Received: 21 May 2020 | Accepted: 6 August 2020

DOI: 10.1111/ijcp.13664

ORIGINAL PAPER

THERAPY AREA: OTHER

THE INTERNATIONAL JOURNAL OF CLINICAL PRACTICE WILEY

Importance of general adiposity, visceral adiposity and vital signs in predicting blood biomarkers using machine learning

Weihong Zhou[1] | Yingjie Wang[1] | Xiaoping Gu[2] | Zhong-Ping Feng[3] | Kang Lee[4] | Yuzhu Peng[1] | Andrew Barszczyk[3]

[1]Health Management Centre, Drum Tower Hospital Affiliated to Nanjing University Medical School, Nanjing, China

[2]Department of Anaesthesiology, Drum Tower Hospital Affiliated to Nanjing University Medical School, Nanjing, China

[3]Department of Physiology, University of Toronto, Toronto, ON, Canada

[4]Dr. Eric Jackman Institute of Child Study, University of Toronto, Toronto, ON, Canada

Correspondence
Yuzhu Peng, Health Management Centre, Drum Tower Hospital Affiliated to Nanjing University Medical School. No. 321 Zhongshan Road, Nanjing 210008, China.
Email: pyz1131@163.com

Andrew Barszczyk, Department of Physiology, University of Toronto, 1 King's College Circle, Rm 3306, Toronto, ON M5S 1A8, Canada.
Email: andrew.barszczyk@utoronto.ca

Abstract

Introduction: Blood biomarkers are measured for their ability to characterise physiological and disease states. Much is known about linear relations between blood biomarker concentrations and individual vital signs or adiposity indexes (eg, BMI). Comparatively little is known about non-linear relations with these easily accessible features, particularly when they are modelled in combination and can potentially interact with one another.

Methods: In this study, we used advanced machine learning algorithms to create non-linear computational models for predicting blood biomarkers (cells, lipids, metabolic factors) from age, general adiposity (BMI), visceral adiposity (Waist-to-Height Ratio, a Body Shape Index) and vital signs (systolic blood pressure, diastolic blood pressure, pulse). We determined the predictive power of the overall feature set. We further calculated feature importance in our models to identify the features with the strongest relations with each blood biomarker. Data were collected in 2018 and 2019 and analysed in 2020.

Results: Our findings characterise previously unknown relations between these predictors and blood biomarkers; in many instances the importance of certain features or feature classes (general adiposity, visceral adiposity or vital signs) differed from their expected contribution based on simplistic linear modelling techniques.

Conclusions: This work could lead to the formation of new hypotheses for explaining complex biological systems and informs the creation of predictive models for potential clinical applications.

1 | INTRODUCTION

Blood biomarkers help characterise physiological and disease states; they are routinely measured as part of clinical blood panels. Extensive study has linked concentrations of routinely measured biomarkers (eg, cell counts, lipids and metabolic factors) with vital signs and adiposity indexes. Vital signs (eg, heart rate, blood pressure) are highly dynamic and reflect the status of a person's cardiovascular system and their fitness. Adiposity indexes capture general adiposity (eg, Body Mass Index) or visceral adiposity (eg, Waist-to-Height ratio, a Body Shape Index) and provide a relatively stable indication of an individual's metabolic state; they are calculated from simple physical characteristics (eg, height, weight and waist circumference). It is highly advantageous to establish the extent to which such easily accessible features predict blood biomarker concentrations because features such as blood pressure, pulse and adiposity are manageable

Weihong Zhou, Yingjie Wang, and Xiaoping Gu are equal contribution.

through medication and lifestyle changes. Such features could therefore form the basis for efficient screening and intervention (where causal relations exist) against unhealthy blood biomarker changes that lead to disease.

