

APPLIED RESEARCH

Synthetization, Distortion, and Geometric Correction of Isoelectric Focusing Gels for Newborn Screening

SANTERI MÄNTYNIEMI^{1,2}, EERO LEHTONEN², (Member, IEEE),
AND AKI KOIVU^{1,2}, (Member, IEEE)

¹PerkinElmer Wallac Oy, 20750 Turku, Finland

²Department of Computing, University of Turku, 20014 Turku, Finland


Corresponding author: Santeri Mäntyniemi (mantyniemi.santeri@gmail.com)

ABSTRACT Hemoglobinopathies are inherited red blood cell disorders. They are commonly screened from newborns early after birth, and if not detected early and treated properly, some of them can be fatal. One of the most common method for screening hemoglobinopathies in newborns is isoelectric focusing (IEF). IEF separates different hemoglobins to distinct bands based on their isoelectric point. Different hemoglobinopathies can be detected from IEF as additional bands from normal hemoglobins and also from quantitative issues in normal hemoglobins. However, distortions are commonly present in IEF gels that hinder the interpretation of the gel. In this study, we showcase a method for straightening distorted images of IEF gels. Since no dataset containing images of IEF gels was present, we created a novel Sig2Img synthetic sample generator. This allowed us to create gel sides with any desired bands and intensities. Artificial distortions were also applied with basic image processing functions, to represent different kinds of distortions commonly found in IEF gels. Fifteen experiments were created to evaluate the method. The experiments showed that the correction method managed to straighten gel sides even with added difficulties, such as within-sample and sample-to-sample distortions, and added Gaussian noise. Experiments also showed that the method retained the relative quantities of the bands accurately. The straightening enables easier visual interpretation of the gel when screening for hemoglobinopathies. It also allows accurate quantitation of bands which was not originally possible from distorted samples. This gives the method clinical significance, as quantitation can be added to the clinical workflow.

INDEX TERMS Artificial neural networks, correction, distortion, isoelectric focusing, newborn screening, synthetization.

I. INTRODUCTION

Hemoglobinopathies [1] cause genetically abnormal or decreased production of globin chains in the human body. These diseases are commonly screened from newborns within the first week of life [2]. *Isoelectric focusing* (IEF) [3] is the gold standard method for *newborn screening* (NBS) of hemoglobinopathies. IEF uses an agarose gel that can separate different hemoglobins of a patient sample, enabling qualitative and semi-quantitative analysis [4]. However, distortions in the gel format can disrupt the relative

The associate editor coordinating the review of this manuscript and approving it for publication was Siddhartha Bhattacharyya .

quantitation of samples, which is why quantitation is commonly not utilized in clinical screening disorder logic of hemoglobinopathies.

In this study, we propose novel methods for synthetization and distortion of IEF gels, and showcase our earlier proposed algorithm [5] for the correction of these distortions. The primary application domain for this method is NBS, which imposes certain restrictions regarding the usage of IEF, gel format and commonly used sample material.

In order to overcome the limited availability of gel image data, a novel Sig2Img generator neural network was fitted for the purpose of creating synthetic IEF samples. Based on prior literature [6], [7], [8], [9], the band patterns of different

samples were used to create synthetic signal vectors, that were then used as input for the Sig2Img model, in order to create realistic sample images.

By using relative quantitation as the metric for our experimentation, we demonstrate the clinically significant performance of our correction method. Multiple cases of distortion are tested with gels mimicking the routine NBS gels.

Our experiments showcase that the correction method is successful even with IEF gels having different kinds of hindrances, such as abnormal hemoglobin variants, within-sample and sample-to-sample distortions, and random noise. The experiments also highlight the correction method's ability to enable relative quantitation for originally distorted bands. The differences in the relative quantities of the bands, before and after correction, show that the method retains the relative quantities accurately.

A. HEMOGLOBINOPATHIES

Hemoglobinopathies are a group of inherited genetic red blood cell disorders [10]. They can be divided into two categories: quantitative disorders in hemoglobin synthesis and qualitative disorders in hemoglobin structure. Quantitative disorders in hemoglobin synthesis result in abnormal quantities of normal hemoglobins. These are called thalassemias, with alpha- and beta-thalassemias having the greatest clinical significance. Qualitative disorders in hemoglobin structure result in abnormal hemoglobin variants. It is estimated that 7% of world population carries some form of hemoglobinopathy [11].

