

OPEN

Machine Learning to Quantitate Neutrophil NETosis

Laila Elsherif^{1,9*}, Noah Sciaky², Carrington A. Metts¹, Md. Modasshir³, Ioannis Rekleitis³, Christine A. Burris⁴, Joshua A. Walker¹, Nadeem Ramadan¹, Tina M. Leisner¹, Stephen P. Holly⁵, Martis W. Cowles⁶, Kenneth I. Ataga^{7,10}, Joshua N. Cooper^{4,12} & Leslie V. Parise^{1,8,11,12}

We introduce machine learning (ML) to perform classification and quantitation of images of nuclei from human blood neutrophils. Here we assessed the use of convolutional neural networks (CNNs) using free, open source software to accurately quantitate neutrophil NETosis, a recently discovered process involved in multiple human diseases. CNNs achieved >94% in performance accuracy in differentiating NETotic from non-NETotic cells and vastly facilitated dose-response analysis and screening of the NETotic response in neutrophils from patients. Using only features learned from nuclear morphology, CNNs can distinguish between NETosis and necrosis and between distinct NETosis signaling pathways, making them a precise tool for NETosis detection. Furthermore, by using CNNs and tools to determine object dispersion, we uncovered differences in NETotic nuclei clustering between major NETosis pathways that is useful in understanding NETosis signaling events. Our study also shows that neutrophils from patients with sickle cell disease were unresponsive to one of two major NETosis pathways. Thus, we demonstrate the design, performance, and implementation of ML tools for rapid quantitative and qualitative cell analysis in basic science.

Machine learning (ML) is a branch of artificial intelligence built on the idea that computers can acquire knowledge through data and observations without explicit programming; they can then learn to generalize from examples and make predictions. ML has been used in analysis of genomics¹, drug discovery^{2,3}, modeling protein structures⁴ and disease diagnosis^{5,6}. Few studies have used automated imaging technologies with ML in non-diagnostic and exploratory research-focused efforts such as cell quantification in animal models. Cell classification is laborious, as it relies on heavily supervised image analysis tools with the need for continuous user interaction from image pre-processing steps through feature extraction to final classification/labeling. Although the reduced error rates and exceptional learning speeds of ML have the potential to transform the field of cellular imaging, they have not been widely adopted in the biological sciences, in part because of a lack of testing and validation of such methods due to shortages of large datasets for training. Convolutional neural networks (CNNs) are deep learning algorithms commonly used in image recognition and classification. Their structure was inspired by the structure of the mammalian visual cortex, whose function is pattern recognition and computing complex object attributes. Here we demonstrate for the first time the feasibility of designing different CNNs to address key questions relating to the timely topic of neutrophil NETosis.

NETosis (cell death by Neutrophil Extracellular Traps) is characterized by chromatin decondensation, nuclear swelling and rupture, followed by DNA release in a net-like structure⁷. It constitutes a response to danger signals and is a vital part of the immune system evolutionarily conserved across multiple organisms^{8–11}; however,

¹Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC, 27599, USA.

²Department of Pharmacology, University of North Carolina, Chapel Hill, NC, 27599, USA. ³Department of Computer Science and Engineering, College of Engineering and Computing, University of South Carolina, Columbia, SC, 29208, USA.

⁴Department of Mathematics, College of Arts and Sciences, University of South Carolina, Columbia, SC, 29208, USA. ⁵Department of Pharmaceutical Sciences, Campbell University, Buies Creek, NC, 27506, USA. ⁶EpiCypher, Inc. Durham, Durham, NC, 27709, USA. ⁷Department of Medicine, University of North Carolina, Chapel Hill, NC, 27599, USA.

⁸Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, 27599, USA. ⁹Present address: Department of Medicine, University of Tennessee Health Science Center, 956 Court Avenue B330, Memphis, TN, 38163-2116, USA.

¹⁰Present address: Department of Medicine, University of Tennessee Health Science Center, 956 Court Avenue D324, Memphis, TN, 38163-2116, USA. ¹¹UNC Blood Research Center, University of North Carolina, Chapel Hill, NC, 27599, USA. ¹²These authors jointly supervised this work: Joshua N. Cooper and Leslie V. Parise. *email: lelsheri@uthsc.edu

¹²These authors jointly supervised this work: Joshua N. Cooper and Leslie V. Parise. *email: lelsheri@uthsc.edu

excessive NETosis is associated with severe inflammation and disease progression, making it an important indicator, endpoint, and target in several clinical trials^{12–19}.

To date, at least two distinct *in vitro* NETosis mechanisms have been identified: one is mediated by the enzyme peptidyl arginine deiminase (PAD4) and another by reactive oxygen species (ROS) generated following protein kinase C (PKC) activation and the granular protease, neutrophil elastase (NE)^{20,21}. To decondense chromatin, PAD4 citrullinates histone arginine residues, causing a decrease in electrostatic interaction between DNA and histones; whereas NE degrades histone proteins. The exact role of ROS in NETosis is unknown. These two NETosis pathways utilize different biochemical mediators, for example histone citrullination is a marker for the PAD4-mediated but not the ROS-mediated pathway^{22,23}. Targeting PAD4 with pharmacological inhibitors is successful in inhibiting NETosis *in vitro*²⁴ and in animal models^{25–27}; however, no targeted therapies for the ROS-pathway exist^{27–31}. DNase-1 has also been used to digest the nuclear backbone of NETs and prevent ischemia-reperfusion (I/R) injury³², lung injury due to cigarette smoke or mechanical ventilation^{33,34}, and deep vein thrombosis³⁵. However, using DNase to dismantle NETs risks enhancing bacterial³⁶, fungal³⁷ and viral pathogen³⁸ virulence. A better strategy to inhibit NETosis would be to prevent chromatin decondensation, the defining characteristic of NETotic cells for both pathways. Chromatin decondensation underlies the change in nuclear shape from multilobulated to spherical/ovoid and is a specific marker for NETotic cells that is distinct from necrosis and apoptosis³⁹. The earliest attempts to quantitate NETosis involved measuring changes in the size and shape of neutrophil nuclei using fully-supervised image processing methods which suffer from the weaknesses of rule-based modeling^{40–43}. Rule-based modeling is the use of fixed, often highly complex, rules crafted by domain experts to make predictions or decisions – and, in contrast with statistical learning, is extremely labor-intensive, highly limited in its ability to generalize, and constrained by human cognitive biases and finite imagination⁴⁴.

There are two general strategies to quantitate NETosis and NETs. One focuses on the quantitation of the final product of NETosis, the extruded NETs and associated proteins by targeting citrullinated histones using ELISA, assessing the activity of DNA-associated NE and, image analysis to estimate the fraction of an area covered by NETs (NET area)^{45–48}. This strategy forms the basis of the two commercially available kits to quantitate NETosis. Coelho *et al.* were the first to introduce a high-throughput version of NET area analysis using supervised machine learning method⁴⁷. A limitation of this strategy is the increased risk for false positives given that DNA release also occurs following other processes such as necrosis. Additionally, in the absence of a specific marker such as citrullinated histone antibodies, quantitation of extruded NETs fails to provide information regarding the NETosis pathway involved.

The other main strategy to quantitate NETosis capitalizes on the discernable change that occurs in nuclear shape (image analysis of fixed and adherent fluorescently-labeled neutrophil DNA^{42,43}, flow cytometry⁴⁹ and imaging flow cytometry⁵⁰), which involves assessing the fraction of NETotic nuclei in a mixed cell population. With this second strategy, investigators have used both adherent and non-adherent neutrophils. Flow cytometry-based methods including imaging flow cytometry are high-throughput and automated but require non-adherent NETotic cells. There is increasing evidence showing that neutrophil adhesion is an essential component of NETosis and that certain inducers require neutrophil adhesion to a matrix before NETosis can occur^{51,52}. Therefore, one potential drawback to flow cytometry-based detection is the underestimation of the amount of NETosing cells. The method presented here falls within the second strategy of the quantitation of the number of NETotic nuclei in a mixed cell population. The improvement we introduce is the use of convolutional neural networks to classify NETotic and non-NETotic adherent neutrophils; efficiently process thousands of cell images captured by an automated microscope system; facilitate graded dose-response analyses; predict which of the two major NETosis pathways is involved; and distinguish NETotic from necrotic cells.