It is well established that blood pressure is associated with concentrations of red blood cells,[1] white blood cells,[2] uric acid[3] and blood glucose.[4] Adiposity indexes predict the diagnostic criteria of metabolic syndrome (a disease of obesity); these diagnostic criteria include triglycerides,[5] low-density lipoprotein (LDL)[5,6] and glucose.[7] They are further associated with concentrations of high-density lipoprotein (HDL),[5,6,8,9] cholesterol,[5] uric acid,[10,11] white blood cells[12] and platelets[13] in the blood.

These previously established links tend to be simplistic linear relations modelled independently of one another. Yet relations in biological systems are frequently non-linear,[14] and predictions often benefit from multiple predictors (additively, or synergistically through interactions). Such complex non-linear relations have been characterised between some anthropometric predictors and blood biomarkers (lipids, glucose) using machine learning techniques,[6,7] but they have yet to be investigated on a wider scale for other blood biomarkers and with more diverse predictors.

Investigating such relations is now becoming possible with the proliferation of large clinical datasets and advanced machine learning algorithms. Large datasets such as electronic health records contain a plethora of potential predictors on a clinically diverse sample of the population,[15] and they provide the necessary statistical power for identifying subtle relations between variables. Advanced machine learning algorithms have advantages over traditional statistical regression techniques that often result in more accurate models. First, they automatically identify and model complex yet-to-be-discovered relations[16,17] that are not feasible to identify with traditional (manual) methods. Second, they are not subject to the assumptions of traditional statistical modelling (eg, homoscedasticity in the data)[18]; such assumptions often do not hold true in nature and thus can hinder accuracy. Third, machine learning constructs (eg, neural networks, decision trees) allow more flexibility in mapping inputs to outputs than traditional statistical modelling that relies on more deterministic mathematical functions. Thus, predictions can potentially be more precise. Furthermore, the relative contribution of each feature can be assessed using one of several techniques (eg, permutation analysis). Thus, advanced machine learning algorithms are ideal for modelling and studying novel interactions among variables.

The present study evaluated the extent to which general adiposity (BMI), visceral adiposity (Waist-to-Height Ratio, a Body Shape Index), vital signs (systolic blood pressure, diastolic blood pressure, pulse) and age predict routinely measured blood biomarkers (cells, lipids, metabolic factors) when they are modelled non-linearly and in combination with one another. We modelled these relations separately in both men and women using an advanced machine learning algorithm from a large annual physical examination dataset of generally healthy adults. We determined the prediction accuracy of these models. We further calculated feature importance to identify the key predictive features for each blood biomarker. We aimed to identify previously unknown relations between these predictors and blood biomarkers.

What's known

- Non-invasive characteristics like certain vital signs, adiposity indexes and physical features exhibit modest predictive relations with blood biomarkers.

What's new

- We have built machine learning–based computational models that predict blood biomarkers from non-invasive features (vital signs, adiposity indexes and physical characteristics) with appreciable accuracy.
- In doing so, we have identified novel relations (including non-linear relations) between multiple interacting features.

2 | METHODS

2.1 | Study population and measures

Generally healthy adults (≥18 years of age) were recruited at the Health Management Centre at Nanjing Drum Tower Hospital (Nanjing, China) in conjunction with routine annual physical examinations. The use of human data in this study was approved by the institutional review board at Nanjing Drum Tower Hospital and subjects provided consent for their participation in the study. A total of 31,701 patient encounters were included for analysis. Their inclusion criteria were having: (i) complete data and (ii) no outliers (defined as ≥4 standard deviations above or below the mean). Complete data involved a routine blood panel, vital signs (blood pressure, pulse), height, weight, waist circumference, age and gender (see Appendix: Supplementary Methods for details). Data were collected in 2018 and 2019 and analysed in 2020.

2.2 | Statistical analysis: Model training, validation and testing

We created predictive models for the following blood biomarkers: red blood cell count, white blood cell count, platelet count, high-density lipoprotein, low-density lipoprotein, total cholesterol, triglycerides, fasting blood glucose and uric acid. The predictors for each model were as follows: age, Body Mass Index (BMI), a Body Shape Index (aBSI), waist-to-height ratio (WaHtR), systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse.