Out of the thousands hemoglobinopathies found, most are clinically insignificant [10]. However, there are hemoglobinopathies that result in severe symptoms and can also be fatal. An example of this is sickle cell disease (SCD). SCD is an umbrella term for disorders caused by the occurrence of hemoglobin S (HbS). SCD causes normally round red blood cells to transform into crescent, sickle-like shape, therefore the name sickle cell. These cells are not as elastic as normal red blood cells, and can actively block proper blood flow. The most severe form of SCD is the homozygous sickle cell disease (SCD-SS), which is caused by receiving the mutated gene from both parents. Having only one mutated gene is referred to as sickle cell trait (SCD-AS) or sickle cell carrier. There are also multiple other versions of SCD, that have different combinations of HbS with thalassemias or other hemoglobin variants. It is approximated that SCD affects 300 000 newborns every year, with 75% of them being from Sub-Saharan Africa [12].

SCD is asymptomatic the first months of life, because fetal hemoglobin (HbF) protects the newborn from symptoms caused by the disease [13]. HbF is the main hemoglobin found in fetuses. It allows the transfer of oxygen from maternal to the fetal circulation, because it has higher oxygen affinity than adult hemoglobin (HbA) [14]. HbF remains as the dominant hemoglobin until three months after birth. Between 3 to 6 months after birth is also when the symptoms caused by SCD start to emerge [13]. Issues such as acute splenic

sequestration, overwhelming septicemias, and acute chest syndrome bring great danger to the second half of the newborns first year of life with a high mortality rate. This is the reason why newborn screening is crucial for detecting these disorders as early as possible to prevent or at least effectively treat them, for example with blood transfusions.

B. NEWBORN SCREENING

NBS for hemoglobinopathies such as SCD is well-established, as related prevention programs have been in place for over 40 years. Commonly, the screening is conducted in centralized laboratories. At a hospital, fresh cord blood or heel pricked dried blood spot (DBS) samples are collected from the newborn, then delivered and tested for abnormalities. For this purpose, technologies such as liquid chromatography (HPLC), tandem mass spectrometry and electrophoresis-based methods such as capillary zone electrophoresis (CZE) are used. From these methods, IEF remains widely adopted, due to high sensitivity and relatively low cost [15].

IEF, a type of gel electrophoresis, can be used to separate different hemoglobins of a patient sample. The gel format contains a pH gradient, and once electric current is applied, the hemoglobins will migrate within the gel to specific regions where the pH is equal to their isoelectric points. This enables sensitive and qualitative measurement of abnormal hemoglobins such as HbH, which is undetectable by other biochemical methods such as HPLC and CZE. For NBS, IEF is considered the gold standard method [3].

C. IEF GEL

Similarly to any NBS test conducted in a centralized laboratory, the gel format is made for measuring multiple patient samples during a run. For example, using a 44-well template on a two-sided gel amounts to 88 sample positions in one gel. In this context, one or more position per side is used for control samples. The AFSC control sample is commonly used due to the Hb pattern it contains [16].

The IEF gel is considered as a one-dimensional format, as the separation of hemoglobins occurs on one axis. Fig. 1 showcases the separation of HbA, HbF, HbS and HbC of a control sample, and an unaffected newborn sample with HbF^{ac}, HbA and HbF. In the context of NBS of hemoglobinopathies, typically the IEF workflow would include full 1st tier screening gels, followed by a confirmatory gel of the screening positive results. Qualitative measurement of patient samples is commonly done by visually comparing the individual bands to the nearest control samples. The presence of any abnormal bands, or the abnormal ratio of HbA and HbF would indicate a screening positive result. Digitalization and relative quantitation [17] of gels is commonly not done, one technical reason for this is that the relative quantitation is affected by distortions routinely found in IEF gels.

While gels used in IEF are robust when compared to the competing technologies used for NBS of

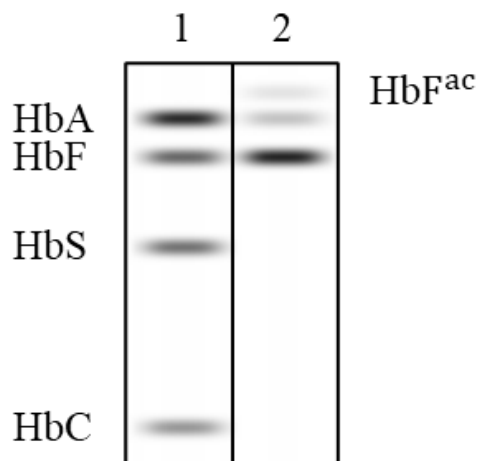


FIGURE 1. Example of an IEF result, where darker bands translate to more intensity. Lane 1 contains an AFSC control sample, which contains individual bands for HbA, HbF, HbS and HbC. Lane 2 contains an unaffected newborn patient sample with bands for HbF^{ac}, HbA and HbF. The separation of hemoglobins in the y-axis provides the means for interpretation.

