2018). In other words, handling immense amounts of raw data, which is oftentimes incomplete or inaccurate, can be problematic. It is equally necessary to consider potential dangers and pitfalls of big data in healthcare. Although big data analytics and machine learning hold great promise in addressing problems in medicine, the ethical ramifications such as data privacy and security, legal challenges, or biases must also be addressed. This project also aims to shed light on how deep learning is affecting the medical sector–for better and for worse. Deep learning holds great promise to democratize medical care, giving the patient a more transparent and understandable insight into the diagnosis and treatment process. At the same

time, deep learning poses an enormous threat to a fair medical and research practice, depending on who collects, curates, and evaluates medical data.The project also deals with ethical questions related to data security and biases in the medical field. As the so-called "big data revolution" progresses, fields such as digital health or medical informatics are flourishing–and it is precisely this combination of computer science and bioscience that holds enormous promise in addressing some of the most pressing problems in medicine.

In recent research such as Esteva et al. (2017) and Tschandl et al. (2018), image processing of clinical photographic images has been identified as an especially promising tool for medical diagnosis. Computational techniques allow health professionals to predict health conditions early on, thus increasing the probability of accurate diagnosis and successful treatment–or eliminating the need for any treatment at all. Deep learning also has very interesting implications for the field of personalized medicine, where medical treatments are tailored to the individual patient. This type of custom treatment could be developed based on personal patient data that is processed using deep learning. Beyond patient care, the machine learning algorithms could also reveal larger trends or developments in public health, thus aiding in medical research on a large scale. In order to fully embrace this new approach to medical treatment, a broader understanding of the underlying machine learning algorithms and visualization techniques is necessary, which is precisely the focus of this work.

2 Background

2.1 Data Analytics

Data collection and analysis have always been an important part of medical and biological research; however, in more recent years, the rapid increase in the amounts of data collected and processed on a daily basis have given data science a new importance in the medical field. Furthermore, wearable devices, smartphones, and mobile health applications which measure and record a variety of different personal information, have only contributed to the prevalence of "big data". The applications of data science in medicine range from research of epidemiology to the care of an individual patient. Factors such as volume, velocity, and variety of data, or the or "the 3 V's", as Sharma et al. (2018) describes them, are the main challenges that can complicate the data analysis process. Therefore, it is imperative that heterogeneous and complex medical data can be processed and evaluated effectively. Machine learning algorithms, for example, can be utilized to gain novel insights into patterns or trends in data and can potentially aid in medical research and in a practical environment, such as clinics or doctor's offices. One specific implementation of "big data" analytics in health are Electronic Health Records (EHRs). In recent years, EHRs have been identified as an important step towards a data-driven health sector. EHRs

allow for a uniform and highly accessible collection of patient data. In their research, Sacristán et al. find that "the real challenge of big data is how to use large-scale population- based analyses to benefit individual patients" (Sacristán et al. 2015). One can extend this idea of predictive analytics for the individual by examining how machine learning on Electronic Health Records (EHRs) can filter important information for a patient.

2.2 Machine Learning Applications in Medicine and Dermatoscopy

The ambitions of using machine learning in the health sector are very diverse: from individual patient care to research projects over longer periods of time. For patients, one of the main benefits of using machine learning is that the diagnosis and classification of disease are instant, allowing for more accurate and more immediate treatment. Image classification, for example, can be used for medical diagnosis and prediction of disease onset. On a larger scale, data-driven medicine holds great promise for research and development because large amounts of patient data can be evaluated to shed light on the effectiveness of existing medical treatments, patterns of disease in certain area, or on possible new medicine or treatment options. One of the large challenges is integrating machine learning-based technology into the workflow of a health practitioner, mainly because the algorithms and techniques need to be understood before they can be used effective. In a medical setting, machine learning can be utilized to combat the challenges of incomplete, unstructured, and inaccurate data. Here, machine learning is useful for disease detection and prediction, for prediction of epidemiological phenomena, and for the development of pharmaceuticals. This machine learning-based approach also supports the notion of personalized medicine, seeing as medical diagnoses and treatments can be tailored to the individual patient using their specific data. Via predictive modeling, data is used to inform which treatments are most suitable for a patient. The field of dermatology lends itself to machine learning systems: As outlined in Tschandl et al. 2018, dermatology oftentimes relies on photographic and dermatographic images for medical diagnosis, making machine learning an ideal approach to improving recognizing and detecting skin conditions and diseases. One common challenge posed by machine learning systems for dermatological diagnosis is that systems are generally limited to a small number of conditions or classes. Moreover, it is notable that prior research "disregarded non-melanocytic pigmented lesions although they are common in practice" (Tschandl et al. 2018). Therefore, more recent has focused on examining a wider array of lesion types. The HAM10000 dataset, which stands for "Human Against Machine with 10000 training images" (Tschandl et al. 2018), provides a an extensive framework for dermatological machine learning implementations in future work. The dataset consists of dermatoscopic images which stem from two different research sites, the Medical University of Vienna in Austria and the medical practice of Cliff Rosendahl in Queensland, Australia. The images from the Medical University of Vienna were recorded before digital photography and were ultimately stored

as digitized diapositives. All images were reviewed for image equality and all images were standardized for analysis purposes. In addition, lesion images were recorded at different magnifications to prevent image quality loss by cropping and zooming into images. As Tschandl et al. (2018) points out, this dataset was designed as an open source resource for researchers "to boost the research on automated diagnosis of dermatoscopic images" (Tschandl et al. 2018). Due to the large number of dermatoscopic images and the standardized format in which the images are presented, the HAM10000 dataset is ideal for an image processing-based project.

2.3 Image Classification Techniques for Cancer Screening and Treatment

In recent literature, artificial neural networks have been used frequently in image processing for diagnosis of different types of cancer. In artificial neural networks (ANNs), the input nodes or neurons consist of observational data. Based on these input values, other nodes are calculated, which form the successive layers of the network. Nodes that operate within the network are referred to as hidden nodes, which ultimately calculate the output nodes. The network operates on a feed-forward mechanism such that a weighted sum of inputs is used to determine the value of the next layer in the network; the outputs become the input of the next layer. The images below from Meyfroidt et al. (2009) show a schematic drawing of an individual node (left) and an artificial neural network of nodes (right). The individual neuron takes inputs x1, x2, . . . , xn, where each input has a distinct weight. A function (.) takes a weighted sum of the inputs to calculate the output y of the neuron. The objective is to produce an output y that is sufficiently close to the target output value t. Back propagation can be used to adjust weights in the neural network in order to minimize the difference between y and t. A drawback with this algorithm is that neural networks are prone to overfitting; however, by adding a stopping iteration, this problem can be mitigated. For ANNs, extensive training data is required for the algorithm to yield accurate results.
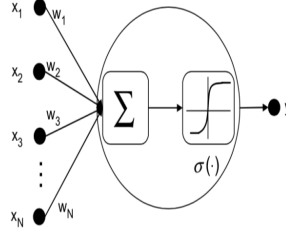
Fig. 4. A single hidden node or neuron in a network, which is a function of nodes (x) in a previous layer to predict the value of a node in the next layer. A weighted sum of the inputs is passed through a function σ(.) forcing the output of the neuron to be within a given interval. The output of a single unit is a thresholded linear combination of the inputs: $y = \sigma(x \cdot w) = \sigma(\sum_{n=1}^{N} w_n x_n)$. If σ is chosen to be a step-function such that the neuron's output can only take on the values +1 or −1, then the neuron is known as a perceptron. This type of neuron is typically used for scenarios with Boolean inputs. Another neuron of choice is the sigmoid unit, where σ is the sigmoid or logistic function. In this case, the output of the neuron can be any continuous value in the interval [0,1]. The weights can be seen as the components of a weight vector with the same dimensionality as the input space. This weight vector defines a decision surface in the input space. More complex and expressive (i.e. nonlinear) decision surfaces can be obtained when using a collection of units organized in a network configuration.
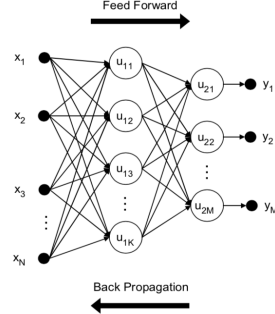

