Rheologica Acta
https://doi.org/10.1007/s00397-023-01402-2

ORIGINAL CONTRIBUTION

Physiology-based parameterization of human blood steady shear rheology via machine learning: a hemostatistics contribution

Sean Farrington[1,2] · Soham Jariwala[1,2] · Matt Armstrong[3] · Ethan Nigro[1,2] · Norman J. Wagner[1,2] · Antony N. Beris[1,2]

Received: 24 February 2023 / Revised: 6 June 2023 / Accepted: 7 June 2023

Abstract

Hemorheology is the study of blood flow and the mechanical stresses and kinematics involved. The Casson constitutive equation is a popular and simple model used to describe the steady shear rheology of blood, with only two parameters that specify an infinite shear viscosity and a yield stress that depend on blood physiology. Previous literature has identified hematocrit and fibrinogen concentration as the two most important physiological factors that affect blood flow, but previous parameterizations of the Casson model may not be reliable due to the use of non-standardized data sets. This study uses machine learning and the largest standardized dataset to improve the parameterization of the Casson model with respect to hematocrit and fibrinogen concentration for healthy individuals. The study also employs machine learning to identify a potential additional factor, the mean corpuscular hemoglobin (MCH), that may affect blood rheology. The proposed approach demonstrates the potential for machine learning to improve the connection between physiology and blood rheology with possible implications in cardiovascular diagnostics.

Keywords Hemorheology · Machine learning · Blood physiology · Casson model · Yield stress · Viscosity

Introduction

Cardiovascular disease is the leading cause of death in the USA, making up a quarter of the national death toll (CDC 2022). Risk factors of cardiovascular diseases are typically diagnosed by classifying anomalies such as high blood pressure or cholesterol (Fryar et al. 2012). Early diagnosis of these diseases improves the possibility of successful treatment (Kyrle and Eichinge 2005; Torpy et al. 2007). Risk factors are usually assessed by routine biochemical blood tests that examine levels such as low-density lipoprotein (LDL) cholesterol, blood cell count, and triglycerides (MayoClinic 2021). Physicians decide to conduct further specific biochemical tests by comparing risk factors to a healthy range. While values outside the healthy range are indicative of potential disease, they do not provide a possible causative link to blood flow, which itself may be important in cardiovascular disease diagnostics and prevention (Dintenfass 1974; Beris et al. 2021). Consequently, numerous researchers are exploring aspects of blood rheology as a potential indicator of cardiovascular health as well as a possible diagnostic of disease (Connes et al. 2016; Hitsumoto 2017; Javadi et al. 2022; Tabesh et al. 2022).

Given that one primary function of blood is to transport oxygen to cells by diffusing through vascular walls, serious damage—often irreversible—occurs to tissues with interrupted flow (Dintenfass 1974). For example, an eye disease called retinopathy occurs when the smallest capillaries lack oxygen and develop a microstroke (Torpy et al. 2007). Hardened red blood cells are a crucial factor behind retinopathy and play a key role in the rheology of blood. Many biochemical markers are identified in the presence of pathological conditions, but blood rheology may provide a more direct link to the disease's manifestation. There is evidence in the literature for correlations between blood rheology and diseases such as sickle cell anemia (Connes et al. 2016),

✉ Antony N. Beris
beris@udel.edu

[1] Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19713, USA

[2] Center for Research in Soft Matter & Polymers (CRiSP), University of Delaware, 150 Academy Street, Newark, DE 19716, USA

[3] Department of Chemistry and Life Science, Chemical Engineering Program, United States Military Academy, West Point, New York 10996, USA

Published online: 27 June 2023

diabetes mellitus (Le Devehat et al. 2004), and hypertension (Chien 1986).

There are many attempts from the literature to recommend blood rheology as an important parameter for routine blood tests because of its relation to cardiovascular disease (Lowe et al. 2000; Baskurt and Meiselman 2008; Lemonne et al. 2014; Hitsumoto 2017; Tabesh et al. 2022). There cannot be a single healthy range for every patient as the mean value of blood viscosity is not sufficient to describe the variability in a healthy population. Therefore, we need a model to reduce this variability. The diagnosis must be personalized to understand if a change in rheology comes from physiology variations or from cardiovascular disease. A physiological parameterization of rheology allows a physician to compare a patient's empirical results against a predicted healthy range for a patient's particular blood rheology. A model that quantifies variations of rheology parameters in healthy blood could also determine deviations that are presented by drugs such as statins or aspirin (Lowe et al. 2000; Rosenson et al. 2004), or disease such as thrombosis or hypertension (Chien 1986; Kyrle and Eichinge 2005). For example, the yield stress may increase within the healthy range of fibrinogen concentration, and it is important to distinguish whether the high yield stress is a result of increased fibrinogen concentration or a disease. In industrial manufacturing rheology is ubiquitous as a measurement for quality control (Gahleitner 1999), which is comparable to how blood rheology is seen as a potential tool to screen for abnormalities in patient's blood flow. We limit our investigation here to the case of the steady shear rheology of blood from healthy donors.

While the prior discussion shows that many disease's manifest in atypical blood rheology, there are already significant variation in blood rheology of healthy individuals (Horner 2020). The potential for using blood rheological measurements to aid in diagnosis of disease states first requires a quantitative connection between blood physiology and rheological properties in healthy individuals as a basis. The work presented here is a step toward developing a better, quantitative relationship between blood physiology and steady shear blood rheology in healthy individuals by exploiting both improvements in experimental data sets for healthy donors (Horner 2020) as well as advanced data science using machine learning (Beris et al. 2021).

The blood is a dense suspension of red blood cells, white cells, and platelets suspended in plasma which is an aqueous solution containing on the order of 1000 different proteins (Benjamin and McLaughlin 2012). Much of the complexity in its bulk rheological behavior is a consequence of the aggregation of red blood cells into rouleaux structures and their deformation (Beris et al. 2021). At low shear rates ($\dot{\gamma} <\sim 0.01\ s^{-1}$), red blood cells aggregate into stacks known as rouleaux. Rouleaux provides structure to the fluid, creating a non-zero apparent yield stress, and their breakage under flow contributes to a shear-thinning viscosity and hysteresis, which makes blood a thixo-elasto-viscoplastic fluid (Beris et al. 2021). The rouleaux structure is dependent on a plasma protein, fibrinogen, which contributes to the cell-cell interaction (Baskurt et al. 2011). In the intermediate shear rate region ($\sim 0.01 < \dot{\gamma} <\sim 10\ s^{-1}$) blood experiences shear thinning from disintegration of the rouleaux stacks (Beris et al. 2021). Further shear thinning occurs at higher shear rates ($\dot{\gamma} >\sim 100\ s^{-1}$) due to red blood cell deformation (Beris et al. 2021).

Prior work on this topic focused on two important physiological factors on the steady shear viscosity: the red blood cell volume fraction (hematocrit) and a large (450 Å long and 90 Å wide) plasma protein responsible for the formation of red blood cell rouleaux, known as fibrinogen (Baskurt et al. 2011). The hematocrit is relevant because the red blood cells are by far the dominant particles in blood by volume fraction (36–51%) and size (7–8 μm) and are expected to increase blood viscosity as with any suspension (Wagner and Mewis 2021). As noted, fibrinogen leads to rouleaux network formation and an apparent yield stress, dominating the low shear viscosity. While it is known that many other cardiovascular risk factors in blood—such as cholesterol, triglycerides, and plasma proteins—have been found to influence blood rheology even for healthy individuals (Moreno et al. 2015; Apostolidis and Beris 2016), we first consider only hematocrit and fibrinogen because of their dominant impact on blood rheology in healthy individuals.

Steady shear blood rheology based on literature data (Chien 1975) with and without fibrinogen shows (a) the existence of a non-zero yield stress, manifested by an unbounded viscosity as the shear rate approaches zero, that is dependent on the presence of fibrinogen and (b) overall shear-thinning behavior (Fig. 1). At high shear rates, the blood is shear thinning even after the rouleaux are

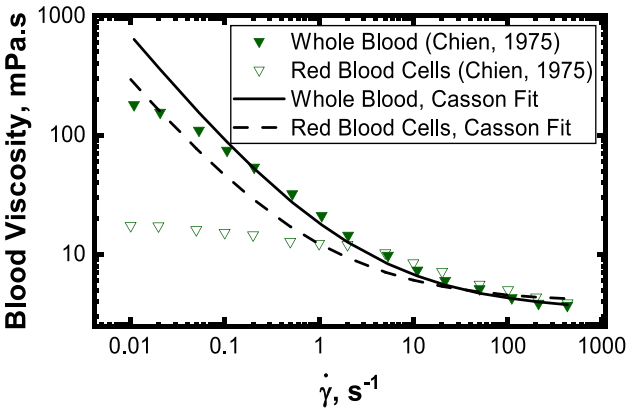

Fig. 1 Casson fit to Chien Data with (closed symbols) and without (open symbols) fibrinogen in log-log coordinates of blood viscosity versus shear rate

disaggregated as the red blood cells deform and shear-align, with both factors depending strongly on hematocrit but not expected to depend on fibrinogen. Fits with the Casson equation, Eq. (1), are also shown (Fig. 1). The Casson equation reproduces the whole blood data reasonably well until shear rates below 0.1 s^{-1}. The description of the RBC only data is poor, indicative of the absence of a viscoplastic behavior for that system. In the following work we will only be using the Casson equation to fit whole blood data.