We began by separating our data by gender so we could create separate models and study male and female predictors

independently. For each gender, we randomly allocated 80% of our dataset for model training, and the remaining 20% as the validation dataset. We trained a separate multilayer perceptron neural network model for predicting each blood biomarker. For each model, we used the full set of predictors. Our rationale for choosing this type of model and model training details are described in Appendix: Supplementary Methods. A network diagram is shown in Figure S1.

After training, we calculated model accuracy using the 20% of the data that was not used in model training and testing (the validation set). Accuracy calculated on this 'unseen' set reflects the model's accuracy when generalised to the population. We calculated accuracy in two ways. First, we determined the explained variance of the model (relative to just predicting the mean) by calculating the coefficient of determination (R^2). Next, we calculated the mean absolute error (MAE) as the absolute difference between reference and predicted values. Our rationale for choosing these measures is discussed in Appendix: Supplementary Methods. We further calculated proportional feature importance to determine the relative predictive ability of each feature within the model (see Appendix: Supplementary Methods for details).

We repeated this entire training and testing process (including randomised subject allocation) 100 times for each model so we could generate statistical estimates of model performance and feature importance that generalise beyond any one specific partitioning of subjects. We assessed model performance and feature importance for each model by calculating the mean and 95% confidence interval for each measure (explained variance, mean absolute error and relative feature importance) across all 100 model iterations. We felt that 100 iterations would ensure that we could estimate the true mean with high precision (narrow confidence intervals) for all models; this facilitates comparisons of model performance and feature importance. For ease of understanding, we also calculated the Pearson correlation coefficient. It was calculated by taking the square root of the mean explained variance and the square root of the bounds of its 95% confidence interval across all 100 model iterations.

3 | RESULTS

We summarised the predictors and blood biomarkers of our study participants by gender (Table 1). Associations between individual predictive features and blood biomarkers are described in Figure S2. Mean feature values and Pearson correlation coefficients between predictors and blood biomarkers generally differed between men and women.

3.1 | Model performance

We quantified the prediction accuracy of each blood biomarker model. For both male and female models, explained variance ranged from 4% to 24%, and all models predicted significantly better than chance ($P < .001$) (Table 2). Correspondingly, all models exhibited a Pearson correlation coefficient that statistically differed from zero ($P < .001$), but these correlations varied widely between $r = .20$ and $r = .49$ depending on the blood biomarker being modelled (Figure 1). For men, the best-performing models were those for red blood cell count ($r = .49$), triglycerides ($r = .38$), high-density lipoprotein ($r = .35$) and glucose ($r = .31$). For female models, they were those for triglycerides ($r = .48$), uric acid ($r = .43$), glucose ($r = .40$), total cholesterol ($r = .38$), low-density lipoprotein ($r = .37$) and high-density lipoprotein ($r = .32$). In all cases, the performance of our combined models exceeded the performance of each model's single best feature (as reported in Figure S2), thus justifying our use of multivariate models.

TABLE 1 Subject characteristics

	Males (n = 21,138)	Females (n = 9,532)
Predictors		
Age, years	49 ± 16	46 ± 17
Body mass index, kg/m^2	25.0 ± 2.9	22.5 ± 2.9
Body shape index, m$^{11/6}$ kg$^{-2/3}$	0.079 ± 0.004	0.075 ± 0.004
Waist-height ratio	0.51 ± 0.05	0.47 ± 0.06
Systolic blood pressure, mm Hg	128 ± 16	120 ± 18
Diastolic blood pressure, mm Hg	83 ± 11	74 ± 11
Pulse, bpm	77 ± 11	79 ± 11
Blood panel: cells		
Red blood cell count, million cells/μL	5.0 ± 0.4	4.5 ± 0.3
White blood cell count, million cells/μL	6.4 ± 1.5	6.0 ± 1.4
Platelet count, million cells/μL	205.9 ± 47.5	227.7 ± 50.5
Blood panel: lipids		
High-density lipoprotein, mmol/L	1.21 ± 0.31	1.49 ± 0.35
Low-density lipoprotein, mmol/L	2.47 ± 0.66	2.43 ± 0.67
Total cholesterol, mmol/L	4.78 ± 0.85	4.78 ± 0.90
Triglycerides, mmol/L	1.63 ± 0.95	1.15 ± 0.69
Blood panel: metabolic factors		
Glucose, mmol/L	5.4 ± 0.8	5.1 ± 0.7
Uric Acid, μmol/L	365 ± 74	269 ± 62

