First Draft

Review article

Target Journal: Clinical Diabetes, Endocrinology, and Metabolism

Clinical Use of Machine Learning in Rare Endocrine Disorders

Efe Y. Akinci[1], Elif A. Oral[1].

[1]Brehm Center for Diabetes Research and Division of Metabolism, Endocrinology and Diabetes, University of Michigan, Ann Arbor, MI, USA.

Short Title (max 50 characters):

Machine Learning in Rare Endocrine Disorders

Word Count: 4424

Tables: 1

Figures: 2

Address correspondence to:
Elif Arioglu Oral, MD
Professor of Medicine; Metabolism, Endocrinology and Diabetes Division
Department of Internal Medicine
Caswell Diabetes Institute
North Campus Research Complex, 25-3696
University of Michigan, Ann Arbor, Michigan 48109-2800
Phone: 734-615-7271
E-mail: eliforal@med.umich.edu

Abstract

Machine learning is an area of study in which computers generate solutions to problems they have not been explicitly programmed to solve. The accelerating pace of developments in machine learning has sparked research into its use in medicine. Previous studies have been able to achieve promising results through machine learning in the areas of patient identification, diagnoses, prognoses, and predicting patient and treatment outcomes. These results make machine learning a worthwhile tool to investigate for its use in the diagnosis and treatment of rare endocrine diseases. In the field of rare endocrine diseases, machine learning presents new possibilities and obstacles in many different contexts. Recent studies involving machine learning have shown promise in combating underdiagnosis and aiding in identifying patients who may have rare endocrine diseases. Additionally, machine learning has been successful predicting patient and treatment outcomes. Machine learning also has the capability to learn from images, a capability that has shown to be effective in diagnosing certain rare diseases. In this review article, we will briefly review the common concept of machine learning in medicine. We then will address the issue of data scarcity in rare endocrine diseases, considering the obstacles it may present to the use of machine learning and exploring possible solutions. Finally, we will have a quick look at lipodystrophy, a disease that presents challenges that machine learning may be suited to solve.

Keywords: Artificial intelligence, machine learning, computer assisted diagnosis, rare endocrine diseases.

Introduction

The technique of machine learning, a process in which a computer generates a solution to a problem through data analysis methods instead of being explicitly programmed to solve the problem, has seen increasing use in the field of medicine. Recent results from studies attempting to use machine learning in medicine have been promising. In particular, machine learning has shown great promise in the areas of diagnosis through medical records and imaging. In this article, we aim to review the current literature on the topic of the use of machine learning in diagnosing rare endocrine diseases and explore possible avenues for further research.

Machine Learning

Machine learning can be defined as “a computational process that uses input data to achieve a desired task without being literally programmed (i.e., “hard coded”) to produce a particular outcome” [1]. In the studies we will be focusing on throughout this review, many different types of machine learning algorithms are widely used. Therefore, we believe that it is important to include a short summary of the terms we will be using to describe and evaluate machine learning methods.

Types of Machine Learning

There are various methods in which a computer can solve problems. In this section, we explore a few common types of machine learning algorithms.

Supervised Learning

To train a supervised machine learning algorithm, we provide the computer with a method of measuring the error of its attempt, whether this be through labeled data or other means. The computer is then able to analyze its mistakes and adjust its solution in order to achieve greater accuracy through methods such as back-propagation. These kinds of algorithms are useful in cases where we have easy access to labeled data. Some examples may include training an algorithm to predict treatment outcomes when we already have access to a dataset that includes past patient outcomes, or an algorithm to diagnose a disease when we have access to health records that already differentiate between those who have the disease and those who do not.

Unsupervised Learning

An unsupervised machine learning algorithm trains on datasets where we do not know the labels and do not have a way of evaluating the error of a given attempt. These kinds of models are useful for exploring and finding structures in data. An example task would be training a model to classify images of tennis balls and oranges without labelling the images. Although the computer does not know which images are apples and which are oranges, it may classify pictures into general categories of “orange things” and “yellow things.”

Semi-Supervised Learning

A semi-supervised machine learning model is a mixture of supervised and unsupervised learning concepts. To train a semi-supervised machine learning model, we use labelled and unlabeled sets of data. Although the performance may not be as impressive as a supervised model training on a fully labeled set of data, the semi-supervised model is still able to use the unlabeled images to extract certain features that can improve performance. Coupled with a way of measuring error from the labeled images, semi-supervised images can be an alternative to other types of learning when our dataset presents obstacles to labeling. A use case of this type of learning would be classifying CT scans for certain types of cancer when it would be too laborious to manually label every image in a dataset. Instead, oncologists could label some of the images, and allow the machine to use the unlabeled images to supplement the labeled images through methods such as feature extraction.

Common Machine Learning Algorithms

Throughout this article, we will review multiple studies that largely use the same few machine learning algorithms. Therefore, we believe it is important to provide a brief summary of a few machine learning algorithms in this section.