As mentioned above, steady shear blood rheology is characterized by a shear-dependent (shear-thinning) viscosity. A typical method for clinical assessment of the steady shear blood viscosity is to choose a high shear rate (~ 1000 s^{-1}) and a low shear rate (~ 1 s^{-1}) for measuring it (Nwose and Richards 2011; Choi et al. 2022; Çınar et al. 2022). The two measurements are meant to capture rheological behavior primarily attributed to rouleaux at the low shear rates and the individual cells at the high shear rates. Equivalently, one can transform these two measurements into rheological parameters through a fit using a rheological model that also has two parameters. Such a model is the Casson constitutive equation for steady shear (Casson 1959) with two parameters for yield stress and viscosity. These Casson constitutive model parameters provide for a more systematic and physical representation of rheology than just low and high shear rate viscosities and can be used to predict the steady flow behavior of blood more generally. Furthermore, the ratio of Casson viscosity and yield stress is a characteristic time for the material that defines the shear rate for when rouleaux and viscous contributions to the viscosity are equivalent. These model parameters enable a more direct connection to the physiology than simply reporting viscosities at fixed shear rates as the rouleaux affect the yield stress and the individual cells affect the infinite shear viscosity.

There are two principal methods used to connect the blood rheology with blood physiology: (1) a correlation of constitutive model parameters to physiology through experiments or (2) microscopic simulations that have resolution to the level of individual red cells. The microscopic simulations have some distinct advantages with a more direct route between the involved microstructure and the macroscopic properties. As such, they have provided good results (Fedosov et al. 2011; Hoore et al. 2018; Yazdani et al. 2021; Javadi et al. 2022) when pathological conditions, such as sickle cell anemia, affect parameters such as the shape of red blood cells that can be easily captured in microscopic simulations. However, these simulations still need to include fitted parameters coming from detailed experiments, such as those characterizing interparticle forces acting between blood cells (Korculanin et al. 2021) as well as the role of the many biochemical components in blood that are not included in the simulation but present in blood plasma. The first approach based on experimental measurements, applied in this work, requires extensive, standardized data sets of both rheological properties and corresponding blood physiology from healthy donors. After selection of a constitutive model for describing aspects of blood rheology as well as the independent biochemical factors, parameterization involves nonlinear fitting (Apostolidis and Beris 2014). While some approaches can be guided by rheological and biophysical models, the complexity of the problem quickly becomes daunting when considering complex rheological equations of state and the extensive range of biochemical factors in blood that are known to affect health (Horner et al. 2019; Armstrong et al. 2022). Therefore, as an alternative approach, we restrict this first study to a simple, but robust and well understood constitutive model for the steady shear rheology of blood (Casson model) and limit ourselves at first to the two dominate factors (hematocrit and fibrinogen) to develop a quantitative parameterization using a machine learning framework and the largest existing standardized blood rheometry dataset (referred to as Horner data hereafter) (Horner 2020). This limited study not only serves as an illustration of the method but provides information already useful for hematologists and those studying blood flow more generally. Importantly, adoption of machine learning methods provides a route to scaling up the research to include much more accurate and necessarily complex thixo-elasto-viscoplastic models for dynamic blood flow (Armstrong et al. 2022; Jariwala et al. 2022) and for dramatically increasing the number and range of biochemical markers that may affect dynamic blood rheology. We also wish to note that the macroscopic approach employed here can benefit from results of the microscopic approaches to synergistically inform and test each other.

Early work on blood rheology by the pioneering scientist Ed Merrill at the Massachusetts Institute of Technology in part motivates the objectives presented here. Merrill began by observing the shear rate dependence on the viscosity of blood (Wells and Merrill 1961) and its dependence on the hematocrit (Wells and Merrill 1962). Merrill was interested in testing the limits of the current viscometry technology by searching for the yield stress of blood down to near-zero shear rates (Merrill et al. 1963a, b). The yield stress was defined asymptotically using the Casson relation (see methods), which was originally developed for pigment oils (Casson 1959) and first introduced for use in blood rheology by the industrial rheologist Scott Blair (Blair 1959). Merrill showed the relationship of temperature, hematocrit, and fibrinogen to the Casson yield stress at low flow (Wells and Merrill 1962; Merrill et al. 1963a, b; Merrill et al. 1966). Merrill also correlated the Casson yield stress with endogenous fibrinogen as a quadratic function at a constant hematocrit of 40% (Merrill 1969).

Apostolidis and Beris (2014) took inspiration from the work of Merrill and developed a parameterization of the

Casson equation with hematocrit and fibrinogen using data available at that time taken from various literature sources. The Apostolidis-Beris parameterization follows a more traditional approach, where a model is developed using physical arguments, and then, adjustable parameters are introduced to fit the experimental data. This parameterization found that the Casson viscosity depends only on the hematocrit and temperature linearly while the Casson yield stress depends on hematocrit, fibrinogen, and temperature nonlinearly (Apostolidis and Beris 2014). The data to develop this model was collected from various literature sources including measurements on blood yield stress in isolation to the Casson model (Merrill et al. 1963a, b). The prediction fits the chosen data well and should extend to new data. However, as it is shown in this work, the Apostolidis-Beris yield stress parameterization does not properly describe more recent, extensive experiments performed under standardized conditions. In this work, machine learning allows qualification of that hypothesis and produces a new correlation that remains to be further validated using additional standardized and protocol consistent data. The Apostolidis-Beris parameterization was extended to include transient rheology from the modified Delaware model (Mujumdar et al. 2002) to make predictions in thixotropic systems (Apostolidis et al. 2015). Apostolidis and Beris attempted to include cholesterol and triglycerides into the parameterization of the Casson equation finding that critical ratios of these variables used in cardiovascular disease diagnostics were also important to blood rheology (Apostolidis and Beris 2016).

The Horner dataset established a protocol for standardizing blood rheometry and avoided effects due to extended storage (Horner 2020). The living fluid blood ages ex vivo, altering blood rheology parameters with time from withdrawal. Many studies on blood rheology use data that are not standardized and do not account for time from withdrawal (Horner et al. 2018). The availability of standardized blood rheology measurements has enabled researchers to build well-determined physiology and rheology relationships by overcoming the challenges associated with non-standardized blood metrology (Horner et al. 2019; Armstrong et al. 2022). Horner and coworkers (Horner et al. 2018) established effective procedures for measuring blood rheometry that ensures repeatability and minimizes the ex vivo aging effects. Significant aging effects may occur in as little as four hours from withdrawal; therefore, the Horner protocol calls for blood samples to be loaded on the rheometer within 15–30 min from withdrawal. The blood is stored at ambient temperature to avoid a hysteretic effect from cooling the sample that lasts even after it reaches the measurement temperature (Horner 2020). Aging affects the rheology of blood namely by adenosine triphosphate (ATP) deprivation which is required for red blood cells to maintain their deformability and aggregability (Horner 2020). These effects are most important to the more complex low shear rate behavior of blood leading to thixotropy and syneresis (Horner 2020). To address syneresis in the low shear rate measurements, the protocol extracts the maximum stress that is a result of a transient thixotropic response followed by a continuous decrease from syneresis (Horner 2020). There are also physical measurement considerations within the protocol by never exceeding a shear rate of 1000 s^{-1} to avoid irreversible damage of the red blood cells (Horner 2020). The samples are conditioned between each measurement by subjecting it to 300 s^{-1} for 30 s (Horner 2020). All samples are measured in an ARES-G2 rheometer with the double wall Couette, which resists the effects of margination and sedimentation (Horner 2020). Although both steady and transient data have been collected, we use only the steady shear data. Each sample in this protocol is physiologically characterized with 25 standard biochemical tests of the blood, gender, and age and are publicly available (see Data availability) (Horner 2020).

Engineering and related fields are being transformed by data science and machine learning techniques that aid in work such as the design and discovery of new materials (Ashraf et al. 2021). Biology and medicine have also benefited from machine learning employed to facilitate the extraction of relevant features to a disease (Zitnik et al. 2019). Formulation science is also a useful field to apply data science for developing new formulas that create desired properties, such as optimal drilling mud properties (Ashraf et al. 2021; Magzoub et al. 2021). Machine learning techniques have been used recently in rheological analysis to develop constitutive models for a thixo-elasto-viscoplastic material (Mahmoudabadbozchelou et al. 2021; Saadat et al. 2022) which can have similar complexities to a biological material. This method shows success because of the many components to thixo-elasto-viscoplastic materials that make them challenging or sometimes impossible to model with physics-based equations. One larger aim of the present work is at the interface of biology, rheology, and data science to emerge in developing an understanding of the blood rheological properties for a healthy person using machine learning techniques in a new field termed hemostatistics. *In this work, we provide a proof of concept application of machine learning in the field of blood rheology by using it to parameterize the modeling of the steady-state case in terms of relevant physiological parameters.*

This paper is structured in the following format. In the "Materials and methods" section, the Casson constitutive model is presented and an existing parameterization in terms of hematocrit and fibrinogen is reviewed, along with the relevant blood data from Horner to be used in this study, and finally, an overview of the machine learning technique employed. The results of these techniques are presented in the "Results" section starting with a validation of Gaussian

process regression with synthetic data derived from the Apostolidis-Beris parameterization. Then, the results of Gaussian process regression parameterization are shown with the standardized blood rheometry from Horner data (Horner 2020). The results are discussed in the "Discussion" section and conclusions presented in the "Conclusions" section.

