OPEN

Machine Learning Analysis for Quantitative Discrimination of Dried Blood Droplets

Lama Hamadeh[1]*, Samia Imran[1], Martin Bencsik [1], Graham R. Sharpe[2], Michael A. Johnson[2] & David J. Fairhurst [1]

One of the most interesting and everyday natural phenomenon is the formation of different patterns after the evaporation of liquid droplets on a solid surface. The analysis of dried patterns from blood droplets has recently gained a lot of attention, experimentally and theoretically, due to its potential application in diagnostic medicine and forensic science. This paper presents evidence that images of dried blood droplets have a signature revealing the exhaustion level of the person, and discloses an entirely novel approach to studying human dried blood droplet patterns. We took blood samples from 30 healthy young male volunteers before and after exhaustive exercise, which is well known to cause large changes to blood chemistry. We objectively and quantitatively analysed 1800 images of dried blood droplets, developing sophisticated image processing analysis routines and optimising a multivariate statistical machine learning algorithm. We looked for statistically relevant correlations between the patterns in the dried blood droplets and exercise-induced changes in blood chemistry. An analysis of the various measured physiological parameters was also investigated. We found that when our machine learning algorithm, which optimises a statistical model combining Principal Component Analysis (PCA) as an unsupervised learning method and Linear Discriminant Analysis (LDA) as a supervised learning method, is applied on the logarithmic power spectrum of the images, it can provide up to 95% prediction accuracy, in discriminating the physiological conditions, i.e., before or after physical exercise. This correlation is strongest when all ten images taken per volunteer per condition are averaged, rather than treated individually. Having demonstrated proof-of-principle, this method can be applied to identify diseases.

The generation of complex and varied patterns as a result of the liquid drying process is a common yet intriguing phenomenon in nature[1]. One of the first studies of drying droplets began with the publication in 1997 by Deegan *et al.*[2] with an explanation of the formation of a ring-like structure commonly called a 'coffee ring' resulting from the evaporation of a droplet containing microparticles[3]. The reason for the formation of such a pattern is that the contact line, where the droplet meets the substrate, is pinned in place, due to the particles in the liquid. Consequently, liquid from the center of the droplet must flow outwards to replenish the liquid that evaporates at the rim bringing the particles with it[4]. The generic form of the coffee ring has been investigated and analysed by many researchers using different liquids including coffee[2,5,6], nanofluids[7], polymers[8,9] and DNA[10]. Interestingly, in the past few decades, the analysis of patterns from dried droplets of biological fluids[11–13] has gained a lot of attention due to applications in fields such as biomedical[14–16] and forensic sciences[16,17], with some reports of successful medical applications of the Litos test for analysing urine droplets[18]. Biological liquids, such as blood, are complex systems containing various components, including macromolecules and cells. When a droplet of blood is placed on a solid substrate to dry, a band of darker red forms at the periphery around the rim, much like the coffee ring. Inside this is a zone called the corona, which often contains clear radial cracks. The central zone is usually paler in colour and contains smaller, more randomly oriented cracks. There can be fractionation of the blood components between the regions of the droplet due to their size or mobility[19]. Among the many dried droplet patterns of biological liquids, those left by dried blood droplets have been investigated in some detail[20,21]. The overall

[1]Department of Physics and Mathematics, School of Science and Technology, Nottingham Trent University, Nottingham, Clifton Campus, NG11 8NS, United Kingdom. [2]Exercise and Health Research Group, Sport, Health and Performance Enhancement (SHAPE) Research Centre, School of Science and Technology, Nottingham Trent University, Clifton Campus, NG11 8NS, United Kingdom. *email: Lama.hamadeh@ntu.ac.uk

dynamics of evaporation is often described by five separate stages[21,22]. Researchers have also investigated crack formation[19,21,23], showing how the number of cracks increases with droplet diameter as well as the effect of substrate[24], humidity[25,26], evaporation rate[4], surface roughness and wettability[16,27] on the final pattern. Recently Smith *et al.*[28] have shown that the drying mechanism of a blood droplet which is dropped onto a surface is influenced by the impact energy, which increases the droplet diameter and redistributes red blood cells within the droplet. Interestingly, blood samples have also consistently proven to be a key source for a wide variety of diagnostic and research purposes. Several studies showed that the patterns observed in dried blood droplets carry important information about the health status of humans and can be used for disease diagnosis[14,21,29]. Although similar drying processes occur during the drying of blood and colloidal suspensions, the variety, complexity and interactions between the various components of a blood droplet mean that much richer patterns are formed including radial and tangential cracks, colour variations, the formation of lobed fingers, fractionation of components and in some cases even spiral cracks. These various features must depend on the physical properties of the blood, which are in turn linked to the health condition of the person, for example to whether the person is healthy or suffers from anaemia or hyperlipidaemia[16]. Linking the observed patterns to the blood chemistry is a complex, multi-faceted challenge, but one that could lead to improvements in diagnosis of disease and other health conditions. Even though these methods have enabled researchers to identify distinguishing patterns in dried blood droplets, objective assessment systems are needed to remove subjective evaluation[20,30]. Recently, some technologies have been designed to overcome human subjectivity, such as acoustical-mechanical impedance (AMI)[31]. However there remains much scope for more stable and accurate technologies to be developed[15,20,32]. In this paper we develop sophisticated image processing routines and a highly-accurate machine learning algorithm to quantitatively and objectively discriminate the patterns from a large number of dried blood droplets, removing much of the subjectivity inherent in previous studies.

Increasingly sophisticated digital image processing algorithms have evolved to exploit the most informative features contained within an image. Over the years, methods have been developed to reduce variations in image quality[33], for example, through spatial filtering[34], edge detection[35] and various types of interpolation; of which bi-linear interpolation[36], cubic convolution[37] and cubic spline interpolation[38] are the most well-known and widely used. Moreover, in many applications, the primary interest is not in specific features of individual images, but rather in the image texture or the degree of regularity governing a pattern formed by multiple images. Such textures are most conveniently analysed not in the spatial domain but in the spatial frequency domain[39]. A very common method used in describing the distribution of an image's pixel intensities as a function of space and frequency is the power spectrum[40,41]. Reducing or excluding the high-frequency components while preserving or choosing the low-frequency components helps in not only extracting and exhibiting the most useful information in an image but also reducing a large amount of noise.

The extraction of implicit, previously unknown, and potentially useful information from image data by using computer programs that automatically sift through a database, seeking regularities or patterns[42] has recently become a popular and effective practice to discover real trends in raw data. One of the tools used to help with such extractions is machine learning[43–46]. In general terms, machine learning is a set of tools that allows users to teach computers how to perform tasks by providing examples of how they should be done. The increasing interest in machine learning methods, driven by the desire to discover trends or irregularities in huge databases, has led to many successful applications of machine learning in various domains, taking advantage of the development of robust and efficient algorithms to process data and the falling cost of computational power[47–50]. Particularly, machine learning algorithms have been used and developed extensively in the domain of image analysis[51,52]. Blood images are governed by a large number of physiological and environmental variables some of which might be correlated. These correlations bring about a redundancy in the information that can be gathered by the dataset of the images. One approach to coping with redundancy is to reduce the dimensionality of the problem by combining data from correlated features. The success of reducing dimensionality is enabled by (i) significant performance gains in computational speed and memory and (ii) the generation of physically interpretable spatial modes that are linked to the underlying physics[53]. Linear combinations between variables are particularly attractive because they are simple to compute and analytically tractable. One of the most ubiquitous methods in dimensionality reduction is principal component analysis (PCA)[54]. PCA is mathematically defined as an orthogonal linear transformation that reconstructs the data to a new coordinate system such that the greatest variance by some projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on[55]. Furthermore, classification methods are often performed in a tailored low dimensional basis extracted as hierarchical features of the data[56]. Linear discriminant analysis (LDA)[57,58] is a supervised learning method that is commonly applied in conjunction with PCA for discrimination tasks. LDA searches for those vectors in the underlying space that best discriminate between classes. More formally, given a number of independent features derived from the data, LDA creates a linear combination of those that yield the largest mean differences between the desired classes[59]. Indeed, PCA-LDA is one of the classic approaches used to introduce machine learning methodologies[54]. PCA and LDA have both been used extensively for biological data discrimination purposes[60,61] and for image processing[56,62].