Support Vector Machines

A support vector machine aims to create a hyperplane or a set of hyperplanes (a plane that is one dimension smaller than the space the data are in, e.g., a line in 2D space, a 2D plane in 3D space) that are able to separate the data into distinct classes. Support vector machines can use kernel mapping to map data features into higher dimensions in cases which it isn’t possible for a hyperplane to classify the data.

Neural Networks

Neural networks use nodes called “neurons” in a system that aims to mimic the structure of a biological brain. These nodes can take an input, and transform that input based on a series of weights and activation functions to output a new output. This output can then be used for classification and regression purposes. Neural networks often have multiple layers of neurons between the input and output layers. In supervised neural networks, the weights that determine the activation and output of the neurons are automatically adjusted over time based on an error, or cost, function that provides a quantitative value to measure the error of the network.

Random Forest

A random forest algorithm is a decision-tree based algorithm that uses many uncorrelated decision trees who all “vote” on the classification. Random forest algorithms are useful in cases where there are many input features or overfitting is a concern. Due to the reliance of the algorithm on uncorrelated decision trees, the features must be uncorrelated. Correlated features increase the error rate of the algorithm.

Logistic Regression

Logistic regression fits data points onto a logistic function. This mapping results in a value for each input that we can classify into either class based on a threshold value. Logistic regression is especially useful when considering probability as the values outputted by logistic regression can be interpreted as probabilities.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are often used in imagine analyses. It would be very computationally expensive to use a regular fully connected neural network to analyze imagines; a regular, 1080p color image would provide $1920 \times 1080 \times 3$ features to analyze when flattened into a regular vector. Instead, CNNs use an approach that is based on various kernels (also called filters) to detect patterns within an image. Pooling layers are then able to extract important features from the previous layer while shrinking the number of features. Multiple convolutional and pooling layers to identify more complex patterns are often used. Finally, the results of the convolutional and pooling layers can be connected to a fully connected layer to classify an object.

Evaluating Machine Learning Performance

Sensitivity, specificity, and the area under the receiver operating characteristic curve are commonly used to evaluate machine learning performance. The sensitivity of a model, also known as the true positive rate (TPR), describes the percentage of positive cases a model is able to identify. This rate is given by the total number of true positives divided by the sum of total true positives and false negatives (FN).

$$TPR = \frac{TP}{TP + FN}$$

The specificity of a model, also known as the true negative rate (TNR) is given by the total number of true negatives (TN) divided by the sum of true negatives and false positives (FP).

$$TNR = \frac{TN}{TN + FP}$$

These two metrics are inversely related. As we reduce our threshold to make our model more sensitive, we are bound to include more false positives in our more sensitive classification. Conversely, when we increase our threshold, we are bound to miss some borderline cases, leading us to include more false negatives in our less sensitive classification.

One additional metric we use is the AUC (or less ambiguously, the AUROC). For this metric, we use the true positive rate and the false positive rate (FPR).

$$FPR = \frac{FP}{TN + FP}$$

TPR and FPR are positively correlated — as we decrease our threshold and our program gets more sensitive, the number of true positives we find will increase, but so will the number of false positives. An ROC curve graphs the TPR and FPR against each other at all possible thresholds.

The AUC is a measure of the area under the ROC curve. An ideal prediction algorithm (an algorithm that would classify correctly every time) would have an AUC of 1.0, while an algorithm that guesses randomly would have an AUC of 0.5.

Machine Learning in Medicine

The rapid expansion of computing power and the increase in accessibility to data-processing utilities have led to an increase in usage of machine learning for various purposes in the field of medicine. For example, personalized medicine has been a strength of machine learning. The ability of machine learning to train models that can easily be used to provide interpretations of data, assist in prognoses and predict treatment outcomes [2-4]. Here, we review several studies reaffirming the utility and effectiveness possible through machine learning in medicine and then we discuss the use of machine learning algorithms in rare endocrine disorders in detail.

Machine learning is a promising tool in assisting in diagnoses. Its ability to find analyze data, including images, to find complex patterns allows it to accurately draw conclusions from large amounts of data. A recent study [5] utilizing logistic regression and support vector machines, two different methods that incorporate machine learning, was able to achieve a sensitivity/specificity rate of 93.2%/ 98.4% in the diagnosis of transition zone prostate cancer. Another study [6] was able to utilize three different methods of machine learning to identify breast cancer from characteristics of cell nuclei. The study was able to achieve AUC values of 0.97, 0.95, and 0.94 using a support vector machine, artificial neural network, and a generalized linear model respectively. Furthermore, machine learning has shown promise in predicting disease and treatment outcomes. One study [7] aimed to utilize various machine learning algorithms in predicting ischemic stroke outcomes. The machine learning algorithms were trained on 1744 data points using 38 variables such as patient demographics, initial National Institutes of Health Stroke Scale scores, and time from onset to admission. The machine learning algorithms were then compared to a baseline conventional method, the Acute Stroke Registry and Analysis of Lausanne (ASTRAL) score. Using the ASTRAL score, the study achieved an AUC value of 0.839. Although this did not differ significantly from a random forest and logistic regression models (AUC values of 0.857 and 0.849 respectively), a deep neural network was able to achieve an AUC value of 0.888, significantly higher than the conventional ASTRAL score. Notably, when the machine learning algorithms were provided solely with the criteria the ASTRAL system used, the performance of all of the machine learning algorithms were comparable with the ASTRAL score, highlighting the capability of deep learning models to find complex relationships between features.