Note: Values are mean ± standard deviation.

3.2 | Feature importance

We further calculated the relative importance of each predictor for each blood biomarker model (Figure 2, with values listed in Tables S1 and S2) and identified the 'top predictors' defined as all features

TABLE 2 Explained variance and mean absolute error of male and female predictive models

	Explained variance, % (95% CI)	Mean absolute error (95% CI)
Male models		
Cells		
Red blood cell count	24 (24-24)***	0.26 (0.26-0.26) million cells/μL
White blood cell count	6 (5-6)***	1.12 (1.12-1.12) million cells/μL
Platelet count	5 (5-5)***	36.18 (36.1-36.26) million cells/μL
Lipids		
High-density lipoprotein	13 (12-13)***	0.23 (0.23-0.23) mmol/L
Low-density lipoprotein	4 (4-4)***	0.51 (0.51-0.51) mmol/L
Total cholesterol	5 (5-5)***	0.67 (0.66-0.67) mmol/L
Triglycerides	14 (14-14)***	0.63 (0.62-0.63) mmol/L
Metabolic factors		
Glucose	10 (10-10)***	0.60 (0.60-0.60) mmol/L
Uric acid	8 (8-8)***	55.62 (55.5-55.74) μmol/L
Female models		
Cells		
Red blood cell count	5 (5-6)***	0.24 (0.24-0.24) million cells/μL
White blood cell count	7 (7-8)***	1.03 (1.03-1.04) million cells/μL
Platelet count	4 (4-5)***	38.98 (38.86-39.1) million cells/μL
Lipids		
High-density lipoprotein	10 (10-10)***	0.27 (0.27-0.27) mmol/L
Low-density lipoprotein	14 (14-14)***	0.50 (0.50-0.05) mmol/L
Total cholesterol	15 (14-15)***	0.66 (0.66-0.67) mmol/L
Triglycerides	23 (23-24)***	0.45 (0.44-0.45) mmol/L
Metabolic factors		
Glucose	16 (15-16)***	0.47 (0.47-0.47) mmol/L
Uric acid	19 (18-19)***	44.48 (44.32-44.63) μmol/L

Note: Values are means (95% confidence intervals) for 100 model iterations. Explained variance was calculated as the coefficient of determination (R^2) against the mean-predicting model.

***$P < .001$ versus the mean predicting model.

contributing to the first 70% of feature importance. We first considered male models. For red blood cell count, age was the top predictor (84% importance). For white blood cell count, WaHtR (visceral adiposity) and pulse were the top predictors (70% combined importance). For platelet count, age (63% importance) was the most important predictor, with WaHtR and pulse also contributing as top features (10% each). For high-density lipoprotein, BMI (general adiposity) was the top predictor (76% importance). For low-density lipoprotein, age (48% importance) and WaHtR (18% importance) were the top predictors. For total cholesterol, age was the most important feature (60% importance), with DBP, pulse and WaHtR contributing an additional 35% importance in approximately equal amounts. For triglycerides, BMI was the best predictor (52% importance). ABSI, age and DBP were also top predictors since they accounted for an additional 40% feature importance in approximately equal amounts. For glucose, age, SBP and pulse were the most important features (79% combined importance). For uric acid, BMI was the most important feature (62% importance). ABSI and WaHtR (both visceral adiposity features) were important as well (25% combined importance).

