

Application of machine learning in understanding atherosclerosis: Emerging insights

Cite as: APL Bioeng. 5, 011505 (2021); doi: 10.1063/5.0028986

Submitted: 8 September 2020 · Accepted: 21 January 2021 ·

Published Online: 16 February 2021



View Online



Export Citation



CrossMark

Eric Munger,^{1,2,3}  John W. Hickey,⁴  Amit K. Dey,¹  Mohsin Saleet Jafri,²  Jason M. Kinser,² 
and Nehal N. Mehta^{1,a)} 

AFFILIATIONS

¹National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

²George Mason University, Fairfax, Virginia 22030, USA

³Johns Hopkins University, Baltimore, Maryland 21208, USA

⁴Stanford University, Stanford, California 94306, USA

^{a)} Author to whom correspondence should be addressed: nehal.mehta@nih.gov. Tel.: 1-301-827-0483. Fax: 1-301-827-0915

ABSTRACT

Biological processes are incredibly complex—integrating molecular signaling networks involved in multicellular communication and function, thus maintaining homeostasis. Dysfunction of these processes can result in the disruption of homeostasis, leading to the development of several disease processes including atherosclerosis. We have significantly advanced our understanding of bioprocesses in atherosclerosis, and in doing so, we are beginning to appreciate the complexities, intricacies, and heterogeneity atherosclerosis. We are also now better equipped to acquire, store, and process the vast amount of biological data needed to shed light on the biological circuitry involved. Such data can be analyzed within machine learning frameworks to better tease out such complex relationships. Indeed, there has been an increasing number of studies applying machine learning methods for patient risk stratification based on comorbidities, multi-modality image processing, and biomarker discovery pertaining to atherosclerotic plaque formation. Here, we focus on current applications of machine learning to provide insight into atherosclerotic plaque formation and better understand atherosclerotic plaque progression in patients with cardiovascular disease.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0028986>

INTRODUCTION

Cardiovascular disease (CVD) including heart attack and stroke is the leading cause of death worldwide and is usually preceded by accelerated atherosclerosis.¹ Atherosclerotic plaque formation, development, and progression are complex and involve many factors including elevated cholesterol, heightened immune activity, smooth muscle cell proliferation, and endothelial dysfunction.² As plaques in the vessel wall build and progress and become high-risk, they may rupture and thus the associated arteries may be occluded distally, leading to impaired flow of oxygen-rich blood to the heart muscle, brain, and other parts of the body.

With such an impact on our health, substantial resources have been allocated to study atherosclerosis in large cohorts across multiple clinical, cellular, and molecular modalities. Consequently, there are a growing number of clinical and experimental datasets available that include multiple variables and risk factors known to be associated with

atherosclerosis as well as many additional variables whose impact on disease is not well established.³ Ideally, the analysis of such datasets should include the development of both statistical and machine learning (ML) models that have minimal statistical bias and accurately account for important prognostic factors. Together, these models can be used to accurately quantify inferences within the data and forecast condition states useful for advanced treatment decisions. In fact, in a recent study designed to assess the benefits of artery-stenting in patients with atherosclerotic renal artery stenosis, the rigorous statistical analysis concluded that, when focused on a composite end point, there was no significant distinction among the treatment regimens.⁴ However, in a follow-up study designed to compare ML methods using the same data, 4 of 5 methods tested showed a small but distinct ability to identify the patients that achieved the same end point.⁵ This critical review aims to highlight the successful analysis of well-characterized datasets through ML methods that are inherently

capable of minimizing statistical bias with a focus on atherosclerotic plaque. First, we briefly discuss the complex process of atherogenesis. Then, we show how ML has been utilized in atherosclerosis and CVD research and then focus on a recent study exploring the role of ML in risk stratifying psoriasis patients at high risk of developing coronary artery disease.⁶ Finally, we include a discussion of potential biological insights and the future impact and direction in the application of ML for novel biomarker discovery in atherosclerosis.

Biology of atherosclerosis

Atherosclerosis is one of the leading causes of mortality and morbidity in developed countries and is widely recognized as a chronic inflammatory disease in addition to being a lipid-associated disease.² It is a complex disease involving many different immune cell types, cytokines, chemokines, hemodynamics, and biomechanics of the vessel walls. Atherosclerosis leads to the formation of plaques that thicken large and medium-sized arterial walls, narrowing the lumen, which obstructs blood flow.^{2,7} Foam cells within plaques activate platelets and endothelium cells, which induces the migration and accumulation of vascular smooth muscle cells (VSMCs) from the media to the intima (layers of the vessel wall). VSMCs proliferate within the intima and secrete extracellular matrix macromolecules.⁸ Subsequently, VSMCs and leukocytes are activated, which in turn increases inflammatory cell infiltration in the atherosclerotic lesion. This process is accompanied by the eventual apoptosis of VSMCs.⁸ The combination of greater inflammatory cell infiltration and apoptotic death together leading to hypoxia-induced necrosis promotes the development of the plaque into the late fibroatheromatous lesion with eventual neovascularization.⁹ These nascent immature micro-blood vessels are inherently leaky and permit extravasation of erythrocytes into the plaque, further contributing to necrotic core enlargement.⁹ The many accumulating cells, including macrophages and activated endothelial cells, promote pro-inflammatory cytokines, matrix metalloproteinases (MMPs), and cathepsins.¹⁰ This, along with Interferon-alpha inhibition of collagen formation by VSMC, weakens the already fragile fibrous cap, increasing the likelihood of rupture.¹¹ This allows simple mechanical agitation from blood flow to cause the atherosclerotic cap covering the inflammation site to rupture, resulting in possible blood coagulation and thrombosis, which leads to decreased blood flow distally and thus a vascular event.