Materials and methods

Constitutive model

The Casson equation (Casson 1959) has been used for correlating the steady shear flow of blood with only two material parameters: yield stress and shear viscosity,

$$\sqrt{\tau} = \sqrt{\tau_{y,C}} + \sqrt{\mu_C \dot{\gamma}}, \tag{1}$$

where τ is the shear stress, $\dot{\gamma}$ is the shear rate, $\tau_{y,C}$ is the Casson yield stress, and μ_C is the Casson viscosity. The parameters of the Casson equation are fit to standardized steady shear data collected by Horner (Horner 2020) using least squares regression. Fitting is done in Casson coordinates as shown in Fig. 2 to calculate the slope and intercept. It is a known problem that the Casson equation does not fit the low shear rate data of blood well because of more complicated behaviors such as viscoelasticity and thixotropy due to rouleaux (Beris et al. 2021). The poor low shear rate fit does not affect the present work because all fits were obtained over a consistent shear rate range. More extensive constitutive equations have been developed to improve the description of the viscoelastic and thixotropic effects at low shear rates (Armstrong et al. 2022), but will not be studied in this work. The equation to calculate the Casson parameters with least squares is as follows,

$$\sqrt{\mu_C} = \frac{N\left(\sum\left(\sqrt{\tau\dot{\gamma}}\right)\right) - \sum\left(\sqrt{\tau}\right)\sum\left(\sqrt{\dot{\gamma}}\right)}{N\left(\sum(\dot{\gamma})\right) - \left(\sum\left(\sqrt{\dot{\gamma}}\right)\right)^2}, \tag{2}$$

and

$$\sqrt{\tau_{y,C}} = \frac{\sum\left(\sqrt{\tau}\right) - \sqrt{\mu_C}\sum\left(\sqrt{\dot{\gamma}}\right)}{N}, \tag{3}$$

where N is the number of data points in the dataset. The Casson equation addresses the steady shear behavior of blood in an approximate fashion, which is not sufficiently accurate at very low shear rates and misses all transient behavior but is useful here as a first approximation and to compare with previous blood flow studies and parameterizations. The fits of all twenty healthy donors to the Casson model are shown in the SI Sec. 1 and the parameters and their uncertainties are summarized in Tables 1 and 2, to be presented and discussed in the "Horner data" section.

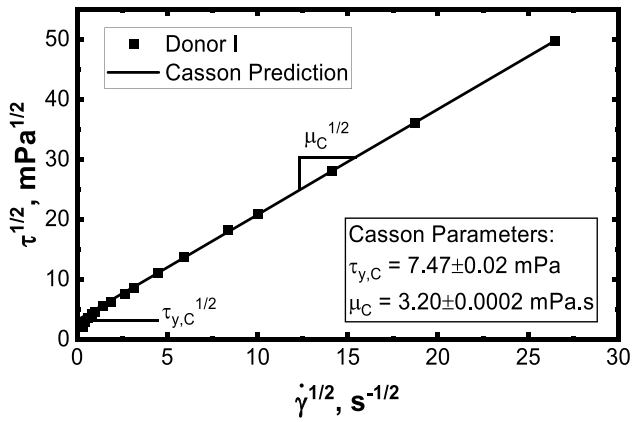

Fig. 2 Least squares fit of steady shear data in Casson coordinates with intercept and slope. The least squares approach calculates a bound on the parameters shown. The data shown is from donor I of the Horner dataset. Fits for all 20 donors are provided in the supplemental information (SI Sec. 1)

Apostolidis and Beris parametric estimation

A parameterization developed by Apostolidis-Beris describes the Casson yield stress and viscosity in terms of hematocrit and fibrinogen (Apostolidis and Beris 2014). The Casson yield stress (dyne/cm^2) and viscosity (dyne s/cm^2) are given as,

$$\tau_{y,C} = \begin{cases} (H - H_c)^2 * (0.5084 c_f + 0.4517)^2, & H > H_c \\ 0, & H \le H_c \end{cases}, \tag{4}$$

And

$$\mu_C = \eta_p\left(1 + 2.0703H + 3.7222H^2\right)\exp\left(-7.0276\left(1 - \frac{T_0}{T}\right)\right), \tag{5}$$

respectively, where H is the hematocrit as a fraction, c_f is the concentration of fibrinogen as g/dL, H_c is the critical hematocrit as a fraction, η_p is the plasma viscosity as dyne s/cm^2, T_0 is a constant reference temperature in Kelvin, and T is the temperature in Kelvin. One dyne s/cm^2 is equivalent to 100 mPa s. The behavior of Casson viscosity is a second-order polynomial depending on hematocrit and an Arrhenius term for temperature adjustment. The Casson yield stress is nonlinear with a piecewise function dependent on a critical hematocrit value. The critical hematocrit is given as,

Table 1. All donors and their Casson viscosity predicted by least squares with the Casson equation, the Apostolidis-Beris parameterization, Gaussian process regression (GPR) with hematocrit and fibrinogen, and GPR with mean corpuscular hemoglobin (MCH) included. The shaded donors are outside the healthy physiological range and not included in training. Values in italics indicate which values for the donors are outside the healthy range

			Casson Viscosity, mPa.s			
Donor	**Hematocrit, %**	**Fibrinogen, mg/dL**	**Least Squares**	**Apostolidis-Beris**	**GPR**	**GPR with MCH**
A	42.6	186	3.0230±0.0004	3.12	3.14±0.179	3.14±0.18
B	41.7	223	3.23265±0.00005	3.06	3.13±0.172	3.12±0.17
C	38.6	249	2.82682±0.00004	2.87	2.93±0.178	2.93±0.18
D	38.9	282	3.03674±0.00005	2.89	3.00±0.176	2.99±0.18
F	35.8	223	2.5428±0.0002	2.71	2.68±0.199	2.73±0.21
G	36.9	271	2.9319±0.0005	2.77	2.83±0.186	2.85±0.19
H	40.8	287	3.3256±0.0003	3.01	3.15±0.172	3.13±0.17
I	41.6	272	3.2022±0.0002	3.06	3.19±0.170	3.17±0.17
J	47	319	3.9964±0.0002	3.41	3.64±0.192	3.65±0.19
K	42.9	142	2.9808±0.0004	3.14	3.10±0.193	3.11±0.19
L	46.2	316	3.3970±0.0003	3.36	3.58±0.187	3.58±0.19
M	43.1	249	3.1308±0.0003	3.15	3.27±0.170	3.26±0.17
N	51.3	247	5.287±0.001	3.71	3.83±0.220	3.89±0.27
O	38.3	333	2.9325±0.0002	2.85	3.02±0.187	3.01±0.19
P	45.2	214	3.3708±0.0002	3.29	3.37±0.176	3.38±0.18
Q	43.6	252	3.0793±0.0002	3.19	3.31±0.170	3.30±0.17
R	43.8	199	3.2398±0.0002	3.20	3.25±0.176	3.25±0.17
S	45.1	210	3.4655±0.0002	3.28	3.36±0.176	3.37±0.18
T	43.4	248	3.0979±0.0002	3.17	3.29±0.170	3.28±0.17
U	43.4	237	3.4155±0.0003	3.17	3.27±0.170	3.27±0.17

Table 2. All donors and their Casson yield stress predicted by least squares with the Casson equation, Apostolidis-Beris parameterization, Gaussian process regression (GPR) with hematocrit and fibrinogen, and Gaussian process regression with mean corpuscular hemoglobin (MCH). The shaded donors are outside the healthy physiological range and not included in training. Values in italics indicate which values for the donors are outside the healthy range

			Casson Yield Stress, mPa			
Donor	**Hematocrit, %**	**Fibrinogen, mg/dL**	**Least Squares**	**Apostolidis-Beris**	**GPR**	**GPR with MCH**
A	42.6	186	5.09±0.04	3.34	8.08±2.58	7.11±1.82
B	41.7	223	5.823±0.005	3.72	8.03±2.53	8.18±1.83
C	38.6	249	9.019±0.003	3.45	7.17±2.61	7.71±1.86
D	38.9	282	9.714±0.005	4.04	7.43±2.59	8.28±1.87
F	35.8	223	4.15±0.02	2.55	6.03±2.81	6.51±2.01
G	36.9	271	4.05±0.05	3.41	6.67±2.70	6.48±1.91
H	40.8	287	8.42±0.03	4.59	8.09±2.54	8.50±2.10
I	41.6	272	7.47±0.02	4.53	8.26±2.52	7.02±1.82
J	47	319	7.76±0.02	7.06	9.85±2.71	7.55±1.90
K	42.9	142	6.95±0.03	2.78	7.89±2.68	13.41±2.18
L	46.2	316	14.33±0.03	6.73	9.68±2.66	13.61±2.29
M	43.1	249	13.80±0.03	4.49	8.58±2.52	10.79±1.95
N	51.3	247	23.5±0.1	6.67	10.18±3.04	8.35±2.07
O	38.3	333	6.14±0.02	4.75	7.45±2.68	6.70±1.94
P	45.2	214	7.56±0.02	4.36	8.93±2.58	7.48±1.85
Q	43.6	252	8.16±0.02	4.67	8.74±2.52	8.54±2.27
R	43.8	199	8.77±0.02	3.79	8.49±2.56	7.78±1.98
S	45.1	210	9.25±0.01	4.27	8.88±2.58	7.20±1.89
T	43.4	248	7.66±0.01	4.54	8.66±2.52	8.31±1.84
U	43.4	237	9.31±0.03	4.35	8.60±2.52	10.80±1.95

$$H_c = \begin{cases} 0.3126c_f^2 - 0.468c_f + 0.1764, c_f < 0.75 \\ 0.0012, c_f \geq 0.75 \end{cases} . \quad (6)$$

When the hematocrit is greater than this critical value, the Casson yield stress exhibits a nonlinear behavior that depends on the square of the difference between hematocrit and critical hematocrit and the square of an adjusted fibrinogen concentration. Below critical hematocrit the Casson yield stress does not exist. The Apostolidis-Beris parameterization finds two uses in this work. First, we develop synthetic datasets by randomly choosing hematocrit and fibrinogen within a healthy range and calculating the Casson properties with the Apostolidis-Beris equations. The synthetic data is then used in the "Validation and evaluation of machine learning with synthetic data" section to train various algorithms to help decide which algorithm would be most suitable to our problem. Second, the parameterization becomes a baseline comparison for our improved parameterization with machine learning and standardized data.

