A computer vision approach to the assessment of dried blood spot size and quality in newborn screening

Nick Flynn[a*], Stuart J Moat[b,c], Sarah L Hogg[a]

a. Biochemical Genetics Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
b. Department of Medical Biochemistry, Immunology & Toxicology, University Hospital Wales, Cardiff, UK
c. School of Medicine, Cardiff University, Cardiff, Wales, UK

*** Corresponding author**

Biochemical Genetics Unit, Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge Biomedical Campus, Hills Road, Cambridge, CB2 0QQ, UK. E-mail: nick.flynn@nhs.net. Tel: +44 1223 596173

Keywords:

Newborn screening, dried blood spot quality, computer vision, machine learning, artificial intelligence

Non-standard abbreviations:

DBS: dried blood spot

NBS: newborn blood spot screening

CV: computer vision

Computer vision tool to assess blood spot acceptability

CLIR: Collaborative Laboratory Integrated Reports

IMD: inherited metabolic diseases

IAC: incorrect application classifier

TSH: thyroid stimulating hormone

IRT: immunoreactive trypsinogen

Phe: phenylalanine

Tyr: tyrosine

Xle: total leucines (includes leucine, isoleucine and alloisoleucine)

LIMS: laboratory information management system

Met: methionine

C5: isovalerylcarnitine

C5DC: glutarylcarnitine

C8: octanoylcarnitine

C10: decanoylcarnitine

%CV: percentage coefficient of variation

Computer vision tool to assess blood spot acceptability

Abstract

Background: Dried blood spot (DBS) size and quality affect newborn screening (NBS) test results. Visual assessment of DBS quality is subjective.

Methods: We developed and validated a computer vision (CV) algorithm to measure DBS diameter and identify incorrectly applied blood in images from the Panthera DBS puncher. We used CV to assess historical trends in DBS quality and correlate DBS diameter to NBS analyte concentrations in 130,620 specimens.

Results: CV estimates of DBS diameter were precise (percentage coefficient of variation <1.3%) and demonstrated excellent agreement with digital calipers with a mean (standard deviation) difference of 0.23mm (0.18mm). An optimised logistic regression model showed a sensitivity of 94.3% and specificity of 96.8% for detecting incorrectly applied blood. In a validation set of images (n=40), CV agreed with an expert panel in all acceptable specimens and identified all specimens rejected by the expert panel due to incorrect blood application or DBS diameter >14mm. CV identified a reduction in unsuitable NBS specimens from 25.5% in 2015 to 2% in 2021. Each mm decrease in DBS diameter decreased analyte concentrations by up to 4.3%.

Conclusions: CV can aid assessment of DBS size and quality to harmonize specimen rejection both within and between laboratories.

1. Introduction

The size and quality of dried blood spot (DBS) specimens affect newborn blood spot screening (NBS) analyte concentrations [1–3], with a recent 20-year review of NBS finding that poor specimen quality contributed to at least five missed cases of cystic fibrosis [4]. DBS acceptance guidelines recommend rejecting specimens where DBS are too small (<8mm), too large (>14mm), compressed, multi-spotted (composed of several small overlapping drops of blood), or where the blood has not completely soaked through the filter paper [1–3,5]. In 2018-19 UK NBS laboratories rejected approximately 13,000 specimens (1.8% of all babies screened) due to poor DBS size and quality [6], of which approximately 80% were due to small or incorrectly applied blood.

DBS specimen quality is currently assessed by visual inspection which is time consuming, requires thorough staff training, and may result in misclassification of specimen quality. There is a lack of objective data on DBS specimen quality which impedes reliable identification of individual specimen collectors or maternity units collecting poor quality specimens. Although a UK NBS programme key performance indicator records the proportion of rejected specimens for each maternity unit, interpretation is complicated by the local laboratory's acceptance criteria and so comparisons between different screening regions within the country may be misleading. Since DBS quality can affect NBS analyte concentrations by up to 35% [2] , objective data on DBS size and quality may be a useful additional parameter to improve performance in predictive models for NBS that use large datasets, such as Collaborative Laboratory Integrated Reports (CLIR) [7,8].