Atherosclerotic plaques may stabilize or reduce in size, lipid content, foam cell content, and macrophage inflammation through a process known as atherosclerosis regression.¹² This process is driven primarily by the Macrophage Reverse Cholesterol Transport (RCT) mechanism, allowing for atherosclerotic plaques to eliminate cholesterol. Enhancing foam cell cholesterol efflux by high-density lipoprotein (HDL) particles is considered a primary activator of the RCT mechanism, which is believed to offer atheroprotection.¹³

Machine learning and atherosclerosis

Levels of low-density lipoprotein (LDL), apolipoprotein B (ApoB), HDL, apolipoprotein A1 (ApoA1), and C-reactive protein (CRP) are significant biomarkers of vascular inflammation and are critical to almost all inflammatory disease processes including atherosclerosis.¹⁴ Yet inefficiencies in biomarker identification, prioritization, and verification complicate the discovery of new biomarkers for

practically all diseases.¹⁵ Traditional approaches attempt to analyze the correlation between a single variable and disease, thus having a narrow approach. This siloed approach has been exhausted in the past, partially to distinguish biological mechanisms and because only a single disease-related biomarker could be measured in plasma relatively easily. However, with an increase in biological metrics and analytes being sampled, bigger and better-curated datasets across not only patients but across scales are becoming available. These data typically include patient history, tissue, and physiological readouts, cell subsets and percentages, and molecular information such as the metabolism of these cells. Machine learning modeling and analysis of such datasets can provide accurate and clinically practical tools for diagnostics and may prove to be very effective for the discovery of dynamic proteomic and metabolic biomarkers in the field of translational medicine (Fig. 1—machine learning modeling can easily convert raw data into deployable models).

Multiple articles have focused on ML for the role of image processing. For example, Alberto *et al.* reported a comparison of the many methods for characterizing plaque components, plaque morphology, and arterial wall measurements using optical coherence tomography (OCT).¹⁶ Nils *et al.* also discussed the application of ML in coronary artery disease using cardiac computed tomography (CT) images for the detection and characterization of atherosclerotic plaque.¹⁷ An important point is that data extracted from the analysis of images as well as the image itself can be utilized by ML to discover novel relationships.

Recent applications of machine learning analysis in atherosclerosis research

Beyond clinical risk assessment, the scientific community has recognized the benefit of mechanistic discovery and prioritizing biomarkers for diseases. As a result, there has been significant work in interpreting ML approaches for ranking variables focused on finding ways to replace the relevance scores with measures that can be interpreted using standard methods. To this end, a prior study concluded that ML methods, while computationally expensive, are capable of identifying truly relevant biomarkers within datasets with multivariate interacting variables,¹⁸ helping clear the path for ML analysis of healthcare data (Table 1).

Machine learning analysis for risk stratification

Studies have shown that ML methods may be better for the prediction of cardiovascular events than traditional risk assessment scales commonly used in clinical practice.^{23,26,27} In 2017, researchers studied the ability of ML techniques to classify six cardiovascular outcomes and compared these results with standard cardiovascular risk scores.¹⁹ Using the MESA (Multi-Ethnic Study of Atherosclerosis) dataset that included 735 variables from imaging and noninvasive tests, questionnaires, and biomarker panels, they applied a random survival forest method to rank the top-20 predictors of each outcome.¹⁹ Interestingly, imaging, electrocardiography, and serum biomarkers were shown to be more represented in the top-20 important variables selected when compared to traditional cardiovascular risk factors in predicting cardiovascular outcomes.¹⁹

More recently, ML approaches are being applied to develop predictive models to classify individuals who are likely to be diagnosed with diseases as a result of atherosclerosis.²⁴ Terrada *et al.* introduce a

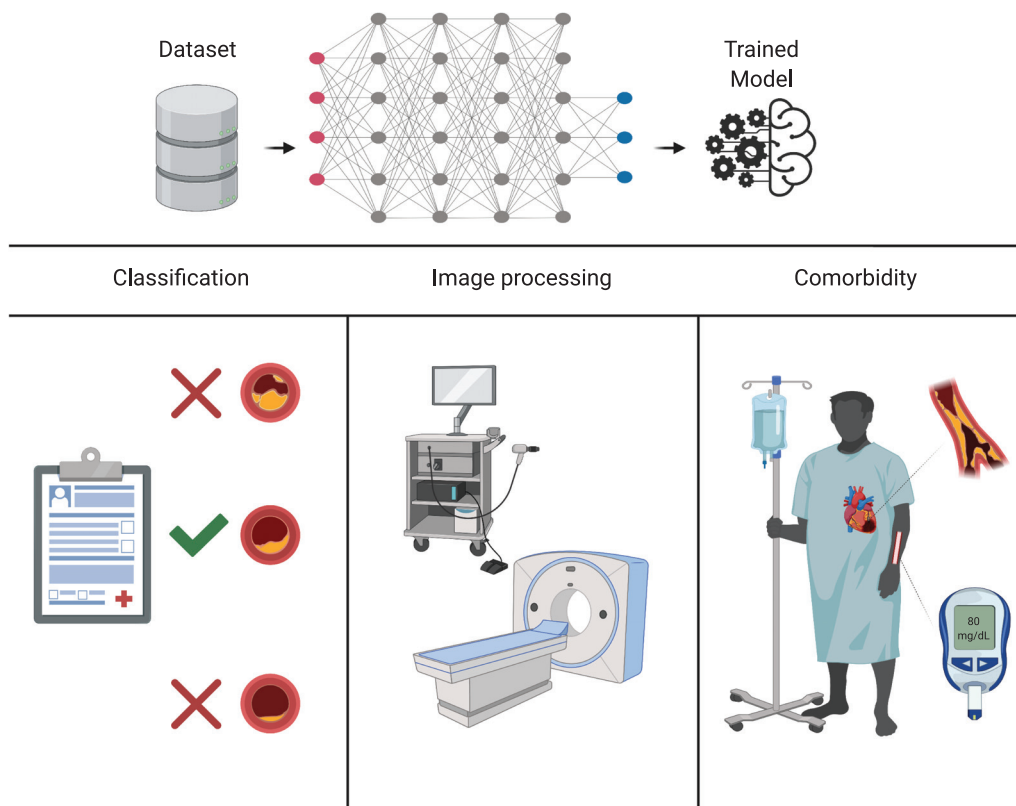


FIG. 1. Machine learning modeling converts raw data into efficient and accurately trained models that can be used for patient risk analysis and medical image processing even when patients exhibit complex comorbidities. Created with BioRender.com.