In this work, we quantitatively and objectively analyse the power spectrum of a dataset of over one thousand images of dried blood droplets. The blood is taken from thirty healthy male volunteers at five time points: before an exhaustive cycling exercise, at peak exertion, and at two, four and six minutes into the post-exercise recovery. Our overall goal is to develop an image processing method to identify and extract the most distinguishing features in the images and apply an advanced multivariate statistical machine learning algorithm to quantitatively and objectively discriminate between the images. This will allow us to identify, from images of dried blood droplets, the physiological states of the cyclists; whether the person is at rest or has performed physical exercise.
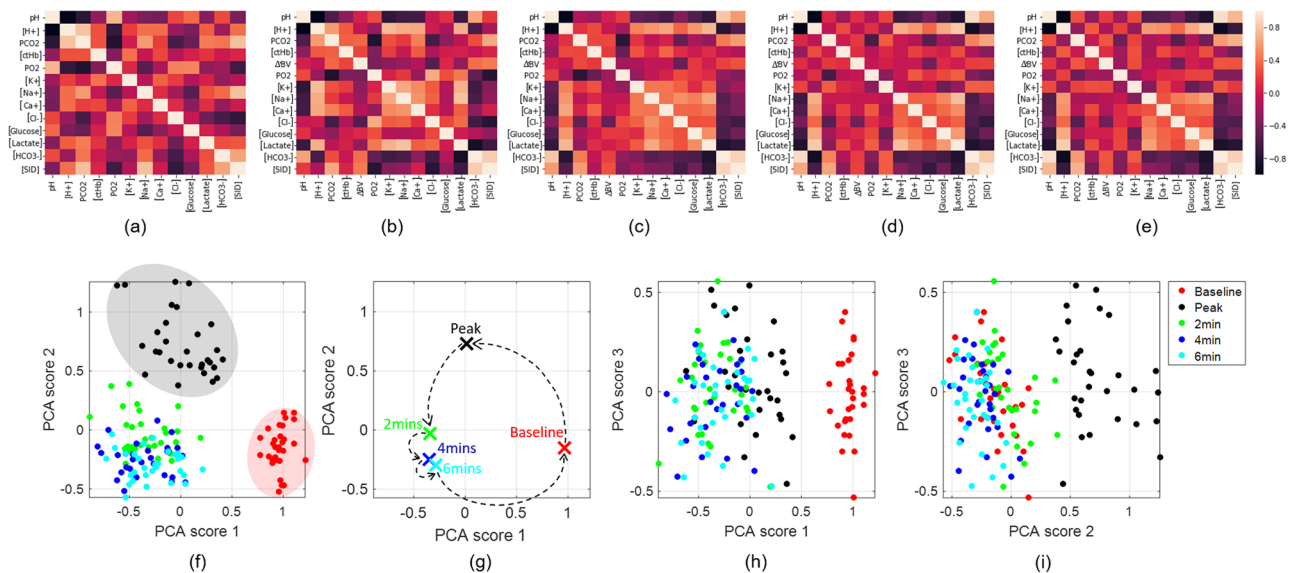
Figure 1. Statistical analysis of the measured blood chemistry properties. Correlation matrices are calculated for the 14 properties measured at (**a**) baseline, (**b**) peak exercise, (**c**) after 2 minutes exercise, (**d**) after 4 minutes exercise and (**e**) after 6 minutes exercise. The first three scores of the unsupervised dimensionality reduction and clustering method, principal component analysis, are shown in separate 2D plots (**f,h,i**) for the chemical properties: $[H^+]$, PCO_2, $[Hb]$, ΔBV, PO_2, $[K^+]$, $[Na^+]$, $[Ca^{2+}]$, $[Cl^-]$, $[Glucose]$, $[Lactate^-]$ and $[HCO_3^-]$, discarding pH and [SID] as they are related to the other quantities. The best discrimination between all five blood conditions is revealed along the first principal component score shown in (**f**). The centroid's behaviour related to the conditions clusters is shown in (**g**) which strongly suggests that the blood condition, from the chemistry point of view, is returnings towards baseline after 6 minutes of exercise.

This paper is organised as follows. First we present the results obtained from optimising our machine learning statistical model when applied to our image-processed dried blood droplet dataset. Next, we discuss these results and the potential use for current and future work. Finally we explain the methods followed in this work.

Results

Exercise responses and blood chemistry analysis.

The gas exchange threshold during the cycling exercise test was 2.81 ± 0.66 L · min^{-1} (250 ± 62 W). Peak cycling power, maximum oxygen uptake and maximum heart rate were 335 ± 64 W, 3.63 ± 0.73 L · min^{-1} (47.7 ± 8.9 mL · kg^{-1} · min^{-1}), and 184 ± 12 bpm, respectively. For all blood parameters, data were analysed in GraphPad Prism (V7.05) using repeated measures analysis of variance and Tukey's multiple comparison post-hoc test. The blood parameters measured in this study included pH, partial pressures of carbon dioxide and oxygen (PCO_2, PO_2), and concentrations of K$^+$, Na$^+$, Ca^{2+}, lactate (La$^-$), Cl$^-$, HCO$_3^-$, glucose and haemoglobin (Hb). Subsequently, $[H^+]$, the strong ion difference ($[SID]$) and changes in blood volume from rest (ΔBV) were calculated. In Fig. 1, we present the correlation matrices between all these chemical properties for all five measurement points. It can be seen that a remarkable correlation, revealed by instances of black or pale pink pixels, between some of these properties occurs when changing the condition of the blood. For instance, for blood taken at rest, pH and PCO_2 show a strong negative correlation illustrated by the third dark square in the top row in Fig. 1a. Exercise reduces this correlation until the final blood sample taken after 6-min recovery Fig. 1e, as shown by the red square which indicates a much lower correlation. These features would be expected based on physicochemical principles of blood acid-base balance[63]. Specifically, changes in pH (or $[H^+]$) and $[HCO_3^-]$ are determined by changes in $[SID]$, the total concentration of weak acids in the blood, and PCO_2[63]. At rest, between-participant differences in blood pH are mainly determined by differences in PCO_2 since $[SID]$ and the total concentration of weak acids are usually kept within narrow limits both within- and between-participants. During exercise, however, $[SID]$ falls due mainly to an increase in $[La^-]$, whereas PCO_2 may change little from rest and actually decline in recovery (as shown in Table 1). Thus, the fall in blood pH during and after maximal exercise is primarily explained by reductions in $[SID]$, with increases in the total concentration of weak acids making an additional, albeit smaller, contribution. These principles are therefore consistent with the reduced correlation between pH and PCO_2 as the experiment progressed.

We then use PCA to uncover linear combinations of these fourteen values which vary between conditions and will ultimately not only significantly reduce the dimensionality of the blood data but also show a possible clustering between the blood conditions. Fig. 1f,h,i show the relationship between pairs from the first three PCA scores demonstrating that, not only have we reduced the dimensionality of the system to a feature subspace, but the first and second PCA scores also give excellent unsupervised clustering of the five conditions as baseline and peak data points fall into two non-overlapping clusters (shaded), with both scores contributing equally to distinguish the physiological states. On the other hand, although there is an overlap between all the three sets of recovery data points, they are noticeably separated from baseline clusters using the first PCA score and from the peak cluster

	Rest	End of Exercise	Recovery (min)		
			2	4	6
$[La^-](mmol \cdot L^{-1})$	1.1 ± 0.4	10.8 ± 3.1**	15.2 ± 2.2**	15.6 ± 2.2	15.3 ± 2.2
$[Na^+](mmol \cdot L^{-1})$	141 ± 1	148 ± 2**	146 ± 1**	144 ± 1**	143 ± 1**
$[K^+](mmol \cdot L^{-1})$	4.0 ± 0.2	5.4 ± 0.6**	4.1 ± 0.2**	3.8 ± 0.2**	3.8 ± 0.2
$[Ca^{+2}](mmol. L^{-1})$	1.23 ± 0.03	1.31 ± 0.04**	1.28 ± 0.03**	1.26 ± 0.03**	1.24 ± 0.03
$[Cl^-](mmol \cdot L^{-1})$	105 ± 1	108 ± 2**	107 ± 2**	106 ± 1	106 ± 1
$[SID](mmol \cdot L^{-1})$	40 ± 2	34 ± 3**	28 ± 2**	$27 \pm 2^*$	26 ± 2
$PCO_2(mmHg)$	44.9 ± 4.1	52.8 ± 9.9**	36.3 ± 3.9**	33.8 ± 3.2	33.2 ± 2.7
$[H^+](nmol \cdot L^{-1})$	40.8 ± 2.2	61.8 ± 8.6**	63.8 ± 7.0	65.4 ± 7.2	65.1 ± 7.1
$[HCO_3^-](mmol \cdot L^{-1})$	25.7 ± 1.1	18.4 ± 2.1**	14.8 ± 1.5**	$14.0 \pm 1.4^*$	13.9 ± 1.5
$[Hb](g \cdot dL^{-1})$	15.4 ± 1.1	16.7 ± 1.0**	16.7 ± 1.0	16.6 ± 1.0	16.4 ± 1.0

Table 1. Measured blood variables at rest, at the end of exercise, and at 2-, 4- and 6-min recovery. Values are mean $\pm$ standard deviation. For all variables there was a significant ($P < 0.001$) change over the duration of the experiment. Significant changes from the preceding value are indicated with * for $P < 0.05$ and ** for $P < 0.001$.

using the second PCA score. The third PCA score does not offer any additional distinction, as shown in Fig. 1h,i. In addition, the behaviour of the centroids of each cluster, indicated by the crosses on Fig. 1g, shows that from a chemical point of view, the condition of the blood is returning towards the baseline status in a loop-like manner. This is manifested by the combination of the first pair of PCA scores, with the first one carrying more information than the second one. Numerically, had we continued to take samples longer into the recovery period, the centroids would eventually return to the baseline position.