We next considered female models. For red blood cell count, age and DBP were the top predictors in women (77% combined importance). For white blood cell count, WaHtR, pulse, age and SBP were the top predictors (90% combined importance). For platelet count, age was most important (60% importance) and pulse was the other top feature (20% importance). For high-density lipoprotein, BMI and age were the top predictors (78% combined importance). For low-density lipoprotein, age was the top predictor (80% importance). For total cholesterol, age was once again the sole top predictor (95% importance). For triglycerides, BMI, WaHtR and age were the top predictors (79% combined importance). For glucose, age and SBP were most important, with BMI and WaHtR being slightly less important top features (84% combined importance). For uric acid, BMI and age were the most important predictors (90% combined importance).

FIGURE 1 Prediction accuracy of male and female predictive models. Pearson correlation coefficient (*r*) (95% confidence interval) for 100 models. All correlations are significant at $P < .001$.

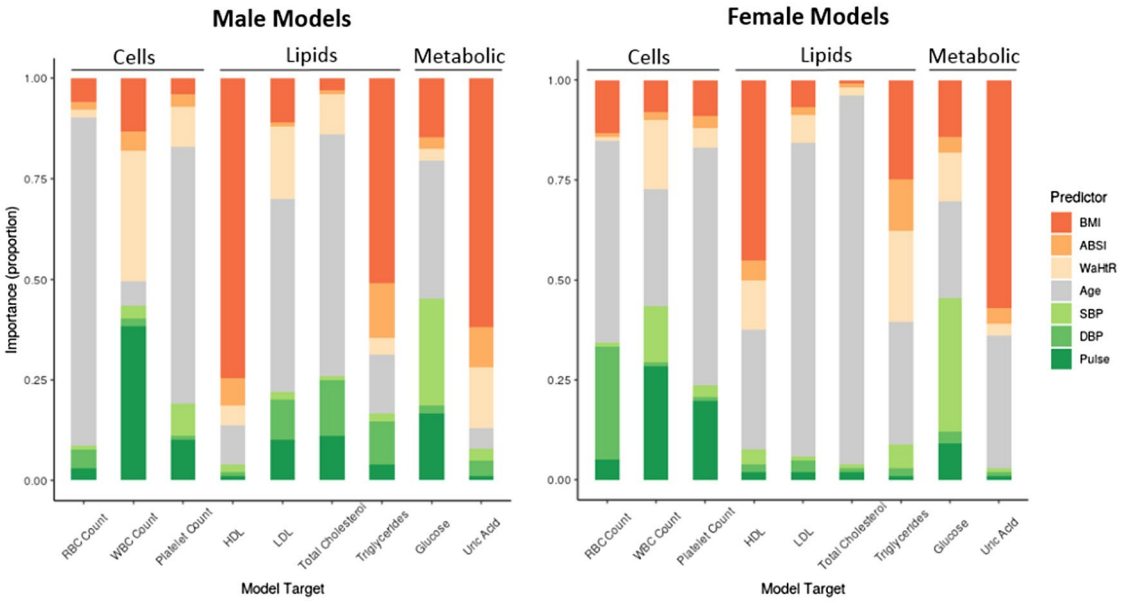

FIGURE 2 Feature Importance for male and female predictive models. Mean proportion of importance for 100 models. Anthropometric adiposity measures are shades of orange, and physiological measures are shades of green

4 | DISCUSSION

Our study set out to determine the extent to which a general adiposity index (Body Mass Index), visceral adiposity indexes (waist-to-height ratio, a Body Shape Index), vital signs (systolic blood pressure, diastolic blood pressure, pulse) and age predict blood biomarkers when they are modelled non-linearly and in combination with one another.