Medical Diagnosis Support Systems (MDSS) after having achieved an accuracy of 98%.²⁴ In development, the MDSS was assessed using seven different classification algorithms: Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), Classification Ensemble (CE), and Discriminant Analysis (DA).²⁴ The MDSS used data from 835 patient medical records who suffer from atherosclerosis, usually caused by coronary artery diseases (CAD), collected from three databases: the Cleveland Heart Disease, Hungarian, and Z-Alizadeh Sani databases.²⁴ The system's input was a combination of variables included in the final dataset and resulted in an accuracy rating of 98%, far more accurate than other approaches.²⁴

While such models represent a significant step forward in large-scale computational clinical diagnostics for atherosclerotic diseases, the results highlight the potential of ML as an approach to clinical analysis as well as some areas where more work is needed. A system such as the MDSS need not be primarily focused on any one disease. Indeed, systems such as this can be developed to efficiently classify and diagnose individuals for many different diseases. The biggest challenge in the development of such systems is proper feature selection.²⁴ While Terrada *et al.* did achieve a significant accuracy, the widespread deployment of the system across other diseases is limited by the use of domain knowledge feature

selection from datasets and feature engineering. The system generated the appropriate features based on external feature engineering because the authors had yet to develop a technique to automate feature selection.²⁴

Machine learning analysis for the characterization of atherosclerotic plaque formation

Imaging methods have significantly advanced in recent decades, allowing for early detection and characterization of atherosclerosis.²⁸ Some studies have used ML for the analysis of imaging data to characterize plaque morphology. A more recent study compared the image segmentation accuracy of three popular ML algorithms, convolutional neural networks (CNNs), random forest (RF), and support vector machine (SVM).²⁹ Based on each algorithm's ability to characterize the very thin second layer of coronary artery (media), the CNN was shown to be very effective when applied as a feature extractor and RF proved to be a very accurate classifier. The system resulted in an overall classification rate up to 96% when applied for the recognition and specification of the media intima from other tissues.²⁹ Of note is the comparison of both traditional ML methods and deep learning. As one of the first applications of a deep learning CNN to coronary artery image segmentation, this offered the opportunity to expand available

TABLE I. Recent applications of machine learning methods in understanding atherosclerosis.

References	Focus	Significant discovery
Ambale-Venkatesh <i>et al.</i> ¹⁹	Event prediction	Machine learning methods applied to deep phenotyped datasets showed high predictive cardiovascular event accuracy. ¹⁹
Chen <i>et al.</i> ⁵	Event prediction	Machine learning methods can be applied to effectively predict risk in patients with severe dilated cardiomyopathy. ⁵
Han <i>et al.</i> ²⁰	Risk stratification/feature selection	Machine learning approaches can be used to identify important features related to quantitative atherosclerosis characterization and patients at risk of rapid coronary plaque progression. ²⁰
Hu <i>et al.</i> ²¹	Risk stratification/feature selection in comorbidity	Machine learning methods can be used to effectively predict prediabetes at risk for rapid atherosclerosis progression. ²¹
Huynh-Thu <i>et al.</i> ¹⁸	Validate machine learning for biomarker discovery	Machine learning approaches leveraged to explore methods of features extraction. ¹⁸
Motwani <i>et al.</i> ²²	Event prediction	Machine learning applied to clinical and Coronary CT data predicted 5-year all-cause mortality better than using clinical or Coronary CT data alone. ²²
Munger <i>et al.</i> ⁶	Regression/biomarker discovery	Machine learning methods can be effective in identifying the top predictors of noncalcified coronary burden in psoriasis. ⁶
Quesada <i>et al.</i> ²³	Risk stratification	Machine learning methods demonstrated better predictive capacity for cardiovascular events and better classification indicators than some traditional risk scores. ²³
Terrada <i>et al.</i> ²⁴	Diagnostic classification	Use of machine learning to develop highly accurate approach toward facilitating large-scale clinical diagnostics for atherosclerosis. ²⁴
van Rosendaal <i>et al.</i> ²⁵	Risk stratification	Risk scoring using machine learning analysis of standard 16 coronary segment stenosis and composition information has better prognostic accuracy than current Coronary CT integrated risk scores. ²⁵
Weng <i>et al.</i> ²⁶	Risk stratification	Machine learning shows greatly improved accuracy of cardiovascular risk prediction. ²⁶
Xie <i>et al.</i> ²⁷	Clustering/regression	Weighted machine learning approaches can improve the accuracy of cardiovascular diseases disease risk. ²⁷

training datasets. This is because images selected to be manually curated would only require that a small (8×8) pixel patch be manually curated. Once the CNN learns the features over this small patch from within the larger image, it can then apply this learned 8×8 pixel feature detector to the rest of the image. Thus, more images can be quickly curated and added to the training set.