hemoglobinopathies, they are certainly not immune to problems arising from routine use. Uneven heat distribution of gel can cause band distortions [18]. Distorted electrical fields, overly high voltage for the application, salts that inhibit conductivity, edge effects during loading the gel and smearing relating to sample concentrations can also cause distortions to manifest [19]. Distortions such as these hinder the relative quantitation, as the sample image is transformed to 1D pixel intensity signal by taking the sum, mean or median of pixel intensities. Fig. 2 showcases the two geometric distortions commonly present in IEF; within-sample distortion causes the 1D transformation to produce unclear signals and quantitation, and sample-to-sample distortion, which in the worst case causes the comparison of patient samples to the nearest controls unfeasible.

By proposing a correction solution to the sample distortions found in NBS IEF gels, adding relative quantitation to the clinical workflow would be feasible. The added sensitivity would improve the gold standard method for NBS of hemoglobinopathies even further, by providing a quantitative measurement instead of a subjective, qualitative result.

D. RELATED WORK

The amount of prior literature relating to computationally correct distortions from one-dimensional gels is modest. This is to be expected, since in research domains where these agarose gels in general are used, the image analysis is commonly ad hoc in nature, and in NBS the qualitative interpretation of the results is commonly done manually by a laboratory worker. In the past, we have published a preliminary investigation that proposes the framework for computational correction, and is reported in the Master's thesis called Geometric Sample Area Correction of an IEF Gel in a Newborn Screening Application Using Image Processing Methods [5], from which this study

is a continuation of. Relevant prior art from other authors is described below.

Koutny and Yeung, proposed a centroids-based algorithm for correcting the distortions from an electrophoresis gel [19]. However, in their example the correction made relative quantitation impossible for highly skewed samples, something which can be evident in our problem domain. Also, resulting corrected sample bands were visually highly synthetic, and not representative of the original bands. Compared to our problem domain, corrected samples would still have to be interpretative with the existing clinical workflow.

In 2001, Bajla *et al.* also proposed an analysis workflow for DNA gel electrophoresis [20]. In particular, they considered horizontal lane distortions that can be present in one-dimensional electrophoresis gel results. Their proposed solution did not involve correcting the image, and focused more on finding optimal band boundaries from images that can contain such distortions.

In 2005, Bajla *et al.* showcased a geometric distortion correction for DNA gel electrophoresis, which involves the user to manually draw or set a distortion grid, for which the gel image is corrected for [21]. While suitable for research purposes, the method involves manual steps and requires sufficient domain knowledge, and is therefore not suitable for routine IEF screening workflow.

Intarapanich *et al.* in 2015 proposed an image-processing tool for DNA gel electrophoresis that featured sample segmentation, band extraction and sample classification [22]. Their proposed cross-correlation adjustment to correct distortions was reported as effective, however the results were poor if the samples in the gel had few bands, which is routinely the case in NBS of hemoglobinopathies when compared to DNA analysis, as most samples contain four or less bands.

It should be noted that imaging analysis of two-dimensional electrophoresis gels is a more popular research topic [23], due to the fact that in DNA and protein research the separation of molecules by isoelectric point and molecule mass are both considered important. These methods do not however translate well to analyzing the one-dimensional IEF gels used in newborn screening, because the patterns of interest are different.

II. METHODS AND MATERIALS

The correction of within-sample and sample-to-sample distortion by our proposed method was tested by defining probable real-world use cases. Full 1st tier screening gels and partially filled confirmatory gel sides were considered. Since NBS can be time-sensitive, running partial gels is common. These gels increase the complexity of the correction problem by having less information that can be utilized by the correction method. Sides full of unaffected newborn samples and sides with SCD samples were considered, as the correction method would have to perform well in the presence of Hb variants. In terms of distortions, Gaussian noise [24] representing a non-ideal imaging setup was also considered,