Fig. 5. Graphical representation of a neural network with N input units (x), K units (u) in a hidden layer, M units (u) in the output layer, corresponding to M outputs (y). For each training example **x** and target **t**: the input **x** is propagated forward through the network. The errors are propagated backward through the network. For each output unit u_{2m} the error term δ_m is calculated: $\delta_m = y_m(1 - y_m)(t_m - y_m)$. For each hidden unit u_{1k} (with output y_k), the error term δ_k is calculated: $\delta_k = y_k(1 - y_k)\sum_{m=1}^{M} w_{km}\delta_m$. Next, all weights w are adjusted according to: $w = w + \Delta w$. Where $\Delta w = \eta \delta x_i$. δ depends on whether the unit belongs to the output or to a hidden layer, x_i is the input value to the unit with weight w, η is the learning coefficient, a constant that regulates the magnitude of Δw, or how much the weights are allowed to change per iteration. This back propagation algorithm is used for training networks of sigmoid neurons, which have a continuous differentiable σ function, a requirement for the computation of the gradient of the error.

Image from Meyfroidt et al. (2009), Individual node (left) and neural networks of nodes(right)

Neural networks have shown to be particularly useful for the analysis of sensor data (including photographic images). One use case discussed in Meyfroidt et al. (2009) is the prediction of survival of ICU patients. Prediction of survival using ANN was compared to prediction via logistic regression. Overall, the authors found that ANNs showed a higher percentage of correctly classified ICU patients with a lower error compared to the logistic regression-based approach, thus underlining the potential of using ANNs in a clinical setting. Overall, neural networks are especially useful for projects and problems requiring image processing, mainly because of the numerous layers within the algorithm, allowing or image recognition.

2.4 The Convolutional Neural Network

A specific example of image processing using neural networks is the research conducted by Esteva et al. 2017. Esteva et al. (2017) examines how convolutional neural networks (CNNs) can be used to detect skin cancer–more accurately and at a faster pace than a dermatologist, as stated by the authors. Cancer screenings are another specific application of machine learning. For example, algorithms that detect skin cancer via image recognition is one highly practical use of machine learning.
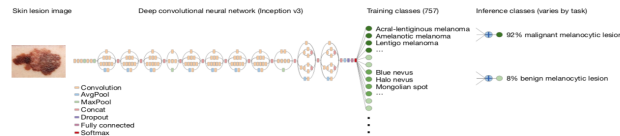

Image from Esteva et al. 2017, Layout of Convolutional Neural Network for Skin Cancer Classification

In this area, Esteva et al. (2017) examined how deep convolutional neural networks (CNNs) can detect and classify skin cancer. The ambition of this research lies in creating a classifier that detects skin cancer at the same level of accuracy as a human dermatologist. For this reason, the authors tested their predictive model against 21 dermatologists. The model was trained with 129, 450 clinical images which included a total of 575 disease classes. From the skin lesion images, the algorithm partitions the dataset into fine-grained subclasses used for training. It is notable that Esteva et al. (2017) counteract the problem of having a "black box" algorithm, as they lay out how coarser disease classes in the inner layers can be obtained from the smaller fine-grained classes. The benefit this machine learning approach offers over conventional treatment methods: Dermascopy requires specialized equipment and histological images require invasive treatment on the patient, while the algorithm is able to classify based on photographic images. In addition, these photographic images are robust to variation because of the high volume of training data. Important parameters to establish the algorithms feasibility are specificity and sensitivity, which allows for comparison to the dermatologists' classifications. As seen in the figure below, the CNN maintains very high specificity and sensitivity, outperforming nearly all of the 21 dermatologists.
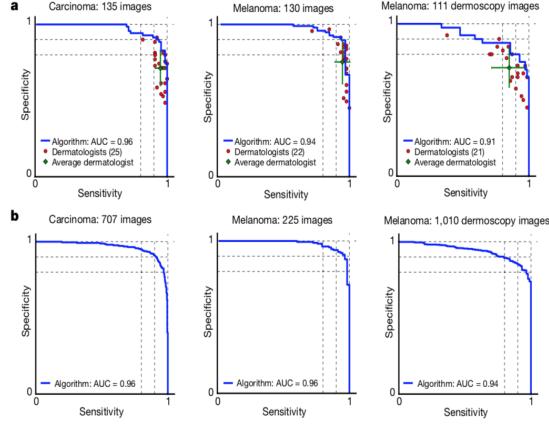
Image from Esteva et al. 2017, Specificity and sensitivity findings for different disease classes

One feature of convolutional neural networks that makes them especially suitable for image processing is their hierarchical structure. CNNs simplify the highly complex image data. CNNs rely on breaking down the complex input layer into matrices of lower dimension (the so-called convolutional layer). Techniques such as Pooling further decrease the dimensionality of the data. One popular CNN is the ResNet50, which Li et al. (2018) rely on in their work. This is ideal for medical imaging–skin lesion images can have very subtle differences, which may not be captured using traditional neural networks. A similar image-based approach the one presented in Esteva et al. (2017) has been adapted in other research as well. CNNs are not limited to identifying skin cancer, but have been successfully employed for the detection of lesions in mammograms (Ribli et al. 2017). A model using CNNs is offered as an improvement to current Computer Aided Detection systems (CAD) in order to provide more accurate results in cancer detection and classification. This technique also aims to minimize human intervention. It is notable that neural networks appear to be one of the more commonly used techniques in recent years for research on diagnosis.Overall, the findings from Ribli et al. (2017), Esteva et al. (2017), and Wozniak et al. (2018) suggest that neural networks could become viable diagnostic tools in cancer research.