Computer vision (CV) refers to how computers can gain high level understanding from digital images or videos. CV can classify images and detect and track objects, and has

found potential applications in various fields of medicine [9]. Images of DBS specimens may be captured at multiple points in the NBS laboratory. The PerkinElmer Panthera punching device used for sampling (‘punching’) DBS specimens in many NBS laboratories world-wide contains an integrated camera that shows a video feed of the DBS specimen and displays annotations indicating the locations of sub-punches. Furthermore, the DBS punching instrument automatically saves images of specimens.

The aim of this study was to investigate the utility of CV as a tool to assess DBS specimen size and quality using images obtained from the DBS punching instrument.

2. Materials and methods

2.1. Images and analyte concentrations

Images of DBS specimens were saved from the integrated camera of the PerkinElmer Panthera Puncher. DBS images saved at punching contain annotations that include the locations of sub-punches. Paired NBS screening analyte concentrations (see Supplemental Methods for methods used) were obtained from the laboratory information management system (LIMS).

2.2. Computer vision algorithm

A CV algorithm for DBS quality was developed in Python (version 3.6.13). In brief, the algorithm uses the OpenCV computer vision library (version 3.4.2) [10] to manipulate the image and detect the edge of DBS (**Figure 1A**). The algorithm then analyses the shape of the detected DBS, calculating the pixel area of the DBS and other properties describing their shape. Pixel area is converted to DBS diameter using an instrument specific conversion factor, assuming the DBS is a perfect circle. Finally, calculated DBS properties are used as parameters for a logistic regression machine learning model to identify incorrectly applied blood (incorrect application classifier, IAC), which is described in further detail in section 2.3.

Examples of processed images are shown in **Figure 1B**. The code is available on GitHub [11] and is described in further detail in the Supplemental Methods.

2.3. Incorrect application classifier (IAC) training, tuning and testing

The workflow for training, tuning, and testing the logistic regression model to identify incorrectly applied blood spots (incorrect application classifier, IAC) is summarised in **Figure 2**. We used images of specimens received between January 2017 and

December 2021 recorded in the LIMS as unsuitable due to incorrect application of blood to the filter-paper collection device. This includes layered, overlapping and multi-spotted specimens. Control DBS specimens were obtained over the same time-period. We reviewed all processed images and excluded those that were incorrectly classified or where CV did not correctly identify the DBS boundary. The final dataset included 430 incorrectly applied blood specimens and 1,561 acceptable control specimens.

A logistic regression model was trained using the open-source machine learning library scikit-learn (version 0.24.2) [12] with the saga solver. To optimise and test the model a train-test split with a test size of 30% was performed. Data were standardised by removing the mean and scaling to unit variance. A grid search with five-fold cross validation was used to tune model hyperparameters to optimise model accuracy with lasso, ridge and elastic net regularisation. The model with optimum hyperparameters was then trained on the entire IAC dataset for the final model.

2.4. Validation of DBS detection and DBS properties

CV identification of the DBS boundary and sub-punch annotations were assessed by visual inspection of 827 consecutive images by one laboratory scientist (NF). Some images contained more than one DBS within the circular punching region. CV estimates of DBS diameter were compared to the average of duplicate measurements with Preciva digital calipers in 69 DBS. Long-term imprecision of DBS diameter and other shape descriptors were assessed by obtaining images of DBS specimens on a set of four DBS controls. Images of controls were obtained once per day on 68 days over a six-month period. The effect of lighting was assessed on paired images of twenty DBS specimens obtained while the blood spot puncher was either under direct

sunlight or artificial lighting. Images collected at barcode scanning and at punching were compared on paired images of 92 DBS. Comparisons were analysed by Deming regression in Analyse-It.