Power spectrum discrimination. We developed an optimised machine learning algorithm that successfully discriminates the logarithmic power spectrum of the volunteer-averaged images. The feature extraction is carefully described in the Methods section later on. We first perform a dimensionality reduction using principal component analysis (PCA) followed by a supervised discrimination method known as linear discriminant analysis (LDA) which linearly combines the PCA scores to optimise the discrimination between the images of two specific different conditions under scrutiny. For each condition per participant, ten to twelve multiple images were acquired. Fig. 2 shows the outcome of the discrimination process between baseline and the other conditions over the average of all the images per participant per condition. It can be seen in Fig. 2d that the error in discrimination is lowest, 5%, between the baseline images and those taken after 6 minutes of recovery. Both classes are clearly discriminated mostly along the first linear discriminant score, LD_{S1}, as shown in Fig. 2d,3 In addition, Fig. 2d,1 shows the first linear discriminant function which exhibits the behaviour of the spatial frequency as a function of radial information. It can be noticed, by a simple projection, that a clear pattern indicates rich discriminating information situated close to the droplet's centre and near its periphery. The bright green and red patches shown in the linear discriminant function provide important fingerprints that can be used to discriminate new images. Although the second linear discriminant score, LD_{S2}, shows a poor projection axis for separating the centroids, in Fig. 2d,2, it reduces misclassification when using an appropriate decision boundary such as the circle exemplified in Fig. 2d,3. These two functions are able to classify the "baseline" and "after 6 mins" conditions of unknown blood droplet images with a high accuracy of $\eta = 95\%$.

We then used the comparison between droplets taken at baseline and after 6 minutes recovery, to investigate the effect of changing the number of averaging images on the final error rate of the statistical model. We find that averaging over all the images per participant per condition gives the best discrimination. Passing the entire database of individual images to our predictive model lowers its accuracy to nearly 20% as demonstrated in Fig. 3a. This error rate is gradually lowered by averaging more images per participant per condition until it reaches the best discrimination accuracy of 95% illustrated in Fig. 2d,3. This suggests that using multiple images is necessary, as it averages out the random variations between all droplets, highlighting the true differences due to blood differences.

Additionally, the behaviour of the centroid of each condition class, indicated by the crosses on Fig. 4, is also analysed. This shows, in clear contrast with the behaviour seen in Fig. 1g, that the centroid of each condition's cluster moves monotonously and linearly only along the LD_{S1} axis over time. Unlike the blood chemistry centroid trajectory presented in Fig. 1g, here we do not see the blood droplet patterns returning towards their original baseline state.

Discussion

Some researchers have used drying blood droplets to diagnose disease. One paper[64] reported attempts to use droplets of blood serum to diagnose cancer as "statistically unreliable", yet the authors were able to qualitatively visualise the profound changes in the phase transition and the phase state of blood plasma proteins in patients with metastatic cancer. However, others have had more success. In ref. [31] images are presented from a blood serum with a range of diseases, including cancers and hepatitis. These droplets are distinguished statistically using the "shape index" of the time evolution of the acousto-mechanical impedance of the droplets, measured using a quartz resonator. The shape index shows good specificity and sensitivity to the various conditions. Further
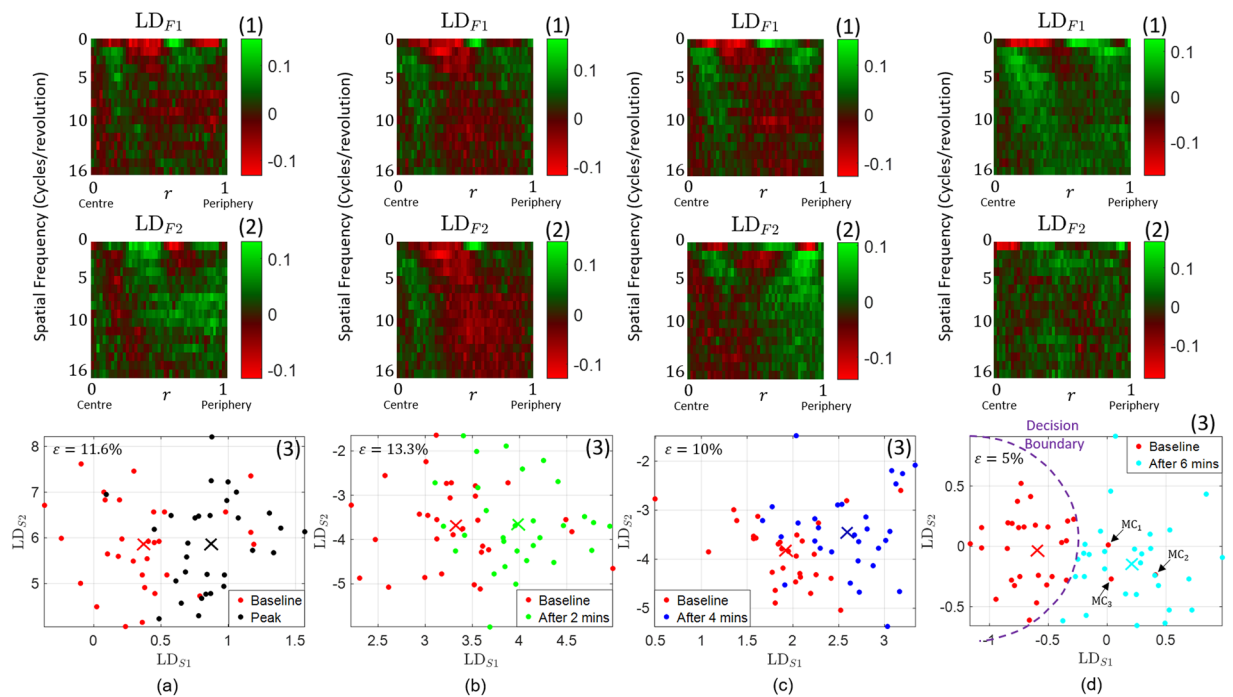
Figure 2. The discrimination outcome of the optimised machine learning algorithm to discriminate the logarithmic power spectrum of two blood conditions: (**a**) baseline against peak, (**b**) baseline against after 2 mins, (**c**) baseline against after 4 mins and (**d**) baseline against after 6 mins. The lowest error rate has been obtained when discriminating baseline and after 6 mins with $e = 5\%$ shown in the scatter plot (d,3) that illustrates the relationship between the first two LDA scores, i.e., LD_{s1} and LD_{s2}. Here MC_i, where $i = 1, 2, 3$, refers to the misclassified baseline points. Figures (d,1) and (d,2) represent the first and second LDA functions, respectively, where each element of each row (frequency) corresponds to the average pixel intensity at a given radius. The spatial frequency of the linear discriminant functions is measured in units of cycles per revolution.
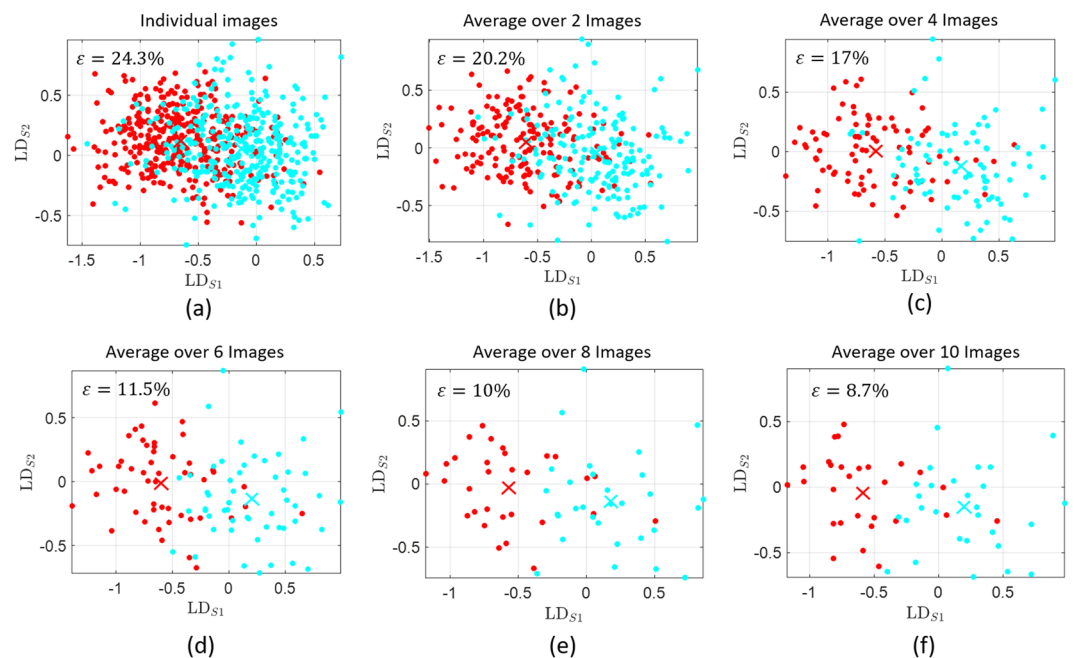