Moreover, being able to clinically distinguish plaque phenotypes macroscopically is critical as it is known there are certain plaques more vulnerable to rupture.³⁰ Molecularly, vulnerable plaques are characterized by an intact thin fibrous cap, necrotic core, neovascularization, and a reduced volume of smooth muscle cells (SMCs).³¹ Macrophages in the fibrous cap release matrix metalloproteinases (MMPs), which, along with high shear stress, promote weakening of the cap.³¹ Additionally, microcalcification, iron accumulation within the fibrous cap, and macrophage cell death contribute to the weakening of the fibrous cap.³¹ Upon rupture, the contents of the necrotic core come into contact with circulating blood, and the coagulation cascade involving platelets is activated in response to the exposure of lipids and tissue factors, which were present in the necrotic core, leading to a thrombus and thus leading to decreased blood flow and a cardiovascular event.³¹ This is a complex process; however, ML methods could be used to extract

valuable feature importance in even complicated clinical processes and complex datasets.²⁷

Intracoronary optical coherence tomography image processing

Intracoronary optical coherence tomography (IOCT) is the application of light-based intracoronary endoscopic imaging along with optical coherence tomography to create cross-sectional images of the artery lumen and wall. Recently, ML guided image processing was applied to analyze IOCT images to generate an automated system to characterize atherosclerotic tissue.³² The authors used a supervised ML algorithm on image pixels to classify each pixel based on textural features and used those results to estimate the value of the optical attenuation coefficient.³² The images were then manually curated, and it was determined that the proposed method obtained an overall classification accuracy of 81.5%.³² While the training and validation dataset was relatively small, their work demonstrated that this application of ML showed potential for clinical use. Additionally, ML-based automated image segmentation and analysis can be used effectively to quantify atherosclerosis plaque deposition.³³ In this study, ML-based OCT image segmentation was used to highlight the borders along the vessel lumen and identify regions of atherosclerotic plaque.³³

Coronary computed tomography angiography image processing

Coronary computed tomography angiography (CCTA) is a non-invasive imaging technique that leverages radiolabeled compounds using computed tomography (CT) to depict the structure and content of the blood vessels including atherosclerotic plaque (Fig. 2). CCTA has proven an effective method for the characterization of coronary plaque burden because CCTA provides a characterization of not only lumen stenosis and arterial remodeling but also plaque subcomponents, including calcified, non-calcified, and high risk features.³⁰ A recent ML application related to CCTA was based on data from the PARADIGM (Progression of Atherosclerotic Plaque Determined by Computed Tomographic Angiography Imaging) registry.²⁰ The PARADIGM data were collected between 2003 and 2015 as part of a multi-center effort. At the time of the analysis, 1083 consecutive patients had undergone serial CCTA and met the inclusion criteria for the analysis.²⁰ After analysis using ML focused on selecting and ranking each feature’s importance for classifying individuals at risk for rapid plaque progression, the authors were able to determine that quantitative atherosclerotic plaque characterization was most influential. This was followed by qualitative plaque characterization by CCTA variables and then clinical and laboratory measures.²⁰

Often, studies incorporating CCTA also collect several non-imaging-based circulating biomarkers and clinical data as well. In one of the first large scale implementations of ML with CCTA data, it was shown that ML algorithms integrating clinical factors as well as coronary computed tomography angiography (CCTA) data demonstrated

superior prediction of 5-year all-cause mortality than clinical or CCTA data alone.²² Another study demonstrated that a risk score developed by using ML algorithms had greater prognostic accuracy for cardiovascular disease risk stratification from CCTA readings than the standard CCTA risk scores.²⁵

Application of machine learning in a human disease model of inflammation

Inflammation is critical to the progression of atherosclerosis. Psoriasis, a chronic inflammatory skin disease that affects nearly 2%–3% of the global population, is associated with accelerated rates of atherosclerosis, especially non-calcified coronary plaques.^{34,35} It is now well known that atherosclerotic plaque deposition is accelerated by untreated chronic inflammation in coronary arteries.³⁶ One recent application of ML used the Psoriasis Atherosclerosis Cardiometabolic Initiative (PACI) dataset to investigate biomarkers associated with this elevated cardiovascular risk in psoriasis. A combination of feature engineering and the application of ML methods has been used to demonstrate the mapping of complex biological data to a specific clinical outcome and that ML methods can be leveraged to identify top predictors of non-calcified coronary burden in patients with psoriasis.⁶ The analysis showed that the top-20 clinically significant biomarkers were body mass index, visceral adiposity, total adiposity, apolipoprotein A1, high-density lipoprotein, erythrocyte sedimentation rate, subcutaneous adiposity, small low-density lipoprotein particle, cholesterol efflux capacity, absolute granulocyte count, total cholesterol, waist-to-hip ratio, apolipoprotein B, very-low-density lipoprotein particle, absolute