2.5. Utility of CV as a decision support tool

To demonstrate the utility of the CV algorithm, we used an independently prepared validation set of images that comprised both the front and back of 40 specimens scaled to original size and printed on card. Images of the glassine envelope were also supplied for some of the specimens. The validation set was designed to contain a challenging range of specimen rejection reasons including incorrect blood application, small or large size, compression, contamination, and incomplete or inappropriate drying. A consensus interpretation based on UK and CLSI guidelines [5] had been agreed for each specimen by an expert panel containing representatives from four UK NBS laboratories. The expert panel used the ability of the Panthera puncher to select three punches from a DBS as a proxy for identifying DBS with a diameter >8mm.

Images were positioned under the puncher camera and a single image taken of the back of the DBS. DBS diameter measured as <8mm or >14mm were classified as small or large, respectively. DBS specimens were assessed by the IAC and DBS classified as uniform (predicted probability, $P<0.25$), IAC-borderline (P: 0.25-0.50), or incorrectly applied ($P>0.50$). Classifications were based on unrounded data.

Specimens were considered acceptable if they had at least two 8-14mm uniform DBS. Specimen quality was flagged for possible rejection if the acceptable criteria were not met but there was either i) one >10mm uniform or IAC-borderline DBS or ii) one 8-10mm uniform DBS *and* one or more 8-10mm IAC-borderline DBS or iii) no uniform

DBS *and* two or more 8-10mm IAC-borderline DBS. Specimens not meeting the above criteria were rejected outright.

2.6. Historical trends and regional variation in DBS quality

To assess historical trends in DBS quality and variation between maternity units within the Cambridge screening region, we analysed the first image of the first routine specimen received for babies screened between February 2015 and August 2021. The proportion of unsuitable specimens by month or maternity unit was compared to historical specimen rejection rates recorded in the LIMS.

2.7. Relationship between DBS diameter and NBS analyte concentration

NBS specimens collected between day 5 and 8 of life and analysed between February 2015 and August 2021 were used to assess the effect of DBS diameter on NBS analyte concentrations. Only specimens with a total of four punches (as detected by CV) were included to prevent inaccurate pairing of images to analyte concentrations in specimens where the initial analysis was repeated (for example due to assay failure or an initial abnormal result). Analyte concentrations for inherited metabolic diseases (IMD) screening, thyroid stimulating hormone (TSH) and immunoreactive trypsinogen (IRT) were paired with the DBS containing the first, second or third punch, respectively, since this is the order in which specimens are processed. Analytes measured for IMD screening in the UK screening programme are phenylalanine (Phe), tyrosine (Tyr), total leucines (Xle), methionine (Met), isovalerylcarnitine (C5), glutarylcarnitine (C5DC), octanoylcarnitine (C8) and decanoylcarnitine (C10). TSH concentrations on specimens analysed before May 2018 were excluded as before this date TSH was only recorded to the nearest integer, rather than to one decimal place. DBS classified as incorrectly applied by the IAC (P>0.50) were excluded. Since

changes in analytical methodology and/or long-term assay drift may confound the relationship between NBS analyte concentrations with DBS size and quality, concentrations were compared to the daily mean concentration of a reference population of 10-12mm DBS, which approximates the diameter of the reference 50µL sub-punch used in previous studies on DBS quality [1,2]. The difference between measured analyte concentrations and the analyte and day specific reference mean was calculated and used for statistical analysis by ordinary least squares in statsmodels (version 0.12.2) [13].

2.8. Research approval

The study was reviewed and approved by Cambridge University Hospitals NHS Foundation Trust Research and Development department (reference A095880).

3. Results

3.1. Detection of DBS boundary and punches

In 827 consecutive images, CV detected all 878 DBS within the punching region of the puncher (some images contained more than one DBS within the punching region). The detected boundary deviated from the true DBS boundary in eight DBS (0.9%), of which seven were due to interference from numbers printed on the filter paper, while the remaining DBS lay in the shadow of the gripping hand at the lower edge of the image (**Supplemental Figure 4A-H**). All deviations were minor and deemed unlikely to significantly affect calculated properties of DBS. No additional artefactual DBS were identified in any image. 615 images contained sub-punch annotations; CV correctly identified the number of sub-punch annotations in 614/615 of these DBS (99.7%) (single exception shown in **Supplemental Figure 4I**).