The two main objectives of a CNN are to extract the features while reducing the dimension of the image, to make the network computationally efficient. The CNN reads in an image as a matrix of pixels. This project employs color photographs, hence the corresponding matrices have depth of three–for the red, green, blue values. Generally, the CNN uses a series of filters to extract important features in the image and it reduces the dimensions of the image simultaneously to make the processing more manageable. This transforms the original matrix into smaller matrices containing the high level features of the original image. Training the network allows for a more precise and nuanced detection of

features, seeing as the weights within the Kernels and the Fully Connected layer are updated with every training iteration. There are different types of layers within the network: The Convolutional Layers, or Kernels, are smaller matrices that traverse across the original image matrix and extract features from the image, creating so-called feature maps. A ReLU activation then normalizes the values of the convolutional matrix, setting any negative entry equal to 0. This ReLU step introduced non-linearity into the network, which accounts for the non-linear input data, namely the original image. Pooling layers, then, reduce the size of this convolutional matrix via Max Pooling or Average Pooling. Max Pooling is the more frequently used variant, and loads the maximum value of a smaller section within the given matrix to form a smaller, more manageable matrix. The three previous steps can be repeated to extract increasingly more complex features, at which point the matrix can be flattened into a vector. The Fully Connected Layer is a vector (fully "flattened"image) that connects each entry to a label. The classification occurs via Softmax classification, which outputs scores (usually via cross-entropy) indicating how closely an entry is related to a label. Based on these scores, a final label can attributed to the image. It is notable, that the training process involved backpropagation and a number of epochs to encapture the image's important features. A general overview of a CNN is provided below.
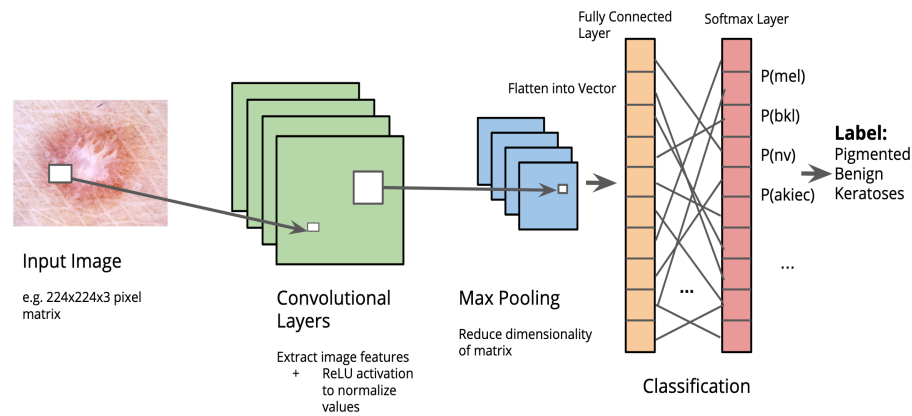
General Architecture of a CNN

For training a CNN model, there are several larger available pre-trained architectures, such as AlexNet or ResNet. These pre-trained models are useful in image processing projects because they provide a network that already has the capability to distinguish many images. Thus, in a project such as this one, the network can be trained on the highly specific and minute details of image, rather than beginning to learn without any prior reference. This process is referred to as transfer learning and allows for more accurate classifications, especially with complex data, such as medical images. Two architectures used in this work are

introduced below.

2.5   AlexNet

AlexNet is a pre-trained convolutional neural network first created in 2011; Named after its creator Alex Krizhevsky, it is pre-trained on approximately one million photos derived from the ImageNet database. AlexNet contains 8 layers; of these, five layers are convolutional layers and 3 layers are fully connected layers with "650,000 neurons" (Krizhevsky et al. 2017). One aspect that sets AlexNet apart from other networks is the fact that it utilizes so-called Dropout layers to mitigate overfitting. Dropout layers cause neurons to be removed from the network; it is used in lieu of regularization. In addition, AlexNet was one of the earliest popular convolution networks to use the ReLUU function as a means for introducing non-linearity instead of the Tanh function. Because of its architecture and high performance, AlexNet is very frequently used for transfer learning problems.

2.6   ResNet-18

Another neural network that has proven popular for transfer learning is ResNet, also referred to as Residual Network. ResNet belongs to the group of Artificial Neural Networks (ANNs). There are different iterations of ResNet, such as ResNet-18 or ResNet-50 (indicating the number of layers). It is notable that a higher number of layers does not automatically correlate to higher performance or prediction accuracy. In general, it is challenging to maintain a high accuracy for very deep networks. ResNet-18 has 18 layers. The ResNet architecture possesses characteristic residual blocks and skip layers. These skip layers mitigate the vanishing gradient of deep networks, and allow for layers to communicate (using activations) with layers other than the next one in the network. Like AlexNet, this network has been pre-trained on over one million images (all from ImageNet).

2.7   Data Visualization

Research in recent years , such as the research conducted by West et al. 2014, suggests a new approach to medicine, which places a greater emphasis on data science and computer science. Ideally, this new approach is more patient-centric than the traditional model: In the traditional model, doctors and healthcare providers are the main acting entities, while the patient assumes a somewhat passive role. Technologies such as machine learning and big data analytics can empower the patients to become more self-sufficient and can improve patients' understanding of medicine.

Research conducted by Ledesma et al. (2016) highlights how visualizations can increase usability of an EHR interface by making the data accessible to

the user. To communicate these ideas to the broader public, software libraries designed specifically for health visualizations prove extremely useful. For example, Ledesma et al. (2016) introduce hFigures, a JavaScript library for health data visualization, which is built on D3.js. In this research, the authors created application and a JavaScript library that is designed for the visualization of biological and health-related data. Creating an application that utilizes a library such as hFigures means considering user experience aspects as well and determining which health parameters are vital to convey a specific message about the data. In the case of hFigures, blood pressure, sleep, fitness, lab testing, among others are considered as primary indicators of a person's general health. It is important that these indicators are chosen with the audience of the visualization in mind–for a patient, this means creating a concise and clear overview of essential data (Ledesma et al. 2016). Another highly popular data visualization tool that places emphasis on interactive, user-friendly charts and graphs is Tableau. Tableau's focus lies on creating graphs instantly from one or multiple data sources (SQL database, Excel file, JSON file, XML, etc.) without required programming knowledge. Via a desktop user interface, the user can import the dataset and manipulate which variables to consider and which chart types to use. Tableau visualizations can then be published to a webpage.
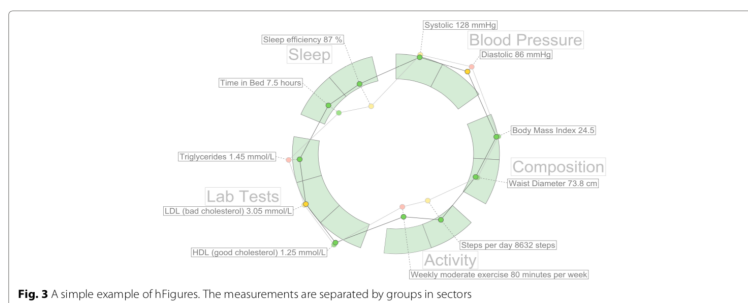

Fig. 3 A simple example of hFigures. The measurements are separated by groups in sectors

Image from Ledesma et al. (2016), Circular Visualization

2.8   Using D3 for Visualization

The open-source D3.js library was first created in 2011 by Mike Bostock. The library provides open-source JavaScript frameworks for different types of visualizations. Using D3 for a project entails viewing the D3 gallery, and establishing the chart types that are suitable for the project at hand. Using the open-source code frameworks one can adjust existing JavaScript chart functions to match the project data. This includes accounting for the type of data , the number of variables, the labels, chart colors, etc. D3 utilizes Scalable Vector Graphics (SVGs) to append the graphs and charts to the HTML page. By appending SVG objects in the HTML page, images of charts are directly added to the webpage. The data for D3 charts is loaded in either through .csv files and/or JSON files, so that data can be dynamically processed and instantly visual-

ized. In contrast to tools such as Tableau, D3 requires extensive knowledge in JavaScript, HTML/CSS, and the DOM (Document Object Model). D3 visualizations are embedded within JavaScript files and are expressed as JavaScript code. Individual shapes required for a visualization (including size, orientation and color of these shapes) are all defined in the JavaScript code. The templates and frameworks provided directly through d3js.org are open to the public and have these shapes written and defined for different chart types.