Results and Discussion

A large set of human-annotated neutrophil nuclei (103,874) were used to train and test two CNNs in identifying three classes of objects. Non-NETotic (class 1) nuclei are identified by the presence of lobules (2 or more) and bright fluorescent signal, whereas NETotic (class 2) nuclei are spherical and have a significantly diffuse fluorescent signal. The CNNs were trained by either pixel-level (PL), end-to-end or object-level (OL) classification techniques (Fig. 1A). The negative class for PL (class 0) was randomly selected pixels that did not have a signal (blank areas). For OL, class 0 denotes all objects not classified as 1 or 2, which may contain areas of spread NETs or nuclei not identifiable as one of the two classes. We have designed PL using an architecture with fewer hyper-parameters, enabling operations on commodity hardware without loss in classification accuracy. PL learns from annotated neutrophil nuclei without additional knowledge of the image or its objects. The first steps in PL training involves the building of a feature map or an activation map which is a representation of the input image (Supplementary Information, Fig. 1A). In contrast, OL requires high performance computing specialized hardware. Prior to training the OL network, all image objects were individually identified (Fig. 1A). The dataset was divided into a training set composed of 80% of the data, and a testing set composed of 20% of the data that served as out-of-sample (never “seen” by the network) examples for evaluation (Table 1). We assessed network accuracy, precision and recall using confusion matrices. We obtained accuracies of 98.9% and 94.2% respectively for PL and OL, signifying that if either CNN was given a new image of a nucleus, its prediction would be correct with the probabilities of 98.9% and 94.2% respectively (Fig. 1B, Matrices 1–2). Precision is a measure of how often the CNN is correct in predicting a positive event (true positive/true positive + false positive). Recall is an indicator of the sensitivity of the model and is calculated by taking the ratio of correct positive predications to the total actual positives (true positive/true positive + false negative). Given that our model is not binary, an example of what would constitute a true positive (TP), true negative (TN), false positive (FP) and false negative (FN) for class 1 is shown in Fig. 1B Matrix 3. PL achieved both high precision and recall values for classes 0–2. Although OL's precision and recall were comparable to PL's for classes 1 and 2, predicting class 0 was neither as precise nor as sensitive leading to the overall lower accuracy of OL. This slightly lower accuracy score is due to the more demanding discrimination task this model is trained to perform; class 0 for OL was a variegated combination of objects that are not labeled as

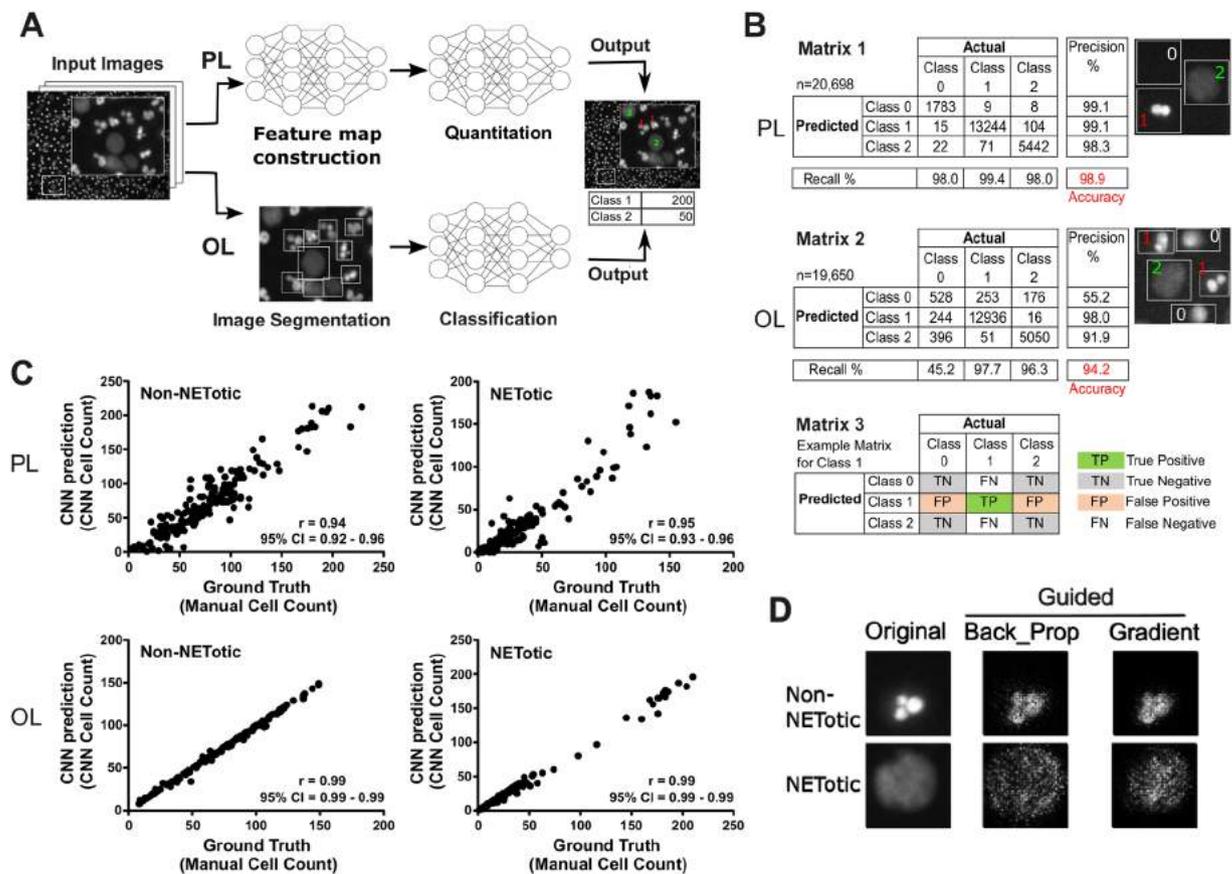


Figure 1. (A) The first panel represents one field (672×512 pixels of a 16-bit image) out of 36 fields in a typical image; the inset contains nuclei with the annotations 1 or 2 corresponding to non-NETotic and NETotic nuclei, respectively. The annotations were generated manually using ImageJ. The pixel-level (PL) classifier was trained by scanning the whole image (a total of 4032×3072 pixels) in 32×32 -pixel patches and classifying each patch as a class 1 or 2 using the annotations found on the image. The object-level (OL) classifier uses drawn bounding boxes of variable dimensions around all objects identified in the image and uses the object in the bounding box as training data. (B) Confusion matrices were used to evaluate model performance on the holdout dataset excluding the training dataset. The n numbers represent the holdout dataset only. The numbers in red denote model accuracy, which is the percentage of total correct predictions by the CNN. Recall is the number of true positives divided by true positive + false negative or the fraction of actual true positive predictions identified correctly. Precision is the number of true positive values divided by true positive + false positive or the fraction of positive identifications that were actually correct. An example of what would constitute true positive, true negative, false positive and false negative for class 1 is shown in matrix 3. Matrices 1 and 2 represent confusion matrices for PL and OL respectively. The nuclei images adjacent to the matrices indicate which nuclei were labeled as class 0, 1 or 2 by the two different CNN. The major difference in training of the CNNs is the class 0 category. (C) Pearson's correlation coefficient (r) was used to compare the quantification of PL and OL (CNN prediction) to that performed manually (ground truth). A total of 186 and 161 images containing hundreds of cells were quantified by PL and OL, respectively, with a confidence interval of 95%, $p < 0.0001$ for all obtained R values. Each dot represents an image that was counted manually and by a CNN for the total number of Non-NETotic (graphs on the left), and NETotic cells (graphs on the right) in an image. (D) Guided backpropagation as well as gradient-weighted class activation mapping were used to generate saliency maps evaluating the relative contributions of each pixel to the CNN's prediction. The brighter a pixel appears on this map, the more salient it is in identifying the phenotype and the more value it has in determining the CNN's prediction.

	Total No. of Nuclei used for:		
	Training	Validation	Testing
PL	83099	—	20774
OL	65323	17209	19650

Table 1. Breakdown of the number of nuclei used for training, validation and testing the PL and OL CNNs.

class 1 or class 2. PL and OL quantitation accuracy were compared to manually annotated images containing several hundred nuclei per image by plotting CNN prediction versus manual cell count and performing a correlation analysis (Pearson's $r \geq 0.94$, $p < 0.0001$ for all obtained r values) (Fig. 1C). Correlation analysis was also performed on quantitation of 44 images using ImageJ. The performance of ImageJ in the quantitation of NETotic nuclei was better than non-NETotic (Pearson's $r = 0.74$, 95% CI = 0.57–0.85 for NETotic and $r = 0.42$, 95% CI = 0.14–0.64 for non-NETotic, $p < 0.01$ for both values) as can be seen in Supplementary Information Fig. 2C. CNNs achieved much greater accuracy in quantitation of NETotic and non-NETotic nuclei as was expected.