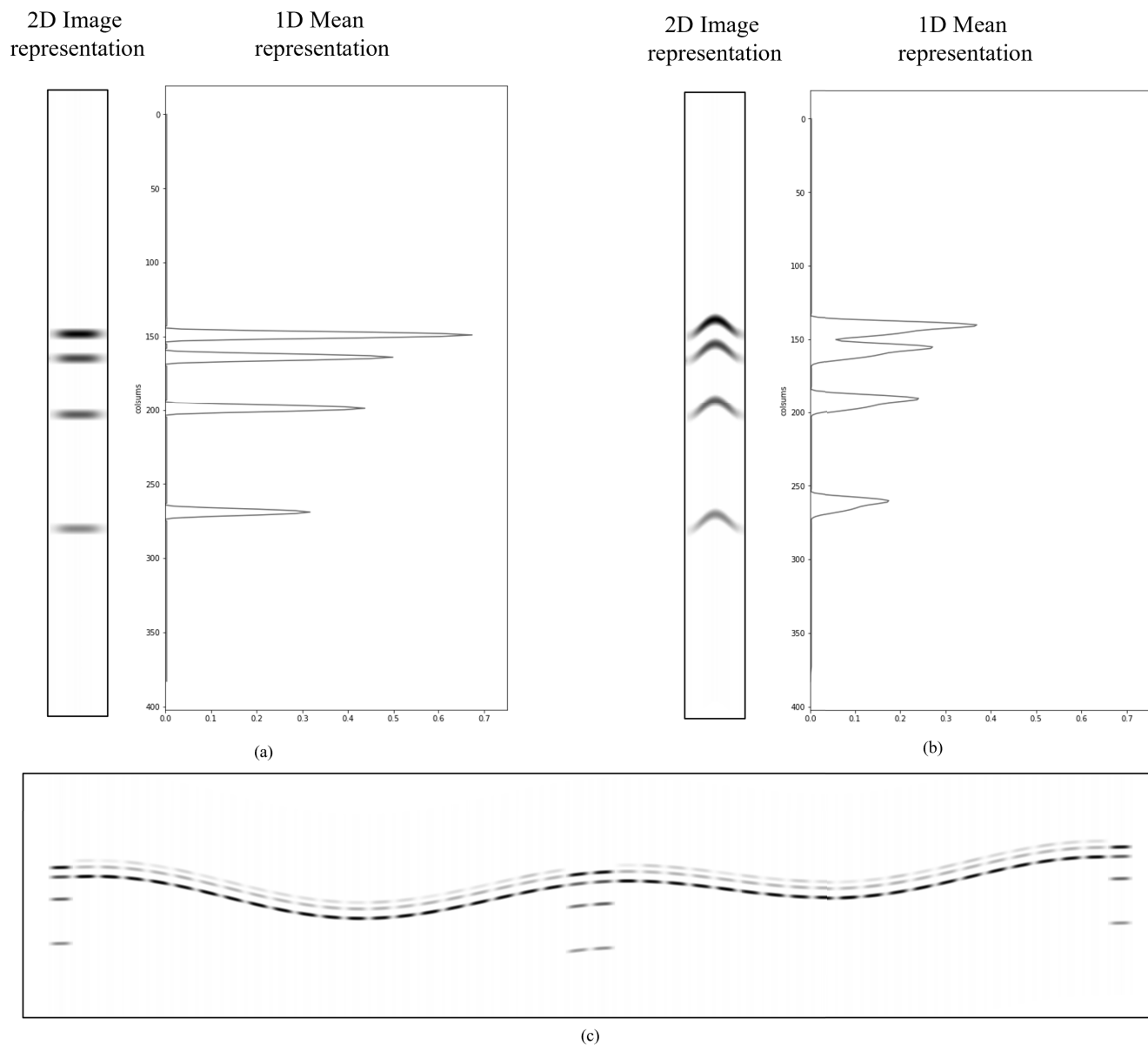


FIGURE 2. Example of within-sample and sample-to-sample distortion. In subfigure (a), the within-sample distortion of the image representation is minimal, which produces distinct AFSC Hb peaks in the signal representation. In subfigure (b), the distortion of the horizontal information in AFSC image representation causes more overlap of bands, which produces non-representative relative quantitation if calculated. Subfigure (c) showcases a gel side containing sample-to-sample distortion. This drifting of the sample area causes problems when highly distorted patient samples are quantitatively compared against the closest control sample.

as added background noise will make the detection and quantitation of Hb bands more difficult.

Based on the aforementioned aspects relating to the correction, the following experiments were designed to be done:

- (1.1) Full 44-well side of unaffected newborns and AFSC controls.
- (1.2) Full 44-well side of unaffected newborns and AFSC controls, sample-to-sample distortion.
- (1.3) Full 44-well side of unaffected newborns and AFSC controls, within-sample distortion.
- (1.4) Full 44-well side of unaffected newborns and AFSC controls, Gaussian noise.

- (2.1) Partial 22-well side of unaffected newborns and AFSC controls.
- (2.2) Partial 22-well side of unaffected newborns and AFSC controls, sample-to-sample distortion.
- (2.3) Partial 22-well side of unaffected newborns and AFSC controls, within-sample distortion.
- (2.4) Partial 22-well side of unaffected newborns and AFSC controls, Gaussian noise.
- (3.1) Full 44-well side of unaffected newborns, 6 SCD newborns and AFSC controls.
- (3.2) Full 44-well side of unaffected newborns, 6 SCD newborns and AFSC controls, sample-to-sample distortion.