3.2. Comparison to manual measurement of diameter

Calculated DBS diameter showed excellent agreement with measurement of DBS diameter by digital calipers (**Figure 3**). The mean (standard deviation) difference was 0.23 (0.18) mm with a maximum observed difference of 0.65 mm (n=69). Deming regression analysis demonstrated a slope (95% CI) of the regression line of 1.015 (0.991 –1.039), intercept (95% CI) of 0.098 (–0.096 to 0.292) and an Sy.x of 0.176.

3.3. Imprecision of CV tool

The percentage coefficient of variation (%CV) for DBS diameter on a set of four DBS controls varied between 0.5% and 1.3%. Parameters used in the IAC showed a %CV of up to 6.3%. Precision profiles are shown in **Supplemental Figure 5**.

3.4. Effect of image type and lighting

DBS diameter showed a small negative bias for images captured at punching (n=92, mean difference -0.09 mm, standard deviation 0.06 mm) compared to images captured during barcode scanning. Direct sunlight resulted in a negative bias for DBS diameter (n=20, mean difference: -0.40 mm, SD 0.16 mm) compared to artificial lighting conditions. Other calculated DBS properties were unaffected by image type or lighting, with a mean difference of <0.01 in all cases (**Supplemental Figures 6-7**).

3.5. Incorrect application classifier (IAC)

Exploratory data analysis of the IAC dataset showed differences in the distribution of blood spot properties for incorrect application DBS and single spot controls (**Figure 4A**). The optimised logistic regression model showed an accuracy of 96.3% on the test set, with a sensitivity for detecting incorrectly applied DBS of 94.3% and specificity of 96.8% (**Figure 4B**). Inaccurate classifications are shown in Supplemental Figures 8 and 9. Four incorrectly applied DBS and seven single spot controls had a predicted probability within a borderline range (predicted probability between 0.25 and 0.50).

3.6. Utility as decision support tool

In the validation set, CV identified all acceptable specimens (n=17) and rejected (or flagged for possible rejection) all specimens rejected by the expert panel due to incorrect application (n=11) or large DBS size (n=1) (**Figure 5A-D**). Other specimens rejected by the expert panel due to incorrect blood application are shown in **Supplemental Figure 10**. Two specimens were rejected by the expert panel due to small DBS but the algorithm identified at least two DBS with diameter >8mm. In one of these specimens (**Figure 5E**) DBS diameters were very close to the 8mm limit, with DBS 4 measuring at 8.006mm by CV but 7.9mm by digital calipers. In the other specimen (**Figure 5F**), measurement of DBS diameter by digital calipers suggested that three DBS (DBS 2-4) met the minimum size requirement of 8mm. The expert panel failed to reach a consensus in two specimens (**Supplemental Figure 11B-C**); measurement of DBS diameter by digital calipers in these specimens supported the CV prediction. CV could not reliably detect other causes of poor DBS quality that the tool had not been trained to detect, such as compression, contamination, or incomplete/inadequate drying (**Supplemental Figure 11D-H**). However, one contaminated specimen was rejected by CV due to irregular shaped DBS (**Supplemental Figure 11-I**).

3.7. Historical trends and regional variation in DBS quality

CV analysis of historical DBS images (n=177,886) demonstrated a marked improvement in DBS quality, with the proportion of unsuitable specimens decreasing from 25% in 2015 to around 2% in 2021 (**Figure 6A**). Improvement was mainly due to a reduction in incorrectly applied specimens (**Figure 6B**) and was accompanied by a

more uniform size distribution of single spot DBS (**Figure 6C**). Unsuitable DBS varied between maternity units from 1.03% to 3.75% in 2021 (**Figure 6D)**. Incorrectly applied specimens were the major cause of unsuitable DBS for all maternity units (**Figure 6E)**. CV identified differences in the size distribution of single spot DBS between maternity units, with median DBS diameter ranging from 10.0 to 10.9 mm (**Figure 6F)**.