Figure 3. The effect of averaging separate images from the same volunteer and the same physiological state, on the final discrimination outcome. In (**a**) individual images with no averaging are used for the discrimination process. The accuracy of this approach is 75.7%. The accuracy increases noticeably with increasing the number of averaging images in (**b**–**f**). As shown previously in Fig. 2d,3, all images (usually ten to twelve) taken per volunteer per condition are averaged and the accuracy has considerably increased to 95%. An optimised training exercise was undertaken for each separate case.
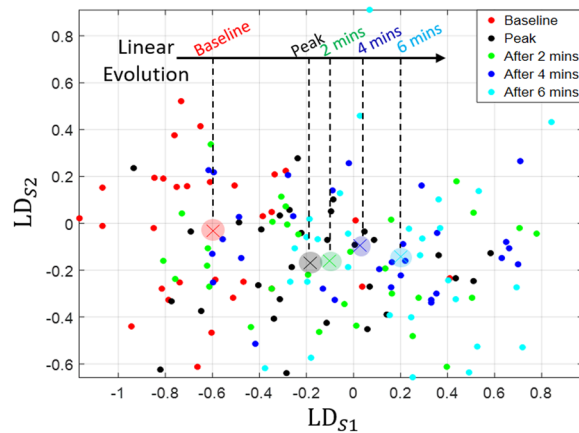
Figure 4. The trajectory of the centroid of the blood condition clusters, with images averaged over each person. The individual volunteer's images, projected onto our DF space, are also shown, to disclose the inter-subject variability.

studies[65] have used manual identification of morphological features in dried droplets (which the authors call 'facias') of blood serum, tears and synovial fluid and present statistically significant correlations between features and conditions. A similar approach has found quantitative discrimination for identifying hepatitis[66] in blood serum, and shown qualitative differences due to anaemia and hyperlipidemia[21].

The presence of various proteins in blood serum is shown to qualitatively affect the microstructure of the dried film[67] with some success in a quantitative analysis by manually categorising the orientation of cracks. A more quantitative approach, in which crack angle and plaque sizes are measured, shows a convincing correlation with relative humidity[68]. An alternative automated approach[69] calculating a "distance" between images of uninfected blood droplets and those with either tuberculsosis or anemia shows some promise.

A recent study[70] is the first to probe the physical mechanisms linking measured blood properties to the observed patterns in whole blood droplets. The authors find that a normalised capillary pressure (which depends on blood viscosity, mean corpuscular volume and haematocrit), controls the properties of the cracks that form in healthy blood compared to that with neonatal jaundice and thalassamia. They also observed that droplet drying times vary with disease conditions.

A recent review article[71] discusses spectroscopy of droplets, and makes conclusions regarding the optimal conditions, suggesting using "a diluted aliquot of serum (1 μL) spotted on a smooth, flat and homogeneous surface, at an elevated temperature or high humidity in order to result in a more uniformly spread sample, with a faster drying time".

Our approach differs from other techniques as we do not impose a discrimination based on human observations of distinguishing features. Instead, we employ a machine-learning optimisation technique which discovers the statistically most significant measure to classify and identify the different conditions. Our approach has the potential for objectively diagnosing medical conditions using fully automated image processing tools.

Our findings show for the first time how quantitative and objective image analysis of patterns left by dried blood droplets can succeed in detecting dominant features and optimally discriminating different physiological conditions, with an error of 5%. This approach has great potential in providing a new, rapid and reliable method for not only discriminating healthy blood conditions, but also in disease detection.

The images have been analysed and prepared before passing them on to the machine learning algorithm for discrimination. This analysis consists of several steps: applying interpolation to homogenise the spatial extent of any structure of each image; removing cracks and discontinuities found to be appearing inconsistently; extracting the most useful information that might be implicit in the image function; converting to polar coordinates; and finally calculating the power spectrum of the angular variations, removing any dependency on droplet absolute orientation.

Our machine learning algorithm is performed first using principal component analysis, as an unsupervised dimensionality reduction method, followed by linear discriminant analysis, as a supervised classification method. These two methods are very common for discrimination purposes. However, our novel contribution lies in developing a time-saving optimisation process to enhance the algorithm's training capability and improve its prediction outcome. This process is applicable to large datasets since it does not search within all possible combinations of participants for the ideal and lowest error rate possibilities. Instead, it systematically 'selects' the training data that we would like the iterative process to be applied to. This significantly lowers the computation time and successfully trains the algorithm. The optimised algorithm yielded a high discrimination accuracy of 95%, typically 10% higher than when using a standard analysis. Note that the numerical search outputs discriminating features that were determined using up to 50% of the PC scores, thereby using details that cannot necessarily be identified by the naked eye. In fact, only averaging numerous images belonging to specific categories allows the user to discriminate them by visual inspection, substantiating the need for numerical discrimination.
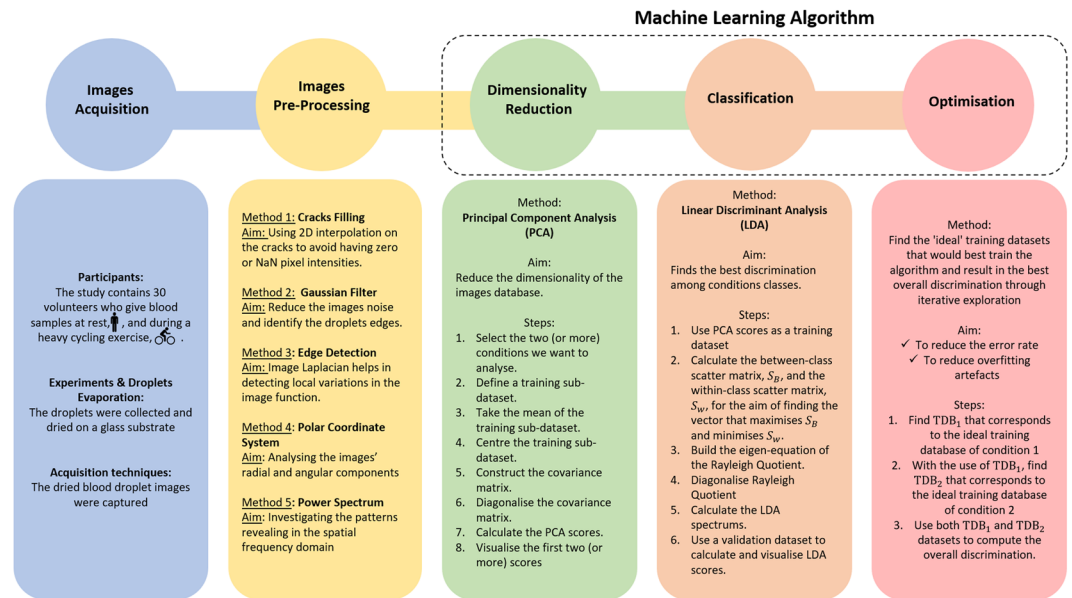
Figure 5. Work methodology. The study begins by building the database of the dried blood images taken from 30 healthy volunteers before and after a cycling exercise. These images are then pre-processed individually using different image analysis techniques; cracks filling, Gaussian filter, edge detection, polar coordinate and power spectrum. The database of the logarithmic power spectrum of the images are later used to be passed on to our machine learning algorithm. This algorithm starts with reducing the dimensionality of the image database using principal component analysis. Using the resulting feature space, i.e., the subspace that contains the highest varying principal components, we can move on to the classification step where we use linear discriminant analysis to find the best line that best separates the conditions of the images. The performance of this algorithm is further enhanced by applying an optimising iterative search to select the 'ideal' few participants who would best train the algorithm and result in the lowest overall error rate possible.