Understanding how and why CNNs achieve such high performances is an active area of research involving approaches such as network output analysis and saliency maps. To investigate what the CNNs were learning that enabled accurate predictions, we obtained examples of image objects with low and high network output activation to gain insight into the models' understanding of the data (Supplementary Information, Fig. 1B). Objects with high network activation are those with high probability of belonging to their assigned class and are more representative of that class. Although useful in visualizing a prototype of each class; this method is not informative as it fails to describe the most identifying feature(s) captured by the network of each class. The saliency method attempts to identify salient regions in a visual object by assigning a feature importance to each pixel. We obtained class saliency maps by comparing the gradient of output class in relation to an input image, essentially determining which pixels/regions in the input image are most responsible for the networks' decision⁵³. Saliency maps are analogous to heat maps and are a mechanism to visualize what the network is learning. On the saliency map, brighter pixels are more critical for object recognition/classification. We used two techniques to generate saliency maps, using 10 representative nuclei of each class (NETotic and non-NETotic) (Fig. 1D and Supplementary Information, Fig. 1C for additional examples of each class). The saliency map of class 1 shows two or more closely clustered pixel groups that strongly resemble the original input image, whereas the class 2 map shows a more dispersed pixel distribution with slight emphasis on one relatively central region in the nucleus. This region could represent the innermost core of the chromatin that has not been completely decondensed.

As mentioned above, two distinct pathways of NETosis have been well identified, PAD4 and ROS, which can be activated by A23187 (a Ca^{2+} -ionophore that activates PAD4), and PMA (a PKC activator that mediates ROS formation) respectively. Histone citrullination is unique to NETosis agonists that can activate PAD4 by causing calcium release (Supplementary Information, Fig. 2A)⁵⁴. We hypothesized that in addition to the differences in their triggering and propagation mechanisms, the two pathways also might yield NETotic nuclei with different phenotypes that could be differentiated and quantified simultaneously by neural networks. This would forgo the need for specific staining for citrullinated histones (PAD4 pathway) or the colocalization of chromatin and neutrophil elastase staining (ROS pathway). We designed a CNN classifier to achieve this goal. The CNN was able to differentiate with 73% accuracy NETotic nuclei induced by either A23187 or PMA, even though nuclei appeared to be indistinguishable by traditional human microscopic image analysis (Fig. 2A, confusion matrix and images below).

We then addressed another important distinction required for accurate classification of neutrophil cell death, that of NETotic versus necrotic nuclei. Necrosis is generally identified as loss in cell membrane integrity and cell/nuclear swelling due to the influx of water and ions. Although specific information regarding chromatin changes during neutrophil necrosis are not available, it appears that chromatin condensation (commonly referred to as clumping) is a common occurrence in both apoptosis and necrosis⁵⁵. Clumping does not occur during NETosis but the final stages of both NETosis and necrosis are similar in that chromatin is released extracellularly. We have chosen to differentiate NETosis and necrosis based on nuclear morphology by treating neutrophils with two different stimuli for necrosis, hypotonic shock and freezing at -80°C ^{56,57}. Although the change in chromatin with hypotonic shock (example image in Supplementary Information, Fig. 1D) was easily distinguishable morphologically from that occurring during NETosis, the changes during freezing were less drastic and more like NETosis. Therefore, CNNs were trained to classify NETotic nuclei induced by A23187 or PMA and necrotic nuclei induced by freezing (Fig. 2B, Matrices 1–2). Remarkably, the two CNNs differentiated between NETotic and necrotic nuclei with $>99\%$ accuracy. Unlike the subtle differences between PMA- and A23187-induced NETotic nuclei, necrosis results in nuclei that are easily distinguishable from NETotic nuclei. Because CNNs can discriminate between PAD4- versus ROS-dependent NETosis and NETosis versus necrosis, they can potentially predict the signaling mechanisms associated with novel chemicals or disease states and facilitate the identification of appropriate inhibitors.

Having established highly consistent and accurate CNN performance, we implemented it to measure the concentration–response relationship of commonly used NETosis agonists in neutrophils isolated from healthy donors. We determined the EC50 values (concentrations resulting in 50% of NETotic nuclei) to be 2.1 nM for PMA and 900 nM for A23187 (Fig. 2C,D). Moreover, we found that DMSO, the vehicle for both agonists induces NETosis at $\geq 0.08\%$ v/v in culture media (Fig. 2E, graph and image), suggesting that treatment with agonists in solutions containing $\geq 0.08\%$ DMSO should be avoided since it could result in an additive or synergistic effect on NETosis. The ability to quantitate EC50 values allows for a more precise assessment of new NETosis inhibitors or agonist potency in a high-throughput platform. In addition to using CNNs to perform straightforward object quantitation, we attempted to understand the spatial distribution of NETotic cells upon treatment with the two agonists, PMA and A23187. One impetus for this analysis was the observation that low PMA concentrations caused the appearance of adjacent NETotic and non-NETotic cells, suggesting the existence of distinct neutrophil populations with differing thresholds for PMA-induced NETosis (Fig. 2F, top image). Low A23187 concentrations did not induce this effect, but did induce NETotic neutrophil clustering, suggesting the possible release of a paracrine factor that potentiates NETosis in neighboring cells (Fig. 2F, bottom image) or activates neutrophil adhesion. To translate this important observation into a quantitative measurement, we capitalized on the ability of our ML model to locate cells by phenotype in microscope images and record their bounding boxes, to derive a list of the NETotic cells' approximate geometric centers. These centers were then processed to compute the “average nearest