Overall, the objective of the data-centric approach is to democratize patient care by allowing the patient to actively participate in their own health. Finding effective ways to integrate machine learning and data science into the workflow of both medical professionals and patients. This work seeks to use skin lesion classification as one specific application of convolutional neural networks; in addition, it explores how the machine learning aspect can be paired with data visualizations to convey important findings.

2.9   Ethical Concerns

This project also aims to address the ethical implications associated with big data analysis and data analytics within the medical field. One common problem is that a dataset that does not truly represent the population–either through selection bias or not enough power (Vayena et al. 2018). This problem is already highly prevalent, and the trend will only continue with growing accessibility to patient data. For this reason, the patient's data literacy–the ability to understand and draw conclusions from one's own data–becomes all the more crucial. In addition, the user must be made familiar with the general mechanism of the analytics process. "Black box" algorithms, where the exact mechanisms of an algorithm are poorly understand, make this aspect difficult (Vayena et al. 2018). Another large ethical issue is the question of blame and liability if an error occurs during diagnosis or treatment. If an error during diagnosis or treatment occurs, it may be difficult to hold a health practitioner accountable if machine learning and artificial systems become the primary form of patient care. Seeing as this data-driven approach is fairly new, the exact ethical ramifications remain to be seen.

2.10   Legislative Measures

More recently, legislative measures have been enacted to address the increasingly prevalent problems caused by big data analytics. Vayena et al. (2018) lists the European General Data Protection Regulation (GDPR) in their work. The GDPR is one of the most recent examples of direct legal measures that are being taken to promote data transparency and data fairness in Europe. Under the GDPR, consent from the users is required to record data; in addition, the user has the right to access all collected data. In the United States, regulations on data usage are starting to take hold as well. The FDA, for example, seeks to address bias in data through their Digital Health Innovation Action Plan.

The plan includes regulation of medical software; exact measures on data bias are yet to come. The American Medical Association (AMA) also has begun to address this changing data-driven medical sector through AI-related policy recommendations. However, as data analytics grows, further regulations and procedures are necessary to ensure data security, fairness and transparency.

3 Methods

3.1 Data Collection and Parsing

This project uses the HAM10000 dataset for all machine learning analyses and visualizations, which was curated for machine learning purposes and for skin cancer research. The HAM10000 dataset, which stands for "Human Against Machine" consists of over 10,000 different images of skin lesions and stems from the Harvard Dataverse. An accompanying csv file maps the image number to one of seven lesion classes (Melanoma, melanocytic nevi, dermatofibroma, vascular lesions, basal cell carcinoma,actinic keratoses/ intraepithelial carcinoma , benign keratosis-like lesions). In addition, this csv file contains metadata about the lesion in the corresponding image, i.e. age, sex, locality of the lesion on the body, etc. The metadata is later used for the D3 visualizations. The images in the dataset are all standardized in size with dimensions 450x600x3. Moreover, all of the lesions are centered in the images and only one lesion is pictured in each image.

To use HAM10000 dataset for machine learning purposes it was important to separate the images by lesion class into respective subfolders. For this step, a Java program was written that moved the images from the original dataset into the seven subfolders for each lesion class. Ultimately, these seven folders contained all of the original images.

3.2 Building the Convolutional Neural Network

The next step was building the convolutional neural networks in MATLAB to classify the images in the HAM10000 dataset. Convolutional neural networks are appropriate for the skin lesion classification because, unlike conventional neural networks, CNNs are able to retain spatial dependencies of the image. This is highly important in order for the model to detect objects and features in the image that set it apart from other classes. All machine learning models were built and tested in MATLAB. Three different CNNs were built for this project: A network with the AlexNet architecture, a model using ResNet-18, and a custom network, which was not pretrained. AlexNet and ResNet-18 were chosen to compare popular transfer learning approaches to a custom "from-scratch" network. AlexNet was chosen for its frequent use in transfer learning problems. ResNet-18 was chosen for its unique residual block structure, and its high performance. ResNet-50 was initially considered, but ultimately not selected for the final three models; ResNet-50 proved to be highly computationally expensive,

especially running on a single CPU. Early stopping via a validation patience of 3 was employed, however the loss converged to a value that was too high for further consideration, compared to the other models (validation loss during training was around 50 percent).

For all three models, the images were stored in a DataStore and divided into test and training datasets with a 20:80 ratio. Then the training set was split for validation purposes. The input for all three models were images from the HAM10000 training set; the output for each of the models was the predicted label (i.e. lesion class) for a given image. Each model was built in a separate MATLAB file and saved for testing purposes. Because training neural networks is computationally costly, early stopping was implemented to stop the training process when the network had sufficiently converged to a specific loss.

The AlexNet model required an input size of 227x227x3; the training and test images were resized accordingly in an Augmented Image DataStore. The general architecture of AlexNet was preserved; however, the fully connected layer was adjusted and replaced to match the number of labels (7 lesion classes). After training with different learning rates, it became evident that the optimal learning rate for AlexNet on the HAM10000 was 0.001–other learning rates such as 0.1,0.01, and 0.0001 converged at a low accuracy. This could be observed over the training progress graph in which one could see the training and validation accuracy as the iterations progressed. ResNet-18 had an input size of 224x224x3, and it required adjusting and replacing the classification layer to match the project labels. The learn rate with the highest accuracy was 0.001 with no specified mini batch size, and 6 epochs (as was the case for the AlexNet model). For the custom network, layers were defined in the training section of the MATLAB file. The optimal initial learn rate for this custom network was 0.1 with no specified mini-batch size. Other learn rates, including 0.001 and 0.01 converged at a lower accuracy during training. The network ultimately had 15 layers, using ReLU activation and Max Pooling. Across all models, parameters were adjusted based on the algorithm's performance during training. All models were trained in a MATLAB section separated from testing; this aids reproducibility, as models can be trained and tested individually. In addition, testing the model does not require a re-training, rather the finished model can loaded and utilized.

## 3.3   Testing and Evaluating the Models

To test the three models, an initial 80:20 split into test data and training data was conducted. Two types of tests were conducted: One test of the model on the entire test set and one test using bootstrap resampling. All three models were tested using bootstrap resampling on the test set. In a separate MATLAB file, the three saved CNN models were loaded and the bootstrap resampling was conducted. The main measure for "success" was the accuracy of the models. Then a paired t-test was conducted using the average accuracy values from

the bootstrap resampling. The bootstrap resampling and t-test were used to evaluate the three models and to compare them to one another. Additional measures in this project to evaluate the success of the algorithm were specificity and sensitivity; these were used as measures for false negatives and false positive. Sensitivity was especially interesting because it was a measure for how many patients with a malignant lesion were actually classified as having a malignant lesion. Specificity, on the other hand, measured how many patients without malignant lesions were classified as having not-malignant lesions.

## 3.4 Creating D3 Visualizations

For the D3 visualizations, the initial step required viewing and comparing different chart types in the D3 Gallery to establish which visualization types would best suit the data for this project. All visualizations were appended to different HTML pages, each with a correpsonding CSS stylesheet. In other words, for each visualization in this project, there is a JavaSCript file (with embedded D3), an HTML file, a CSS file, and one or more data sources (.csv file or JSON file). The framework for a data visualization is written in JavaScript and is available on D3js.org as open-source code, which can be used for used for one's own project and visualizations. This code is then modified to meet the criteria of this project. For D3, it is necessary to host visualizations on a web server over an HTML page. This is mainly because D3 makes use of the Document Object Model, in which data is read in dynamically over the browser. For this project, all visualizations were run over a local MAMP server on localhost. MAMP must be installed for this; other web servers can be used as well. To view a visualization, one opens the HTML file in the web browser over localhost.