Our work highlights the importance of averaging over a sufficient number of droplet images for each individual as this improves the final discrimination accuracy. It is also interesting to note that the centroids representing the blood images move in a linear direction, and have not started to show any sign of returning to baseline after 6 minutes of recovery. In contrast, the blood chemistry centroids seem to have almost recovered, by that time. A summary of statistical analyses of the blood chemical properties is given in Table 1 and clearly shows that there is no variable that continues to move away from the resting value. For future studies, we would recommend allowing recovery to continue for much longer after exercise, possibly up to one hour. This observation also suggests that some other property of the blood is affecting the droplet patterns than any of the parameters that were measured.

Methods

In this section, we present and explain the methods adopted in this work with the aim of quantitatively and objectively discriminating the conditions of the dried blood droplet images. The experimental protocols (exercise test and blood sampling) were approved by the Nottingham Trent University Human Ethical Review Committee. All methods were performed in accordance with the relevant guidelines and regulations of the standards set by the Declaration of Helsinki. There are no competing interests associated with this work. The main stages followed in this work can also be seen in Fig. 5.

Participants, exercise protocol and blood sampling. Thirty healthy, recreationally active, non-smoking, males (age: 24 ± 6.9 years; height: 178 ± 5.5 cm; body mass: 76 ± 9 kg) provided written informed consent to take part in the study. Participants performed a maximal incremental cycling ramp test on an electromagnetically braked cycle ergometer (Excalibur Sport; Lode, Groningen, the Netherlands). Participants performed 3 minutes of unloaded cycling followed by an incremental ramp protocol ($30 \text{ W} \cdot \text{min}^{-1}$), at their preferred cadence, until the limit of tolerance or task failure (cadence below 60 rpm). Thereafter, participants remained seated on the cycle ergometer for 6 minutes. During exercise participants wore a facemask (model 7940; Hans Rudolph, Kansas City, MO) connected to a flow sensor (ZAN variable orifice pneumotach; Nspire Health, Oberthulba, Germany) that was calibrated using a 3-L syringe. Gas concentrations were measured using fast responding laser diode absorption spectroscopy sensors which were calibrated using gases of known concentration (5% CO_2, 15% O_2, balance N_2), and ventilatory and pulmonary gas exchange variables were determined breath-by-breath (ZAN 600USB; Nspire Healh). The gas exchange threshold was determined using the V-slope method[72] and the maximum oxygen uptake was taken as the highest 10 second mean value. Heart rate was measured using short-range telemetry (Polar S160, Polar, Kempele, Finland). Arterialised venous blood (2 mL) was drawn from a heated (using an infra-red lamp) dorsal hand vein using an indwelling 21-G cannula and a syringe containing dry electrolyte-balanced heparin (*safe*PICO, Radiometer, Copenhagen, Denmark). Blood samples were taken at rest, at the end of exercise, and after
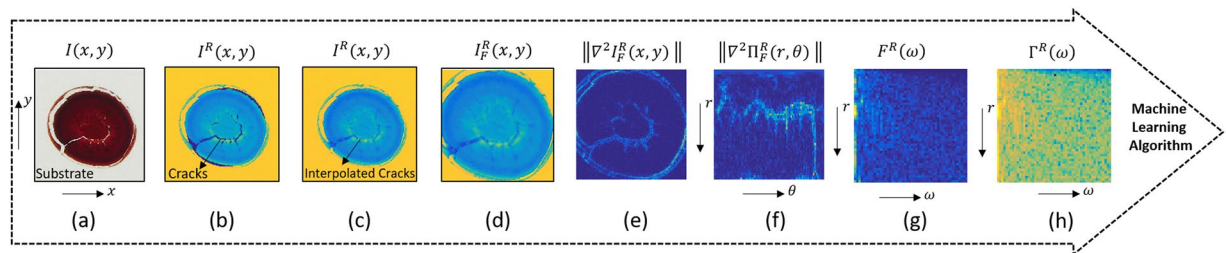
Figure 6. Image pre-processing methods where (**a**) shows the raw blood image, (**b**) represents the red channel of the image, (**c**) illustrates the 2D interpolated cracks, (**d**) shows a 2D Gaussian filtered image, (**e**) shows the Laplacian of the image function, (**f**) is the polar coordinate version of the image, (**g**) is the 1D power spectrum of the image previously shown and (**h**) is the logarithmic power spectrum.

2-, 4- and 6-min recovery. Blood was analysed immediately (ABL90 FLEX; Radiometer) for pH, PCO_2, $[HCO_3^-]$, $[K^+]$, $[Na^+]$, $[Ca^{2+}]$, $[La^-]$, $[Cl^-]$, [glucose], and [Hb]. The remaining blood was transferred into a heparin-coated Eppendorf tube for subsequent blood droplet analysis. The hydrogen ion concentration ($[H^+]$) was calculated from the measured pH value. The strong ion difference ([SID]) was calculated as the difference between the sum of the strong cations and the strong anions: $[SID] = ([Na^+] + [K^+] + [Ca^{2+}]) - ([Cl^-] + [La^-])$[63]. We calculated ΔBV from resting levels using changes in [Hb][73].

Droplet preparation and evaporation. Significant care was required to deposit the droplets repeatably. A micro-pipette was used to produce a 10 μL droplet at the tip of the pipette, small enough so that gravity did not cause it to fall. The pipette was slowly lowered until the droplet came into contact with the glass slide, when it would transfer to the substrate with minimal impact velocity. The glass slides measured 76.2 mm $\times$ 25.4 mm and were first cleaned with an air duster. For each blood sample, at least 12 droplets were pipetted on each of two slides in two rows of six. The spacing between droplets was around four times the radius, reducing any possible effects of crowding. Immediately after pipetting, the droplets were very mobile so the slides were left for between 10 to 12 minutes. After this time, the very edge of the droplet had dried sufficiently to pin the liquid to the glass, enabling the slides to be moved without risk of disturbing their circular shape. The slides were then transferred to an air-tight Bel-Art transparent desiccator with dimensions 337mm wide by 254mm deep by 216mm tall, and left there overnight.

The chamber was fitted with a mechanical hygrometer to measure the relative humidity inside the chamber, but we also used a humidity probe to monitor temperature and humidity during drying. The chamber had a movable bottom shelf which contained regular holes allowing for air exchange. Beneath the shelf we placed a Petri dish containing a supersaturated solution of NaCl salt, which fixed the relative humidity in the chamber in the range 70% to 75%[74]. At this humidity, the evaporation rate was slow, taking around 24 hours to fully dry, and reducing the influence of neighbours. The chamber had a capacity to accommodate 16 to 18 glass slides.

Image acquisition. Droplet images were recorded using a CASIO EX-Z1000 digital camera in HDR super-macro mode with fixed magnification, with resolution 150 px/mm mounted 9cm above a white flat back-light (The Imaging Source DCL/BK.WI 5070/EU). Reflections from other lights in the room were eliminated using screens. As the dried deposits had a tendency to flake or peel, the slides were carefully transferred from the drying chamber and placed directly on top of the backlight. Images were cropped manually to contain one complete droplet per image. The dataset of the dried blood images can be accessed via https://doi.org/10.6084/m9.figshare.11373948.v1.

Image pre-processing. After recording, documenting and cropping the raw blood images, image processing routines must be implemented before quantitatively analysing the data. Here, we present all the image pre-processing techniques that we have applied on our blood raw images for the aim of (i) reducing their noise, (ii) enhancing their quality and (iii) extracting the most important information to fully prepare them for the classification process. This improves the accuracy of the subsequent statistical analysis and makes the discrimination process well-trained. Fig. 6 illustrates the image analysis techniques applied on each image of our database before proceeding to the discrimination algorithm.