Horner data

The data for every donor in the Horner dataset, including the predicted values from Gaussian process regression and the standard deviation, are in Tables 1 and 2. The healthy range for hematocrit is 36–47% for females and 41–51% for males (Padilla and Abadie 2021). The healthy range for fibrinogen concentration is 150–350 mg/dL for both male and female (Padilla and Abadie 2021). Three donors were excluded from this study because they were outside one of these normal ranges but are useful for testing whether the model that is developed can detect such deviations.

The Apostolidis-Beris parameterization does not show good agreement with the Casson yield stress of Horner data (Horner 2020) because of the non-standardized dataset used for the Apostolidis-Beris regression (Apostolidis and Beris 2014). A plot of the Apostolidis-Beris predicted yield stress versus measured yield stress is presented in the "Approximation method implementation of machine learning" section (Fig. 5). This poor agreement is a consequence of the non-standardized datasets used, which involved multiple geometries, temperatures, shear rates, resuspension techniques, and did not consider time from withdrawal (Apostolidis and Beris 2014). Some measurements were made to extract the yield stress at low shear rates independent of the Casson model, the yield stress in those measurements is not comparable to the Casson yield stress over a larger shear rate region. The time from withdrawal is a crucial aspect to blood rheology because of the ex vivo aging of the living fluid, blood (Horner et al. 2018). Aging will increase the red blood cell density, decrease the surface charge of red blood cells, and alter membrane properties with unpredictable behavior (Horner et al. 2018). These microstructural changes have a significant effect on the rheological properties. These effects on blood rheology emphasize the connection between physiology and rheology must be made with the standardized dataset created by Horner and coworkers.

Machine learning

This work employs supervised machine learning algorithms (Pedregosa et al. 2011) to parameterize the Casson equation with physiology. Supervised learning is a technique where all inputs and outputs are labeled, and a function is generated to predict the outputs. The hyperparameters are adjusted to best fit the training data, and then, the model predicts the testing data, which the model has not been exposed to. The accuracy is assessed through a chosen metric comparing predicted versus actual data for both training and testing, such as root mean squared error or mean absolute percentage error.

The first method for machine learning is through MATLAB to validate the approach with synthetic data made from the Apostolidis-Beris parameterization. The chosen machine learning algorithms are stepwise linear regression, fine tree, support vector machine, Gaussian process regression, and neural network. The commercially available MATLAB framework uses automated methods for hyperparameter tuning to create an appropriate fit of these models. The purpose of this preliminary study with MATLAB is to screen machine learning techniques and decide on a method to further study our data with. MATLAB is a standard available tool to perform a preliminary unbiased analysis exploring the default optimization options. For further analysis, other more extensive tools were used and more systematic interrogation of hyperparameter tuning on the model performance. The Scikit-learn package (Pedregosa et al. 2011) through Python was used for this further analysis with the Horner dataset (Horner 2020). The only algorithm applied in Python uses Gaussian process regression from the Scikit-learn package as this was found through the MATLAB work to be the most successful. Aside from the obvious improvement in terms of the prediction accuracy, Gaussian process regression also offers a standard deviation for each prediction. Belonging to the family of non-parametric Bayesian methods, Gaussian processes have well-defined continuous analytical solutions for the mean and standard deviation functions for the chosen training set (Rasmussen and Williams 2006). There are two strategies to produce results in this work, called approximation and generalization. The approximation method uses all the data ($N = 17$ healthy donors) to produce the best model. In the approximation approach, the hyperparameters of the kernel are obtained through the log-marginal likelihood optimizer in Scikit-Learn. The chosen kernel is the radial basis function plus a white noise term as it was found to have the best generalizability across

donors and produced physically interpretable behaviors. We visualize the predictions as a surface plot to ensure there is no overfitting, but we have no metric to address overfitting in the approximation approach. The approximation is useful for comparing the root mean squared error to the existing parameterization. The generalization method is used to determine the quality of the Gaussian process regression fit by using K-fold cross validation to split the training data into K subsets to validate how the model performs on different subsets of data. The optimal hyperparameters for each subset are calculated with the internal log-marginal likelihood optimizer through Scikit-Learn. The Gaussian process regression model is then compared to the validation data to observe how well the parameterization predicts data unseen in the training. The results are produced with a constant random seed of 1743, the year University of Delaware was founded, to provide a blinded train/validation split.

Results

Validation and evaluation of machine learning with synthetic data

The MATLAB machine learning user interface is used to validate the approach and compare the performance of multiple algorithms for initial screening. The data used in this section was generated by randomly selecting hematocrit and fibrinogen values within the physiological range and calculating yield stress and viscosity from the Apostolidis-Beris parameterization. The algorithms tested are shown in Fig. 3 on a plot of the predicted Casson yield stress versus the Apostolidis-Beris Casson yield stress. Figure 3 qualitatively explains each algorithm's description of the nonlinearities in the Apostolidis-Beris parameterization of Casson yield stress. The worst performing algorithm is the "Fine Tree" method which is more suited for discrete systems instead of continuous ones. Table 3 shows the root mean squared error and R^2 for the training and testing sets of each algorithm. An order of magnitude smaller root mean squared error in Gaussian process regression makes the method a good choice moving forward. We benchmarked the machine learning algorithms in MATLAB and later implemented them in Python for reproducibility and automation.

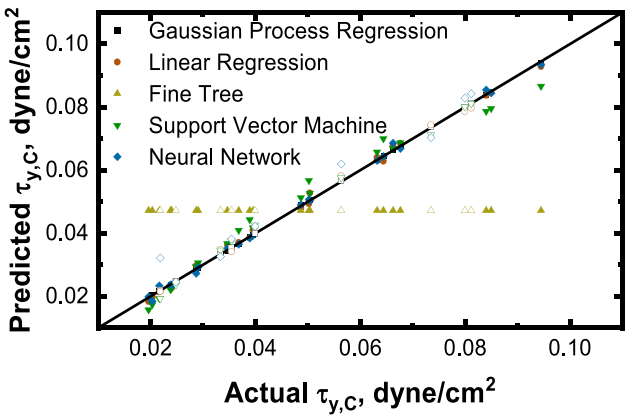

Fig. 3 Predicted versus actual Casson yield stress for various algorithms employed in MATLAB on Apostolidis synthetic data. Closed symbols are data used for training and open symbols are data used for testing

Gaussian process regression on the Horner dataset

Approximation method implementation of machine learning

Gaussian process regression is applied to all available data to describe the Casson parameters with hematocrit and fibrinogen. Figure 4 shows a surface of the Gaussian process regression behavior for (a) Casson yield stress with an R^2 of 0.208 and (b) Casson viscosity with an R^2 of 0.687 according to hematocrit and fibrinogen with one standard deviation. The surfaces are visualized here to understand the two-dimensional input nature of our problem. However, three-dimensional visualizations can be unintuitive for drawing quantitative insight. Therefore, multiple metrics and two-dimensional figures are used to describe the quality of these fits. The residuals of the fits are provided in supplemental information (SI Sec. 2). The predicted values for each donor are in Tables 1 and 2. To summarize these data better, the predicted Casson parameters versus actual Casson parameters are plotted (Fig. 5a, b). The Casson viscosity (Fig. 5b) predictions follow the ideal line well, as demonstrated by the R^2 as well, with some data overpredicted and some data underpredicted within a small tolerance. The Casson yield stress (Fig. 5a) overpredicts the small values in the data and underpredicts the larger values meaning Gaussian process regression found a predictor that is close to the mean and accounts for a large amount of noise.

Table 3 Root mean squared error (RMSE) and R^2 for all algorithms training and testing with Apostolidis synthetic data

Algorithm	Training		Testing	
	RMSE*10^3	R^2	RMSE*10^3	R^2
GPR	1.87	0.99	0.627	1
Linear	2.27	0.98	1.8	0.99
Tree	15.2	0.17	11.3	0.67
SVM	2.27	0.98	2.15	0.99
NN	5.97	0.87	6.29	0.9

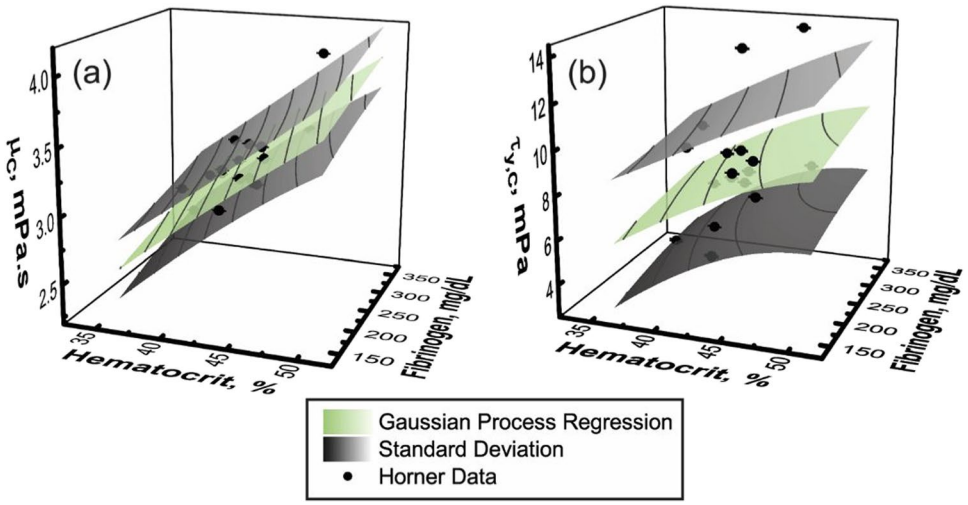

Fig. 4 Approximation of **a** Casson viscosity and **b** Casson yield stress from Horner data with the Gaussian process regression and the standard deviation