Machine Learning in Rare Endocrine Diseases

Recently, machine learning has been shown to be very promising in the area of rare disease research. Common problems such as underdiagnosis and misdiagnosis due to obscurity, identification, and diagnosis are issues that the large data-processing capabilities of machine learning are tailored to solve.

Identification of Patients with Rare Diseases

The difficulty of identifying patients with rare diseases continues to be a large obstacle in rare disease research. According to a survey in 2013, the average patient with a rare disease visits 8 doctors and receives 2 to 3 misdiagnoses over the course of 5-8 years before being correctly diagnosed [8]. The difficulty in diagnosing rare diseases, in combination with their low prevalence rates, makes them challenging to study. Even the most experienced of clinicians may not have encountered a given rare disease in their careers. It is simply not possible for any single clinician to be able to easily diagnose of a majority of rare diseases. Therefore, when a patient presents with a rare disease, it is possible, if not likely, that a clinician will not consider the rare disease until they have exhausted more familiar diseases and conditions, even if the information available strongly aligns with the characteristics of the rare disease at hand [9]. Machine learning, however, has the capacity to avoid this problem. Once models to identify rare diseases have been built, computers can use multiple models to predict whether the patient is likely to have any of the tested diseases at once without obscurity being a factor in how strongly each condition is considered. Once a list of probable conditions has been built, the clinician can then use their expertise to pinpoint a condition and begin an appropriate treatment. Therefore, models that can identify and diagnose diseases can provide a substantial improvement in patient care.

The following example give various examples of machine learning in this use case. Hypophosphatasia, a rare disease caused by mutations in the ALPL gene that encodes tissue nonspecific alkaline phosphatase causing loss-of-function in the enzyme is a suitable candidate for a relatively simple machine learning model. A study [10] utilized the biomarkers alkaline phosphatase (ALP) and pyridoxal 5′-phosphate (PLP) to construct two machine learning models to identify the disease. Using a support vector machine that utilized both of the biomarkers, the study was able to achieve an AUC value of 0.936 in identifying hypophosphatasia.

Additionally, improvements in analyzing “big data,” datasets too large to be analyzed by traditional analysis methods, mean that machine learning models can take advantage of feature sets so large, manual analysis could not feasibly find the same complex relationships between variables machine learning models could. One promising use of this capability of machine learning in rare disease research is the analysis of health records. Electronic health records of patients with rare diseases may uncover previously unknown relationships between the disease and various health parameters. Additionally, these relationships could then be used to identify other patients that may have the same rare disease. Underdiagnosis is a large obstacle in rare disease research and treatment, and the powerful data analysis capabilities of machine learning could provide a solution. Indeed, machine learning has previously been successfully used to predict different medical events such as in-hospital mortality and discharge diagnoses and to analyze patient outcomes from electronic health records [11, 12].

In the context of endocrine diseases, previous studies also show promise. Although the endocrine diseases they focused on were not rare, the following studies show successful results in identifying endocrine diseases through health records. One study [13] that focused on diagnosis through electronic health records aimed to identify type-2 diabetes patients using multiple methods of machine learning. The electronic health records contained information such as previous diagnoses, medications taken, and lab reports. The models were tested using 3 different feature sets. All machine learning models were able to achieve AUC values ≥ 0.88 on all feature sets, with a Naïve-Bayes based model achieving AUC values of 0.98, 1.00, and 1.00 on the 3 feature sets. In comparison, a state-of-the-art algorithm designed by an expert in the field had an AUC value of 0.71. It is important to note that the expert-algorithm had a specificity of 1.00, while the Naïve-Bayes classifier had a specificity of 0.93, 0.88, and 0.94. This reaffirms the role of machine learning in medicine as a tool rather than a decision maker. Additionally, another study [14] from 2009 utilized a support vector machine to assess type-1 diabetes risk in individuals using genome-wide association studies (GWAS). The study was conducted by training an SVM using singe nucleotide polymorphisms selected with various thresholds on a number of GWAS datasets. Although the precise AUC values varied slightly depending on which dataset the model was trained and tested on, the SVM was able to consistently achieve AUC values ≥ 0.8.