- Crack filling: Cracks often appear in dried droplet patterns and can be wavy, straight, arched, spiral, circular, and three-armed cracks[75–81]. Significant work has been done on explaining the reasons behind crack-formation and the parameters affecting their formation in dried droplets, and specifically in blood droplets[4,20,21]. Cracks appear as white pixels in the image with maximum pixel intensity, and as others have reported "cracked surfaces across a drop can generate spectral distortions often in the form of baseline discrepancies"[71]. In our situation, the cracks deteriorate the study quality as they appear inconsistently in our data set. Therefore, we decided to automatically 'airbrush' the cracks by performing a 2D interpolation reducing this discontinuity. Since cracks usually have low luminance and due to the fact that most useful information in the blood droplets images are concentrated in the red channel, our crack detection criterion focuses on the red-channel of the images with a condition that the other two colour channels; blue and green, are above a certain value,

i.e., $I^R(x, y)$[if: $I^{B,G}(x, y) > \alpha$], where $I^R(x, y)$ and $I^{B,G}(x, y)$ are the image matrix functions in the Cartesian coordinate system for the red and blue-green colour channels, respectively, and α is the value that best reveals the cracks, as shown in Fig. 6b. This enhances the clarity of the cracks where presented in an image. Once we identify the position of the cracks, a 2D linear scattered interpolant method is applied to interpolate their values, as shown in Fig. 6c.

- Gaussian filter: To consistently detect the boundaries of the blood droplet, ignoring marks on the substrate, edge detection must be preceded by the use of a 2D Gaussian filtering, also smoothing out boundaries originating from within the droplet itself[82],

$$G(x, y) = \exp\left(-\frac{x^2 + y^2}{\sigma^2}\right),$$

(1)

where σ denotes the width (standard deviation) of the bell-shaped function. For a very large value of σ, hardly any filtering is recorded. However, with decreasing the value of σ, the Gaussian filter starts to attenuate the high frequency components, removing those responsible for noise and other small scale (quickly varying) information. In our case, we have set the value of $\sigma = 0.01$ cycles per pixel (or one-hundredth of a wavelength per pixel) throughout this project. The pixel at the centre of the Fourier space receives the maximum weight, and the remaining coefficients drop off smoothly with increasing distance from the centre. Each image is filtered by multiplying its Fast Fourier Transform, i.e., $I_F^R(\omega, \nu) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I^R(x, y)e^{-i(\omega x + \nu y)}dxdy$, with the Gaussian function defined in Eq. 1. Next, to access the spatial information of the filtered image, we perform the inverse Fourier Transform, i.e., $I_F^R(x, y) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I_F^R(\omega, \nu)e^{i(\omega x + \nu y)}d\omega d\nu$. Figure 6d shows the cropped image of the red-channel blood droplet as a result of the filtering.

- Edge detection: Edges are common features in an image and are identified by rapid local variations in the image function. Edge detection is a process of identifying and locating sharp discontinuities in an image[83]. During the history of image processing, a variety of edge detection approaches have been devised which differ in their purpose and in their mathematical and algorithmical properties[84]. The location of edges in a 2D image are normally determined either by (i) finding the image gradient extrema (maximum or minimum) or by (ii) finding the zero-crossings of the Laplacian of the image. The Laplacian filter is a second derivative commonly used for edge detection and image enhancement in digital images. Second-order derivatives have a stronger response than first derivatives to fine detail, such as thin lines, weak edges and isolated points and that is due to the fact that first derivatives may not be large enough to distinguish the edge points, and, therefore, a weak edge can go undetected by such methods. Taking the second derivative of the points or Laplacian amplifies the changes in the first derivative and, therefore, increases the chances of detecting a weak edge[85]. Mathematically, the Laplacian of our 2D image $I_F^R(x, y)$, is given by,

$$\nabla^2 I_F^R(x, y) = \begin{bmatrix} I_F^R(x, y) \\ I_F^R(x, y) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial^2 I_F^R(x, y)}{\partial x^2} \\ \dfrac{\partial^2 I_F^R(x, y)}{\partial y^2} \end{bmatrix},$$

(2)

and its magnitude, as shown in Fig. 6e, can be written as,

$$\left\|\nabla^2 I_F^R(x, y)\right\| = \sqrt{\left(\frac{\partial^2 I_F^R(x, y)}{\partial x^2}\right)^2 + \left(\frac{\partial^2 I_F^R(x, y)}{\partial y^2}\right)^2}.$$

(3)

- Polar Coordinate System: In this step, since the rich information in a circular dried blood droplet image is highly located in the radial or angular areas in reference to the centre point $r = 0$, we convert the magnitude of the image Laplacian, i.e., $\left\|\nabla^2 I_F^R(x, y)\right\|$, from Cartesian to Polar Coordinate System, i.e., $\left\|\nabla^2 \Pi_F^R(r, \theta)\right\|$ as shown in Fig. 6f. Using this technique is very important since it helps exhibit the symmetry of the pattern distribution in the blood droplet.

- Power spectrum: In the final pre-processing step prior to applying our machine learning algorithm, we apply the absolute value of the Fast Fourier Transform on the angular component of the logarithmic polar form of the image Laplacian of the red channel of each droplet image to calculate their power spectra, as illustrated in Fig. 6g. By applying the power spectrum on the angular component of the polar image the absolute measure of angle is removed, i.e., the remaining information becomes immune to rotations about the centre of the droplet. This extracts the spectral information and exhibits the pixel intensity within the spatial frequency domain for each image. This can be written as:

$$F^R(\omega) = \left| \text{FFT}(\log \|\nabla^2 \mathbf{\Pi}_F^R(r, \theta)\|) \right| = \left| \int_{-\infty}^{+\infty} \log \|\nabla \mathbf{\Pi}_F^R(r, \theta)\| e^{-i\omega\theta} \mathrm{d}\theta \right|. \tag{4}$$

Furthermore, Fig. 6h shows that we take the logarithm of the power spectrum of the image, i.e., $\Gamma^R(\omega) = \log(F^R(\omega))$. This (i) enhances small pixel intensity and (ii) it works better than linear power spectra (see supplementary material).

Machine learning algorithm.   At this stage, our images are ready to be passed on to our machine learning algorithm for discrimination purposes. The input matrix $\mathbf{\Gamma}^R$ of each image is made of 48 rows and 103 columns, where each column represents the frequency of the spectral image, and each row represents radial component of the image. We have seen that the first row, i.e., $r = 1$, carries negligible radial information that could be useful in our subsequent steps. Similarly, the second half of the full frequency bandwidth, i.e, $c \geq 103/2 \geq 52$, exhibits minimal important features. Therefore, it is plausible to discard the first row and truncate the full frequency bandwidth to preserve only the first half, i.e., the first 52 columns. Hence, the input matrix $\mathbf{\Gamma}^R$ in this case consists of 47 rows and 52 columns (see supplementary material).

The input matrix $\mathbf{\Gamma}^R$ is originally a 4D matrix with the first dimension referring to the radial information within all the frequencies that make up the spectral image stacked up in a single column vector. The second, third and fourth dimensions represent the number of blood conditions, people and images taken from each person per condition, respectively, i.e.,$[\mathbf{\Gamma}^R] = 2444 \times 5 \times 30 \times 12$. This matrix has been rearranged to be a 2D matrix after performing: (i) averaging over all, or some of, the images per volunteer (ii) normalisation of the image pixel intensities where minimum values are set to zero and maximum values are one, i.e.,$[\mathbf{\Gamma}_i^R] = 2444 \times 30$, where i refers to the five blood conditions: baseline, peak, and after 2, 4, and 6 minutes.

This algorithm has three main steps to fully complete the discrimination process. It starts with an unsupervised learning that uses a dimensionality reduction method to extract the most varying components. For this step we use principal component analysis (PCA). Afterwards, a supervised learning method, linear discriminant analysis (LDA), is introduced which uses a linear combination of half of the largest PCA scores to find the vector that best separates the conditions of the blood images (see supplementary material). All the previous stages produce a machine learning algorithm that yields a fairly acceptable discrimination prediction outcome. However, the accuracy of this model can be significantly enhanced and the possibility of overfitting can be greatly lowered and hence the discrimination outcome will be noticeably improved by performing an optimising process which we shall describe next.