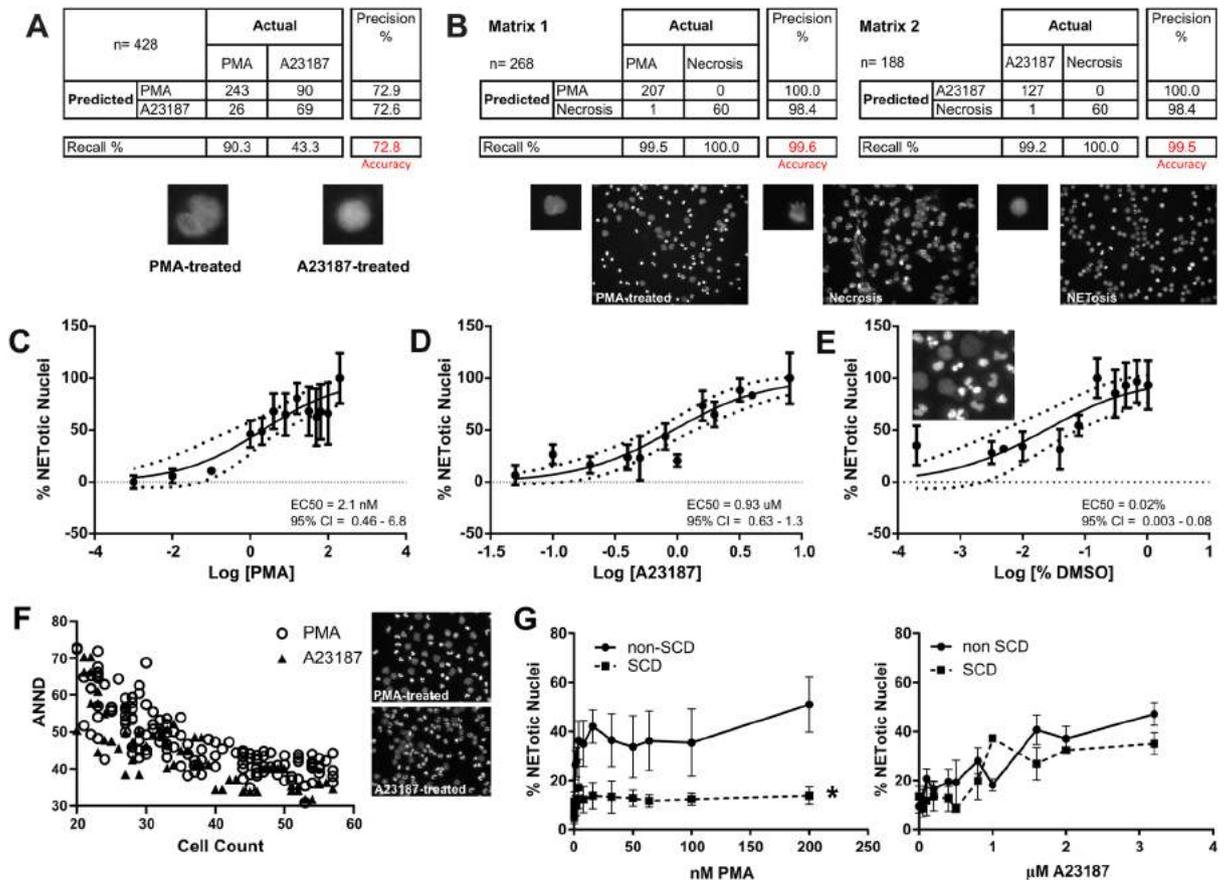


Figure 2. (A) A CNN trained on a total of 1286 individual NETotic nuclei (807 and 479 nuclei produced as a result of PMA or A23187 treatment, respectively). Treatment time was 120 min and we analyzed images where agonist concentration would result in 50% NETotic nuclei. Nuclei were cropped out of images in a bounding box of a fixed dimension and used for CNN training. The confusion matrix demonstrates CNN's ability to differentiate between the two nuclei with an accuracy of ~73%. The dataset used for testing ($n = 428$) represents 25% of the total number of images and is a subset that was excluded from the training set. An example of a PMA- and A23187-treated nucleus is shown below the matrix. (B) Necrosis was induced by freezing neutrophils at -80°C for 80 min on tissue culture plates. Comparison between necrotic and NETotic nuclei was performed similarly to that between PMA- and A23187-induced NETotic nuclei. A total of 244 necrotic, 506 A23187-induced and 833 PMA-induced NETotic nuclei were used for training the CNN. Although the number of necrotic nuclei used for training was relatively small, the CNNs achieved exceptionally high performance accuracies in differentiating between necrotic and NETotic nuclei as is seen in matrix 1 for PMA-treatment versus necrosis and matrix 2 for A23187-treatment versus necrosis. Images below the confusion matrices demonstrate the clear difference in appearance of NETotic and necrotic nuclei. (C) PMA was used to induce ROS-dependent NETosis and EC_{50} values were calculated as described in Methods. The percentage of DMSO in the highest agonist concentration was used as the vehicle control. EC_{50} was determined to be 2.1 nM for PMA and the 95% CI interval = 0.5–6.8, ($n = 11$). The 95% confidence bands are the dashed lines on the plot. (D) A23187 was used to induce PAD4-dependent NETosis and EC_{50} was determined to be 930 nM for A23187 with a 95% CI interval of 0.6–1.3, ($n = 8$). The percentage of DMSO in the highest agonist concentration was used as the vehicle control. The 95% confidence bands are the dashed lines on the plot. (E) DMSO is a widely used solvent for many NETosis agonists and inhibitors and we show that it acts as a NETosis agonist, reaching a maximum response of 30% at 0.16% DMSO ($n = 3$ –5 data points for each concentration used), 95% CI for $\text{EC}_{50} = 0.003$ –0.085. The dash lines represent 95% confidence bands. (F) Dispersion and clustering of NETotic nuclei differs depending on the type of treatment as can be seen in the images (right panels). CNNs were used to calculate and compare the clustering characteristics using Average Nearest Neighbor Distances which reveal significant differences ($p \leq 0.025$) between PMA and A23187 treatments. Images chosen for analysis were those containing similar numbers of NETotic nuclei. The two sets of ANND values were compared using Kolmogorov-Smirnov test, which was run 10 times to obtain the following p-values: 0.0000022, 0.018, 0.0050, 0.00082, 0.0048, 0.000046, 0.025, 0.000077, 0.016, 0.0063. The average p-value is 0.0076. (G) The same cell isolation and agonist conditions were used to treat neutrophils from patients with SCD at steady-state (squares, dashed lines, $n = 7$) and non-SCD (circles, solid line, $n = 11$). The number of nuclei analyzed following PMA treatment is 752,824 and 410,238 for non-SCD and SCD groups respectively. Following A23187 treatment the number of nuclei analyzed was 314,037 and 94,759 for non-SCD and SCD groups respectively. Neutrophils from patients with SCD responded poorly to PMA treatment (two-way ANOVA, $p < 0.05$) suggesting impairment in the ROS-dependent NETotic pathway (graph on the left); in contrast no significant difference was observed between SCD and non-SCD donors in response to A23187 treatment (graph on the right), suggesting that PAD4-mediated NETosis is unimpaired in SCD.

neighbor distance” (ANND), a well-known metric for gauging degrees of clustering or dispersion (dating back to at least 1954, see Clark-Evans 1954)⁵⁸. In general, if the ANNDs is small, the points (and therefore the cell centers) are more clustered whereas larger ANNDs indicate greater dispersion (Fig. 2F). Once the ANND statistics were computed for PMA- and A23187-treated NETotic cells, the two sets of numbers were compared using a general non-parametric, two-sample Kolmogorov-Smirnov⁵⁹ test, to determine whether two sets of values were drawn from the same distribution. The average p-value was 0.0076 for the null hypothesis that the two cell sets were equally clustered/dispersed, with no results exceeding a p-value of 0.025 (Fig. 2F). From this result we conclude that using PMA or A23187 as a NETosis agonist results in NETotic nuclei with different clustering/dispersion patterns.

We are interested in the potential for CNNs to study NETosis in patient populations. We chose patients with sickle cell disease (SCD) given that chronic inflammation and hypercoagulability are well known complications in SCD^{60,61} and that NETs are thought to be important for both processes^{13,62,63}. Previous studies report finding soluble components of DNA and nucleosomes in plasma from both SCD patients and humanized SCD mice¹⁶. Furthermore, plasma from SCD patients was found to cause NET production in neutrophils from healthy individuals¹⁶ leading to the conclusion that NETosis is associated with SCD pathophysiology. However, both studies use surrogate indicators for NETosis and did not measure NETosis directly in neutrophils isolated from SCD patients. Our technology allows us to quantitatively assess NETosis in neutrophils of patients with SCD at steady state (episodes where patients are not experiencing pain and other symptoms associated with crisis). SCD neutrophils responded similarly to neutrophils from non-SCD donors to A23187, indicating functional PAD4-dependent NETosis (Fig. 2G). In contrast, SCD neutrophils were insensitive to PMA (Fig. 2G), suggesting impairment in their ROS-dependent NETosis potential. Our results agree with previous studies reporting that neutrophils from SCD patients have reduced capacity for oxidative burst and therefore a reduction in ROS^{64–68}. However there are also reports of increased oxidative burst and ROS generation in neutrophils in patients with SCD⁶⁹. The decreased potential for NETosis in neutrophils from patients with SCD could explain their increased susceptibility to certain invasive bacterial infections as their neutrophils are unable to NETose and trap and kill bacteria efficiently. Our results could also reflect the response of neutrophils from patients with SCD to hydroxyurea or other pharmacological interventions. Additional studies with a larger patient cohort are certainly warranted and could elucidate important aspects in SCD disease mechanisms.

The method described herein is a major improvement over other available microscopy-based methods in NETosis research; it is far less laborious than manual object detection and classification, much more robust than semi-automated image analysis approaches, and substantially more accurate than the use of surrogate biomarkers. The simple (PL) and advanced (OL) classification models performed similarly well, suggesting that more complex CNNs are not needed for comparable classification tasks. In addition to straightforward quantitation of the percentage of NETotic nuclei in images, a task requiring substantial human effort and time, we demonstrate the ability of CNNs to perform tasks that would be very challenging if not impossible for the human eye. We demonstrate the application of CNNs in addressing both basic science and pre-clinical questions such as dose-response relationships of common NETosis inducers and the potential of patient neutrophils to NETose. Our analysis is especially useful in the quantitation of NETotic nuclei *in vitro* for graded dose-response relationships whereas analyses of the NET area, defined as the fraction of the regions in an image covered by NETs is more suitable for NET degradation and an all-or-none response⁴⁷. A significant fraction of neutrophils have to extrude their DNA for any meaningful reading of NET area estimation which leads to decreased sensitivity of the method. Furthermore, the fact that extruded DNA can result from necrosis as well as NETosis, the measurements of extracellular DNA lack specificity for NET DNA. Specificity is vastly improved by using NET-specific markers such as an antibody targeting citrullinated histones; however, not all NETosis agonists cause histone citrullination^{22,23}. Investigators have used traditional flow cytometry and imaging flow cytometry systems such as ImageStream to quantitate NETotic nuclei that are in suspension and not adhered onto a plate or a matrix^{49,50,70}. There is a high risk of underestimating the percentage of NETotic neutrophils using imaging flow cytometry due to the reduced ability of non-adherent neutrophils to NETose after stimulation with certain agonists. It was shown that the bacterium *Acinetobacter baumannii* inhibits NETosis in human neutrophils by reducing neutrophil adhesion to glass slides⁵¹. Furthermore, it appears that some NETosis inducers such as LPS are dependent on neutrophil adhesion and substrate elasticity whereas others such as PMA are not⁵². Treating human neutrophils with a function blocking antibody to the integrin Mac-1 ($\alpha M\beta_2$) results in reduced NET deployment; whereas mice lacking the β_2 subunit of the integrin adhesion receptor have a reduced NETosis response to hantavirus infection^{71,72}. Similarly blocking of the integrin LFA-1 ($\alpha L\beta_2$) inhibits NETosis following endotoxin challenge⁷³. DNA Area and NETosis Analysis (DANA) is an area-based analysis that is the most similar to the one proposed in the current work as it detects both NETotic nuclei and NET area⁴¹. DANA uses ImageJ and Java to automate image segmentation and nuclei quantitation. As discussed in the introduction, ImageJ analysis falls into the category of rule-based modeling and is highly dependent on operator-defined features and continuous operator supervision and adjustments.