We modelled these relations using advanced machine learning techniques, determined the accuracy of these models and characterised the importance of each of these features within the model. Our work supports the use of multivariate models for improving prediction accuracy over and above using a single best feature for biomarker prediction. Below, we compared our results against findings from other multivariate studies that also used non-invasive and easily accessible features (eg, adiposity indexes, vital signs, age) to predict these blood biomarkers.

We first considered cell counts. To the best of our knowledge, our work is the first to investigate these features as predictors of red blood cell count. We found diastolic blood pressure and age to be major predictors of red blood cell count in women. Only one past study has used stepwise multiple regression to predict white blood cell count. It found that age and BMI were major predictors, whereas WaHtR made a minor contribution and SBP and DBP had no significant effect.[12] By comparison, our model puts significant emphasis on WaHtR (visceral adiposity) rather than BMI (general adiposity), and newly identifies pulse as a highly predictive feature. This discrepancy suggests the importance of considering non-linear interactions among multiple predictors and their relation to white cell counts, which stepwise linear regression fails to do. We are not aware of any multivariate studies predicting platelet count. Single-variable correlations between platelet count and adiposity indexes have been described here (Figure S2) and elsewhere[13] and they are small. Adiposity features are not important in our predictive models for platelet count either. Rather, our work newly identifies pulse as a top predictive feature in combination with age.

We next considered lipids. Studies have predicted HDL concentration using stepwise multiple regression.[11,19] A study in Chinese adults study found that BMI, waist circumference, gender and age were similarly important predictors. By contrast in our models, BMI was the most important predictor of HDL concentration for Chinese men and women in our sample, and waist circumference-based measures were not informative. Age was highly important in the female model despite being poorly correlated with HDL on its own. This suggests that age exhibits a non-linear relation with HDL or that its importance comes from its interaction with BMI.

A study predicting LDL concentration in Chinese adults from WC, BMI, age and gender using stepwise multiple regression found that age was a major predictive factor and that WC was also predictive but to a much lesser degree.[19] Our models also identify age as the most important predictor and WC as a minor predictor; we further identified DBP and pulse as important predictors (>20% combined importance).

Studies have predicted total cholesterol concentration using stepwise multiple regression.[5,11] A study in Chinese adults showed that demographics (age, gender, geographic region, smoking, drinking, family income, education, diet and sedentary activity) were the most important feature, and BMI and WC contributed very little.[5] Our models identify age as a major predictor and we show for the first time that DBP and pulse are important predictors (>25% combined feature importance) in men.

Studies have predicted triglyceride concentration using stepwise multiple regression.[5,11,19] A Chinese study examining WC, BMI, Age and gender as predictors found that WC was important (along with age).[19] Our female models identify aBSI and WaHtR (both visceral adiposity indexes) as making up the majority of the adiposity contribution, but in men BMI (general adiposity) was most important. Another Chinese study found that BMI and WC explained 4.8% and 4.7% of additional variation (5.2% together) in triglycerides over and above the variation explained by a demographic base model (18.2%).[5] This even split of BMI and WC are somewhat consistent with our results in women, but again not with the massive importance of general adiposity (BMI) in men.

Finally, we considered metabolic factors. A previous study predicted glucose concentration from a set of adiposity features (BMI, fat percentage, WC and WaHtR), gender and age using stepwise multiple regression.[11] WaHtR was the only important feature (age produced a tiny improvement). Our glucose models rely minimally on WaHtR. Rather, vital signs (especially systolic blood pressure) make up almost half of the prediction importance and age is a major feature as well. The high importance of age in our models suggests that it has a non-linear relation with glucose concentration or that it interacts with vital signs to predict glucose.

A past study predicted uric acid concentration from a set of adiposity features (BMI, fat percentage, WC and WaHtR),[11] gender and age using stepwise multiple regression. It reported that WC was the only major predictor. By contrast in our models, BMI accounted for the majority of feature importance in both our male and female models.