Optimisation process.   The novel idea of our optimisation process is a new strategy to find and use 'ideal' training datasets that would best train the algorithm and result in the best overall discrimination, allowing the most predominant, generic discriminating features to be identified between the five experimental conditions. One approach to find these 'ideal' data/participants is to systematically explore all possible combinations of 20, for example, out of 30 participants and compute the error rate for each individual iteration and the one yielding the combination of participants with the lowest error rate reveals the strongest generic discriminating features[86]. However, applying this method on our large image database would result in a total number of iterations of 30, 045, 015 which would take months to fully explore. Therefore, we have redesigned the structure of this optimisation process to be applicable to large datasets with less computational running times. Suppose the two blood condition matrices that we would like to optimise their discrimination are denoted by A and B, then the optimisation method is constructed in two sets of iterations:

- First set of iterations: The main idea here is to find the ideal few participants in the first condition matrix A that best train our algorithm. Each iteration in this step builds a training dataset which is a concatenation of a 'selected' slice of matrix A with matrix B. This slice initially contains two participants from A: 1 and i, where $i = 2, 29$. At the end of each iteration tested on the validation dataset of ALL images in A and B, the error rate is calculated as follows: each measurement (comprising of half of the largest PCA scores) is projected onto a two-dimensional LDA space, yielding a set of two LDA scores. In this space, the coordinates of the two centroids are calculated, and for each iteration, the Euclidean distances to both centroids are further calculated. The ratio of these two distances is used to assess whether the trial ends up in one of the blood conditions. The iterations ending up with the incorrect membership are expressed as a percentage error rate. The iteration that corresponds to the lowest error rate, j, is added to the following searching iterative process to the training slice of A. Here we take 1,j and i which is the remaining 28 participants and the same process is repeated. After scanning all the possibilities with the lowest error rate, we now take participant 2 and i where $i = 1, 3, \ldots, 29$ and the same iterative process is done again. By the end of these two main explorations, we choose the combinations that correspond to the lowest error rate and they will ultimately form the ideal choices to build the optimal training dataset of matrix A that corresponds to the first blood condition.
- Second set of iterations: The main idea here is to find the ideal participants in the second condition matrix B that best train our algorithm by using the ideal choices of A from the previous step. Taking the ideal training choices of the first blood condition, we perform the same iterative explorations on matrix B calculating the error rate values by the end of each iterations. Similarly to the first set of iterations, the lowest error rates obtained will build the optimal training dataset of matrix B.

By the end of these two sets of iterations, we end up with two ideal training datasets corresponding to both blood conditions, i.e., A and B. Passing these two training datasets on to our classifier highly enhances the

discrimination outcome. It can be noticed that since we take only participants 1 and 2 from each matrix to search for the lowest error rate possibilities, this significantly decreases the optimisation time and accordingly enhances the overall discrimination process.

Approval. The experimental protocols (exercise test and blood sampling) were approved by the Nottingham Trent University Human Ethical Review Committee.

Accordance. The study conformed to the standard set by the Declaration of Helsinki.

Informed consent. Informed consent was obtained from all participants/volunteers.

Data availability

All relevant data are within the paper and its Supporting Information files. The dried blood images dataset generated and analysed during this study is available in the Figshare repository, https://doi.org/10.6084/m9.figshare.11373948.v1.

References
1. Zang, D., Tarafdar, S., Tarasevich, Y. Y., Choudhury, M. D. & Dutta, T. Evaporation of a droplet: From physics to applications. *Physics Reports* (2019).
2. Deegan, R. *et al.* Capillary flow as the cause of ring stains from dried liquid drops. *Nature* **389**, 827–829 (1997).
3. Larson, R. Transport and deposition patterns in drying sessile droplets. *Transport Phenomena and Fluid Mechanics* **60**, 1538–1571 (2014).
4. Zeid, W., Vicente, J. & Brutin, D. Influence of evaporation rate on cracks' formation of a drying drop of whole blood. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **432**, 139–146 (2013).
5. Deegan, R. *et al.* Contact line deposits in an evaporating drop. *Phy. Rev. E* **756** (2000).
6. Deegan, R. Pattern formation in drying drops. *Phys. Rev. E* **61** (2000).
7. Parsaab, M., Harmandab, S. & Sefiane, K. Mechanisms of patter formation from dried sessile drops. *Advances in Colloid and Interface Science* **254**, 22–47 (2018).
8. Msambwa, Y., Shackleford, A., Ouali, F. & Fairhurst, D. Controlling and characterising the deposits from polymer droplets containing microparticles and salt. *The European Physical Journal E* **39** (2016).
9. Baldwin, K. & Fairhurst, D. The effects of molecular weight, evaporation rate and polymer concentration on pillar formation in drying polyethelen oxide droplets. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **441**, 867–871 (2014).
10. Smalyukh, I., Zribi, O., Butler, J., Lavrentovich, O. & Wong, G. Structure and dynamics of liquid crystalline pattern formation in drying droplets of dna. *Physical Review Letters* **96** (2006).
11. Yakhno, T. Salt-induced protein phase transitions in drying drops. *Journal of Colloid and Interface Science* **318**, 225–230 (2008).
12. Annarelli, C., Fornazero, J., Bert, J. & Colombani, J. Crack patterns in drying protein solution drops. *The European Physical Journal E* **5**, 599–603 (2001).
13. Carreon, Y., Gonzalez-Gutierrez, J., Perez-Camacho, M. & Mercado-Uribe, H. Patterns produced by dried droplets of protein binary mixtures suspended in water. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **161**, 103–110 (2018).
14. Sefiane, K. On the formation of regular patterns from drying droplets and their potential use for bio-medical applications. *Journal of Bionic Engineering* **7**, S82–S93 (2010).
15. Yakhno, T. *et al.* Drying drop technology as a possible tool for detection leukemia and tuberculosis in cattle. *Journal of Biomedical Science and Engineering* **8**, 1–23 (2015).
16. Smith, F. & Brutin, D. Wetting and spreading of human blood: Recent advances and applications. *Current Opinion in Colloid and Interface Science* **36**, 78–83 (2018).
17. Peschel, O., Kunz, S., Rothschild, M. & Mutzel, E. *Blood stain pattern analysis* **7**, 257–270 (2011).
18. Sefiane, K. On the formation of regular patterns from drying droplets and their potential use for bio-medical applications. *Journal of Bionic Engineering* **7**, S82–S93 (2010).
19. Ahmed, G., Tash, O. A., Cook, J., Trybala, A. & Starov, V. Biological applications of kinetics of wetting and spreading. *Advances in colloid and interface science* **249**, 17–36 (2017).
20. Chen, R., Zhang, L., Zang, D. & Shen, W. Blood drop patterns: Formation and applications. *Advances in Colloid and Interface Science* **231**, 1–14 (2016).
21. Brutin, B. & Sobac, D. Loquet, B. and J., S. Pattern formation in drying drops of blood. *J. fluid Mech* **667**, 85–95 (2011).
22. Sobac, B. & Brutin, D. Structural and evaporative evolutions in desiccating sessile drops of blood. *Phys. Rev. E* bf84 (2011).
23. Sobac, B. & Brutin, D. Desiccation of a sessile drop of blood: Cracks, folds formation and delamination. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **448**, 34–44 (2014).
24. Brutin, D., Sobac, B. & Nicloux, C. Influence of substrate nature on the evaporation of a sessile drop of blood. *Journal of Heat Transfer* **134** (2012).
25. Zeid, W. & Brutin, D. Influence of relative humidity on spreading, pattern formation and adhesion of a drying drop of whole blood. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **430**, 1–7 (2013).
26. Zeid, W. & Brutin, D. Effect of relative humidity on the spreading dynamics of sessile drops of blood. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **456**, 273–285 (2014).
27. Smith, F., Buntsma, N. & Brutin, D. Roughness influence on human blood drop spreading and splashing. *Langmuir* **34**, 1143–1150 (2017).
28. Smith, F. R., Nicloux, C. & Brutin, D. Influence of the impact energy on the pattern of blood drip stains. *Phys. Rev. Fluids* **3** (2018).
29. Sefiane, K. Patterns from drying drops. *Advances in Colloid and Interface Science* **206**, 372–381 (2014).
30. Chen, R., Zhang, L., Zang, D. & Shen, W. Understanding desiccation patterns of blood sessile drops. *Journal of Materials Chemistry B* **5**, 8991–8998 (2017).
31. Yakhno, T. A. *et al.* The informative-capacity phenomenon of drying drops. *IEEE Engineering in Medicine and Biology Magazine* **24**, 96–104 (2005).
32. Kokornaczyk, M. *et al.* Self-organized crystallization patterns from evaporating droplets of common wheat grain leakages as a potential tool for quality analysis. *The Scientific World Journal* **11**, 1712–1725 (2011).
33. Robertson, S., Azizpour, H., Smith, K. & Hartman, J. Digital image analysis in breast pathology–from image processing techniques to artificial intelligence. *Translational Research* **194**, 19–35 (2017).