3.8. Relationship between DBS diameter and NBS analyte concentrations

There was a clear relationship between DBS diameter categories and analyte distributions for Phe, Tyr, Xle, Met, C8 and C5DC (n =130,620), TSH (n=66,440), and IRT (n=129,640) (**Supplemental Figure 12**), with ordinary least squares analysis showing that DBS diameter was a significant predictor of analyte concentration ($P<0.01$). However, a significant relationship with DBS diameter was not observed for C5 ($P=0.33$, n=130,620) or C10 ($P=0.12$, n=130,620). At the median analyte concentration of the reference population, ordinary least square analysis showed that each mm decrease in DBS diameter resulted in a decrease in analyte concentration of up to 4.31% (**Table 1**).

4. Discussion

This study reports on the utility of a CV tool for assessing DBS size and quality using images obtained using a DBS puncher. The tool has a high-throughput and can process approximately 2500 images per minute.

Objective methods to assess DBS quality have previously been reported. A web-based application used smartphone photos to assess DBS quality [14]. Designed for research studies to provide feedback on specimen quality shortly after collection, images of the specimen must be manually aligned, and the position of DBS indicated on the app. Another approach used the CardScan optical-scanning device which was separate from other laboratory equipment and could assess DBS quality in approximately 100 specimens per hour [15]. In contrast to these methods, our CV algorithm uses images obtained from an existing NBS laboratory device and automatically detects DBS within the punching region of the instrument. The algorithm requires no special positioning of specimens (beyond the normal positioning of specimens beneath the puncher) or special equipment, making it easier to integrate into NBS laboratory workflows. A similar approach using images from the Panthera puncher and an algorithm in R to measure DBS size was recently reported [16]. However, the R algorithm study did not report validation against manual measurement of DBS size or expert interpretation, and does not detect incorrect blood application which is a major cause of poor DBS quality [6].

While our logistic regression model showed high sensitivity and specificity for detecting incorrectly applied blood, achieving a perfect detection rate is challenging. To reduce the risk of inadvertently using poor quality DBS, a lower threshold could be used to rule out potentially unsuitable DBS quality. Implementing a borderline range with predicted probability of incorrectly applied blood between 0.25 and 0.50 would have

detected around a quarter of missed incorrectly applied blood DBS in the IAC test set. Our algorithm allows customisable reject and borderline thresholds for DBS diameter and incorrect blood application probability, and can annotate images with red (reject), amber (borderline/warning) or green (acceptable) bounding boxes to indicate the predicted classification. Since NBS specimens usually contain multiple DBS, laboratories could prioritise acceptable DBS, while avoiding unsuitable or borderline DBS. If a DBS was within the borderline range and there were no other acceptable DBS to use, the specimen could be subject to further manual review to determine the correct quality classification. Since thresholds for DBS diameter and incorrect blood application probability can be customised, laboratories and screening programmes could optimise cut-offs according to their own risk assessments and the relative values placed on the need to reduce repeat samples and manual review, against the risk of failing to detect poor DBS quality.

In a validation set, our algorithm correctly identified all specimens deemed acceptable by an expert panel and identified all samples rejected due to incorrect blood application or large DBS size. There were two discrepancies in specimens rejected by the expert panel due to small DBS size. The expert panel used the ability of the puncher to select three punches as a proxy for DBS size, which is a more indirect measure than CV which calculates diameter based on the pixel area of DBS. Measurement of DBS diameter by digital calipers suggested that DBS diameters were very close to 8mm in one specimen. In the other discrepant specimens, DBS diameter by digital calipers identified three DBS that met the minimum size requirement suggesting that the specimen should have been accepted in line with the CV prediction. Similarly, in two specimens where the expert panel failed to reach a consensus on whether DBS met the minimum size requirement, CV predictions

agreed with measurement of DBS diameter by digital calipers, suggesting that CV was more reliable at identifying small DBS.