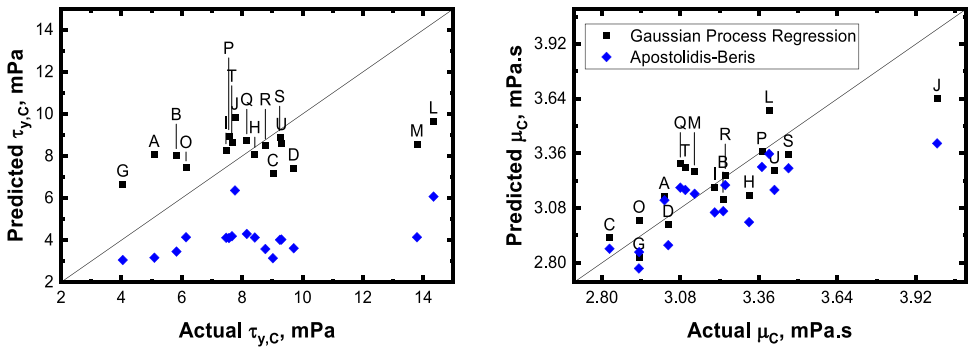

Fig. 5 Predicted versus actual **a** Casson yield stress and **b** viscosity from Gaussian process regression with the approximative approach (black squares) and the Apostolidis-Beris parameterization (blue diamonds). The straight line in these figures indicates full agreement

Generalization method implementation of machine learning

The next step is to understand how well the Horner data physiological parameterization of Casson rheology can be generalized with machine learning. This work uses a K-fold technique for validation where the data is randomly split into folds of data as described previously. The metrics used in the validation are mean absolute percentage error (MAPE) and root mean squared error (RMSE). Similar error in both training and validation verifies that the model is not overfitting and is generalizable to data unseen in training. The exact metrics for training and validation of machine learning depends on the training and validation split of the data; therefore, a K-fold validation technique to split the data "K" times is a more representative metric analysis. The K-fold splitting results are summarized in Table 4 and Table 5.

The Casson viscosity model is generalizable showing both metrics to be on the same order. The Casson yield stress predictions from Gaussian process regression are not as generalizable as Casson viscosity because the validation metrics are greater than the training metrics. Visual inspection of Fig. 4a shows that the prediction surface accounts for noise and does not fit each data point perfectly. The reason for this generalization problem in Casson yield stress is lack of data that are representative of the entire physiological range. Performing a test-train-validation split can bias the model toward the data included in the training set and limit its generalizability to data that lie outside the chosen range.

Application of machine learning to extend physiological inputs

Other physiological characteristics are used to improve the parameterization, mainly the Casson yield stress, found in

Table 4 K-fold validation of the Gaussian process regression generalization performance for Casson yield stress with mean absolute percentage error (MAPE) and root mean squared error (RMSE)

K	MAPE, %		RMSE, mPa	
	Training	Validation	Training	Validation
1	16.6	43.2	2.13	2.96
2	19.9	46.6	1.92	3.11
3	6.63	41.8	0.57	5.45
4	26.4	11.7	2.53	1.13
5	5.75E-3	15.9	5.37E-4	1.51
Average	13.9	31.9	1.43	2.83

Table 5 K-fold validation of the Gaussian process regression generalization performance for Casson viscosity with mean absolute percentage error (MAPE) and root mean squared error (RMSE)

K	MAPE, %		RMSE, mPa s	
	Training	Validation	Training	Validation
1	3.88	3.73	0.164	0.124
2	2.99	7.13	0.111	0.320
3	3.12	7.97	0.125	0.270
4	3.99	3.42	0.154	0.148
5	2.34	4.63	0.101	0.189
Average	3.26	5.38	0.131	0.210

the previous section. Horner data includes a list of many physiology values that could improve the parameterization and demonstrate how machine learning can identify which physiological variables significantly affect the rheology of blood. An additional factor is included with hematocrit and fibrinogen and a new R^2 value for the predictions of Casson yield stress is calculated (Fig. 6). The mean corpuscular hemoglobin (MCH) improved the R^2 significantly more than the other physiological factors. The MCH was found as a direct result of this machine learning study and was not expected to affect the blood rheology a priori.

The additional factor of mean corpuscular hemoglobin (MCH) is used in Gaussian process regression to predict the Casson yield stress in an approximative pattern. Visualizing the dependence of Casson parameters on all the physiological variables is not possible, but the predicted versus actual Casson yield stress is compared (Fig. 7). There is a clear improvement in the predictions of the Casson yield stress compared to Fig. 5a with an R^2 of 0.67, showing a higher quality fit over that considering only hematocrit and fibrinogen ($R^2 = 0.208$).

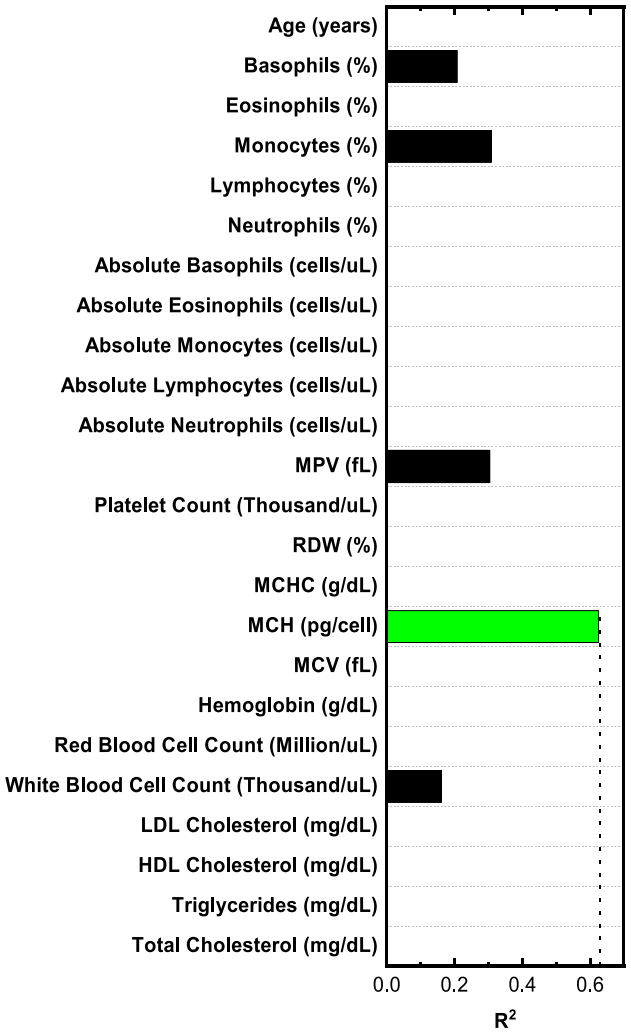

Fig. 6 The R^2 of Casson yield stress prediction when including a single additional factor at a time to the input space. An R^2 of zero in this figure implies that the machine learning algorithm could not generate a meaningful validated correlation. The results for the Casson viscosity are shown in SI Sec. 3

Discussion

Comparison of Gaussian process regression with Apostolidis-Beris

A useful metric to compare two models is the root mean squared error to establish which model is closer to the given data (Ratner 2009; Chicco et al. 2021). The Casson yield stress approximation from Gaussian process regression and Apostolidis-Beris are shown as surface plots (Fig. 8a). The Apostolidis-Beris parameterization surface shows a clear underprediction compared to the Horner data for Casson yield stress with a root mean squared error of 5.70 mPa. The machine learning predictions are more accurate than the Apostolidis-Beris parameterization and improve the root mean squared error to 2.29 mPa accounting for noise in the measurements. The Casson viscosity approximation from Gaussian process regression and Apostolidis-Beris are shown as surface plots (Fig. 8b). The root mean squared error for Casson viscosity is again decreased from 0.401 mPa s with Apostolidis-Beris to 0.151 mPa s with Gaussian process regression. The machine learning predictions of Casson viscosity depends on both hematocrit and fibrinogen, while Apostolidis-Beris only considered hematocrit (Fig. 8b).

To better visualize the variations in Casson yield stress, the Gaussian process regression and the Apostolidis-Beris parameterization are compared with the viscosity flow curve of every donor from the Horner data with the least squares fit to data, Gaussian process regression predictions, and

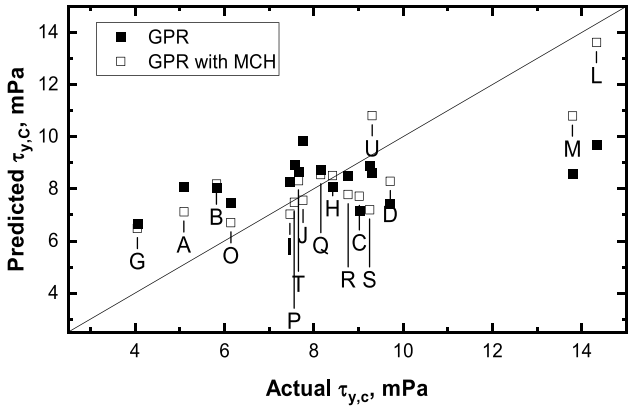

Fig. 7 Predicted versus actual Casson yield stress when using the additional factor of mean corpuscular hemoglobin (MCH) in Gaussian process regression

Apostolidis-Beris predictions (Fig. 9). There is a systematic increase of the low shear viscosity in the predictions from Gaussian process regression over the Apostolidis-Beris parameterization. This systematic increase is due to the data used for the Apostolidis-Beris parameterization, which focused on finding the yield stress at low shear independent of the Casson yield stress. In some instances, the Apostolidis-Beris parameterization more closely predicts the yield stress when compared to the actual data. However, the Gaussian process regression more accurately models the least squares Casson fit to the data, meaning that these predictions are more generalizable. An important insight from this analysis is that the Casson yield stress is not a physically realizable parameter where there are significant deviations at low shear rates from the true behavior (Fig. 9). These deviations combined with the rich dynamic transients of blood necessitates using a more detailed rheological constitutive model such as the tensorial-Enhanced Structural Stress Thixotropic Viscoelastic (t-ESSTV) one developed by Armstrong and coworkers (Armstrong et al. 2022). In future work, the machine learning approach will be used to parameterize the t-ESSTV model with physiology.