For example, to open the stacked radial bar chart showing the prevalence of lesion classes across different age groups, one would start the web server and run http://localhost/radial/index.html. This should yield the following image:

15
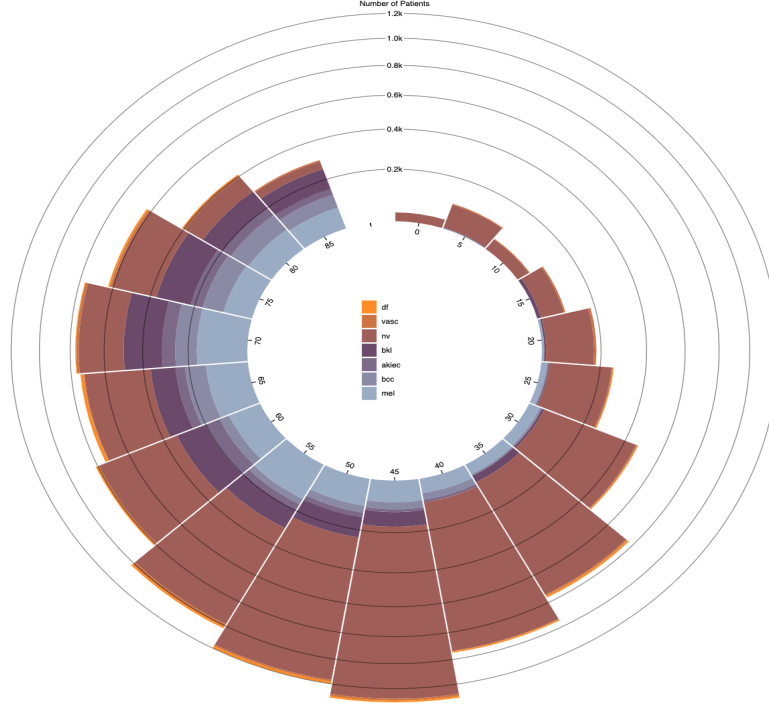
Radial Stacked Bar Chart showing lesion class frequency across age groups

This radial bar chart draws from the radial approach to conveying medical data outlined in Ledesma et al. 2016. The different bars represent different age groups: Using the age provided in the HAM10000 dataset, a new .csv file was created using a Java program that found the frequencies of the different ages in the dataset. The new .csv file was used as input for the D3-based JavaScript file. The colors in the image represent different lesion classes, for which a legend is provided in the center of the graph. The height of each bar indicates the number of patients of a particular age group that have a lesion belonging to a given class. This allows the user to see how frequent certain lesions are across at different ages. A further visualization is used to convey the output of the softmax layer for a test image is the spider chart. The spider chart shows how closely related a skin lesion is to the different classes, based on the probabilities from the neural network's softmax layer. The lesion classes each hold one edge in the "spider web" of the graph. One such Spider Chart is shown below for reference.
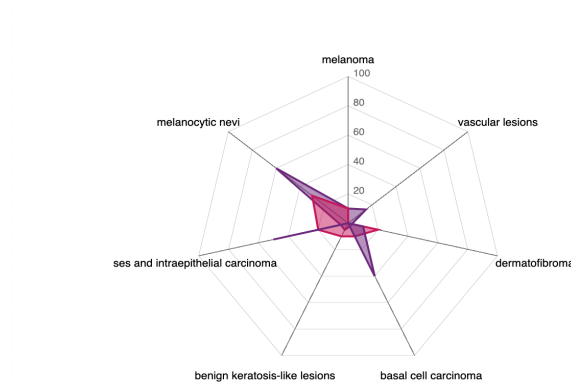
Example of a Spider Chart generated in D3

Another visualization was a so-called Bubble Chart which conveyed where on the body benign keratosis-like lesions are most likely to occur. Each possible locality is represented as a circle (i.e. one circle for "face", another for "back", etc.) One should note that the labels for the bubbles were added in manually after creating the visualization. As a supplement to the D3 approach, Tableau charts were created as a means for comparison (no coding required for Tableau, just downloading Tableau Desktop and importing the HAM10000 .csv file to create the visualizations). The process of creating the circles (i.e. "bubbles") and labeling them is illustrated below.
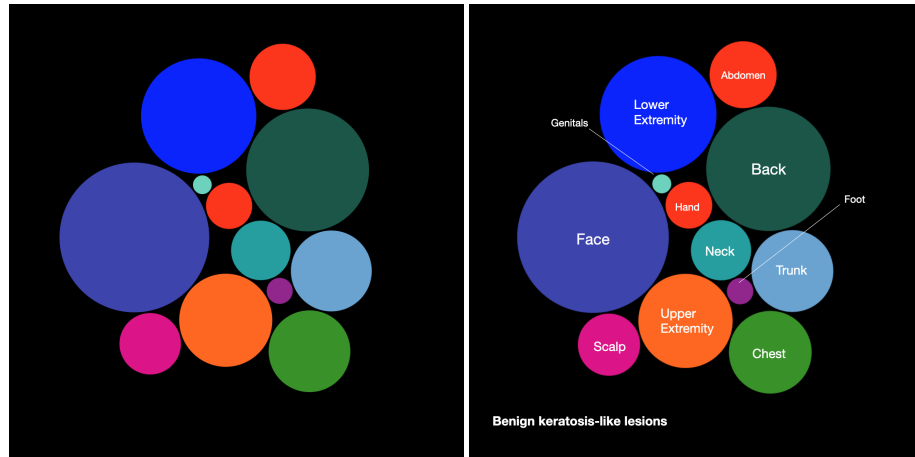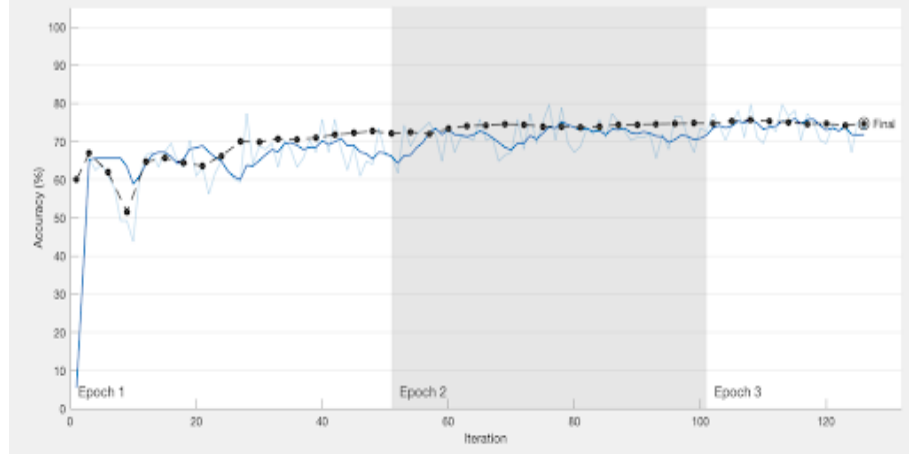