We believe that the results of aforementioned studies and ongoing efforts may help remedy certain difficulties in the identification of rare diseases in individuals. In cases where the lack of

identification is due to the obscurity of a rare disease rather than a lack of data, we believe that scanning of electronic health records and analyzing genome data using machine learning algorithms may help to locate and treat patients with rare diseases.

Aiding in Diagnosis

Machine learning has shown great strength in aiding diagnoses. In the context of rare endocrine diseases where obscurity presents a great challenge, the capability of computers to easily recall large amounts of data makes them uniquely suited to diagnose rare endocrine diseases. Indeed, using computerized decision support systems have previously shown success in assisting to diagnose rare diseases [15]. Additionally, ability of machine learning to easily use a large number of features to estimate outcomes has led to it often outperforming conventional tests when offered with a feature set such that the features that a conventional test would use are a subset of the machine learning algorithm's feature set.

Wallace et al. [16] utilized machine learning in diagnosing phaeochromocytomas and paragangliomas. The study utilized linear discriminant analysis and metabolic profiling on 186 samples to identify paragangliomas, achieving a sensitivity/specificity ratio of 93.2%/99.2% with an AUC of 0.982. In comparison, evaluating the succinate/fumarate ratio (SFR) using conventional means yielded a sensitivity/specificity ratio of 88.1%/99.2% with an AUC of 0.96. However, a second model that included ten instead of four metabolites and used formalin-fixed and/or paraffin embedded tissue in addition to freshly frozen tissue yielded slightly different results with a sensitivity/specificity ratio of with an AUC of. The study noted that the discrepancy may be explained by metabolite levels differing due to stromal contaminants in the tissue used. In the context of rare diseases, standardization of data is a relevant discussion, as data are often collected from various sources. The lack of standardization that may result from such circumstances may pose an obstacle in training machine learning models on rare diseases. Another study [17] used machine learning to identify long intergenic noncoding RNAs (lincRNAs) that may serve as biomarkers in these rare neuroendocrine tumors.

Machine learning has also proven to be useful in differentiating rare diseases. One study [18] attempted to differentiate adrenocortical carcinomas and adrenocortical adenomas. The study used steroid excretion analysis by mass spectrometry to obtain measurements on certain steroid markers. The study then used generalized matrix learning vector quantization and

trained the model using 32 steroid markers as features. The model was able to achieve an AUC value of 0.965 in differentiating the two conditions.

Aiding in Prognosis and Risk Prediction

A future in which patients are able to easily check their risk status for different diseases would allow them to proactively make healthier decisions, seek earlier treatment, and be more observant of symptoms for diseases they are at risk for. Recently, biobanks, repositories that store health information that can be used to retroactively analyze health histories and outcomes, have seen use in estimating underdiagnosis and risk prediction [19, 20]. Through its capability to establish complex relationships in large datasets and various biobanks, machine learning presents itself as a suitable tool to establish patient risk for certain diseases and to aid in prognoses of diseases.

One study [21] utilized the UK Biobank to gather data on individuals with no known previous cardiovascular diseases (CVD) at baseline. The study then developed machine learning models, some using a model design algorithm called “AutoPrognosis,” which were then trained on the database to identify whether any of the individuals developed CVD in any of the follow-ups. The models were then compared to two conventional CVD risk assessment methods: the Framingham Score and Cox proportional hazards (PH) models based on a number of different variables. The Framingham score, used as a baseline, achieved an AUC value of 0.724, while the PH model achieved an AUC value of 0.758 using all variables. In contrast, all machine learning models except an SVM based approach outperformed the Framingham score. Additionally, the AdaBoost, gradient boosting, and AutoPrognosis created model approaches all outperformed the PH score, with the AutoPrognosis created model performing best with an AUC value of 0.774 when utilizing all variables.

Machine Learning in Medical Imaging

A future in which clinicians could get second opinions on various imaging procedures at the click of a button, or a future in which they could analyze images for diseases they may have never heard of would serve to reduce the effects that misdiagnoses and underdiagnoses of rare disease cause. Recent machine learning research and the exponential increase in graphics processing power have made convolutional neural networks (CNNs) a very strong candidate for this task. Convolutional neural networks work by identifying patterns in images that are related to the outcome. This ability of CNNs makes them suited to analyze medical images to aid in

identification and diagnosis of rare diseases. Imaging based machine learning models have already been extensively used in various areas such as cardiology, dermatology, and oncology [22-24].

Although examples in conventional diseases achieve very impressive results, when considering CNN based methods in the context of rare endocrine diseases, we encounter a problem. For example, although Esteva, et al. [25] was able to achieve excellent results in classifying different types of skin cancer through clinical imaging, the study used 129,450 images to train its machine learning algorithm. CNNs often require large datasets of images to train on, which presents an obstacle for their use in rare disease research. However, there are various methods to adapt CNNs to overcome data scarcity issues that may be present in rare disease research.