A limitation of our CV algorithm is that it can only detect some types of poor DBS quality, namely small, large, or incorrectly applied (multi-spotted, overlapping or layered) DBS, since it only considers the size and shape of the DBS. It therefore failed to detect five cases of unacceptable DBS quality in the validation set due to compression, contamination, or incomplete/inadequate drying. Our algorithm can therefore currently only complement, rather than completely replace, human visual inspection. It is possible that further analysis of pixels within or around the DBS boundary could identify other causes of poor DBS quality. One possible approach would be to use a convolutional neural network trained using a large and diverse labelled dataset[9]. Unfortunately, most available images for the development of our algorithm contained punch annotations that obscured part of the DBS, limiting their use for training a robust supervised neural network model. The algorithm has been available on GitHub for others to improve and potentially expand to include other causes of poor DBS quality [11].

Markings on the filter-paper can affect detection of the DBS boundary; the importance of this issue will depend on local DBS filter paper collection device design. The algorithm was affected by lighting, with images captured under bright direct sunlight demonstrating a negative bias compared to artificial lighting. Laboratories implementing CV technology should introduce quality control procedures to monitor CV outputs and detect any significant changes caused by variation in lighting or image quality.

Our CV algorithm can be applied to stored images to assess historical trends and regional variation in specimen quality. We detected a marked improvement in DBS

quality with the proportion of unsuitable specimens decreasing from 25.5% in 2015 to 2% in 2021. This change was not reflected in the historical laboratory rejection rate which declined from a maximum of 4.75% in 2015 before stabilising around 1% in 2017. The failure of laboratory rejection rates to identify DBS quality improvement after 2017 demonstrates their unreliability for monitoring DBS quality and the subjectivity of DBS specimen acceptance criteria which may vary over time and between laboratories. Our algorithm offers an alternative objective approach to assist in identifying poor performing units or individuals to target for further training.

By analysing up to 130,620 paired images and analyte concentrations, our study demonstrates that DBS size affects analyte concentrations in DBS collected from newborn babies, confirming results of previous studies using contrived DBS specimens. At median analyte concentrations, for every mm decrease in DBS diameter there was a decrease in analyte concentration of up to 4.31%. Since the diameter of a 20µL and 50µL DBS are approximately 7.5 and 11.5mm respectively [2], our study predicts a difference in analyte concentrations of up to around 15% between 20µL and 50µL DBS, which agrees well with the 10-15% difference observed in previous studies on contrived DBS [1–3].

The effect of DBS size was more evident for amino acids (Phe, Tyr, Xle, Met), TSH and IRT than for acylcarnitines (C8, C10, C5 and C5DC). The smaller apparent effect for acylcarnitines is likely due to decreased assay performance at the low concentrations observed in the predominantly healthy population studied. In addition, our laboratory only records analyte concentrations to the number of significant figures specified in national NBS pathway. This is especially relevant for C5 (clinical cut-off 2.0µmol/L) as most specimens had a recorded concentration of 0.1µmol/L; rounding error may have caused an underestimation of the effect of DBS size.

Punch location is also known to affect analyte concentrations, with higher concentrations on a peripheral punch compared to a central punch [3]. On average we found that sub-punches taken from smaller DBS specimens were more likely to be closer to the centre of the DBS, which may have exaggerated the relationship between analyte concentration and DBS size.

If integrated into laboratory equipment, CV assessments of DBS size and quality could be included in NBS pathways. This could involve using size-adjusted analyte concentrations, or using a different analytical threshold to rule out a condition in the event of suboptimal sample quality. Such strategies may reduce the number of repeat samples requested, avoiding unnecessary repeat blood collections in newborn babies. The optimum strategy including any thresholds used requires further study.The UK NBS programme uses fixed analyte cut-off values. However, multivariate pattern recognition software such as CLIR adjusts patient results by covariates such as birth weight and age at collection [7,8]. Variables such as DBS diameter may improve the performance of predictive models.

In conclusion, this study demonstrates that CV using images obtained during routine analysis of NBS specimens can reliably measure DBS diameter and identify small or large DBS, or specimens where blood has been incorrectly applied to the filter paper collection device. CV could be integrated into existing laboratory equipment as a decision support tool, assist service evaluation in NBS programs, and produce parameters to correct analytical results for DBS size and quality in NBS protocols.