Bubble Chart in D3, before and after labeling circles

4 Results

4.1 Training Progress and Validation

All training progress graphs show network accuracy plotted against the number of iterations. Validation accuracy is shown as the black dotted line while training accuracy is represented by the blue continuous line.
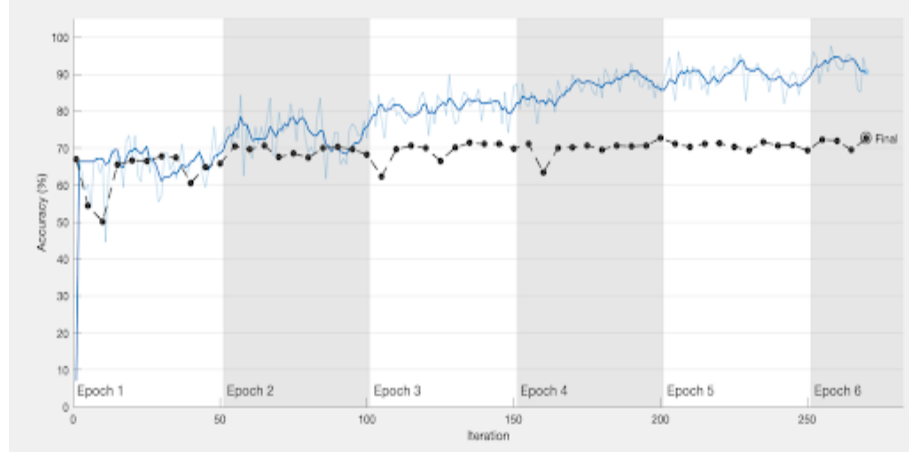

Training progress for AlexNet shown as accuracy against iterations

For the AlexNet model, the validation accuracy was 74 percent and the accuracy on the entire test set was 75.29 percent. Observing the training progress, one sees that the validation accuracy and training accuracy remain nearly identical as the iterations progress. This model had a sensitivity of 0.4 and a specificity of 0.9721.
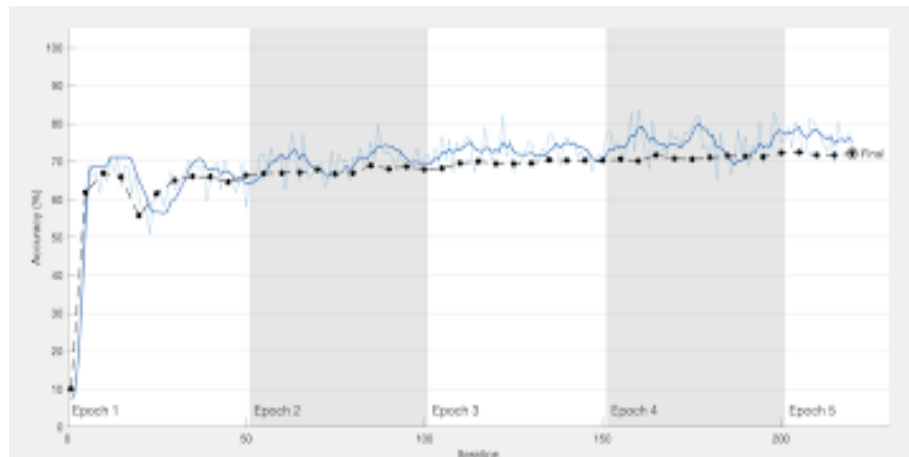
Training progress for ResNet-18 shown as accuracy against iterations

The ResNet-18 model achieved a validation accuracy of 72.10 percent; the accuracy on the entire test set was 75.89 percent. Similarly to the AlexNet model,the difference between validation and training accuracy remains the same as the iterations progress. In addition, the ResNet-18 model had a specificity of 0.9737 and a sensitivity of 0.6505.

Training progress for Custom Network shown as accuracy against iterations

The custom network showed a validation accuracy of 72.72 percent and a training accuracy of 85 percent. Unlike the other two models, the custom model's training accuracy continued to increase while the validation accuracy remained close to 70 percent. The sensitivity was 0.3689 and the specificity was 0.9884.

4.2   Bootstrap-Resampling and Model Comparison

Other resutls stem from the bootstrap-resampling and the paired t-tests. The AlexNet model had an accuracy of 76.77 percent on the boostrap sample, the ResNet-18 model had an accuracy of 75.96 percent on the boostrap sample, and the custom network had an accuracy of 73.74 on the boostrap sample. The paired t-test between AlexNet and ResNet-18 showed a p-value of 0.6285. Between AlexNet and the custom network, the p-value was 0.0955. Finally, the paired t-test between ResNet-18 and the custom network had a p-value of 0.1066. The significance level was set at 0.05.

5 Discussion

5.1 Discussion of Results

Based on the results, especially from the paired t-test, one can conclude that the three models performed very similarly to one another. This may indicate that the ResNet model is not fully optimized; one would have expected this model to outperform the other two. One interesting finding is that the highest accuracy on the test set stems from the custom network. Moreover, when looking at the training progress for the custom network, one sees that the difference between the training set accuracy and the validation set accuracy increases with the number of iterations. This would suggest overfitting: Although the test set was separated from the training set and only used at the very end for testing purposes, one must consider that these images all stem from one large set of similar and standardized images. There may have been unbalanced training/validation/test data, because although the images were shuffled and assigned to these datasets randomly, the number of each lesion class in the HAM10000 was not equal. For example, the number of images for the nevus class was larger than the number of images of melanoma lesions. This could serve as a possible explanation for the overfitting. Overall, the highly accuracy on the test set for the custom network was an unanticipated finding. In the boot strap sample, the accuracy for this custom network is very similar to the other two models. When evaluating the accuracy of the three models, one must take the image data used for training into account. Overall, all three models show an expected graph for the training progress; all them converge to a peak accuracy (and converge to a loss). Considering the complexity of the skin lesion images in the dataset as well as how minute differences between images from different lesion classes were, the accuracy is fairly high and the sensitivity is justifiable given the overall accuracy of approximately 75 percent. In a larger context, the models would not be suitable to replace a doctor's diagnosis in a medical practice, considering the accuracy of the three models. The relatively low sensitivity supports this conclusion that the models are not "industry-ready", although the ResNet-18 model had the highest sensitivity, as expected due to the skip layers in the architecture. Even with future work, the machine learning-based diagnosis could be seen as a tool and complement to the doctor's own expertise, rather than replacing the doctor's diagnosis.

5.2 Future Work

For future work, one could expand the dataset to include images outside of the HAM10000 dataset. One could also increase the number of images to train the models on. This would aid in generalizing the neural network model–and make the model more suitable for actual medical practice. This could also help increase sensitivity and recall–alongside increasing the accuracy of the models, reducing the number of false negatives is paramount. In addition, it is imperative to include other lesion classes other than the ones in the HAM10000 dataset

and to add labels to account for images that do not contain a skin lesion. For example, it would be problematic if a model failed to recognize cuts, freckles, birth-marks, etc. and if the model labeled these as (malignant) lesions.

For training the models in MATLAB, one could train the model using higher computational power to fully leverage all neural network optimization possibilities and to reduce time spent training. Future work could also include creating more visualizations in D3 or Tableau. Tableau would provide quicker insights into the data, seeing as the steps required to create a chart are far less than in D3; in addition, Tableau does not require external (HTML) files or a web server. Rather, the visualizations are stand-alone products–and are easier to manipulate and create. This makes data exploration faster and easier for the general public. Another alternative to D3 that one may consider in future approaches is R or the interactive R Shiny. Using these tools would provide more conventional graphs and charts; it would be of interest to compare a user's response to these new charts compared to some of the more unconventional D3 approaches.