In this section the Casson parameter parameterizations are shown to improve with machine learning, especially the Casson yield stress. The predictions provided in this section show a good approximation to all provided data and they are better than the existing parameterization. However, the Casson viscosity dominates the overall behavior of the Casson flow curve. The data required to make this prediction generalizable is discussed in the next section.

Data requirements for generalization

The typical data requirement for machine learning is ten times the input dimension. The lack of data for the multidimensional physiology space is relevant to two aspects of this study which should be addressed in future work. Overfitting may occur when there are not enough features, i.e., physiology, to describe the targets, i.e., Casson rheology. Including cholesterol, triglycerides, or some other physiology into the fit may improve the results. However, more inputs require more data to obtain accurate predictions. The limitation to the connection of physiology and Casson rheology is that the Casson constitutive model is not an appropriate choice for blood rheology at lower shear rates. The Casson equation typically over predicts the stress at low shear rates and deviates significantly from the true yield stress. This deviation means the Casson yield stress does not always scale with fibrinogen as expected. A perfect example of this scaling issue can be observed in Table 2 where donors T and U have the same hematocrit and donor T has a higher fibrinogen concentration, yet the Casson yield stress of donor T is lower than that of donor U. This result shows the limitations of the current approach with machine learning the

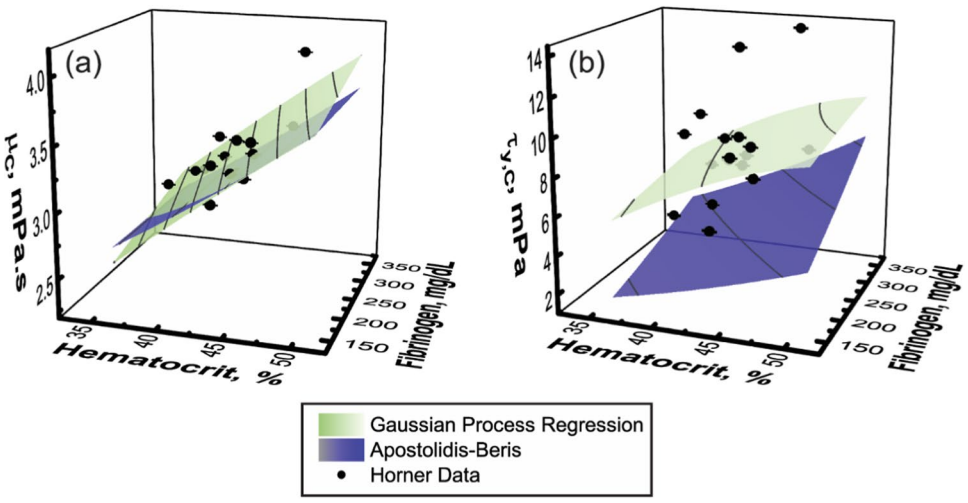

Fig. 8 Gaussian process regression and Apostolidis-Beris parameterization predictions for **a** Casson viscosity and **b** Casson yield stress as a function of hematocrit and fibrinogen

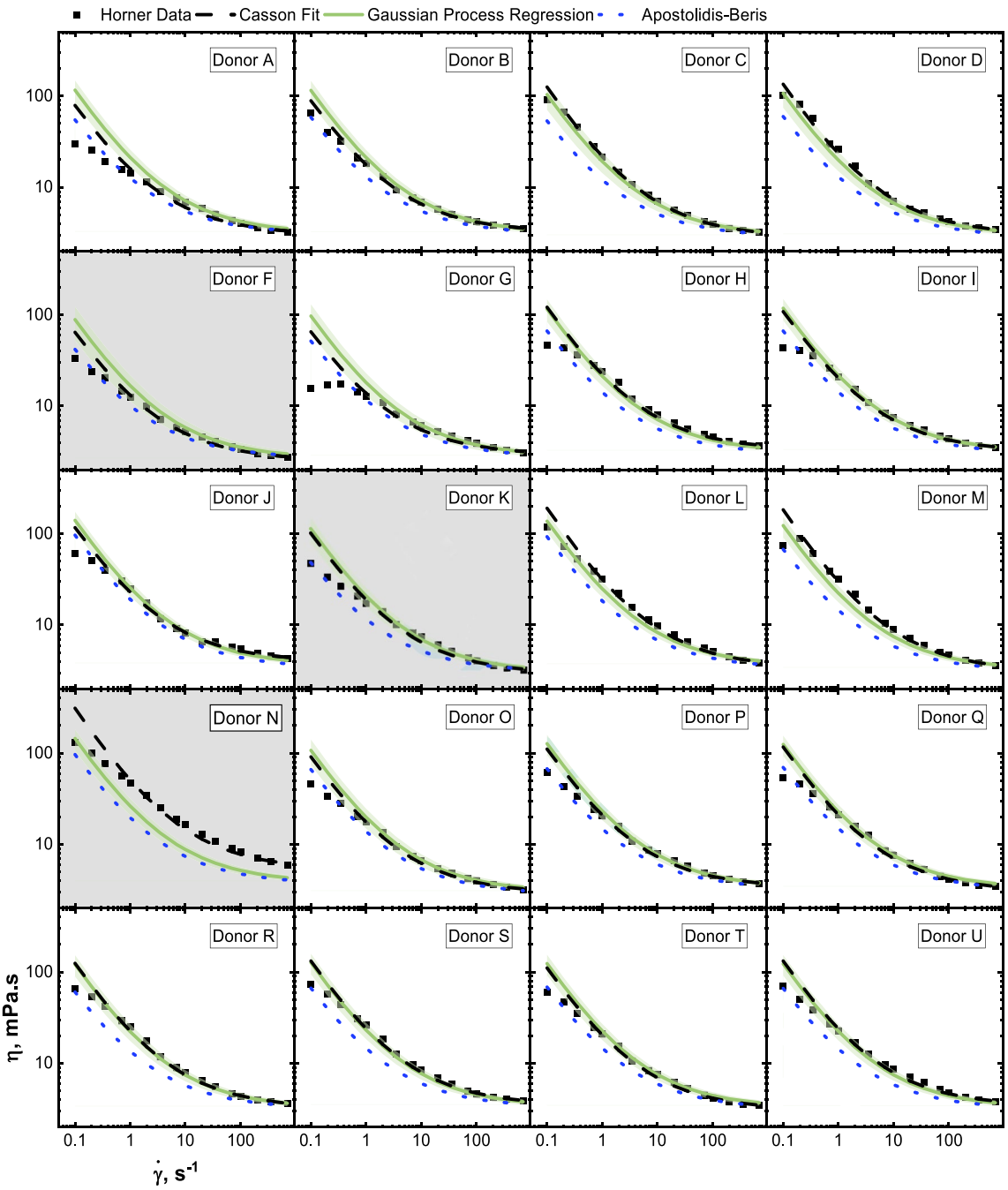

Fig. 9 Viscosity flow curve for every donor from the Horner dataset with Casson predictions from least squares (dashed line), Gaussian process regression (solid line), and the Apostolidis-Beris parameterization (dotted line). The gray plots are donors that were not included in training because they are outside healthy human range of hematocrit or fibrinogen

connection between physiology and Casson rheology. More data is required for a fully confident model, but the current approach has shown promise using the largest existing dataset. These issues are expected to resolve with more experiments or with the development of a high-throughput device for measuring blood rheology.

Although the Gaussian process regression algorithm is improved from the Apostolidis-Beris predictions, the model is still not sufficiently accurate for practical application as demonstrated from the K-fold validation for Casson yield stress. The connection of hematocrit, fibrinogen, and Casson yield stress through Gaussian process regression proved to

be less generalizable than Casson viscosity although measures, such as K-fold validation, were taken to prevent such a situation. This overfitting issue demonstrates why more data is required to include more aspects of physiology to develop a more robust connection between rheology and physiology. Despite the limitations in dataset size, the preliminary identification of MCH as the next most important physiology to the Casson yield stress is discussed in the next section.