One method is referred to as “data augmentation”. In this method, the training dataset is augmented through image manipulation such as rotating, cropping, or changing the zoom levels of images. One recent study [26] utilized data augmentation techniques to train a CNN to detect tuberous sclerosis complex (TSC) on 138 images (69 TSC 69 control). When tested on an unaugmented test dataset (50 images 25 TSC 25 control) the CNN had not been trained with, the CNN was able to achieve a sensitivity/specificity rate of 95%/95% with an AUC of 0.99.

Another machine learning method that can be used to adapt CNNs for rare disease research is transfer learning. Transfer learning is a method in which machine learning models that have been previously trained on other sets of data are adapted into a new model for the task at hand. A study [27] in 2016 attempted to use transfer learning to develop a model to classify lung tissue patterns from CT scans into different interstitial lung diseases. Through transferring layers from a neural network trained on a texture recognition dataset, the study was able to achieve an absolute performance increase of 2%. Similarly, another study [28] that aimed to classify different inherited retinal diseases through fundus autofluorescence imaging (FAF) trained and validated a CNN using 389 FAF images. The study used transfer learning to build a machine learning model that was first trained on the ImageNet dataset, and then fine-tuned on the 389 FAF images to classify them into different groups. The CNN was then tested on a test set of 94 FAF images with an AUC ≥ 0.989 for all tested diseases.

Recent algorithms have also shown promise in learning from fewer images and data points. A recent study [29] attempted to classify retinal diseases from the EyePACS database with the

aim of observing performance differences between various machine learning algorithms as the number of samples used to train the models changed. As a baseline, the study used a traditional fine-tuned ResNet algorithm (RES-FT). The study then compared the baseline to various algorithms that utilized a network model called Augmented Multiscale Deep InfoMax, which is based on the concept of self-supervision. Using 5120 samples, all of the algorithms scored similarly, with an AUC value of 0.8330 for the baseline and 0.8348 for the Deep InfoMax (DIM) algorithm. However, at 160 samples, the baseline saw a dramatic drop to an AUC value of 0.6585, while the DIM algorithm was able to achieve an AUC value of 0.7467. This trend continued for sample sizes of 40 and 10, with the RES-FT algorithm achieving AUC values of 0.5671 and 0.5178 respectively, while the DIM algorithm achieved AUC values of 0.6760 and 0.5778 respectively.

These studies illustrate multiple methods that can be used to overcome obstacles that may occur when using machine learning models on smaller amounts of data. Using these methods, we believe that CNNs continue to show promise in identifying rare endocrine diseases through medical imaging.

Lipodystrophy – A Promising Application of Machine Learning in Rare Diseases

Lipodystrophy (LD) syndromes are characterized by the lack of adipose tissue in the body, causing complications such as insulin resistance, ectopic steatosis, and hyperlipidemia [30]. Lipodystrophy syndromes are considered rare diseases. The combination of a low prevalence rate and the difficulty in diagnosing rare diseases such as lipodystrophy makes it exceedingly difficult to conduct studies on this disease, as patients are scarce and often undiagnosed.

As an example, Gonzaga-Jauregui, et al. estimates the prevalence of LD among the population to be near 47.3 cases per 1,000,000 people [30]. In contrast, Chiquette et al. [31] estimates the prevalence of LD among the general population to be 3.07 cases per 1,000,000 people. The variance between these two figures is most likely due to the criteria used when determining whether a patient could be considered a case of LD. Chiquette, et al. used database-dependent inclusion-exclusion criteria to identify cases of LD, while Gonzaga-Jauregui, et al. observed the expected comorbidities associated with LD to achieve the same task.

The significance of the selected criteria when attempting to identify lipodystrophy from health records is clear from the preceding studies. To obtain a more accurate prevalence value of LD

in the general population, we propose a machine learning model similar to Zheng, et al. The study was able to accurately identify type-2 diabetes through electronic health records (EHRs) that included previous medical conditions, and we believe that a similar approach may yield promising results in identifying LD. Indeed, it is well known that lipodystrophy is associated with a multitude of notable comorbidities and signs such as insulin resistance and dyslipidemia that healthcare providers may not immediately connect to LD, but that will nevertheless be included in EHRs. Furthermore, as LD is known to be an underdiagnosed disease, a successful machine learning algorithm to identify cases through EHRs will have further implications in research and treatment of the disease.

Conclusions

Current literature on machine learning in rare endocrine disorders suggests that machine learning may hold great promise in overcoming major issues in rare disease research. We believe that the diagnostic and identificatory capabilities of machine learning can combat the misdiagnoses and underdiagnoses that present obstacles for patients seeking care and researchers who wish to study rare endocrine disorders. Additionally, the prognostic capabilities of machine learning can help clinicians better understand the course rare endocrine disorders will take and decide on the best treatment for a specific case. Finally, we believe that through methods of remedying problems caused by data scarcity, machine learning will continue to play an increasing role in all aspects of rare endocrine disease research.