Overall, our work identified novel predictor combinations for common blood biomarkers. When it came to feature importance, WaHtR and aBSI predicted somewhat differently despite both being visceral adiposity indexes. These two measures do indeed differ slightly (and are somewhat decorrelated) because they account for body size to different degrees; WaHtR only considers height, while aBSI considers height *and* BMI. This makes visceral adiposity estimates using aBSI less dependent on body size and more dependent on body shape.[20] This distinction appears to be relevant for predicting blood biomarker concentrations. Furthermore, while the predictive ability of adiposity indexes remained relatively consistent across genders, the relative importance of vital signs showed large variations between men and women. A key finding of our work is that vital signs are generally highly important predictors of blood contents, and sometimes more than adiposity indexes.

4.1 | Implications of our findings

The new predictive links identified by our work could spur novel hypotheses describing how various predictors are biologically

interrelated in their prediction of blood biomarkers. It also makes the case for using non-linear multivariate models in identifying and studying such relations. Clinically, our work could support the eventual creation of non-invasive tools for predicting blood contents. Blood collection is inconvenient, uncomfortable and expensive, making it impractical for situations where frequent, rapid or broad testing is required (eg, monitoring response to therapy or large-scale screening). Predictive models based on these and other easily accessible features could non-invasively and inexpensively identify people who are at risk of abnormal blood biomarker concentrations and recommend them for follow-up with a blood test. It is important to verify results to ensure accuracy as well as to potentially gain additional precision that could be diagnostically useful. In some cases, predictive models could be used to suggest lifestyle interventions (assuming such links are causally related). For instance, adiposity is a fair predictor (>20% combined importance) of WBC count, HDL, triglycerides, uric acid and glucose in both genders, and LDL in men only. Modifying one's lifestyle to reduce body fat percentage might positively impact those biomarkers. Similarly, vital signs are fair predictors of WBC count and glucose in both genders, LDL and total cholesterol in men, and RBC count and platelet count in women. Once again, lifestyle modification or drugs to lower blood pressure, or increasing fitness to lower resting pulse rate could help mitigate unhealthy levels of blood biomarkers.

4.2 | Limitations and future work

Future studies could expand on this work by segmenting subjects according to disease status, medication status, ethnicity and lifestyle factors (eg, alcohol, smoking and physical activity) to account for possible confounders and investigate predictive relations under different conditions. It will be important to determine how well our findings generalise to other populations having different genetic and sociocultural characteristics; past studies suggest there are both similarities and differences in terms of which features are most important for predicting certain blood biomarkers. Additional predictive features should be investigated as well; this is facilitated by the recent proliferation of large clinical datasets with novel types of information. In particular, longitudinal features (information collected at multiple time points) could contain information about feature trends and variability; such features have proven useful in physiology-based predictions of cardiovascular disease risk.[21] Finally, our study investigated just one machine learning algorithm. Different types of models (eg, neural network, random forest) may produce different results and this should be investigated.

5 | CONCLUSIONS

In conclusion, we have characterised several novel relations between easily accessible features (adiposity indexes, vital signs, age) and blood biomarkers. This work could lead to the formation of new hypotheses for explaining human physiology. Predictive models utilising these features to determine blood biomarker concentrations could be helpful for identifying and managing health risks in the population while minimising the need for collecting blood, which is traditionally invasive, expensive and time consuming.

DISCLOSURE

None to declare.

ORCID

Yuzhu Peng https://orcid.org/0000-0002-2905-2336
Andrew Barszczyk https://orcid.org/0000-0001-5341-9655