Machine learning introduces much needed impartiality in image analysis by allowing the machine to determine all the features specific to NETotic and non-NETotic nuclei. By choosing to focus on the unique nuclear shape change that occurs in NETosis, we eliminate the need for a universal NETosis-specific marker, and we also facilitate analysis of graded pharmacological responses. To ensure the success of the current method, NETosing neutrophils should only be treated for the minimum time required to cause nuclear shape change. This requirement is easily met given that on average the NETosis process takes 4 hrs to reach completion (extrusion of NETs) *in vitro*. Longer treatment times or very potent stimuli may lead to the release of DNA before nuclear shape change is captured. Recent findings reveal that many scientific discoveries made using machine learning have been difficult to reproduce due to the limited volume of data used for machine learning training. We believe that our models performed exceptionally well due to the large volume of annotated data used for training the CNNs⁷⁴. When such data volumes are not attainable, data augmentation techniques such as subtle image rotations can be employed⁷⁵.

Methods

Sample collection. All subjects gave written informed consent in accordance with the Declaration of Helsinki. Consent was obtained by trained research personnel. The study was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill (UNC) under IRB # 17-3148. Patients with SCD at steady state were recruited from the SCD clinic at UNC. Blood samples were collected via venipuncture from healthy volunteers in the research laboratory of LVP by trained personnel/phlebotomist or donated by the UNC Healthcare Blood Donation Center located at the NC Cancer Hospital. Approximately 5 mL of blood was collected in vacutainers coated with EDTA (BD 367863). Inclusion and exclusion criteria for both SCD and non-SCD volunteers are outlined below.

SCD. Inclusion:

- Diagnosis of SCD (genotype SS, SC or S/ β^0 thalassemia).
- Non-crisis at time of enrollment and no acute pain episodes in the previous 4 weeks.

Exclusion:

- Inability to provide informed consent based on the judgment of study personnel. <18 years of age.
- Anticoagulation therapy
- Recent history of hemoglobin levels of <6 g/dL
- Less than 3 months have transpired since last transfusion.
- Pregnancy

Non-SCD. Inclusion:

- >18 years of age.
- Healthy and of either sex

Exclusion:

- Inability to provide informed consent based on the judgment of study personnel. <18 years of age.
- Anticoagulation therapy
- Taking Tylenol and/or Advil within 3–4 days prior to blood draw
- Pregnancy

Neutrophil isolation and treatment. Following collection, the blood sample was handled under sterile conditions throughout treatment and up to cell fixation to prevent activation during the isolation process. Neutrophil isolation was performed using EasySep Direct Human Neutrophil Isolation Kit (STEMCELL Technologies, British Columbia, Canada, 19666) based on negative selection from whole blood. In brief, unwanted cells are cross-linked to magnetic particles via a tetrameric antibody complex. The isolation procedure is rapid (approximately 30 min). An average of 4.4×10^7 and 5.5×10^7 total neutrophils can be isolated from 5 mL of blood from non-SCD and SCD donors, respectively. Neutrophil isolation from SCD patients by density gradient centrifugation is challenging due to the altered rheology of SCD blood and the ineffectiveness of RBC sedimentation by dextran due to the low RBC aggregation index and rate in SCD patients.

Neutrophils were plated at a density of 3×10^6 /well in 96-well plates (Greiner Bio-One, Germany, 655986). This seeding density is optimal for allowing adequate spacing between nuclei once chromatin decondensation takes place and facilitates object identification and image analysis by CNNs. Cells were suspended in RPMI 1640 L-glutamine media (Gibco/Thermo Fisher Scientific, New York, USA, 11875-093) supplemented with 0.5% FBS (Gibco/Thermo Fisher Scientific, Grand Island, USA, 10438018) and incubated at 37°C, 5% CO₂. Cells would adhere during the first hr followed by treatment with vehicle or agonist for an additional 2 hrs; the total time neutrophils spend in culture is 3 hrs. Each assay condition was run in duplicate and for each well treated with a designated concentration of agonist, a neighboring well was treated with vehicle only. This design allowed for the determination of the effect of vehicle alone on NETosis. Cells were treated with lower than or equal to the maximum percentage of DMSO that did not cause significant NETosis. Necrosis was induced using freezing at –80°C or hypotonic shock. Following agonist incubation, the supernatant was removed, and neutrophils were fixed with 4% paraformaldehyde for 20 min before staining.

Cell staining and imaging. Following fixation, nuclei were stained using 0.05 μ M of the nucleic acid stain SYTOX Green (Invitrogen/Molecular Probes, Oregon, USA, S7020). The BD Pathway Bioimaging 855 system was used to obtain 36 total adjacent images/well. Acquired images were stitched together by BD AttoVision 1.6 acquisition software to yield one image of 4032 \times 3072-pixel dimensions that represents approximately 16% of the total surface area of the well in a 96-well plate. Staining for citrullinated histone H3, myeloperoxidase (MPO) and neutrophil elastase (NE) were performed using ab5103, ab9535, and ab68672, respectively (Abcam, Cambridge, MA, USA). A goat anti rabbit Alexa 594 secondary antibody was used for visualization. Image acquisition for double staining was performed using an Olympus IX81 inverted wide field microscope.

Cell annotation. A total of 103,874 individual nuclei were manually annotated using ImageJ by 4 different investigators blinded to the treatment groups. Approximately 80% of this annotated dataset was used to train the developed CNN classifiers. Annotated neutrophil nuclei were designated as type 1 or non-NETotic if clear normal nuclear segmentation is present whereas type 2 or NETotic nuclei were labeled based on lack of clear nuclear segmentation and diffuse SYTOX Green stain. Nuclei that could not be identified as belonging to either of the two classes or spread NETs were labeled with either 4 or 5, respectively. The CNN classifiers were not trained and did not learn to identify classes 4 and 5 as we opted to train the classifiers on objects that could easily be differentiated.

ImageJ analysis. Image processing of wide-field image sets was accomplished using automated scripts using a freely available software platform ImageJ (version 1.52o). For ImageJ, each FITC image in a directory was opened sequentially, background subtraction applied to correct uneven illumination, and primary objects were defined using Isodata segmentation. Small objects (having areas under 100 pixels) and cells touching the image periphery were removed from data set. The resulting objects were divided into two groups by defining large diffuse cells having a solidity value (Area/Convex Area) of above 5.677 and smaller cells below that threshold (Supplementary Information, Fig. 2C).

PL CNN analysis. Note: A glossary of key terms used in describing the CNN methodology is included in Supplementary Information.

The Pixel-Level (PL) convolutional neural network (CNN) analysis proceeded in two stages.

Stage 1. Cell recognition/discrimination training. Data Preprocessing: First, patches of each 512×672 image indicating NETotic, non-NETotic, or negative class (neither phenotype) were extracted by centering a 32×32 box at each annotation mark (for the positive classes) and a random 32×32 box not containing an annotation mark (for the negative class). The result was 28865 negative class images and 75009 positive class images, for a total of 103874 images, which were then split evenly into 80% training and 20% testing (hold-out) data.

Network Description: The proposed network used an architecture inspired by VGGNet⁶ although substantially shallower to allow for training on commodity laptop hardware and to limit model complexity. There were four convolutional layers (64 filters, each 3×3), with 2×2 max-pooling and 25% dropout used between the second and third layers. The last convolutional layer was followed by a 256-node fully connected layer, and finally by a 3-node fully connected layer to indicate the three class probabilities. All activation functions were rectified linear (“ReLU”) except for the last layer, which was softmax. Weights were initialized randomly before training.

Training: The training dataset did not use augmentation, and categorical cross-entropy was used as the loss function. The network was trained with a learning rate of 0.001, a decay of $1e-7$, and a momentum of 0.9.

Stage 2. Training for cell quantitation. Data Preprocessing: The original, unannotated images (training and testing) were then processed by applying the network trained in Stage 1 to every 32×32 patch of the image, which generated $480 \times 640 \times 3$ feature maps: 740 for training and 186 for testing. The generated feature maps were associated with counts by phenotype from the original annotations.

Network Description: The proposed network again used an architecture inspired by VGGNet, but even shallower than in Stage 1 in order to limit model complexity because of the small training set size. There were two convolutional layers (32 filters, each 3×3), with 2×2 average-pooling used between them and 2×2 max-pooling used after them. This was followed by a 256-node fully connected layer, and finally by a 2-node fully connected layer to indicate the two object counts. All activation functions were rectified linear (“ReLU”).

Training: Mean absolute error was used as the loss function. The network was trained with a learning rate of 0.001, a decay of $1e-7$, and a momentum of 0.9. The network was trained for 30 epochs.

Overall testing: The 186 feature maps generated from test data in Stage 2 were processed by the cell quantitation CNN generated in Stage 2, and the CNN-generated counts (estimated counts) were compared to the manually annotated images (ground-truth).

OL CNN analysis. The pipeline consists of preprocessing both annotations and images and training the proposed network to classify the extracted patches of the cells.

Dataset preprocessing. Repeated erosion and dilation operations remove existing noise and clarify the boundary between the cells. Afterwards, all the connected regions and their centers are detected. At this point, a 112×112 patch is utilized, and the patch was labeled according to the annotation point closest to the center of the patch. The entire dataset was split into 65:15:20 ratio to create training, validation and testing datasets respectively. As part of image preprocessing, all images were normalized by subtracting the mean of image matrix from image matrix and dividing the result by the standard deviation of image matrix.