References

1. El Naqa I, Murphy MJ: **What Is Machine Learning?** In: *Machine Learning in Radiation Oncology: Theory and Applications.* edn. Edited by El Naqa I, Li R, Murphy MJ. Cham: Springer International Publishing; 2015: 3-11.
2. Dilsizian SE, Siegel EL: **Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment**. *Curr Cardiol Rep* 2014, **16**(1):441.
3. Jiang R, Abbott CC, Jiang T, Du Y, Espinoza R, Narr KL, Wade B, Yu Q, Song M, Lin D *et al*: **SMRI Biomarkers Predict Electroconvulsive Treatment Outcomes: Accuracy with Independent Data Sets**. *Neuropsychopharmacology* 2018, **43**(5):1078-1087.
4. Montazeri M, Montazeri M, Montazeri M, Beigzadeh A: **Machine learning models in breast cancer survival prediction**. *Technol Health Care* 2016, **24**(1):31-42.
5. Wu M, Krishna S, Thornhill RE, Flood TA, McInnes MDF, Schieda N: **Transition zone prostate cancer: Logistic regression and machine-learning models of quantitative ADC,**

shape and texture features are highly accurate for diagnosis. *J Magn Reson Imaging* 2019, **50**(3):940-950.
6. Sidey-Gibbons JAM, Sidey-Gibbons CJ: **Machine learning in medicine: a practical introduction**. *BMC Med Res Methodol* 2019, **19**(1):64.
7. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH: **Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke**. *Stroke* 2019, **50**(5):1263-1265.
8. **Rare Disease Impact Report** [https://globalgenes.org/wp-content/uploads/2013/04/ShireReport-1.pdf]
9. Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE: **Cognitive and system factors contributing to diagnostic errors in radiology**. *AJR Am J Roentgenol* 2013, **201**(3):611-617.
10. Garcia-Carretero R, Olid-Velilla M, Perez-Torrella D, Torres-Pacho N, Darnaude-Ortiz MT, Bustamate-Zuloeta AD, Tenorio JA: **Predictive modeling of hypophosphatasia based on a case series of adult patients with persistent hypophosphatasemia**. *Osteoporos Int* 2021.
11. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, Liu PJ, Liu X, Marcus J, Sun M *et al*: **Scalable and accurate deep learning with electronic health records**. *NPJ Digit Med* 2018, **1**:18.
12. Miotto R, Li L, Kidd BA, Dudley JT: **Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records**. *Sci Rep* 2016, **6**:26094.
13. Zheng T, Xie W, Xu L, He X, Zhang Y, You M, Yang G, Chen Y: **A machine learning-based framework to identify type 2 diabetes through electronic health records**. *Int J Med Inform* 2017, **97**:120-127.
14. Wei Z, Wang K, Qu HQ, Zhang H, Bradfield J, Kim C, Frackleton E, Hou C, Glessner JT, Chiavacci R *et al*: **From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes**. *PLoS Genet* 2009, **5**(10):e1000678.
15. Ronicke S, Hirsch MC, Turk E, Larionov K, Tientcheu D, Wagner AD: **Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study**. *Orphanet J Rare Dis* 2019, **14**(1):69.
16. Wallace PW, Conrad C, Bruckmann S, Pang Y, Caleiras E, Murakami M, Korpershoek E, Zhuang Z, Rapizzi E, Kroiss M *et al*: **Metabolomics, machine learning and immunohistochemistry to predict succinate dehydrogenase mutational status in phaeochromocytomas and paragangliomas**. *J Pathol* 2020, **251**(4):378-387.
17. Ghosal S, Das S, Pang Y, Gonzales MK, Huynh TT, Yang Y, Taieb D, Crona J, Shankavaram UT, Pacak K: **Long intergenic noncoding RNA profiles of pheochromocytoma and paraganglioma: A novel prognostic biomarker**. *Int J Cancer* 2020, **146**(8):2326-2335.
18. Arlt W, Biehl M, Taylor AE, Hahner S, Libe R, Hughes BA, Schneider P, Smith DJ, Stiekema H, Krone N *et al*: **Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors**. *J Clin Endocrinol Metab* 2011, **96**(12):3775-3784.
19. Dickey AK, Quick C, Ducamp S, Zhu Z, Feng YA, Naik H, Balwani M, Anderson KE, Lin X, Phillips JE *et al*: **Evidence in the UK Biobank for the underdiagnosis of erythropoietic protoporphyria**. *Genet Med* 2021, **23**(1):140-148.