REFERENCES

1. Wan J, Ristenpart WD, Stone HA. Dynamics of shear-induced ATP release from red blood cells. *Proc Natl Acad Sci U S A*. 2008;105:16432-16437.
2. Orakzai RH, Orakzai SH, Nasir K, et al. Association of white blood cell count with systolic blood pressure within the normotensive range. *J Hum Hypertens*. 2006;20:341-347.
3. Kansui Y, Ohtsubo T, Goto K, et al. Association of serum uric acid with blood pressure in Japanese men: cross-sectional study in work-site group. *Circ J*. 2011;75:2827-2832.
4. Cambien F, Warnet JM, Eschwege E, Jacqueson A, Richard JL, Rosselin G. Body mass, blood pressure, glucose, and lipids. Does plasma insulin explain their relationships? *Arteriosclerosis*. 1987;7:197-202.
5. Du S-M, Ma G-S, Li Y-P, et al. Relationship of body mass index, waist circumference and cardiovascular risk factors in Chinese adult. *Biomed Environ Sci*. 2010;23:92-101.
6. Lee BJ, Kim JY. Identification of the best anthropometric predictors of serum high- and low-density lipoproteins using machine learning. *IEEE J Biomed Heal Informatics*. 2015;19:1747-1756.
7. Lee BJ, Ku B, Nam J, Pham DD, Kim JY. Prediction of fasting plasma glucose status using anthropometric measures for diagnosing Type 2 diabetes. *IEEE J Biomed Heal Informatics*. 2014;18:555-561.
8. Milman N, Kirchhoff M. Relationship between serum ferritin and risk factors for ischaemic heart disease in 2235 Danes aged 30–60 years. *J Intern Med*. 1999;245:423-433.
9. Hwang JH, Kam S, Shin J-Y, et al. Incidence of metabolic syndrome and relative importance of five components as a predictor of metabolic syndrome: 5-year follow-up study in Korea. *J Korean Med Sci*. 2013;28:1768-1773.
10. Mangge H, Zelzer S, Puerstner P, et al. Uric acid best predicts metabolically unhealthy obesity with increased cardiovascular risk in youth and adults. *Obesity*. 2013;21:71-77.
11. Bosy-Westphal A, Geisler C, Onur S, et al. Value of body fat mass vs anthropometric obesity indices in the assessment of metabolic risk factors. *Int J Obes*. 2006;30:475-483.
12. Dixon JB, O'Brien PE. Obesity and the white blood cell count: changes with sustained weight loss. *Obes Surg*. 2006;16:251-257.
13. Farhangi MA, Keshavarz SA, Eshraghian M, Ostadrahimi A, Saboor-Yaraghi Dr AA. White blood cell count in women: relation to inflammatory biomarkers, haematological profiles, visceral adiposity, and other cardiovascular risk factors. *J Heal Popul Nutr*. 2013;31:58-64.
14. Sharp TA, Grunwald GK, Giltinan KEK, King DL, Jatkauskas CJ, Hill JO. Association of anthropometric measures with risk of diabetes and cardiovascular disease in Hispanic and Caucasian adolescents. *Prev Med*. 2003;52:402-409.
15. Obermeyer Z, Emanuel EJ. Predicting the future-big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216-1219.

16. Mullainathan S, Spiess J. Machine learning: An applied econometric approach. *J Econ Perspect*. 2017;31:87-106.
17. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. Vol. 103. New York, NY: Springer, New York; 2013.
18. Bzdok D, Altman N, Krzywinski M. Points of significance: statistics versus machine learning. *Nat Methods*. 2018;15:233-234.
19. Xu C, Yang X, Zu S, Han S, Zhang Z, Zhu G. Association between serum lipids, blood pressure, and simple anthropometric measures in an adult Chinese population. *Arch Med Res*. 2008;39:610-617.
20. Krakauer NY, Krakauer JC. A new body shape index predicts mortality hazard independently of body mass index. *PLoS ONE*. 2012;7:e39504.
21. Zhao J, Feng QP, Wu P, et al. Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. *Sci Rep*. 2019;9:1-10.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zhou W, Wang Y, Gu X, et al. Importance of general adiposity, visceral adiposity and vital signs in predicting blood biomarkers using machine learning. *Int J Clin Pract*. 2021;75:e13664. https://doi.org/10.1111/ijcp.13664