- (3.3) Full 44-well side of unaffected newborns, 6 SCD newborns and AFSC controls, within-sample distortion.
- (3.4) Full 44-well side of unaffected newborns, 6 SCD newborns and AFSC controls, Gaussian noise.
- (3.5) Full 44-well side of unaffected newborns, 6 SCD newborns and AFSC controls, sample-to-sample distortion, within-sample distortion and Gaussian noise (worst-case). Experiment repeated 50 times.
- (4.1) Partial 22-well side of unaffected newborns, 6 SCD newborns and AFSC controls.
- (4.2) Partial 22-well side of unaffected newborns, 6 SCD newborns and AFSC controls, sample-to-sample distortion, within-sample distortion and Gaussian noise (worst-case). Experiment repeated 50 times.

Each experiment follows the same procedure of create a synthetic gel image, calculate relative quantitation per sample, apply some or none distortions or Gaussian noise, use the correction method, and lastly calculate the quantitation again. These before and after quantitation results can be compared to see, if the corrected image resembles the ground truth or not. These differences are calculated for every sample and every Hb band, and then they are investigated per band using mean difference. This is done in order to see if the correction method uniformly processes different Hb bands.

The relative quantitation of the bands is calculated from the peak areas in the 1D mean representation. The peak area is calculated from the width at half-height multiplied by the height of the peak. The percentage of each peak can be calculated from the total area, which is the sum of all the peak areas.

Experiments 1.1 to 1.4 demonstrate a common NBS situation where Hb variant results are not included, due to their rare incidence. The most common Hb variant, SCD, was included in experiments 3.1 to 3.4. One of the areas in the world with the largest incidence of SCD is Uganda, where roughly 13.3% or 1 in 7 babies are born with a SCD trait [25]. Based on this, the worst-case example was estimated as 6 SCD positive patient results within one gel area.

Experiments 2.1 to 2.4 utilize a partial gel side with all patient samples being unaffected newborn, while experiment 4.1 replaces 6 of these samples with SCD samples. Lastly, experiment 3.5 and 4.2 showcase the worst-case situation, where SCD samples are present, there is sample-to-sample distortion, within-sample distortion and Gaussian noise added. These worst case experiments were also repeated 50 times with new samples and distortions generated at each iteration, while the other experiments were calculated once, i.e. singletons.

A. GEL MATERIAL AND SAMPLES

One completed RESOLVETM Hemoglobin gel [16] containing 68 AFSC and 4 FAS control samples was provided by PerkinElmer Wallac Oy (Turku, Finland), which was digitized by an office scanner. The samples were individually partitioned as $72 \times 39 \times 384$ pixel grayscale images. No patient

samples were used in this study. These sample images were used to train our sample image generator model.

The unavailability of publicly open IEF gel image data sets forced us to utilize synthetization, where findings of Hb samples from prior literature were translated to sample images similar to our IEF gel format. However, the proposed correction method was previously evaluated using a separate proprietary real-life gel image dataset in the Master's thesis [5]. In the thesis, the correction method achieved 139 or 97% proper corrections from a total of 142 gel sides.

B. SAMPLE IMAGE SYNTHETIZATION

The example gel and the 72 control sample images extracted from it were not enough for the development and evaluation of a distortion correction method, that is supposed to have clinical relevance during NBS. The gel did not contain any patient samples, and the amount of geometric distortion was low. In order to produce more routine-like data for the study, synthetic gel side images containing control and patient samples were calculated using a novel neural network we call Sig2Img.

Sig2Img is a deep fully-connected feedforward neural network [26], which takes a 1D signal representation of a sample, and reconstructs it as a 2D IEF sample image. The network architecture contains an input layer of 384 nodes, two hidden -and batch normalization layer blocks with 500 nodes each, and an output layer of shape 384,39. ReLU activations were used in hidden nodes. Mean squared error [27] and RMSprop [28] were used to fit the network, with a batch size of 256 and 50 epochs. The relatively simple architecture which was sufficient for our task highlights the difficulty of the task, as peaks in the 1D signal need to be translated to proper band ellipses in 2D space. Subfigure (a) of Fig. 3 showcases the architecture and the training process.