Network description. The proposed network was inspired by DenseNet⁷⁷, in particular, the 121-layer configuration. The last layer of the DenseNet-121 was replaced by a fully connected layer with 3 hidden units in order to output 3 classes (Class 0, 1 and 2) of nuclei. All layers except the last fully connected layer were initialized with weights pre-trained on ImageNet and the last layer was randomly initialized.

Training. The training dataset was augmented by applying random rotations between -10 degrees to $+10$ degrees and horizontal/vertical flips in addition to original patches. Finally, each image was resized to 112×112 pixels, which was the input of the designed CNN training network. The network was trained with a learning rate of 0.001 and momentum of 0.9. The learning rate was reduced by $1e-5$ for every 10 iterations. The network was trained until it starts to over fit, *i.e.*, the validation loss starts increasing when the training loss is decreasing.

Testing. For testing, images not used in training and never seen by the CNN were used. The images were only normalized and fed to the network. No data augmentation was applied in the testing phase.

Table 1 summarizes the breakdown of the number of data points used for training/validation and testing of each model.

Phenotype discrimination convolutional neural network (CNN). *Data preprocessing.* First, 100×100 patches of each image indicating the location of a cell were extracted by manually cutting out a box around each patch. A total of 1,076 PMA-treated cell images and 638 A23187-treated cell images were each split evenly into 75%/25% training and testing (hold-out) data.

Network description. The proposed network used an architecture inspired by VGGNet⁷⁶, although substantially shallower to allow for training on commodity laptop hardware and to limit model complexity. There were nine convolutional layers (64 filters, each 3×3), with 2×2 max-pooling and 25% dropout used after each three layers. The last convolutional layer was followed by a 256-node fully connected layer, and finally by a 2-node fully connected layer to indicate the two class probabilities. All activation functions were rectified linear (“ReLU”) except for the last layer, which was softmax. Weights were initialized randomly before training.

Training. The training dataset used no image augmentation, and categorical cross-entropy was used as the loss function. The network was trained with a learning rate of 0.001, a decay of $1e-7$, and a momentum of 0.9. The network was trained for 100 epochs.

Testing. The 428 test images were processed by the trained CNN, and the class probabilities were set at a threshold of 50% to construct a confusion matrix.

We have posted the source code at <https://github.com/Modasshir/cell-classification-journal>. The source code contains training and testing code along with instructions. The trained weights for the entire dataset can also be downloaded to verify the accuracy on a sample dataset included in the project on the GitHub. The source code has been validated by an independent investigator.

Saliency maps. We employed techniques described previously to generate saliency maps, a method for visualizing the internal “thought process” of a CNN⁷⁸. Saliency maps assign to each pixel of an input image a value indicating the amount of influence on the final classification decision that pixel had; thus, highlighting regions of the image that were important in the decision-making process.

We used two methods: “Guided Backpropagation Class Activation Mapping” (“Guided Backprop”) and “Guided Gradient-weighted Class Activation Mapping” (“Guided Grad-CAM”) to calculate saliency maps. The former method identifies pixels which were highly influential for the final classification probability by tracing derivatives of node activations backwards through the network from the positive class node while discarding negative influences (*i.e.*, derivatives). The latter method adds additional information to this process by also measuring the extent to which pixels occurred in an image region with a cumulatively positive effect on the final classification probability. Guided Grad-Cam is one extra step after Guided Backprop. After backpropagation, we receive the contribution of each image pixel both positive and negative. However, in the grad-cam step, negative contributing pixels are removed.

Analysis of dose-response data. Datapoints were plotted as log [agonist concentration] versus response (% NETotic nuclei). All data were normalized to a starting response of 0 and a plateau of 100%, which were defined by the smallest and the largest means of the dataset, respectively. The 95% confidence bands are represented by the dashed lines and the 95% confidence interval are displayed on the plot. The range of concentrations used for DMSO were 0.0002% to 1.04% in media, ($n = 3-5$); PMA was used at 0.001–200 nM, ($n = 11$); and A23187 was used at 0.05–8 μ M, ($n = 8$). The percentage of DMSO found in the highest agonist concentration was used as the vehicle control for PMA and A23187 groups. Comparison between SCD ($n = 7$) and non-SCD ($n = 11$) dose-response curves was performed using 2-way ANOVA using the mixed-effects model due to the presence of missing datapoints. The total number of nuclei analyzed for the non-SCD control group is 1,066,861 and for the SCD group is 504,997. Interactions between dose and genotype were deemed significant if $p < 0.05$. GraphPad Prism version 8.1 was used for generating graphs and statistical analysis.

Computation of average nearest neighbor distance. Average nearest neighbor distance was calculated to quantitate agonist-induced neutrophil clustering. Because all images are the same size, no normalization is needed for this statistic, simply the average, across all points, of the ordinary Euclidean distance to their nearest neighbor. However, because the number of cells in an image will affect the ANND, the smaller of the two data sets (A23187) was resampled using the larger dataset (PMA) so that the cell-count marginals of the two-dimensional distribution of (cell-count, ANND) pairs were the same.

For ANND analysis, 24 PMA images and three A23187 images (4032×3072 resolution), were each cut into 6×7 rectangles for processing, resulting in 1008 A23187 images and 125 PMA images, since some images had no cells visible. These smaller images were then processed by the CNN to perform instance segmentation and extract the locations of the cell centers for each phenotype. Any images containing fewer than 20 NETotic cells were discarded to reduce noise, resulting in 709 A23187 images and 48 PMA images. We then computed the average nearest neighbor distances of NETotic cell centers in each image efficiently using the `NearestNeighbors.fit()` function from the Python package `scikit-learn`. At this point, we had a pair of numbers for each image: a cell count and an ANND. The pairs of PMA output numbers were culled so that it had approximately same n -size as the much smaller A23187 image set, uniformly at random with probability equal to the ratio of the image set sizes. The maximum of each set’s

cell count was computed, the minimum of the two results was calculated, then all pairs (cell-count, ANND) from the two sets with cell-count larger than this max were removed. We resample the A23187 point set S in order to ensure its cell-count (empirical) marginal distribution matches that of the PMA point set. In particular, we create a new point set S' from S by taking each point P from the PMA set, finding a point Q in S whose cell count is as close as possible to that of P (ties broken randomly), and then inserting Q into S.

Declaration of interests. EpiCypher is a technology development company that does not intend to commercialize the present technology currently.

Data availability

The datasets generated during and/or analyzed as part of the current study are available from the corresponding author on reasonable request.

Received: 5 July 2019; Accepted: 27 October 2019;