Additionally, it would be of interest to further pursue D3 for visualizations and to create more complex and interactive charts. For this, a more extensive background in JavaScript is required.

5.3   Data Security and Biases in Medical Data

Big data analytics generally holds great promise for the medical field; however, there are dangers and pitfalls to this technology. The danger of bias holds especially true for medical and health data, which is especially personal and sensitive. One example is the potential of data abuse and theft. Another problem in data analytics is bias, which can introduced at any stage of the data analysis–from the data collection to the presentation of machine learning results. Ultimately, biases of developers and health professionals are directly reflected in the algorithms and analysis methods they create. One of the primary objectives of this project is to demonstrate how big data can serve and empower the patient or user–and to uncover where possible dangers and pitfalls with this technology lie. Therefore, factors such as understandability and transparency are paramount in my research. After reviewing literature on this topic, it has become evident that there is a deficit in the communication to the individual. In other words, measures need to be be taken to make health data presentable and understandable. This allows the user to make full use of the analyzed data, and also improves data transparency and fairness. Ultimately, these findings should be integrated into the workflow of the health professional or into the daily life of the patient.

Selection bias also could impact how the machine learning algorithm is trained, which in turn would lead classifications that are specific to that faulty dataset. Thus, it is paramount that the dataset itself is chosen carefully and with the possible bias in mind. Although it is stated that the HAM10000 dataset consists of data from different populations, it is unclear whether dataset actu-

ally provides an accurate representation of skin lesion images. In Tschandl et al. (2018), it is noted that the dataset stems from two primary sites in Austria and in Australia. Therefore, it remains unclear whether the HAM10000 dataset provides an accurate representation of different populations. In future work, one could perform similar image analyses on a variety of different datasets to minimize the possibility of bias within the used datasets. The difficulty with this lies mainly in the image preparation and the pre-processing time required to handle data from different sets and sources. Nonetheless, further work using different datasets could yield more refined, accurate, and fair results. In addition, future work could place greater emphasis on the predictive aspect of medical research rather than mere recognition. Another problem with machine learning is the "black box" nature of algorithms; the exact intermediate mechanism between input and output is sometimes poorly understood. Although the "black box" nature of certain machine algorithms may not be counteracted directly, clear visualizations can at the very least shed light on how results are understood and perceived.

This project examines the feasibility of utilizing deep learning for medical diagnosis and disease prediction. Through machine learning algorithms, concrete health conditions can be predicted for the individual patient; this approach also supports the emerging field of personalized medicine where personal data can be leveraged to create custom medical treatments. Current research on deep learning in medicine makes a compelling case for the implementation of machine learning techniques in medical practices–provided that the possible challenges are recognized and addressed accordingly. One notable shortcoming of many current research papers is that the exact data analysis mechanisms are often not explained in detail–or omitted entirely. Another aspect which requires further research is the consumer-end of the data analytics; the data must be understandable and usable to the user. This communicative aspect of deep learning is a further focus of this work–to use visualizations to clarify medical data and communicate its importance to the user. The larger goal of this research topic is to increase health data literacy among patients and practitioners, and to promote new self-sufficiency and self-empowerment through medical data. Fully leveraging the potential of medical data and making this information widely accessible, in turn, could have a profound impact on how patients view and treat their own health.

6   Works Cited

Ajani, Rafiq, et al. "How a Pharma Company Applied Machine Learning to Patient Data." Harvard Business Review Digital Articles, Oct. 2018, pp. 2–5.

Alsallakh, Bilal, et al. "Visual Methods for Analyzing Probabilistic Classification Data." IEEE Transactions on Visualization and Computer Graphics, vol. 20, Dec. 2014, pp. 1703–12. ResearchGate, doi:10.1109/TVCG.2014.2346660.

Amarasingham, Ruben, et al. "Implementing Electronic Health Care Predictive Analytics: Considerations And Challenges." Health Affairs, vol. 33, no. 7, July 2014, pp. 1148–54. healthaffairs.org (Atypon).

Bostock, Mike. "Data-Driven Documents." D3.Js, d3js.org/.

, Noel, et al. "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)." ArXiv:1902.03368 [Cs], Feb. 2019. arXiv.org.

Cruz, Harry, et al. Early Detection of Acute Kidney Injury with Bayesian Networks. p. 8.

Cruz, Joseph A., and David S. Wishart. "Applications of Machine Learning in Cancer Prediction and Prognosis." Cancer Informatics, vol. 2, Jan. 2006, p. 117693510600200. DOI.org (Crossref).

D3. "Radial Stacked Bar Chart." Observable, Observable, 6 Nov. 2019, observablehq.com/
@d3/radial-stacked-bar-chart.

Davies, Alec, et al. "Using Machine Learning to Investigate Self-Medication Purchasing in England via High Street Retailer Loyalty Card Data." PLoS ONE, vol. 13, no. 11, Nov. 2018, pp. 1–14. EBSCOhost.

Drell, Lauren, and Julie Davis. "Getting Started with Predictive Analytics." Marketing Health Services, vol. 34, no. 3, Fall 2014, pp. 22–27.

Esmaeilzadeh, Soheil, et al. "End-To-End Alzheimer's Disease Diagnosis and Biomarker Identification." ArXiv:1810.00523 [Cs], vol. 11046, 2018, pp. 337–45. arXiv.org.

—. "End-To-End Alzheimer's Disease Diagnosis and Biomarker Identification." ArXiv:1810.00523 [Cs], vol. 11046, 2018, pp. 337–45.

Esteva, Andre, et al. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks." Nature, vol. 542, no. 7639, Feb. 2017, pp. 115–18.

Foster, Kenneth R., et al. "Machine Learning, Medical Diagnosis, and Biomedical Engineering Research–Commentary." BioMedical Engineering OnLine, vol. 13, no. 1, July 2014, pp. 94–102. EBSCOhost.

Hammer, Marilyn J. "Research Ethics in Big Data." Oncology Nursing Forum, vol. 44, no. 3, May 2017, pp. 293–95. EBSCOhost.

Jurkiewicz, Carole L. "Big Data, Big Concerns: Ethics in the Digital Age." Public Integrity, vol. 20, Jan. 2018, pp. S46–59. EBSCOhost.

Kim, Seong Jae, et al. "Development of Machine Learning Models for Diagnosis of Glaucoma." PLoS ONE, vol. 12, no. 5, May 2017, pp. 1–16. EBSCOhost.

Kotsiantis, S. B., et al. "Machine Learning: A Review of Classification and Combining Techniques." Artificial Intelligence Review, vol. 26, no. 3, Nov. 2006, pp. 159–90. Crossref.

Krizhevsky, Alex, et al. "ImageNet Classification with Deep Convolutional Neural Networks." Communications of the ACM, vol. 60, no. 6, May 2017, pp. 84–90. DOI.org (Crossref), doi:10.1145/3065386.

kumar, Gagan, and Rohit Kalra. "A Survey on Machine Learning Techniques in Health Care Industry." International Journal of Recent Research Aspects, vol. 3, no. 2, June 2016, pp. 128–32.

Ledesma, Andres, et al. "Health Figures: An Open Source JavaScript Library for Health Data Visualization." BMC Medical Informatics Decision Making, vol. 16, Mar. 2016, pp. 1–19. EBSCOhost.