34. Cannistraci, C., Abbas, A. & Gao, X. Median modified wiener filter for nonlinear adaptive spatial denoising of protein nmr multidimensional spectra. *Translational Research* **5** (2015).
35. Nausheen, N., Seal, A., Khanna, P. & Halder, S. A fpga based implementation of sobel edge detection. *Microprocessors and Microsystems* **56**, 84–91 (2018).
36. Annadurai, S. *Fundamentals of Digital Image Processing* (Pearson Education India, 2007)
37. KEYS, R. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **29**, 1153–1160 (1981).
38. Hou, H. & Andrews, H. Cubic splines for image interpolation and digital filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **26**, 508–517 (1978).
39. OGorman, L., Sammon, M. & Seul, M. *Practical Algorithms for Image Analysis with CD-ROM* (Cambridge University Press, 2008).
40. Siewerdsen, J. H., Cunningham, I. A. & Jaffray, D. A framework for noise-power spectrum analysis of multidimensional images. *Radiation Imaging Physics* **29**, 2655–2671 (2002).
41. Kirby, J. Which wavelet best reproduces the fourier power spectrum? *Computers and Geoscience* **31**, 846–864 (2005).
42. Witten, I., Frank, E., Hall, M. & Pal, C. J. *Data Mining: Practical Machine Learning Tools and Techniques* (Elsevier Inc., 2017).
43. Briscoe, G. & Caelli, T. *A compendium of Machine Learning* (Albex Publishing Corp., Norwood, NJ, 1996).
44. Langley, P. *Elements of Machine Learning* (Morgan Kaufmann Publishers, Inc., 1996).
45. Michalski, R., Carbonell, J. & Mitchell, T. M. *Machine Learning An Artificial Intelligence Approach* (Morgan Kaufmann Publishers, Inc., 1983).
46. Mitchell, T. *Machine Learning* (WCB/McGraw-Hill, 1997).
47. Kerepesi, C., Daroczy, B., Sturm, A., Vellai, T. & Benczur, A. Prediction and characterization of human ageing-related proteins by using machine learning. *Scientific Reports* **8** (2018).
48. Fong, R., Scheirer, W. & Cox, D. Using human brain activity to guide machine learning. *Scientific Reports* **8** (2018).
49. Elton, D. C., Boukouvalas, Z., Butrico, M., Fuge, M. & Chung, P. Applying machine learning techniques to predict the properties of energetic materials. *Scientific Reports* **8** (2018).
50. Toth, T. *et al*. Environmental properties of cells improve machine learning-based phenotype recognition accuracy. *Scientific Reports* **8** (2018).
51. Wang, Y., Fan, Y., Bhatt, P. & Davatzikos, C. High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables. *NeuroImage* **50**, 1519–1535 (2010).
52. Wernick, M., Yang, Y., Brankov, J., Yourganov, G. & Strother, S. Machine learning in medical imaging. *IEEE Signal Processing Magazine* **27**, 25–38 (2010).
53. Kutz, J. N. Deep learning in fluid dynamics. *Journal of Fluid Mechanics* **814**, 1–4 (2017).
54. Duda, R., Hart, P. & Stork, D. G. *Pattern Classification, Second Edition: 1* (A Wiley -Interscience Publication, 2000).
55. Jolliffe, I. *Principal Component Analysis* (Springer Science + Business Media New York, 1986).
56. Brunton, B., Brunton, S., Proctor, J. & Kutz, J. Optimal sensor placement and enhanced sparsity for classification. *SIAM J. Appl. Math.* **76**, 2009–2122 (2016).
57. Bishop, C. M. *Pattern recognition and machine learning* (Springer, New York, 2006).
58. Rao, C. R. The utilisation of multiple measurements in problems of biological classification. *Journal of the royal Statistical Society. Series B (Methodological)* **10**, 159–203 (1948).
59. Martinez, A. & Kak, A. Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, 228–233 (2001).
60. Xiangquan, Z., Xin, L., Yi, X. & Mingjian, H. Pca-lda analysis of human and canine blood based on non-contact raman spectroscopy. *Chemical Journal of Chinese Universities* **38**, 575–582 (2017).
61. Lin, J. *et al*. Raman spectroscopy of human haemoglobin for diabetes detection. *Journal of innovative Optical Health Sciences* **7** (2014).
62. Bai, Z. *et al. Whither Turbulence and Big Data in the 21st Century* (Springer international Publishing Switzerland, 2017).
63. Stickland, M. K., Lindinger, M. I., Olfert, I. M., Heigenhauser, G. J. & Hopkins, S. R. Pulmonary gas exchange and acid-base balance during exercise. *Comprehensive Physiology* **3**, 693–739 (2011).
64. Rapis, E. A change in the physical state of a nonequilibrium blood plasma protein film in patients with carcinoma. *Technical Physics* **47**, 510–512 (2002).
65. Shabalin, V. & Shatokhina, S. Diagnostic markers in the structures of human biological liquids. *Singapore medical journal* **48**, 440 (2007).
66. Martusevich, A. K., Zimin, Y. & Bochkareva, A. Morphology of dried blood serum specimens of viral hepatitis. *Journal of Hepatitis Monthly* **7**, 207–210 (2007).
67. Buzoverya, M., Shcherbak, Y. P., Shishpor, I. & Potekhina, Y. P. Microstructural analysis of biological fluids. *Technical Physics* **57**, 1019–1024 (2012).
68. Zeid, W. B., Vicente, J. & Brutin, D. Influence of evaporation rate on cracks' formation of a drying drop of whole blood. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **432**, 139–146 (2013).
69. Sikarwar, B. S., Roy, M., Ranjan, P. & Goyal, A. Automatic pattern recognition for detection of disease from blood drop stain obtained with microfluidic device. In *Advances in Signal Processing and Intelligent Recognition Systems*, 655–667 (Springer, 2016)
70. Bahmani, L., Neysari, M. & Maleki, M. The study of drying and pattern formation of whole human blood drops and the effect of thalassaemia and neonatal jaundice on the patterns. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **513**, 66–75 (2017).
71. Cameron, J. M., Butler, H. J., Palmer, D. S. & Baker, M. J. Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying. *Journal of biophotonics* **11**, e201700299 (2018).
72. Beaver, W. L., Wasserman, K. & Whipp, B. J. A new method for detecting anaerobic threshold by gas exchange. *Journal of applied physiology* **60**, 2020–2027 (1986).
73. Harrison, M. H. Effects on thermal stress and exercise on blood volume in humans. *Physiological Reviews* **65**, 149–209 (1985).
74. O'brien, F. The control of humidity by saturated salt solutions. *Journal of Scientific Instruments* **25**, 73 (1948).
75. Goehring, L., Clegga, W. & Routh, A. Wavy cracks in drying colloidal films. *Soft Matter Journal* **7**, 7984–7987 (2011).
76. Lazarus, V. & Pauchard, L. From craquelures to spiral crack patterns: influence of layer thickness on the crack patterns induced by desiccation. *Soft Matter Journal* **7**, 2552–2559 (2011).
77. Pauchard, L., Adda-Bedia, M., Allain, C. & Couder, Y. Morphologies resulting from the directional propagation of fractures. *Phys. Rev. E* **67** (2003).
78. Jing, G. & Ma, J. Formation of circular crack pattern in deposition self-assembled by drying nanoparticle suspension. *J. Phys. Chem. B* **116**, 6225–6231 (2012).
79. Sendova, M. & Willis, K. Spiral and curved periodic crack patterns in sol-gel films. *Applied Physics A* **76**, 957–959 (2003).
80. Neda, Z., Leung, K.-t., Jozsa, L. & Ravasz, M. Spiral cracks in drying precipitates. *Phys. Rev. Lett.* **88** (2002).
81. Goehring, L., Conroy, R., Akhter, A., Cleggb, W. J. & Routh, A. Evolution of mud-crack patterns during repeated drying cycles. *Soft Matter Journal* **6**, 3562–3567 (2010).
82. Burger, W. & Burge, M. *Digital Image Processing: An Algorithmic Introduction Using Java* (Springer-Verlage London, 2016).
83. Gonzalez, C. & Woods, R. E. *Digital Image Processing, Second Edition* (Pearson Education, 2008).

84. Ziou, D. & Tabbone, S. Edge detection techniques. International. *Journal of Pattern Recognition and Image analysis* **8**, 537–559 (1998).
85. Najarian, K. & Splinter, R. *Biomedical Signal and Image Processing, Second Edition* (Taylor and Francis Group, LLC, 2012).
86. Bisele, M., Bencsik, M., Lewis, M. C. & Barnett, C. T. Optimisation of a machine learning algorithm in human locomotion using principal component and discriminant function analyses. *PLOS ONE* **12**, 1–19 (2017).

Author contributions

S.I., M.A.J., G.R.S. and D.J.F. designed and carried out the experiments. L.H., M.B. and D.J.F. analysed the images and developed the machine learning algorithms. All authors contributed to writing, editing and approving the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-59847-x.

Correspondence and requests for materials should be addressed to L.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.