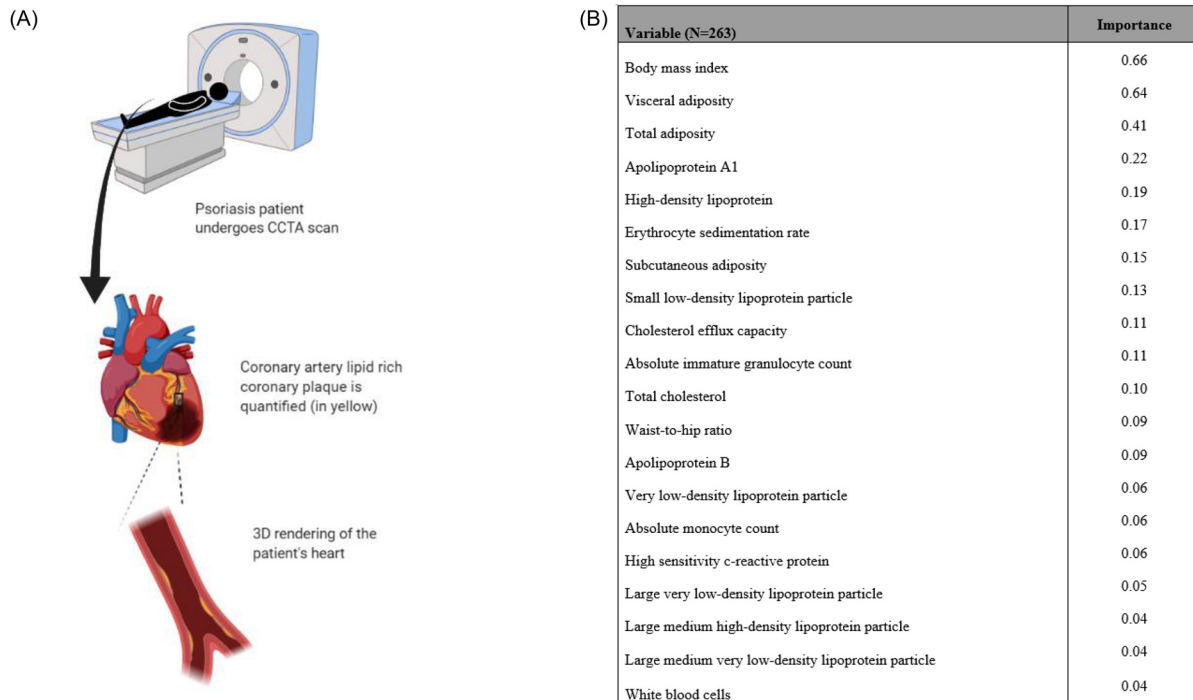


FIG. 2. (a) CCTA is used to quantify atherosclerotic plaque in coronary arteries by obtaining a 3D rendering of heart and mapping lipid-rich non-calcified coronary plaque (in yellow). (b) The Top predictors of non-calcified coronary artery plaque burden are generated using ML.

monocyte count, high-sensitivity C-reactive protein (hs-CRP), large very-low-density lipoprotein particle, large medium high-density lipoprotein particle, large medium very-low-density lipoprotein particle, and white blood cells (Fig. 2—use of coronary computed tomography angiography in psoriasis).⁶ These biomarkers are known to be related to obesity, dyslipidemia, and inflammation, which are relevant to the progression of atherosclerosis in this unique cohort.⁶ Particularly, in psoriasis patients, systemic inflammation is believed to affect a change in the inner lining of arteries causing an expression of adhesion molecules by endothelial cells. This change increases the capture and translocation of leukocytes into the intima.³⁷ Additionally, high-density lipoprotein (HDL) is converted to oxidized HDL in inflammatory states such as psoriasis, thus losing its protective effect.³⁸

This conversion decreases the protective function of HDL, leading to a more atherogenic profile and decreased cholesterol efflux ability, all of which promote plaque formation.³⁸ Indeed, patients with psoriasis are at increased risk of suffering a cardiovascular event compared with the general population, presumably due to accelerated atherogenesis.³⁹ One study demonstrated that severe psoriasis confers an additional 6.2% absolute risk of a 10-year rate of a major cardiovascular event.⁴⁰

This application of ML to atherosclerosis, within the psoriasis group, gives us a unique perspective and window into studying atherosclerosis since people suffering from inflammatory diseases such as psoriasis have a high prevalence of cardiovascular disease. Because of the extenuated inflammation, the progression of the disease, altered molecular networks, disrupted cellular communication, and aggravated skin disease pathophysiology in psoriasis, this biological amplification of biomarkers seen in psoriasis could be used to understand the complex process of atherosclerosis. These boosted biological signals

from a condition that magnifies atherosclerotic disease could increase the signal to noise ratio to elucidate the key players and their relationships amongst the many biological players and processes known (Fig. 3—concept of feature extraction from raw data).

Future directions and conclusions

Cumulatively, our review demonstrates that valuable insights can be extracted when ML is applied to clinical datasets. As a direct result of the recent application of ML in atherosclerosis, the number of known molecular, cellular, and physiological factors affecting atherosclerosis has dramatically increased. Most of these factors have been studied individually, leaving questions as to how many of these factors are interrelated and correlated. Adding environmental factors and broad lifestyle choices to this type of analysis could be particularly interesting from an interventional public health point of view. Thus, ML analysis of broad, well-characterized datasets will lead to a better understanding of how such factors are interconnected and impact cardiovascular health.

Furthermore, given that studies show the physiological pathways involved in atherogenesis are shared by other diseases, studying these diseases in tandem could be fruitful. A significant complication of atherosclerosis is the fact that there are a few symptoms until a cardiovascular event occurs.⁴¹ We imagine a deeper understanding of the condition could be achieved by using ML to analyze a targeted, well-characterized dataset of many variables important to atherosclerosis in combination with other conditions/comorbidities that may serve as surrogates for gauging the severity of underlying atherosclerosis.

We expect four major areas of advancement from applying ML to atherosclerosis. First, being able to identify many salient variables