AFSC and FAS control sample images gathered from the one gel were augmented in order to enrich the training data used for fitting the Sig2Img model. For each image, 20 variations were created using height shift, width shift, zoom of 10% and horizontal flip. This amounted to 1512 images, from which 80% was used as training data, and 20% as the validation data. Since the purpose of Sig2Img model was to capture the underlying signal-to-image function present in our one gel image, its generalizability to other gel images could not be tested due to unavailability of out-of-sample test data. From each sample data, a 1D signal representation was calculated with a mean over every pixel column, min-max normalization and Savitzky and Golay smoothing of the signal [29]. This process is highlighted in subfigure (b) of Fig. 3.

After fitting the model, a signal creation function was made, which would take individual band indexes and their intensities, and produce a 384-element vector. By utilizing prior literature, band locations and their intensities in other studies could be used to produce an artificial 1D signal representation of a sample, using the following algorithm:

Algorithm 1 Algorithm for Synthetic Signal Creation**Input:** bandIntensities (%), bandWidths (px), bandCoordinates (px coordinate)**Output:** synthSignal*Initialisation:*

1: Empty 384-element output vector synthSignal

2: Number of bands bandN

*For loop:*3: **for** $i = 1$ to bandN **do**4: startCoord = bandCoordinates[i] - (bandWidths[i]/2)5: endCoord = bandCoordinates[i] + (bandWidths[i]/2)6: synthX = *Range*(startCoord, endCoord)7: synthY = *Gaussian*(synthX)8: synthY = *Scale*(synthY, 0, bandIntensities[i])

9: synthSignal[startCoord:endCoord] = synthY

10: **end for**11: **return** synthSignal

where *Range* is a function that produces an empty vector of band width length, *Gaussian* generates a Gaussian bell curve and *Scale* scales the values of a vector between a certain min-max range. These synthetic signals were then given to the Sig2Img model as input in order to produce a feasible, life-like IEF sample images. This process is described in subfigure (c) of Fig. 3.

The band patterns and relative intensities were based on prior literature. Chandrakasan and Kamat demonstrated that in their patient population, an unaffected newborn has mean $22.5\% \pm 2.5\%$ HbA and $77.5\% \pm 2.5\%$ HbF [6]. The acetylated form of HbF or HbF^{ac} is also commonly present in such samples, Shiao and Ou estimated the mean for their population being $10.5\% \pm 2.28\%$ [7]. For our purpose, the band intensities for unaffected newborn samples were estimated as being 20% HbA, 10% HbF^{ac} and 70% HbF. Similarly, SCD (SS version, also referenced as FS) was approximated as 0% HbA, 10% HbF^{ac}, 76% HbF and 14% HbS from prior literature [8], and for AFSC control the intensities were approximated as 36% HbA, 25% HbF, 23% HbS and 16% HbC [9]. However, a random multiplier from the range of $\pm 7\%$ calculated from intensity percentages is applied to each individual band in order to give variety to the synthetic samples, while keeping the sum of each samples' band intensities equal to 100%. Band pixel widths and distances between them were estimated from Fig. 3 of Frömmel's work [15].

By utilizing the Sig2Img model, synthetic gel side images were created that would approximate the real-world application of NBS. 44 sample positions in total, each side starting and ending with an AFSC control, and also two controls at the middle. This amounts to 40 positions used for patient samples. In experiments 1.1 to 1.4 and 2.1 to 2.4 where a full and a partial gel side was used, 40 and 19 unaffected newborn samples were created, along with 4 controls. In experiments 3.1 to 3.5 and 4.1 to 4.2 where SCD samples were also

introduced, 6 SCD samples replaced unaffected newborn samples. This data is available upon reasonable request.

C. ARTIFICIAL DISTORTIONS

Image processing functions were developed to simulate sample-to-sample and within-sample distortions in the synthetic gel side images. These functions are comprised of two parts: the generation of a one-dimensional distortion profile curve, and the distortion of the image according to the profile curve.

The sample-to-sample distortion profile generator works as follows. First, three sinusoidal signals are generated with random amplitude, phase and frequency as

$$f_i(x) = a_i \sin(2\pi(\omega_i x + p_i)), \quad x \in [0, 1] \quad (1)$$

where $a_i, p_i \in [0, 1]$ are uniformly distributed random variables, and $\omega_i = i$ for $i = 1, 2, 3$. Notice that since all the frequencies ω_i are integers, it follows that $f_i(0) = f_i(1) = 0$ for all i . These sinusoidal signals are then summed together, and the sum is normalized to have the maximum absolute value of $a_{\max} = 75$:

$$s(x) = a_{\max} \frac{\sum f_i(x)}{\max(|\sum f_i(x)|)}. \quad (2)$$

Next, pixel coordinates $c_x \in [0, W - 1]$ are mapped to the interval $x \in [0, 1]$, where W denotes the width of the image in pixels. This mapping allows to generate the discrete sum curve vector \mathbf{s} , whose values correspond to $s(x)$ at the mapped pixel coordinate values. The index corresponding to maximum absolute value,

$$j = \operatorname{argmax}(|\mathbf{s}|), \quad (3)$$

is used to slice the sum vector to obtain a distortion vector \mathbf{d}_0 whose end points have zero gradient:

$$\mathbf{d}_0 = \mathbf{s}[j : \mathbf{s}[j]], \quad (4)$$

where Python-style slicing notation is used. Finally, a linear gradient can be added to the distortion vector \mathbf{d} as

$$\mathbf{d} = \mathbf{d}_0 + b c_x, \quad (5)$$

where b is a constant 0.02.