Mean corpuscular hemoglobin

Systematically incorporating physiological variables one at a time, as done in the "Application of machine learning to extend physiological inputs" section, leverages the power of the machine learning framework toward discovering important factors in determining the rheology. In the course of this work, MCH was identified as an important factor controlling the Casson yield stress. MCH has not been used in previous modeling studies, although there are some literature sources that indicate hemoglobin's influence on blood rheology (Gustavsson et al. 1994; Coppola et al. 2000). One work demonstrates a positive correlation of blood viscosity with hemoglobin concentration at various shear rates without a constitutive model in both healthy blood and, to a lesser effect, in the blood with coronary artery disease (Gustavsson et al. 1994). The correlation found here with MCH is different because MCH describes the amount of hemoglobin per cell, whereas the hemoglobin concentration is per volume. The hemoglobin affects the density and size of the red blood cells. In this work, we observe an inverse correlation between the Casson yield stress and MCH, which is also observed in the literature (Hutton 1979, Von de Pette et al. 1986). We speculate that a higher MCH value reduces the tendency to form rouleaux, thus decreasing the yield stress. Further detailed studies are required to address the effect of MCH on the cytosol viscosity and how this affects the bulk rheology. We expect that a stiffer red blood cell increases the modulus and decreases the extent of shear thinning. Such a correlation may be physically rational as MCH indirectly describes the size and density of the red blood cells. If we compared two donors with the same hematocrit and different MCH, the one with a higher MCH will have fewer or denser red blood cells, which means less rouleaux structure and a lower yield stress. This analysis may explain the discrepancy noted for donor T as compared to donor U, which had the same hematocrit, but the Casson yield stress behaved inversely from the expected behavior based on the fibrinogen values. Donor T has a higher MCH, and this additional factor could potentially result in a lower Casson yield stress. Hutton (1979) was the first to recognize a significant correlation of whole blood viscosity to iron deficiency (i.e., decreasing MCH) and Van de Pette et al. (1986) confirmed the results. To our knowledge, there is no recent results on the inverse correlation of MCH and blood viscosity and in this work the correlation is done with a constitutive model, which suggests the effect occurs mainly at low shear rates with the Casson yield stress.

An important caveat to this result is that it is only an indication of a correlation between blood rheology and MCH. The Horner dataset is the largest and most detailed dataset of its kind on blood rheology, but there is not enough data to be certain of the three factor (hematocrit, fibrinogen, and MCH) parameterization found in the "Application of machine learning to extend physiological inputs" section. For a statistical basis of the two parameterizations presented in this work, the χ^2 test and the F-test for variability were performed (Table 6). The χ^2 parameter decreases with each parameterization showing the improvement of the fit. The null hypothesis of the F-test states that the variances of two populations are equivalent. Therefore, a p-value below 0.05 rejects the null hypothesis and declares that the chosen regression model significantly reduced the variability of the data.

The results shown in Table 6 indicate that both parameterizations significantly reduce the variability of the Casson viscosity and that the three parameter model significantly reduces the variability of the Casson yield stress, while the two parameter model does not. The addition of MCH certainly improves the mean squared error of the predictions with the Horner dataset; however, the correlation could be a statistical artifact and more data is needed to investigate the underlying physical basis. A common heuristic for machine learning accuracy is to require ten times the input dimension worth of data; therefore, 30 donors would be needed for an accurate parameterization with three physiological factors. However, this method will be valuable when determining which factors are most significant for the ten parameters in the t-ESSTV constitutive model. We suggest the addition of MCH into microscopic simulations in an indirect way by modifying the size and density of the red blood cells. Microscopic simulations could provide insight on the correlation of MCH and Casson yield stress found in this work.

Table 6 Statistical analysis for the two parameterizations of the Casson yield stress and Casson viscosity against the mean

Parameterization	χ^2 Casson yield stress	F statistic	p
Mean	13.4	0	1
H & c_f	10.5	1.93	0.104
H, c_f, & MCH	4.58	9.82	0.0000265
	Casson viscosity		
Mean	0.374	0	1
H & c_f	0.112	11.6	7.51E-06
H, c_f, & MCH	0.112	12.5	5.19E-06

Personalized healthy range

The chief benefit of a connection between blood physiology and rheology is expected to be in diagnostics. To use blood rheology as a diagnostic, the healthy range of rheology parameters must be established and for personalized medicine; this should be known for an individual. The average value of blood viscosity is not important in diagnostics because of the variation among donors. We use the parameterization developed here to reduce this variation and improve personalized diagnostics. This work does not provide a tight correlation to a specific disease, rather a healthy range of blood rheology parameters that are unique among patients. Predicting Casson parameter dependencies on hematocrit and fibrinogen was the focus of previous sections, but as will be shown here, the resultant model predictions with standard deviation can be used to determine a healthy range of Casson parameters personalized to an individual patient. It is important to include the standard deviation calculated by Gaussian process regression so there is a bound on the predictions. A bound on the rheology predictions emulates the healthy ranges used in standard physiological diagnostics. In this section, the machine learning model is used to predict some data that was not included in training or testing for diagnostic purposes.

A physician could implement the following analysis to use this blood rheology parameterization in clinical diagnostics. The surface from Gaussian process regression approximation predictions is shown with an upper and lower standard deviation (Fig. 10). There are three donors plotted (Fig. 10) which were excluded from training because they are outside the healthy ranges of physiology reported by Merck (Padilla and Abadie 2021). The steady shear behavior of donors F, K, and N is shown in the shaded regions of Fig. 9. The low hematocrit (donor F) and low fibrinogen (donor K) donors are close to the machine learning healthy range. The high hematocrit donor (donor N) is more interesting because the prediction is many standard deviations below the actual Casson yield stress. This observation displays how the Gaussian process regression predictions and standard deviations are valuable for diagnostics. Since the Casson yield stress for this donor is far greater than what the healthy range model predicts, there is likely some other factor, such as disease, contributing to the rheology. The model has few data points and from the overfitting problem observed previously it is likely that another factor of physiology is responsible for deviations in the blood rheology. When the model is trained with more donors to account for other physiological factors, the deviations from the predicted healthy range may prove to represent effects from cardiovascular disease or drugs, such as stroke, statins, or aspirin. The results of this initial study suggest that the machine learning framework presented here along with high fidelity data provides a promising route for developing a rheological-based screening diagnostic of cardiovascular diseases or the influence of drugs.

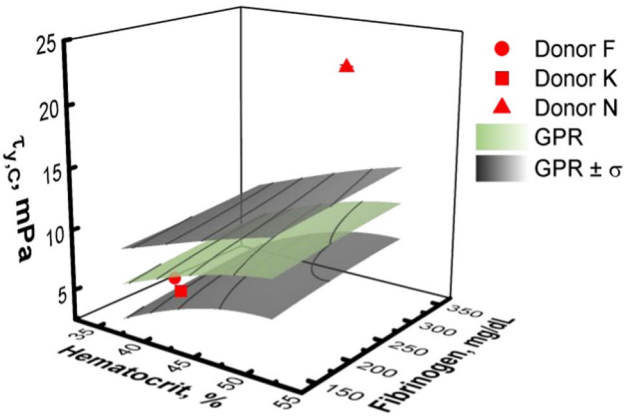

Fig. 10 Approximation of the personalized healthy range of Casson yield stress predicted using Gaussian process regression. Three data points not included in training are presented because they are outside healthy physiological range

Conclusions

The present work demonstrates the usefulness of a machine learning approach for both developing a connection between physiology and rheology of a standardized blood rheometry dataset within the context of the Casson model for steady blood flow as well as for identifying additional physiological factors of interest. Through training and comparison against synthetic data, following the Apostolidis and Beris parameterization, the Gaussian process regression method is found to be the most effective machine learning algorithm for our purpose. Comparison of the Apostolidis-Beris parameterization with more extensive and recent data sets of Horner (Horner 2020) shows that the prior correlation underestimated the Casson yield stress due to the use of non-standard input data. Using the machine learning algorithm, we provide an improved correlation between the Casson yield stress and the Casson viscosity and the blood physiological properties of hematocrit and fibrinogen, available to the community in the form of the machine learning algorithm (see Data availability) and graphical correlation. Furthermore, we demonstrate how the new model has potential to be used in personalized medicine. Another outcome of the work is the identification of mean corpuscular hemoglobin (MCH) as an additional physiological factor affecting the Casson yield stress although more data are needed for a positive confirmation. The results of this study suggests a promising route for the development of a broader and more inclusive connection between blood rheology and physiology that may find application in diagnostics as well as modeling of blood

flow using more complex blood rheology constitutive models and a broader range of physiological properties.

Acknowledgements We would like to acknowledge the NRT Midas program at the University of Delaware for assistance with data science in soft materials. Funding from the National Science Foundation (NSF) through the award CBET 1804911 is also gratefully acknowledged.

Data availability A GitHub repository contains the code used for predictions made in this paper at: https://github.com/FarringtonSM/Hemostatistics_BloodRheology_and_MachineLearning. The data (Horner 2020) used in this work is found at: https://sites.udel.edu/wagnergroup/files/2016/06/Horner-Thesis-Data.xlsx

Declarations

Ethical approval This article does not contain any studies with human or animal participants performed by any of the authors.

Conflict of interest The authors declare no competing interests.