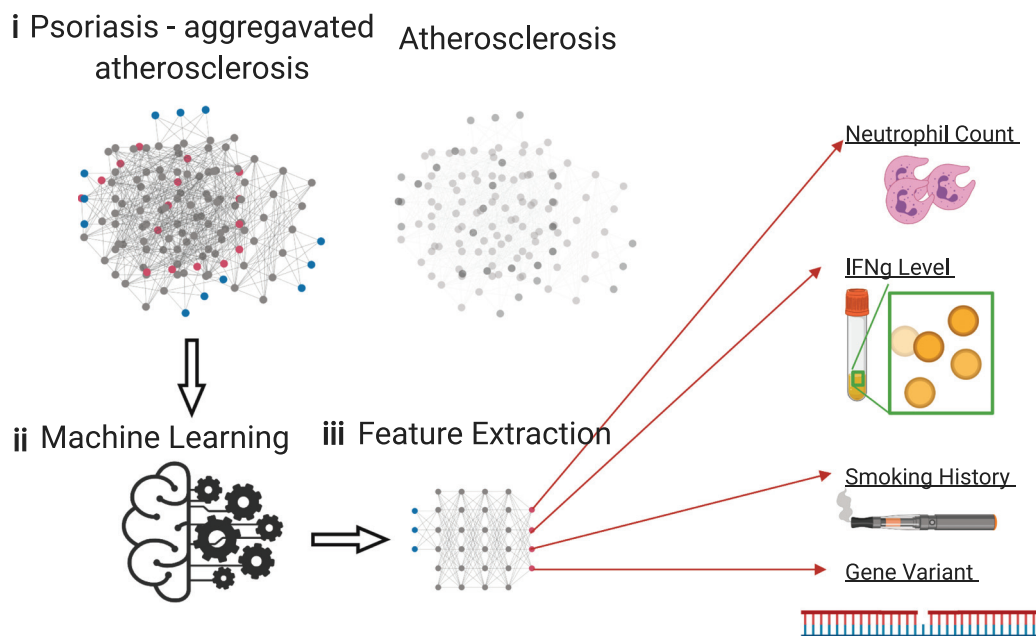


FIG. 3. (i) Psoriasis offers an extenuated view into atherosclerosis. (ii) The biologically meaningful parameters' signals will be amplified and easier to analyze with machine learning. (iii) This will enable significant extraction of potential biomarkers that could be used concomitantly and also explored for a biological mechanism for atherosclerosis so that they can be used as druggable targets. Created with BioRender.com.

will establish new biomarkers that could be used clinically. In a recent evaluation of the feasibility of creating a simple and clinically useful diagnostic panel for heart failure with reduced ejection fraction, the authors took a combined approach using untargeted metabolomics and ML.⁴² It was concluded that the combination of untargeted metabolomics and ML algorithms was a promising tool for the diagnostic workup in heart failure with reduced ejection fraction.⁴² However, large scale, multi-center studies of the methods are crucial for conformation and clinical utility of these tools.

Second, the ability to look at many variables simultaneously may inform us of the actual biological mechanisms and interactions across many different levels of biology from molecules to tissues. It is known that biomarkers play a significant role in the management, diagnosis, prognostication, and screening of diseases such as atherosclerosis. However, the utility of novel, emerging biomarkers is less established. Machine learning analysis allows for a multimodal approach that can simultaneously include many biomarkers, which can help identify multiple different pathophysiological pathways. This would provide a more integrated and informed assessment of the state of the patient, which would provide an improved measurement of risk compared to traditional risk scores.⁴³ In a recent clinical trial focusing on the treatment of preserved cardiac function heart failure patients with an aldosterone antagonist, 49 plasma biomarkers were measured in 379 trial participants.⁴⁴ The aim was to understand the relationship between these biomarkers and the risk of all-cause death or heart failure-related hospital admission. It was concluded that a multimarker ML approach was a promising strategy for accurate risk stratification.⁴⁴

Third, this may increase our current understanding of the relationship of multiple diseases and disease states, allowing a better understanding of complex “multimorbidity” and opening the opportunity for common interventions. Understanding complex multimorbidity is a significant challenge, and as a result, many diseases are studied in isolation. Machine learning provides tools that can be applied to overcome the many challenges in multimorbidity, but only a small percentage have been used for the study of multimorbidity.⁴⁵ Hu *et al.* demonstrated that ML can effectively predict prediabetics at risk for rapid atherosclerosis progression.²¹

Finally, the information these studies provide will highlight the pathway of collecting the next set of datasets for further iterations of ML discoveries. Machine learning techniques are capable of identifying clinically relevant patterns hidden amongst an abundance of potentially irrelevant information.⁴⁶ In a recent review of the application of ML in autoimmune disorders, the authors established that novel ML methods were being applied to data on multiple sclerosis, rheumatoid arthritis, and inflammatory bowel disease. Datasets used often included different data types in the modeling process.⁴⁷ This work shows that complex predictive models may be improved through the integration of specific data types and that an understanding of the best data to include in such models may not be the same as the data traditionally collected.

One limitation in the application of ML in current datasets is that the prospective follow-up data are largely incomplete. We feel that these prospective data will permit longitudinal analyses across phenotypes of interest, thereby accelerating understanding of potential biologic pathways. This has made predicting the value of surrogate endpoints of disease states very difficult. To solve this problem, researchers will need to continue to fill and expand datasets from

prospective studies while learning to adapt many of the traditionally reliable ML algorithms to evaluate high dimensional time-series data. Another limitation is that cross-validation using separate testing and validation datasets for robust model evaluation is not always an option because, while many of the best practices of validation, cross-validation, and independent testing of ML models have been discovered, very few well-characterized datasets related to atherosclerosis are available. Finally, it is well known that the appropriate application of these methods in solving complex problems in the domain of natural language processing or image and speech recognition has been extremely successful. We and others have successfully applied these methods on available high-quality datasets to highlight complicated relationships between multiple diseases. With continued increases in computing power and the vast expansion of well-characterized time-series biological and clinical data expected to be available very soon, now is the time for the application of ML methods to subclinical and interrelated diseases. In conclusion, here we provide a review of how ML can accelerate our understanding of atherosclerosis and some of the key applications of ML to these datasets for characterizing novel biomarkers of atherosclerotic disease progression. Much larger studies on well-characterized datasets will be possible as technology and analytical understanding improve in the future.

ACKNOWLEDGMENTS

We would like to acknowledge the Professionals at Johns Hopkins Engineering for Professionals Program and National Heart, Lung, and Blood Institute that initialized this interaction, leading to this project.

The PACI study was supported by the National Heart, Lung, and Blood Institute (NHLBI) Intramural Research Program (HL06193-07 and HL06235-03). The funding sources had no role in the design and conduct of the review; preparation or approval of the manuscript; and decision to submit the manuscript for publication.