Within-sample distortion generation works as above, but instead of three sinusoidal signals we used only one,

$$f(x) = a_{\max} \sin(\pi x), \quad (6)$$

where the amplitude a_{\max} is set to 4 for convex and -4 for concave banana-shaped distortion profiles.

These distortion profiles can then be used to distort an input image (which can be a synthetic gel side image or an image of a sample) by displacing its pixel columns vertically by the amount of $\mathbf{d}(c_x)$, where non-integer displacements can be implemented using linear interpolation of pixel values. Sample-to-sample distortion is applied between 100 x-coordinates from the start and 100 x-coordinates before the end of the image. Within-sample distortion is applied

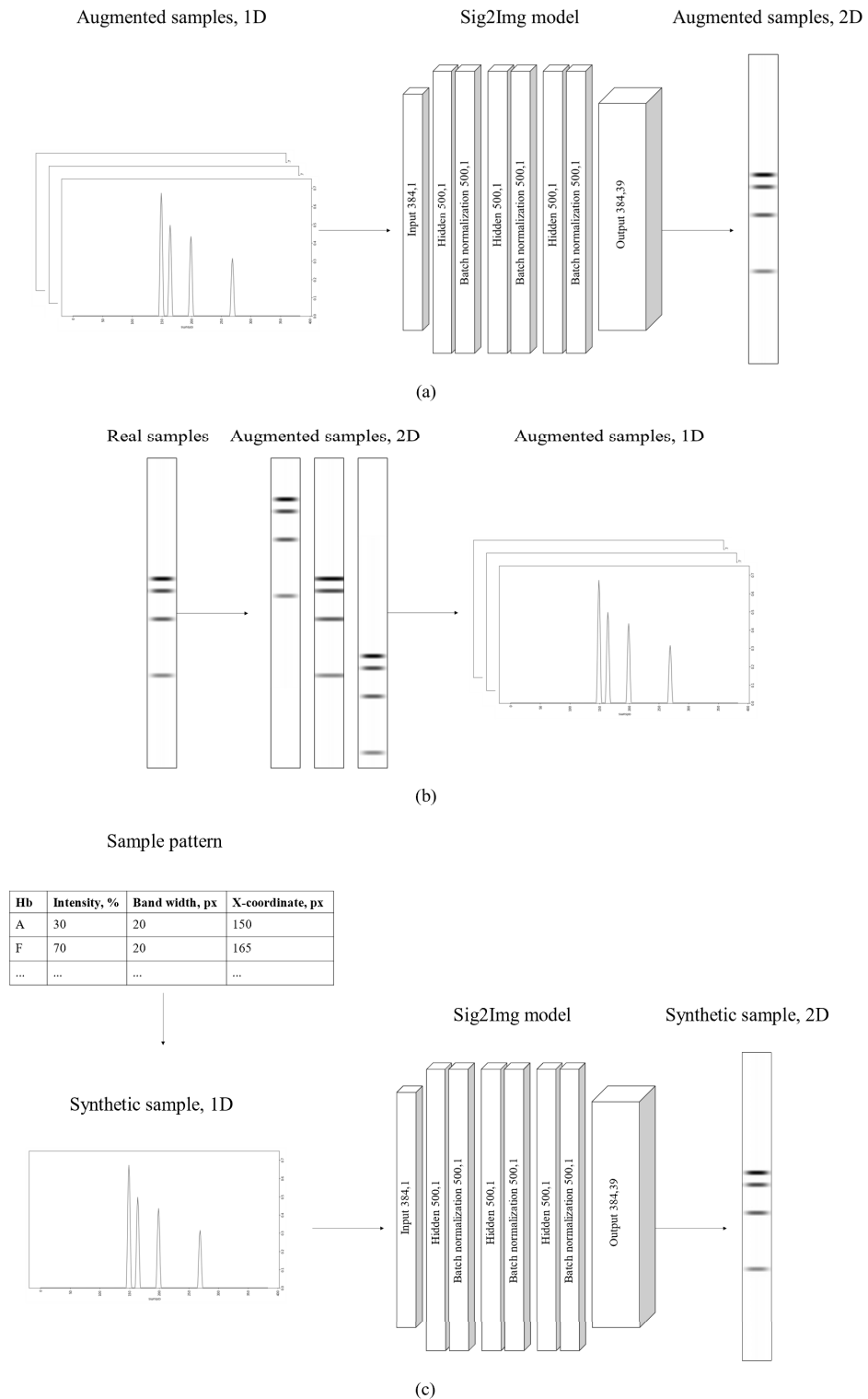


FIGURE 3. In (a), mean signals would be used as the input for Sig2Img model training, and the augmented images as the output. Subfigure (b) showcases this augmentation process. Subfigure (c) demonstrates the inference usage of the Sig2Img model, where sample patterns from prior literature would be used to create a synthetic mean signal, which is fed to the model in order to produce a synthetic sample image.

to all pixels in a sample except for the leftmost and the rightmost pixel columns. To obtain a fully distorted synthetic gel side image, the individual samples are first distorted with

the within-sample distortion profile. In our experiments, the within-sample distortions are applied to an individual sample with a probability of 10%, and the distortion profile alternates

between convex and concave banana-shapes for each applied distortion. Then, the resulting image is further distorted using the sample-to-sample distortion profile. Examples of within-sample and sample-to-sample distortions are depicted in Fig. 2.

We justify the use of only three sinusoidal signals in the creation of the sample-to-sample distortion profile by our experience with prior literature [18], [19], [30] and real-world gel images, which do not typically contain higher frequency distortions.

The magnitude of the Gaussian noise applied in experiments 1.4, 2.4, 3.4, 3.5, and 4.2, was estimated of being five times of background noise, when the office scanner was used.

D. GEOMETRIC CORRECTION

The geometric correction method can be divided into two main phases: the correction of within-sample distortions by straightening individual bands and the correction of sample-to-sample distortions by aligning all samples. These phases and all the steps taken in the process are summarized in Fig. 4.

The correction method requires the position of each sample. The locating of samples can be done manually, but with our synthetic data we already have this information, since we have determined the position for each sample when creating the gel side. With this knowledge, the correction process starts by straightening the curved bands within each individual sample. First step is to locate one band in the sample, and then shifting each pixel column up or down until the band is straight. These are also listed as first and second step in Fig. 4. Since all the bands within a single sample are curved the same way, by locating one band and shifting the pixel columns to make it straight, it results in straightening the whole sample.