Lemon, Stephenie C., et al. "Classification and Regression Tree Analysis in Public Health: Methodological Review and Comparison with Logistic Regression." Annals of Behavioral Medicine, vol. 26, no. 3, Dec. 2003, pp. 172–81. Crossref.

Leslie, Hannah H., et al. "Health System Measurement: Harnessing Machine Learning to Advance Global Health." PLoS ONE, vol. 13, no. 10, Oct. 2018, pp. 1–15. EBSCOhost.

Li, Xiaoxiao, et al. Skin Lesion Classification Via Combining Deep Learning Features and Clinical Criteria Representations. preprint, Bioinformatics, 1 Aug. 2018. DOI.org (Crossref).

Lin, Yu-Kai, et al. "Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach." MIS Quarterly, vol. 41, no. 2, June 2017, pp. 473-A3.

Lucas, Peter J. F., et al. "Bayesian Networks in Biomedicine and Health-Care." Artificial Intelligence in Medicine, vol. 30, no. 3, Mar. 2004, pp. 201–14. Crossref.

Lyon, Aurore, et al. "Computational Techniques for ECG Analysis and Interpretation in Light of Their Contribution to Medical Advances." Journal of The

Royal Society Interface, vol. 15, no. 138, Jan. 2018, p. 20170821. (Crossref).

Maastricht University, NL, and Annika Richterich. The Big Data Agenda: Data Ethics and Critical Data Studies. University of Westminster Press, 2018. Crossref.

Mackinnon, Murray J., and Ned Glick. "Data Mining and Knowledge Discovery in Databases – An Overview." Australian New Zealand Journal of Statistics, vol. 41, no. 3, Sept. 1999, p. 255.

Mader, Kevin. "Skin Cancer MNIST: HAM10000." Kaggle, 20 Sept. 2018, www.kaggle.com/kmader/skin-cancer-mnist-ham10000.

Manisha, and Pawan Lathwal. "Exploring Classification Clustering Techniques for Predictive Analytics." International Journal of Recent Research Aspects, vol. 3, no. 2, June 2016, pp. 76–78.

Martens, David, et al. "Mining Massive Fine-Grained Behavior Data to Improve Predictive Analytics." MIS Quarterly, vol. 40, no. 4, Dec. 2016, pp. 869–88.

Mccarthy, John F., et al. "Applications of Machine Learning and High-Dimensional Visualization in Cancer Detection, Diagnosis, and Management." Annals of the New York Academy of Sciences, vol. 1020, no. 1, 2004, pp. 239–62. Wiley Online Library.

Menger, Vincent, et al. "Transitioning to a Data Driven Mental Health Practice: Collaborative Expert Sessions for Knowledge and Hypothesis Finding." Computational Mathematical Methods in Medicine, Aug. 2016, pp. 1–11. EBSCOhost.

Meyfroidt, Geert, et al. "Machine Learning Techniques to Examine Large Patient Databases." Best Practice Research Clinical Anaesthesiology, vol. 23, no. 1, Mar. 2009, pp. 127–43. Crossref.

Nadkarni, Girish N, et al. "Development and Validation of an Electronic Phenotyping Algorithm for Chronic Kidney Disease." AMIA Annual Symposium Proceedings, vol. 2014, Nov. 2014, pp. 907–16.

Nadkarni, Girish N., et al. "Development and Validation of an Electronic Phenotyping Algorithm for Chronic Kidney Disease." AMIA ... Annual Symposium Proceedings. AMIA Symposium, vol. 2014, 2014, pp. 907–16.

Ortiz, Justin R., et al. "Monitoring Influenza Activity in the United States: A Comparison of Traditional Surveillance Systems with Google Flu Trends."

PLOS ONE, vol. 6, no. 4, Apr. 2011, p. e18687. PLoS Journals.

Panahiazar, Maryam, et al. "Using EHRs and Machine Learning for Heart Failure Survival Analysis." Studies in Health Technology and Informatics, vol. 216, 2015, pp. 40–44.

Parikh, Ravi B., et al. "Making Predictive Analytics a Routine Part of Patient Care." Harvard Business Review Digital Articles, Apr. 2016, pp. 2–5. Prediction Modeling Using EHR Data: Challenges, Strategies, and a Comparison of Machine Learning Approaches. p. 9.

Predictive Analytics in Information Systems Research. Vol. 35, no. 3, 2011, p. 21. "PREDICTIVE ANALYTICS: Using Data to Enhance Clinical Quality." HHN: Hospitals Health Networks, vol. 90, Health Forum, 2016, pp. 49–59. EBSCOhost.

"Radar Chart Redesign." Popular Blocks, bl.ocks.org/nbremer/21746a9668ffdf6d8242. Ribli, Dezső, et al. "Detecting and Classifying Lesions in Mammograms with Deep Learning." Scientific Reports, vol. 8, Mar. 2018. PubMed Central

Sacristán, José A., and Tatiana Dilla. "No Big Data without Small Data: Learning Health Care Systems Begin and End with the Individual Patient." Journal of Evaluation in Clinical Practice, vol. 21, no. 6, Dec. 2015, pp. 1014–17. EBSCOhost.

Shameer, Khader, et al. "Machine Learning in Cardiovascular Medicine: Are We There Yet?" Heart, vol. 104, no. 14, July 2018, pp. 1156–64.

Sharma, Neha, et al. "Big Data Analytics in Health Care: A Literature Survey." International Journal of Recent Research Aspects, vol. 5, no. 1, Mar. 2018, pp. 127–32.
Soma, Jonathan irst name of the creator. "Making an interactive bubble chart with D3 v4" Youtube, created by Jonathan Soma, Nov 4, 2016, https://www.youtube.com/watch?v=lPr60pexvEM.

Tattersall, Andy, and Maria J. Grant. "Big Data - What Is It and Why It Matters." Health Information Libraries Journal, vol. 33, no. 2, June 2016, pp. 89–91. EBSCOhost.

Tschandl, Philipp, et al. "The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions." Scientific Data, vol. 5, Aug. 2018, p. 180161.

Vayena, Effy, et al. "Machine Learning in Medicine: Addressing Ethical Challenges." PLoS Medicine, vol. 15, no. 11, Nov. 2018, pp. 1–4. EBSCOhost.

Waller, Matthew A., and Stanley E. Fawcett. "Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management." Journal of Business Logistics, vol. 34, no. 2, 2013, pp. 77–84. Wiley Online Library.

Wang, Weiqi, and Eswar Krishnan. "Big Data and Clinicians: A Review on the State of the Science." JMIR Medical Informatics, vol. 2, no. 1, Jan. 2014. PubMed Central.

Weiss, Jeremy C., et al. "Machine Learning for Personalized Medicine: Predicting Primary Myocardial Infarction from Electronic Health Records." AI Magazine, vol. 33, no. 4, Dec. 2012, pp. 33–33.

West, V. L., et al. "Innovative Information Visualization of Electronic Health Record Data: A Systematic Review." Journal of the American Medical Informatics Association, Oct. 2014. Crossref.

Wozniak, Justin M., et al. "CANDLE/Supervisor: A Workflow Framework for Machine Learning Applied to Cancer Research." BMC Bioinformatics, vol. 19, no. 18, Dec. 2018, pp. 59–69. EBSCOhost.

Wu, Jionglin, et al. "Prediction Modeling Using EHR Data: Challenges, Strategies, and a Comparison of Machine Learning Approaches." Medical Care, vol. 48, no. 6, 2010, pp. S106–13. JSTOR.

7 Acknowledgements