Published online: 15 November 2019

References

- Zou, J. *et al.* A primer on deep learning in genomics. *Nat Genet* **51**, 12–18, <https://doi.org/10.1038/s41588-018-0295-5> (2019).
- Klambauer, G., Unterthiner, T., Mayr, A. & Hochreiter, S. DeepTox: Toxicity prediction using deep learning. *Toxicol Lett* **280**, S69–S69, <https://doi.org/10.1016/j.toxlet.2017.07.175> (2017).
- Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. *J Chem Inf Model* **55**, 263–274, <https://doi.org/10.1021/ci500747n> (2015).
- AlQuraishi, M. End-to-End Differentiable Learning of Protein Structure. *Cell Syst* **8**, 292–301 e293, <https://doi.org/10.1016/j.cels.2019.03.006> (2019).
- Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118, <https://doi.org/10.1038/nature21056> (2017).
- Guncar, G. *et al.* An application of machine learning to haematological diagnosis. *Sci Rep* **8**, 411, <https://doi.org/10.1038/s41598-017-18564-8> (2018).
- Brinkmann, V. *et al.* Neutrophil extracellular traps kill bacteria. *Science* **303**, 1532–1535, <https://doi.org/10.1126/science.1092385> (2004).
- Fuchs, T. A. *et al.* Novel cell death program leads to neutrophil extracellular traps. *J Cell Biol* **176**, 231–241, <https://doi.org/10.1083/jcb.200606027> (2007).
- Hawes, M. C. *et al.* Extracellular DNA: the tip of root defenses? *Plant Sci* **180**, 741–745, <https://doi.org/10.1016/j.plantsci.2011.02.007> (2011).
- Altincicek, B., Stotzel, S., Wygrecka, M., Preissner, K. T. & Vilcinskas, A. Host-derived extracellular nucleic acids enhance innate immune responses, induce coagulation, and prolong survival upon infection in insects. *J Immunol* **181**, 2705–2712, <https://doi.org/10.4049/jimmunol.181.4.2705> (2008).
- Goldmann, O. & Medina, E. The expanding world of extracellular traps: not only neutrophils but much more. *Front Immunol* **3**, 420, <https://doi.org/10.3389/fimmu.2012.00420> (2012).
- Clark, S. R. *et al.* Platelet TLR4 activates neutrophil extracellular traps to ensnare bacteria in septic blood. *Nat Med* **13**, 463–469, <https://doi.org/10.1038/nm1565> (2007).
- Fuchs, T. A. *et al.* Extracellular DNA traps promote thrombosis. *Proceedings of the National Academy of Sciences* **107**, 15880–15885, <https://doi.org/10.1073/pnas.1005743107> (2010).
- Yoo, D. G. *et al.* Release of cystic fibrosis airway inflammatory markers from Pseudomonas aeruginosa-stimulated human neutrophils involves NADPH oxidase-dependent extracellular DNA trap formation. *J Immunol* **192**, 4728–4738, <https://doi.org/10.4049/jimmunol.1301589> (2014).
- Khandpur, R. *et al.* NETs are a source of citrullinated autoantigens and stimulate inflammatory responses in rheumatoid arthritis. *Sci Transl Med* **5**, 178ra140, <https://doi.org/10.1126/scitranslmed.3005580> (2013).
- Chen, G. *et al.* Heme-induced neutrophil extracellular traps contribute to the pathogenesis of sickle cell disease. *Blood* **123**, 3818–3827, <https://doi.org/10.1182/blood-2013-10-529982> (2014).
- Hakkim, A. *et al.* Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. *Proc Natl Acad Sci USA* **107**, 9813–9818, <https://doi.org/10.1073/pnas.0909927107> (2010).
- Vitkov, L., Klappacher, M., Hannig, M. & Krautgartner, W. D. Extracellular neutrophil traps in periodontitis. *J Periodontol Res* **44**, 664–672, <https://doi.org/10.1111/j.1600-0765.2008.01175.x> (2009).
- Saitoh, T. *et al.* Neutrophil Extracellular Traps Mediate a Host Defense Response to Human Immunodeficiency Virus-1. *Cell host & microbe* **12**, 109–116, <https://doi.org/10.1016/j.chom.2012.05.015> (2012).
- Wang, Y. *et al.* Histone hypercitrullination mediates chromatin decondensation and neutrophil extracellular trap formation. *J Cell Biol* **184**, 205–213, <https://doi.org/10.1083/jcb.200806072> (2009).
- Papayannopoulos, V., Metzler, K. D., Hakkim, A. & Zychlinsky, A. Neutrophil elastase and myeloperoxidase regulate the formation of neutrophil extracellular traps. *J Cell Biol* **191**, 677–691, <https://doi.org/10.1083/jcb.201006052> (2010).
- Kenny, E. F. *et al.* Diverse stimuli engage different neutrophil extracellular trap pathways. *Elife* **6**, <https://doi.org/10.7554/eLife.24437> (2017).
- Neeli, I. & Radic, M. Opposition between PKC isoforms regulates histone deimination and neutrophil extracellular chromatin release. *Front Immunol* **4**, <https://doi.org/10.3389/fimmu.2013.00038> (2013).
- Lewis, H. D. *et al.* Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation. *Nature Chemical Biology* **11**, 189–+, <https://doi.org/10.1038/nchembio.1735> (2015).
- Biron, B. M. *et al.* Cl-Amidine Prevents Histone 3 Citrullination and Neutrophil Extracellular Trap Formation, and Improves Survival in a Murine Sepsis Model. *J Innate Immun* **9**, 22–32, <https://doi.org/10.1159/000448808> (2017).
- Novotny, J. *et al.* Histological comparison of arterial thrombi in mice and men and the influence of Cl-amidine on thrombus formation. *Plos One* **13**, <https://doi.org/10.1371/journal.pone.0190728> (2018).
- Kusunoki, Y. *et al.* Peptidylarginine Deiminase Inhibitor Suppresses Neutrophil Extracellular Trap Formation and MPO-ANCA Production. *Front Immunol* **7**, 1–7, <https://doi.org/10.3389/fimmu.2016.00227> (2016).
- Chumanevich, A. A. *et al.* Suppression of colitis in mice by Cl-amidine: a novel peptidylarginine deiminase inhibitor. *Am J Physiol Gastrointest Liver Physiol* **300**, G929–938, <https://doi.org/10.1152/ajpgi.00435.2010> (2011).
- Knight, J. S. *et al.* Peptidylarginine Deiminase Inhibition Reduces Vascular Damage and Modulates Innate Immune Responses in Murine Models of Atherosclerosis. *Circ Res* **114**, 947–956, <https://doi.org/10.1161/Circresaha.114.303312> (2014).
- Knight, J. S. *et al.* Peptidylarginine Deiminase Inhibition Prevents Neutrophil Extracellular Trap Formation, Modulates Innate Immune Responses, and Reduces Vascular Damage In Mice. *Arthritis Rheum* **65**, S490–S490 (2013).