20. Muller DC, Johansson M, Brennan P: **Lung Cancer Risk Prediction Model Incorporating Lung Function: Development and Validation in the UK Biobank Prospective Cohort Study**. *J Clin Oncol* 2017, **35**(8):861-869.
21. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M: **Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants**. *PLoS One* 2019, **14**(5):e0213653.
22. Martin-Isla C, Campello VM, Izquierdo C, Raisi-Estabragh Z, Baessler B, Petersen SE, Lekadir K: **Image-Based Cardiac Diagnosis With Machine Learning: A Review**. *Front Cardiovasc Med* 2020, **7**:1.
23. Thomsen K, Iversen L, Titlestad TL, Winther O: **Systematic review of machine learning for diagnosis and prognosis in dermatology**. *J Dermatolog Treat* 2020, **31**(5):496-510.
24. Yassin NIR, Omran S, El Houby EMF, Allam H: **Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review**. *Comput Methods Programs Biomed* 2018, **156**:25-45.
25. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S: **Dermatologist-level classification of skin cancer with deep neural networks**. *Nature* 2017, **542**(7639):115-118.
26. Sanchez Fernandez I, Yang E, Calvachi P, Amengual-Gual M, Wu JY, Krueger D, Northrup H, Bebin ME, Sahin M, Yu KH *et al*: **Deep learning in rare disease. Detection of tubers in tuberous sclerosis complex**. *PLoS One* 2020, **15**(4):e0232376.
27. Christodoulidis S, Anthimopoulos M, Ebner L, Christe A, Mougiakakou S: **Multisource Transfer Learning With Convolutional Neural Networks for Lung Pattern Analysis**. *IEEE J Biomed Health Inform* 2017, **21**(1):76-84.
28. Miere A, Le Meur T, Bitton K, Pallone C, Semoun O, Capuano V, Colantuono D, Taibouni K, Chenoune Y, Astroz P *et al*: **Deep Learning-Based Classification of Inherited Retinal Diseases Using Fundus Autofluorescence**. *J Clin Med* 2020, **9**(10).
29. Burlina P, Paul W, Mathew P, Joshi N, Pacheco KD, Bressler NM: **Low-Shot Deep Learning of Diabetic Retinopathy With Potential Applications to Address Artificial Intelligence Bias in Retinal Diagnostics and Rare Ophthalmic Diseases**. *JAMA Ophthalmol* 2020, **138**(10):1070-1077.
30. Gonzaga-Jauregui C, Ge W, Staples J, Van Hout C, Yadav A, Colonie R, Leader JB, Kirchner HL, Murray MF, Reid JG *et al*: **Clinical and Molecular Prevalence of Lipodystrophy in an Unascertained Large Clinical Care Cohort**. *Diabetes* 2020, **69**(2):249-258.
31. Chiquette E, Oral EA, Garg A, Araujo-Vilar D, Dhankhar P: **Estimating the prevalence of generalized and partial lipodystrophy: findings and challenges**. *Diabetes Metab Syndr Obes* 2017, **10**:375-383.

Tables

Table 1. A summary of commonly used types of machine learning algorithms

Type of Machine Learning	Description	Uses
Supervised Learning	To train a supervised machine learning algorithm, we provide the computer with a method of measuring the error of its attempt. This means that for every input, we should know what output we should get. The computer is then able to analyze its mistakes and adjust its solution in order to achieve greater accuracy through methods such as back-propagation.	These kinds of algorithms are useful in cases where we have easy access to labeled data. Some examples may include training an algorithm to predict treatment outcomes when we already have access to a dataset that includes past patient outcomes, or an algorithm to diagnose a disease when we have access to health records that already differentiate between those who have the disease and those who do not.
Unsupervised Learning	An unsupervised machine learning algorithm trains on datasets where we do not know the labels and do not have a way of evaluating the error of a given attempt. Because we do not have a way of evaluating whether are predictions are "correct," unsupervised machine learning aims to find patterns and structure in data.	An example task would be training a model to classify images of tennis balls and oranges without labelling the images. Although the computer does not know which images are apples and which are oranges, it may classify pictures into general categories of "orange things" and "yellow things."
Semi-Supervised Learning	A semi-supervised machine learning model is a mixture of supervised and unsupervised learning concepts. To train a semi-supervised machine learning model, we use labelled and unlabeled sets of data. Although the performance may not be as impressive as a supervised model training on a fully labeled set of data, the semi-supervised model is still able to use the unlabeled images to extract certain features that can improve performance. Thanks to these properties, semi-supervised machine learning can be an alternative to other types of learning when our dataset presents obstacles to labeling.	A use case of this type of learning would be classifying CT scans for certain types of cancer when it would be too laborious to manually label every image in a dataset. Instead, oncologists could label some of the images, and allow the machine to use the unlabeled images to supplement the labeled images through methods such as feature extraction.

Figures

Figure 1. Support vector machines.

Figure 1A shows a randomly scattered array of points that have been classified into two groups based on their positions (2x < y). Figure 1B shows the hyperplane a support vector machine has generated to separate the two groups.