Dr. Mehta is a full time U.S. government employee and has served as a consultant for Amgen, Eli Lilly, and Leo Pharma receiving grants/other payments; as a principal investigator and/or investigator for AbbVie, Celgene, Janssen Pharmaceuticals, Inc, and Novartis receiving grants and/or research funding; and as a principal investigator for the National Institutes of Health receiving grants and/or research funding.

All other authors declare no conflicts of interest with the work presented in this manuscript.

DATA AVAILABILITY

No human data or animal data are included in this review and thus data sharing agreement does not apply to this article.

REFERENCES

- 1M. Heron, “Deaths: Leading causes for 2016,” *Natl. Vital Stat. Rep.* **67**(6), 1–77 (2018).
- 2A. R. Fatkhullina, I. O. Peshkova, and E. K. Koltsova, “The role of cytokines in the development of atherosclerosis,” *Biochemistry* **81**(11), 1358–1370 (2016).
- 3D. E. Bild, D. A. Bluemke, G. L. Burke *et al.*, “Multi-ethnic study of atherosclerosis: Objectives and design,” *Am. J. Epidemiol.* **156**(9), 871–881 (2002).
- 4C. J. Cooper, T. P. Murphy, D. E. Cutlip *et al.*, “Stenting and medical therapy for atherosclerotic renal-artery stenosis,” *N. Engl. J. Med.* **370**(1), 13–22 (2014).