31. Madhi, R., Rahman, M., Taha, D., Morgelin, M. & Thorlacius, H. Targeting peptidylarginine deiminase reduces neutrophil extracellular trap formation and tissue injury in severe acute pancreatitis. *J Cell Physiol* **234**, 11850–11860, <https://doi.org/10.1002/jcp.27874> (2019).
32. Wang, S. K. *et al.* DNase-1 Treatment Exerts Protective Effects in a Rat Model of Intestinal Ischemia-Reperfusion Injury. *Sci Rep-Uk* **8**, <https://doi.org/10.1038/s41598-018-36198-2> (2018).
33. King, P. T. *et al.* Deoxyribonuclease 1 reduces pathogenic effects of cigarette smoke exposure in the lung. *Sci Rep-Uk* **7**, <https://doi.org/10.1038/s41598-017-12474-5> (2017).
34. Yildiz, C. *et al.* Mechanical Ventilation Induces Neutrophil Extracellular Trap Formation. *Anesthesiology* **122**, 864–875, <https://doi.org/10.1097/Aln.0000000000000605> (2015).
35. Brill, A. *et al.* Neutrophil extracellular traps promote deep vein thrombosis in mice. *J Thromb Haemost* **10**, 136–144, <https://doi.org/10.1111/j.1538-7836.2011.04544.x> (2012).
36. Sumbly, P. *et al.* Extracellular deoxyribonuclease made by group A Streptococcus assists pathogenesis by enhancing evasion of the innate immune response. *Proc Natl Acad Sci USA* **102**, 1679–1684, <https://doi.org/10.1073/pnas.0406641102> (2005).
37. Park, H. J. *et al.* A DNase from a Fungal Phytopathogen Is a Virulence Factor Likely Deployed as Counter Defense against Host-Secreted Extracellular DNA. *Mbio* **10**, <https://doi.org/10.1128/mBio.02805-18> (2019).
38. Schattgen, S. A., Gao, G. P., Kurt-Jones, E. A. & Fitzgerald, K. A. Cutting Edge: DNA in the Lung Microenvironment during Influenza Virus Infection Tempers Inflammation by Engaging the DNA Sensor AIM2. *J Immunol* **196**, 29–33, <https://doi.org/10.4049/jimmunol.1501048> (2016).
39. Neubert, E. *et al.* Chromatin swelling drives neutrophil extracellular trap release. *Nat Commun* **9**, <https://doi.org/10.1038/s41467-018-06263-5> (2018).
40. Hakkim, A. *et al.* Activation of the Raf-MEK-ERK pathway is required for neutrophil extracellular trap formation. *Nat Chem Biol* **7**, 75–77, <https://doi.org/10.1038/nchembio.496> (2011).
41. Rebernick, R. *et al.* DNA Area and NETosis Analysis (DNA): a High-Throughput Method to Quantify Neutrophil Extracellular Traps in Fluorescent Microscope Images. *Biol Proced Online* **20**, 7, <https://doi.org/10.1186/s12575-018-0072-y> (2018).
42. Mohanty, T., Sorensen, O. E. & Nordenfelt, P. NETQUANT: Automated Quantification of Neutrophil Extracellular Traps. *Front Immunol* **8**, 1999, <https://doi.org/10.3389/fimmu.2017.01999> (2017).
43. Brinkmann, V., Goosmann, C., Kuhn, L. I. & Zychlinsky, A. Automatic quantification of *in vitro* NET formation. *Front Immunol* **3**, 413, <https://doi.org/10.3389/fimmu.2012.00413> (2012).
44. Walsh, J. O. M. *et al.* In *Computer Vision Conference*. (Las Vegas, Nevada, United States, 2019).
45. Santos, A. *et al.* NETs detection and quantification in paraffin embedded samples using confocal microscopy. *Micron* **114**, 1–7, <https://doi.org/10.1016/j.micron.2018.07.002> (2018).
46. Filio-Rodriguez, G. *et al.* *In vivo* induction of neutrophil extracellular traps by Mycobacterium tuberculosis in a guinea pig model. *Innate Immun* **23**, 625–637, <https://doi.org/10.1177/1753425917732406> (2017).
47. Coelho, L. P. *et al.* Automatic determination of NET (neutrophil extracellular traps) coverage in fluorescent microscopy images. *Bioinformatics* **31**, 2364–2370, <https://doi.org/10.1093/bioinformatics/btv156> (2015).
48. Thalín, C. *et al.* Validation of an enzyme-linked immunosorbent assay for the quantification of citrullinated histone H3 as a marker for neutrophil extracellular traps in human plasma. *Immunol Res* **65**, 706–712, <https://doi.org/10.1007/s12026-017-8905-3> (2017).
49. Zharkova, O. *et al.* A Flow Cytometry-Based Assay for High-Throughput Detection and Quantification of Neutrophil Extracellular Traps in Mixed Cell Populations. *Cytometry A* **95**, 268–278, <https://doi.org/10.1002/cyto.a.23672> (2019).
50. Ginley, B. G. *et al.* Computational detection and quantification of human and mouse neutrophil extracellular traps in flow cytometry and confocal microscopy. *Sci Rep* **7**, 17755, <https://doi.org/10.1038/s41598-017-18099-y> (2017).
51. Kamoshida, G. *et al.* Pathogenic Bacterium Acinetobacter baumannii Inhibits the Formation of Neutrophil Extracellular Traps by Suppressing Neutrophil Adhesion. *Front Immunol* **9**, 178, <https://doi.org/10.3389/fimmu.2018.00178> (2018).
52. Erpenbeck, L. *et al.* Effect of Adhesion and Substrate Elasticity on Neutrophil Extracellular Trap Formation. *Front Immunol* **10**, <https://doi.org/10.3389/fimmu.2019.02320> (2019).
53. Simonyan K., Vedaldi A. & Deep, A. Z. Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv:1312.6034* **20 Dec 2013**, <https://arxiv.org/abs/1312.6034> (2013).
54. de Bont, C. M., Koopman, W. J. H., Boelens, W. C. & Pruijn, G. J. M. Stimulus-dependent chromatin dynamics, citrullination, calcium signalling and ROS production during NET formation. *Biochim Biophys Acta Mol Cell Res* **1865**, 1621–1629, <https://doi.org/10.1016/j.bbamcr.2018.08.014> (2018).
55. Trump, B. F., Berezsky, I. K., Chang, S. H. & Phelps, P. C. The pathways of cell death: oncosis, apoptosis, and necrosis. *Toxicol Pathol* **25**, 82–88, <https://doi.org/10.1177/019262339702500116> (1997).
56. Ting-Beall, H. P., Needham, D. & Hochmuth, R. M. Volume and osmotic properties of human neutrophils. *Blood* **81**, 2774–2780 (1993).
57. Malinin, T. I. Injury of human polymorphonuclear granulocytes frozen in the presence of cryoprotective agents. *Cryobiology* **9**, 123–130 (1972).
58. Clark, P. J. & Distance, F. C. E. to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology* **35**, 445–453, <https://doi.org/10.2307/1931034> (1954).
59. Knuth, D. E. *The Art of Computer Programming, volume 1 3rd ed.*: *Seminumerical Algorithms*. 3 edn, Vol. 2 650 (Addison-Wesley, 1998).
60. Ataga, K. I., Cappellini, M. D. & Rachmilewitz, E. A. beta-thalassaemia and sickle cell anaemia as paradigms of hypercoagulability. *Br J Haematol* **139**, 3–13, <https://doi.org/10.1111/j.1365-2141.2007.06740.x> (2007).
61. Zhang, D. C., Xu, C. L., Manwani, D. & Frenette, P. S. Neutrophils, platelets, and inflammatory pathways at the nexus of sickle cell disease pathophysiology. *Blood* **127**, 801–809, <https://doi.org/10.1182/blood-2015-09-618538> (2016).
62. Hahn, J. *et al.* Neutrophils and neutrophil extracellular traps orchestrate initiation and resolution of inflammation. *Clin Exp Rheumatol* **34**, S6–S8 (2016).
63. Fuchs, T. A., Brill, A. & Wagner, D. D. Neutrophil Extracellular Trap (NET) Impact on Deep Vein Thrombosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **32**, 1777–1783, <https://doi.org/10.1161/atvbaha.111.242859> (2012).
64. Evans, C. *et al.* Impairment of neutrophil oxidative burst in children with sickle cell disease is associated with heme oxygenase-1. *Haematologica* **100**, 1508–1516, <https://doi.org/10.3324/haematol.2015.128777> (2015).
65. Qari, M. H. & Zaki, W. A. Flow cytometric assessment of leukocyte function in sickle cell anemia. *Hemoglobin* **35**, 367–381, <https://doi.org/10.3109/03630269.2011.571329> (2011).
66. Mollapour, E., Porter, J. B., Kaczmarek, R., Linch, D. C. & Roberts, P. J. Raised neutrophil phospholipase A2 activity and defective priming of NADPH oxidase and phospholipase A2 in sickle cell disease. *Blood* **91**, 3423–3429 (1998).
67. Humbert, J. R., Winsor, E. L., Githens, J. M. & Schmitz, J. B. Neutrophil dysfunctions in sickle cell disease. *Biomed Pharmacother* **44**, 153–158 (1990).
68. Amer, J. & Fibach, E. Chronic oxidative stress reduces the respiratory burst response of neutrophils from beta-thalassaemia patients. *Br J Haematol* **129**, 435–441, <https://doi.org/10.1111/j.1365-2141.2005.05463.x> (2005).
69. Hofstra, T. C., Kalra, V. K., Meiselman, H. J. & Coates, T. D. Sickle erythrocytes adhere to polymorphonuclear neutrophils and activate the neutrophil respiratory burst. *Blood* **87**, 4440–4447 (1996).

70. Zhao, W., Fogg, D. K. & Kaplan, M. J. A novel image-based quantitative method for the characterization of NETosis. *J Immunol Methods* **423**, 104–110, <https://doi.org/10.1016/j.jim.2015.04.027> (2015).
71. Neeli, I., Dwivedi, N., Khan, S. & Radic, M. Regulation of extracellular chromatin release from neutrophils. *J Innate Immun* **1**, 194–201, <https://doi.org/10.1159/000206974> (2009).
72. Raftery, M. J. *et al.* beta2 integrin mediates hantavirus-induced release of neutrophil extracellular traps. *J Exp Med* **211**, 1485–1497, <https://doi.org/10.1084/jem.20131092> (2014).
73. Yipp, B. G. & Kubes, P. NETosis: how vital is it? *Blood* **122**, 2784–2794, <https://doi.org/10.1182/blood-2013-04-457671> (2013).
74. Allen, G. In *AAAS Annual Meeting Feb 14–17, 2019* (Washington DC, 2019).
75. Perez, L. & Wang, J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *CoRR* **abs/1712.04621**, <https://arxiv.org/abs/1712.04621> (2017).
76. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556*, September 2014, <https://arxiv.org/abs/1409.1556> (2014).
77. Gao Huang, Z. L., Laurens van der Maaten, K. & Weinberger, Q. in *Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Honolulu, HI, USA, 2016).
78. Selvaraju R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *arXiv:1610.02391*, 7 Oct 2016, <https://arxiv.org/abs/1610.02391> (2016).

Acknowledgements

We would like to thank Tom Niesh and all personnel at the UNC Health Care Blood Donation Center for meeting much of our routine needs of fresh blood samples from healthy volunteers. We also thank Jessica Shen, Emaraghi David, Alex Chung, and Nidhi Gera for their unremitting support and input. We are grateful to David Wichlan, Brigid Scullin, Dell Strayhorn, and Susan Jones for their help in the SCD clinic. This work was supported by Grant # 2013124 from the Doris Duke Charitable Foundation (LVP, LE), IBM Junior Faculty Development Award 2016 from the Office of the Executive Vice Chancellor and Provost (LE) and NIAID STTR # 1R41AI131840-01A1.

Author contributions

L.E., L.V.P., J.N.C. and N.S. planned the work and designed experiments. L.E., J.N.C., C.A.B., J.A.W., N.R. and C.A.M. conducted the experiments and helped with data analysis. M.M., I.R., and J.N.C. designed, trained and tested the CNNs. K.I.A. helped with human studies involving patients with SCD. S.P.H. and T.M.L. helped with experimental design, writing and editing the manuscript. M.W.C. helped with editing the manuscript and L.E. wrote the manuscript with help and input from all co-authors.

Competing interests

US Patent 10,303,923. Applicants: University of North Carolina at Chapel Hill and University of South Carolina at Columbia. Inventors: Leslie V. Parise, Noah Sciaky, Laila Elsherif and Joshua N. Cooper. Status: Allowed. Aspects of the manuscript covered in patent: customizing convolutional neural networks for the quantification of NETosis in human neutrophils *in vitro*. C.A.M., M.M., I.R., C.A.B., J.A.W., N.R., T.M.L., S.P.H., M.W.C. and K.I.A. declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-53202-5>.

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

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019