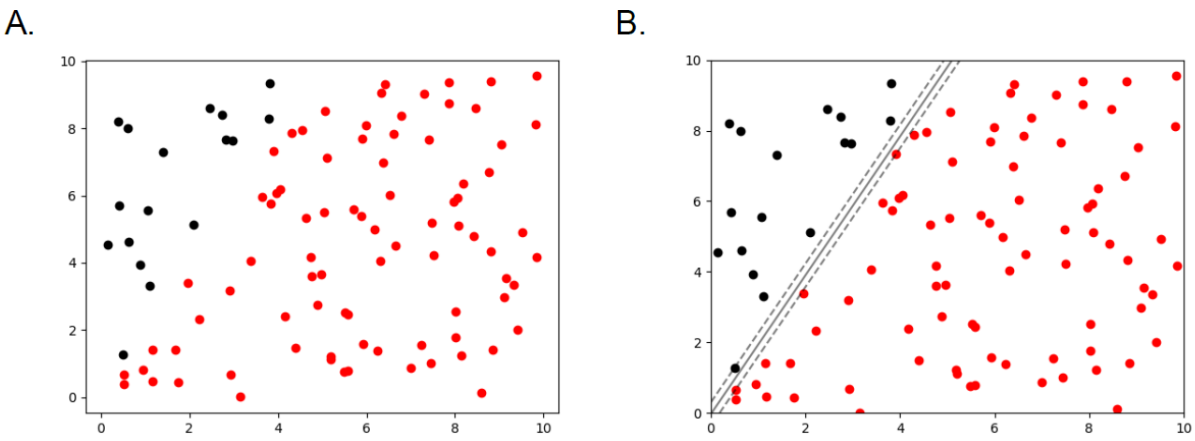
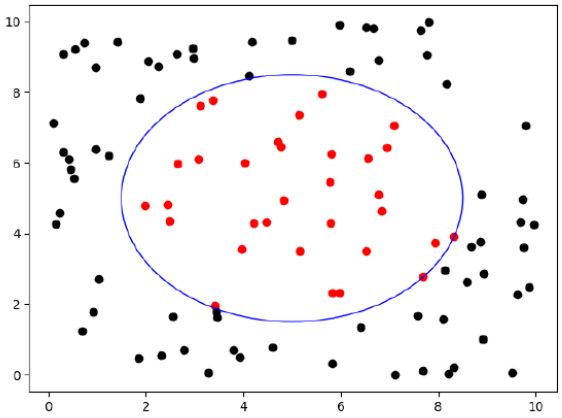


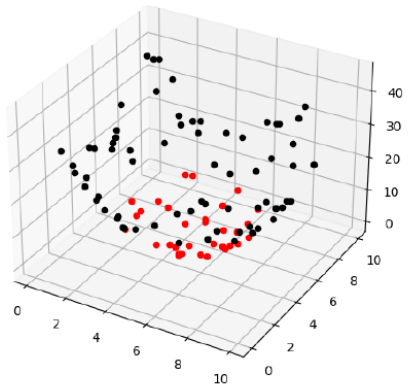
Figure 2. Kernel method.

Figure 2-A shows a randomly scattered array of points that have been classified into two groups based on whether they are more than 3.5 units away from the center of the plot (5, 5). Figure 2-B shows a kernel mapping so that a hyperplane (Z=12.25) could linearly separate the two groups.

A.



B.



Declarations

Ethics approval and consent to participate

No formal Institutional Review Board (IRB) approval is required as the manuscript reviews published literature.

Consent for publication

NA.

Availability of data and materials

The datasets analyzed during this study are included in this published article.

Competing interests

EYA has nothing to disclose. EAO reports the following conflicts: Grant support: Aegerion Pharmaceuticals (now Amryt Pharmaceuticals), Ionis Pharmaceuticals, Akcea Therapeutics, Gemphire Therapeutics, GI Dynamics (current), AstraZeneca (2015-2017). Consultant or Advisor: AstraZeneca, Thera Therapeutics, and BMS (past), Aegerion Pharmaceuticals (now Amryt Pharmaceuticals), Akcea Therapeutics, Ionis Pharmaceuticals, Regeneron Pharmaceuticals (current). Drug support: Aegerion Pharmaceuticals (now Amryt Pharmaceuticals), Akcea Therapeutics, Rhythm Pharmaceuticals (all current). Other support: Aegerion Pharmaceuticals (now Amryt Pharmaceuticals), Regeneron Pharmaceuticals (current).

Funding

The work was partially supported by the Lipodystrophy Fund at University of Michigan.

Authors' contributions

EYA gathered the data and wrote the manuscript. EAO reviewed and edited the manuscript. EAO is responsible for the integrity of the data and the conduct of the study. Both authors read and provided critical input on different versions of the manuscript and approved the final version.

Acknowledgements

None.