Computer vision tool to assess blood spot acceptability

Funding

This research was supported by the NIHR Cambridge Biomedical Centre (BRC 1215 20014).

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

[1] S.J. Moat, C. Dibden, L. Tetlow, C. Griffith, J. Chilcott, R. George, L. Hamilton, T.H. Wu, F. MacKenzie, S.K. Hall, Effect of blood volume on analytical bias in dried blood spots prepared for newborn screening external quality assurance, Bioanalysis. 12 (2020) 99–109. https://doi.org/10.4155/bio-2019-0201.

[2] R.S. George, S.J. Moat, Effect of dried blood spot quality on newborn screening analyte concentrations and recommendations for minimum acceptance criteria for sample analysis, Clin. Chem. 62 (2016) 466–75. https://doi.org/10.1373/clinchem.2015.247668.

[3] A.J. Lawson, L. Bernstone, S.K. Hall, Newborn screening blood spot analysis in the uk: Influence of spot size, punch location and haematocrit, J. Med. Screen. 23 (2016) 7–16. https://doi.org/10.1177/0969141315593571.

[4] I. Doull, C.W. Course, R.E. Hanks, K.W. Southern, J.T. Forton, L.P. Thia, S.J. Moat, Cystic fibrosis newborn screening: The importance of bloodspot sample quality, Arch. Dis. Child. 106 (2021) 253–257. https://doi.org/10.1136/archdischild-2020-318999.

[5] CLSI standard NBS01. Dried Blood Spot Specimen Collection for Newborn Screening, 7th edition, Clinical and Laboratory Standards Institute (CLSI), 2021.

[6] Public Health England, Newborn blood spot screening data collection and performance analysis report 1 April 2018 to 31 March 2019, (2021). https://www.gov.uk/government/publications/newborn-blood-spot-screening-data-collection-and-performance-analysis-report/newborn-blood-spot-screening-data-collection-and-performance-analysis-report-1-april-2018-to-31-march-2019 (accessed September 6, 2022).

[7] D.K. Gavrilov, A.L. Piazza, G. Pino, C. Turgeon, D. Matern, D. Oglesbee, K. Raymond, S. Tortorelli, P. Rinaldo, The combined impact of CLIR post-analytical tools and second tier testing on the performance of newborn screening for disorders of propionate, methionine, and cobalamin metabolism, Int. J. Neonatal Screen. 6 (2020) 33. https://doi.org/10.3390/ijns6020033.

[8] A.D. Rowe, S.D. Stoway, H. Åhlman, V. Arora, M. Caggana, A. Fornari, A. Hagar, P.L. Hall, G.C. Marquardt, B.J. Miller, C. Nixon, A.P. Norgan, J.J. Orsini, R.D. Pettersen, A.L. Piazza, N.R. Schubauer, A.C. Smith, H. Tang, N.P. Tavakoli, S. Wei, R.H. Zetterström, R.J. Currier, L. Mørkrid, P. Rinaldo, A novel approach to improve newborn screening for congenital hypothyroidism by integrating covariate-adjusted results of different tests into CLIR customized interpretive tools, Int. J. Neonatal Screen. 7 (2021) 23. https://doi.org/10.3390/ijns7020023.

[9] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, R. Socher, Deep learning-enabled medical computer vision, Npj Digit. Med. 4 (2021) 5. https://doi.org/10.1038/s41746-020-00376-2.

[10] G. Bradski, The OpenCV Library, Dr Dobbs J. Softw. Tools. (2000).

[11] N. Flynn, nf260/nbscv: Blood spot quality computer vision, GitHub. (n.d.). https://github.com/nf260/nbscv (accessed October 3, 2022).

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[13] S. Seabold, J. Perktold, statsmodels: Econometric and statistical modeling with python, in: 9th Python Sci. Conf., 2010.