References

Apostolidis AJ, Armstrong MJ, Beris AN (2015) Modeling of human blood rheology in transient shear flows. J Rheol 59(1):275–298
Apostolidis AJ, Beris AN (2014) Modeling of the blood rheology in steady-state shear flows. J Rheol 58(3):607–633
Apostolidis AJ, Beris AN (2016) The effect of cholesterol and triglycerides on the steady state shear rheology of blood. Rheol Acta 55(6):497–509
Armstrong M, Pincot A, Jariwala S, Horner J, Wagner N, Beris A (2022) Tensorial formulations for improved thixotropic viscoelastic modeling of human blood. J Rheol 66(2):327–347
Ashraf C, Joshi N, Beck DAC, Pfaendtner J (2021) Data science in chemical engineering: applications to molecular science. Annu Rev Chem Biomol Eng 12(1):15–37
Baskurt O, Neu B, Meiselman HJ (2011) Red Blood Cell Aggregation. CRC Press
Baskurt OK, Meiselman HJ (2008) RBC aggregation: more important than RBC adhesion to endothelial cells as a determinant of in vivo blood flow in health and disease. Microcirculation 15(7):585–590
Benjamin RJ, McLaughlin LS (2012) Plasma components: properties, differences, and uses. Transfusion 52:9S–19S
Beris AN, Horner JS, Jariwala S, Armstrong MJ, Wagner NJ (2021) Recent advances in blood rheology: a review. Soft Matter 17(47):10591–10613
Blair GWS (1959) An equation for the flow of blood, plasma and serum through glass capillaries. Nature 183(4661):613–614
Casson N (1959) A flow equation for pigment-oil suspensions of the printing ink type. Rheol Dis Sys:84–104
CDC. (2022). "Heart disease facts." Retrieved 03/01/2022, from https://www.cdc.gov/heartdisease/facts.htm.
Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7:e623
Chien S (1975) Biophysical behavior of red cells in suspensions. Red Blood Cell 2:1031–1133
Chien S (1986) Blood rheology in myocardial-infarction and hypertension. Biorheology 23(6):633–653
Choi D, Waksman O, Shaik A, Mar P, Chen Q, Cho DJ, Kim H, Smith RL, Goonewardena SN, Rosenson RS (2022) Association of blood viscosity with mortality among patients hospitalized with COVID-19. J Am Coll Cardiol 80(4):316–328
Çınar T, Hayıroğlu Mİ, Selçuk M, Çiçek V, Doğan S, Kılıç Ş, Yavuz S, Babaoğlu M, Uzun M, Orhan AL (2022) Association of whole blood viscosity with thrombus presence in patients undergoing transoesophageal echocardiography. Int J Cardiovasc Imaging 38(3):601–607
Connes P, Alexy T, Detterich J, Romana M, Hardy-Dessources MD, Ballas SK (2016) The role of blood rheology in sickle cell disease. Blood Rev 30(2):111–118
Coppola L, Caserta F, De Lucia D, Guastafierro S, Grassia A, Coppola A, Marfella R, Varricchio M (2000) Blood viscosity and aging. Arch Gerontol Geriatr 31(1):35–42
Dintenfass L (1974) Blood rheology as diagnostic and predictive tool in cardiovascular diseases: effect of ABO Blood groups. Angiology 25(6):365–372
Fedosov DA, Pan WX, Caswell B, Gompper G, Karniadakis GE (2011) Predicting human blood viscosity in silico. Proceedings of the National Academy of Sciences of the United States of America 108(29):11772–11777
Fryar CD, Chen TC, Li X (2012) Prevalence of uncontrolled risk factors for cardiovascular disease: United States, 1999-2010. NCHS Data Brief 103
Gahleitner M (1999) Rheology as a quality control instrument. J Macromol Sci A 36(11):1731–1741
Gustavsson CG, Persson SU, Larsson H, Persson S (1994) Blood-viscosity in relation to blood hemoglobin concentration in healthy-subjects and in patients with different cardiovascular-diseases. Clin Hemorheol 14(5):677–683
Hitsumoto T (2017) Relationship between cardiovascular risk factors and hemorheology assessed by microchannel method in patients with type 2 diabetes mellitus. Diabetol Int 8:316–322
Hoore M, Yaya F, Podgorski T, Wagner C, Gompper G, Fedosov DA (2018) Effect of spectrin network elasticity on the shapes of erythrocyte doublets. Soft Matter 14(30):6278–6289
Horner JS (2020) An experimental and theoretical investigation of blood rheology. University of Delaware
Horner JS, Armstrong MJ, Wagner NJ, Beris AN (2019) Measurements of human blood viscoelasticity and thixotropy under steady and transient shear and constitutive modeling thereof. J Rheol 63(5):799–813
Horner JS, Beris AN, Woulfe DS, Wagner NJ (2018) Effects of ex vivo aging and storage temperature on blood viscosity. Clin Hemorheol Microcirc 70(2):155–172
Hutton RD (1979) The effect of iron deficiency on whole blood viscosity in polycythaemic patients. British J Haematol 43(2):191–199
Jariwala S, Wagner NJ, Beris AN (2022) A thermodynamically consistent, microscopically-based, model of the rheology of aggregating particles suspensions. Entropy 24(5):717
Javadi E, Li H, Gallastegi AD, Frydman GH, Jamali S, Karniadakis GE (2022) Circulating cell clusters aggravate the hemorheological abnormalities in COVID-19. Biophys J 121(18):3309–3319
Korculanin O, Kochetkova T, Lettinga MP (2021) Competition between red blood cell aggregation and breakup: depletion force due to filamentous viruses vs. shear flow. Front Phys 9:721368
Kyrle P, Eichinge S (2005) Deep vein thrombosis. Lancet 365:1163–1174
Le Devehat C, Vimeux M, Khodabandehlou T (2004) Blood rheology in patients with diabetes mellitus. Clin Hemorheol Microcirc 30(3-4):297–300
Lemonne N, Lamarre Y, Romana M, Hardy-Dessources MD, Lionnet F, Waltz X, Tarer V, Mougenel D, Tressieres B, Lalanne-Mistrih ML, Etienne-Julan M, Connes P (2014) Impaired blood rheology plays a role in the chronic disorders associated with sickle cell-hemoglobin C disease. Haematologica 99(5):74–75
Lowe G, Rumley A, Norrie J, Ford I, Shepherd J, Cobbe S, Macfarlane P, Packard C (2000) Blood rheology, cardiovascular risk factors,

and cardiovascular disease: the West of Scotland Coronary Prevention Study. Thromb Haemost 84(10):553–558
Magzoub MI, Kiran R, Salehi S, Hussein IA, Nasser MS (2021) Assessing the relation between mud components and rheology for loss circulation prevention using polymeric gels: a machine learning approach. Energies 14(5):1377
Mahmoudabadbozchelou M, Caggioni M, Shahsavari S, Hartt WH, Em Karniadakis G, Jamali S (2021) Data-driven physics-informed constitutive metamodeling of complex fluids: a multifidelity neural network (MFNN) framework. J Rheol 65(2):179–198
MayoClinic. (2021). "Heart disease: diagnosis & treatment." Retrieved 07/01/2022, from https://www.mayoclinic.org/diseases-conditions/heart-disease/diagnosis-treatment/drc-20353124.
Merrill EW (1969) Yield stress of normal human blood as a function of endogenous fibrinogen. J Appl Physiol 26(1):1–3
Merrill EW, Gilliland ER, Cokelet G, Shin H, Britten A, Wells RE (1963a) Rheology of human blood, near and at zero flow: effects of temperature and hematocrit level. Biophys J 3(3):199–213
Merrill EW, Wells RE, Cokelet GC, Britten A (1963b) Non-Newtonian rheology of human blood-effect of fibrinogen deduced by subtraction. Circ Res 13(1):48
Merrill EW, Gilliland ER, Lee TS, Salzman EW (1966) Blood rheology: effect of fibrinogen deduced by addition. Circ Res 18(4):437–446
Moreno L, Calderas F, Sanchez-Olivares G, Medina-Torres L, Sanchez-Solis A, Manero O (2015) Effect of cholesterol and triglycerides levels on the rheological behavior of human blood. Korea Aust Rheol J 27(1):1–10
Mujumdar A, Beris AN, Metzner AB (2002) Transient phenomena in thixotropic systems. J Nonnewton Fluid Mech 102(2):157–178
Nwose EU, Richards RS (2011) Whole blood viscosity extrapolation formula: Note on appropriateness of units. North American Journal of Medical Sciences 3(8):384–386
Padilla O, Abadie J (2021) Blood tests: normal values. https://www.merckmanuals.com/professional/resources/normal-laboratory-values/blood-tests-normal-values. Accessed 09/01/2022
Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
Rasmussen CE, Williams CKI (2006) Gaussian Processes for Machine Learning. The MIT Press
Ratner B (2009) The correlation coefficient: Its values range between +1/−1, or do they. J Target Meas Anal Mark 17(2):139–142
Rosenson RS, Wolff D, Green D, Boss AH, Kensey KR (2004) Aspirin. J Thromb Haemost 2(2):340–341
Saadat M, Mahmoudabadbozchelou M, Jamali S (2022) Data-driven selection of constitutive models via rheology-informed neural networks (RhINNs). Rheol Acta 61(10):721–732
Tabesh H, Poorkhalil A, Akbari H, Rafiei F, Mottaghy K (2022) Phenomenological characterization of blood's intermediate shear rate: a new concept for hemorheology. Phys Eng Sci Med
Torpy JM, Glass TJ, Glass RM (2007) Retinopathy. JAMA 298(8):944
Von de Pette JEW, Guthrie DL, Pearson TC (1986) Whole blood viscosity in polycythaemia: the effect of iron deficiency at a range of haemoglobin and packed cell volumes. British J Haematol 63(2):369–375
Wagner NJ, Mewis J (2021) Theory and Applications of Colloidal Suspension Rheology. Cambridge University Press, Cambridge
Wells RE, Merrill EW (1961) Shear rate dependence of the viscosity of whole blood and plasma. Science 133(3455):763–764
Wells RE, Merrill EW (1962) Influence of flow properties of blood upon viscosity-hematocrit relationships*. J Clin Investig 41(8):1591–1598
Yazdani A, Deng Y, Li H, Javadi E, Li Z, Jamali S, Lin C, Humphrey JD, Mantzoros CS, Em Karniadakis G (2021) Integrating blood cell mechanics, platelet adhesive dynamics and coagulation cascade for modelling thrombus formation in normal and diabetic blood. J R Soc Interface 18(175):20200834
Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM (2019) Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. Inf Fusion 50:71–91