- ⁵T. Chen, P. Brewster, K. R. Tuttle *et al.*, "Prediction of cardiovascular outcomes with machine learning techniques: Application to the cardiovascular outcomes in renal atherosclerotic lesions (CORAL) study," *Int. J. Nephrol. Renovasc. Dis.* **12**, 49–58 (2019).
- ⁶E. Munger, H. Choi, A. K. Dey *et al.*, "Application of machine learning to determine top predictors of noncalcified coronary burden in psoriasis: An observational cohort study," *J. Am. Acad. Dermatol.* **83**, 1647–1653 (2019).
- ⁷B. Mackness and M. Mackness, "Anti-inflammatory properties of paraoxonase-1 in atherosclerosis," *Adv. Exp. Med. Biol.* **660**, 143–151 (2010).
- ⁸A. Rognoni, C. Cavallino, A. Veia *et al.*, "Pathophysiology of atherosclerotic plaque development," *Cardiovasc. Hematol. Agents Med. Chem.* **13**(1), 10–13 (2015).
- ⁹P. R. Moreno, K. R. Purushothaman, E. Zias, J. Sanz, and V. Fuster, "Neovascularization in human atherosclerosis," *Curr. Mol. Med.* **6**(5), 457–477 (2006).
- ¹⁰P. Libby, J. E. Buring, L. Badimon *et al.*, "Atherosclerosis," *Nat. Rev. Dis. Primers* **5**(1), 56 (2019).
- ¹¹P. K. Shah, E. Falk, J. J. Badimon *et al.*, "Human monocyte-derived macrophages induce collagen breakdown in fibrous caps of atherosclerotic plaques. Potential role of matrix-degrading metalloproteinases and implications for plaque rupture," *Circulation* **92**(6), 1565–1569 (1995).
- ¹²M. Ouimet, T. J. Barrett, and E. A. Fisher, "HDL and reverse cholesterol transport," *Circ. Res.* **124**(10), 1505–1518 (2019).
- ¹³G. Catalano and M. Guerin, "HDL and reverse cholesterol transport," in *The HDL Handbook* (Elsevier, 2010), pp. 61–76.
- ¹⁴T. Soeki and M. Sata, "Inflammatory biomarkers and atherosclerosis," *Int. Heart J.* **57**(2), 134–139 (2016).
- ¹⁵H. Larson, E. Chan, S. Sudarsanam, and D. E. Johnson, "Biomarkers," *Methods Mol. Biol.* **930**, 253–273 (2013).
- ¹⁶A. Boi, A. D. Jamthikar, L. Saba *et al.*, "A survey on coronary atherosclerotic plaque tissue characterization in intravascular optical coherence tomography," *Curr. Atheroscler. Rep.* **20**(7), 33 (2018).
- ¹⁷N. Hampe, J. M. Wolterink, S. G. M. van Velzen, T. Leiner, and I. Išgum, "Machine learning for assessment of coronary artery disease in cardiac CT: A survey," *Front. Cardiovasc. Med.* **6**, 172 (2019).
- ¹⁸V. A. Huynh-Thu, Y. Saeys, L. Wehenkel, and P. Geurts, "Statistical interpretation of machine learning-based feature importance scores for biomarker discovery," *Bioinformatics* **28**(13), 1766–1774 (2012).
- ¹⁹B. Ambale-Venkatesh, X. Yang, C. O. Wu *et al.*, "Cardiovascular event prediction by machine learning: The multi-ethnic study of atherosclerosis," *Circ. Res.* **121**(9), 1092–1101 (2017).
- ²⁰D. Han, K. K. Kolli, S. J. Al'Aref *et al.*, "Machine learning framework to identify individuals at risk of rapid progression of coronary atherosclerosis: From the PARADIGM registry," *J. Am. Heart Assoc.* **9**(5), e013958 (2020).
- ²¹X. Hu, P. D. Reaven, A. Saremi *et al.*, "Machine learning to predict rapid progression of carotid atherosclerosis in patients with impaired glucose tolerance," *EURASIP J. Bioinform. Syst. Biol.* **2016**(1), 14.
- ²²M. Motwani, D. Dey, D. S. Berman *et al.*, "Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis," *Eur. Heart J.* **38**(7), 500–507 (2017).
- ²³J. A. Quesada, A. Lopez-Pineda, V. F. Gil-Guillén *et al.*, "Machine learning to predict cardiovascular risk," *Int. J. Clin. Pract.* **73**(10), e13389 (2019).
- ²⁴O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "A novel medical diagnosis support system for predicting patients with atherosclerosis diseases," *Inf. Med. Unlocked* **21**, 100483 (2020).
- ²⁵A. R. van Rosendaal, G. Maliakal, K. K. Kolli *et al.*, "Maximization of the usage of coronary CTA derived plaque information using a machine learning based algorithm to improve risk stratification; Insights from the confirm registry," *J. Cardiovasc. Comput. Tomogr.* **12**(3), 204–209 (2018).
- ²⁶S. F. Weng, J. Reys, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," *PLoS One* **12**(4), e0174944 (2017).
- ²⁷J. Xie, R. Wu, H. Wang, Y. Kong, H. Li, and W. Zhang, "A novel weight learning approach based on density for accurate prediction of atherosclerosis," in *Intelligent Computing Theories and Application: 15th International Conference, ICIC 2019, Nanchang, China, 3–6 August, 2019*, Lecture Notes in Computer Science Vol 11644, edited by D.-S. Huang, K.-H. Jo, and Z.-K. Huang (Springer International Publishing, Cham, 2019), pp. 190–200.
- ²⁸J. W. Groenendyk and N. N. Mehta, "Applying the ordinal model of atherosclerosis to imaging science: A brief review," *Open Heart* **5**(2), e000861 (2018).
- ²⁹A. Abdolmanaf, L. Duong, N. Dahdah, and F. Cheriet, "Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography," *Biomed. Opt. Express* **8**(2), 1203–1220 (2017).
- ³⁰Y. A. Elnabawi, A. K. Dey, A. Goyal *et al.*, "Coronary artery plaque characteristics and treatment with biologic therapy in severe psoriasis: Results from a prospective observational study," *Cardiovasc. Res.* **115**(4), 721–728 (2019).
- ³¹K. Sakakura, M. Nakano, F. Otsuka, E. Ladich, F. D. Kolodgie, and R. Virmani, "Pathophysiology of atherosclerosis plaque progression," *Heart Lung Circ.* **22**(6), 399–411 (2013).
- ³²G. J. Ughi, T. Adriaenssens, P. Sinnaeve, W. Desmet, and J. D'hooge, "Automated tissue characterization of *in vivo* atherosclerotic plaques by intravascular optical coherence tomography images," *Biomed. Opt. Express* **4**(7), 1014–1030 (2013).
- ³³L. S. Athanasiou, C. V. Bourantas, G. Rigas *et al.*, "Methodology for fully automated segmentation and plaque characterization in intracoronary optical coherence tomography images," *J. Biomed. Opt.* **19**(2), 026009 (2014).
- ³⁴M. P. Schön and T. Ruzicka, "Psoriasis: The plot thickens," *Nat. Immunol.* **2**(2), 91–91 (2001).
- ³⁵M. A. Lowes, A. M. Bowcock, and J. G. Krueger, "Pathogenesis and therapy of psoriasis," *Nature* **445**(7130), 866–873 (2007).
- ³⁶F. Nijm and L. Jonasson, "Inflammation and cortisol response in coronary artery disease," *Ann. Med.* **41**(3), 224–233 (2009).
- ³⁷L. Eder and D. D. Gladman, "Atherosclerosis in psoriatic disease: Latest evidence and clinical implications," *Ther. Adv. Musculoskeletal Dis.* **7**(5), 187–195 (2015).
- ³⁸H. K. Siddiqi and P. M. Ridker, "Psoriasis and Atherosclerosis," *Circ. Res.* **123**(11), 1183–1184 (2018).
- ³⁹L. Eder, A. Dey, A. A. Joshi, W.-H. Boehncke, N. N. Mehta, and A. Szentpetery, "Cardiovascular diseases in psoriasis and psoriatic arthritis," *J. Rheumatol. Suppl.* **95**, 20–27 (2019).
- ⁴⁰N. N. Mehta, Y. Yu, R. Pinnelas *et al.*, "Attributable risk estimate of severe psoriasis on major cardiovascular events," *Am. J. Med.* **124**(8), 775 (2011).
- ⁴¹M. Rafieian-Kopaei, M. Setorki, M. Dousti, A. Baradaran, and H. Nasri, "Atherosclerosis: Process, indicators, risk factors and new hopes," *Int J. Prev. Med.* **5**(8), 927–946 (2014).
- ⁴²M. Marcinkiewicz-Siemion, M. Kaminski, M. Ciborowski *et al.*, "Machine-learning facilitates selection of a novel diagnostic panel of metabolites for the detection of heart failure," *Sci. Rep.* **10**(1), 130 (2020).
- ⁴³B. Ky, B. French, W. C. Levy *et al.*, "Multiple biomarkers for risk prediction in chronic heart failure," *Circ.: Heart Failure* **5**(2), 183–190 (2012).
- ⁴⁴J. A. Chirinos, A. Orlenko, L. Zhao *et al.*, "Multiple plasma biomarkers for risk stratification in patients with heart failure and preserved ejection fraction," *J. Am. Coll. Cardiol.* **75**(11), 1281–1295 (2020).
- ⁴⁵A. Hassaine, G. Salimi-Khorshidi, D. Canoy, and K. Rahimi, "Untangling the complexity of multimorbidity with machine learning," *Mech. Ageing Dev.* **190**, 111325 (2020).
- ⁴⁶F. Jiang, Y. Jiang, H. Zhi *et al.*, "Artificial intelligence in healthcare: Past, present and future," *Stroke Vasc. Neurol.* **2**(4), 230–243 (2017).
- ⁴⁷I. S. Stafford, M. Kellermann, E. Mossotto, R. M. Beattie, B. D. MacArthur, and S. Ennis, "A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases," *npj Digital Med.* **3**, 30 (2020).