[14] H. Veenhof, R.A. Koster, R. Brinkman, E. Senturk, S.J.L. Bakker, S.P. Berger, O.W. Akkerman, D.J. Touw, J.W.C. Alffenaar, Performance of a web-based application measuring spot quality in dried blood spot sampling, Clin. Chem. Lab. Med. (2019) 1846–1853. https://doi.org/10.1515/cclm-2019-0437.
[15] P.D. Dantonio, G. Stevens, A. Hagar, D. Ludvigson, D. Green, H. Hannon, R.F. Vogt, Comparative evaluation of newborn bloodspot specimen cards by experienced laboratory personnel and by an optical scanning instrument, Mol. Genet. Metab. 113 (2014) 62–66. https://doi.org/10.1016/j.ymgme.2014.07.007.
[16] R. Groh, L.M. Weiss, M. Börsch-Supan, A. Börsch-Supan, Effects of spot size on biomarker levels of field-collected dried blood spots: A new algorithm for exact dried blood spot size measurement, Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. 34 (2022) e23777. https://doi.org/10.1002/ajhb.23777.

Tables

Table 1 - Percentage decrease in analyte concentration per mm decrease in DBS diameter at the median analyte concentration of the reference population. n = 130,620 for Phe, Tyr, Xle, Met, C8, C10, C5 and C5DC; n = 66,440 for TSH; n = 129,640 for IRT.

Analyte	Median concentration of reference population	Percentage decrease in analyte concentration per mm decrease in DBS diameter (95% confidence interval)	Probability
Phe	55 µmol/L	1.67 % (1.57 - 1.77)	< 0.01
Tyr	98 µmol/L	1.44 % (1.18 - 1.69)	< 0.01
Xle	157 µmol/L	2.75 % (2.63 - 2.87)	< 0.01
Met	17 µmol/L	1.88 % (1.76 - 2.01)	< 0.01
C8	0.04 µmol/L	1.14 % (0.94 - 1.35)	< 0.01
C10	0.06 µmol/L	0.18 % (-0.04 - 0.4)	0.12
C5	0.1 µmol/L	0.13 % (-0.13 - 0.39)	0.33
C5DC	0.11 µmol/L	0.92 % (0.77 - 1.06)	< 0.01
TSH	1.3 mU/L	4.31 % (3.8 - 4.82)	< 0.01
IRT	17 µg/L	3.08 % (2.88 - 3.29)	< 0.01

Figure captions

Figure 1 - CV algorithm summary and processed images. A**.** A color-coded bounding box highlights the detected DBS, with calculated DBS diameter (mm) and predicted probability of incorrect DBS application shown in the upper left and lower right corners, respectively. Roundness, elongation, circular extent, solidity and convexity are defined in Supplemental Table 1. B. Processed Images, showing examples of unacceptable DBS quality.

Figure 2 - Flowchart for selection of images used to train and test the logistic regression model to identifier incorrectly applied DBS.

Figure 3 - Comparison of CV DBS diameter and diameter measured with digital calipers. A. Deming regression analysis B. Difference plot with mean difference (dashed line) +/- 1.96 standard deviations (dotted line) (n=69).

Figure 4 – IAC dataset and model performance. A. pair plot showing pairwise scatterplots for parameters in the IAC dataset (Incorrect application: n=430, controls: n=1,561). The diagonal shows the distribution of data in each category. B. Confusion matrix for agreement between IAC predictions and the true label in the IAC test set (Incorrect application: n=123, controls: n=475)

Figure 5 – Comparison of CV predictions to expert panel consensus. A. Agreement between expert panel and CV by expert panel classification (n=40). B-D. Example acceptable, incorrect blood application and large DBS. E-F. Specimens rejected by the expert panel due to small DBS size but accepted by CV. The upper right insert shows DBS diameter measured by digital calipers.

Figure 6 – Historical trends and regional variation in DBS quality. A. CV and laboratory rejection rates (n=177,886). B. CV rejection reasons (n=177,886). C. DBS diameter distribution in 2015 (n=21,060) and 2021 (n=16,577). D. CV and laboratory rejection rates by

maternity unit (n=16,208, range 1,141-3,316). E. CV rejection reasons by maternity unit (n=16208). F. DBS diameter distribution for three maternity units (n=6,161)