This locating of a single band starts by creating a 1D signal from each pixel column of the sample. In this study, the synthetic samples represent an ideal case where by storing the position of the highest peak in each pixel columns' signal, the stored locations would trace the band with the greatest intensity. This is because in the synthetic samples the bands are good quality with a clear distinction of the intensities between the bands. However, in real IEF gels, this may not be the case, since the highest peaks can alter between multiple bands in adjacent columns, or a dust particle with greater intensity can even cause the highest peak to be completely outside of the bands.

To match the needs of a real IEF gel, the single band is being traced by storing the location of the four highest peaks in each pixel columns' signal. The process starts by selecting the peak from the first pixel column that is closest to the median of the first five columns' highest peaks. Then it iterates through each pixel column, comparing which of the four highest peaks is closest to the previously selected peak, and selects that one as the next point. An assumption is made that even with a sample containing multiple bands with equally distributed intensities, one band would always contain one of the four highest peaks in each pixel column. In a rare case, if none of the four highest peaks are inside the

same band in the next column, then the previous point that was in the band is selected again. Fig. 5 shows an example of this process with a dust particle affecting the tracing of a single band. First image shows the original image. Second image shows the four highest peaks found in each pixel column, where red represents pixel with greatest intensity, with blue, green, and, purple representing the second, third, and fourth greatest intensity, respectively. As this example shows, the dust particle has the greatest intensity in a few pixel columns, but the distance comparison finds the second highest peak in those pixel columns to be the closest to the previously selected points, which causes the method to trace a single band successfully, as shown in the last image.

The last procedure before straightening the bands is applying a 6th degree polynomial function to the chosen points. This results in more accurate tracing of the band and therefore the outcome is much smoother after shifting the pixel columns up and down, especially for bands that have very large slope.

Fig. 6 shows the straightening of a single sample. The first image shows the original curved sample, from which a single band has been traced with the method described earlier. Each pixel column is then shifted up or down to the median of the chosen points so that they form a straight horizontal line.

The alignment of samples is done after all the bands have been straightened. The first control samples' HbF band is the reference, to which the other samples are aligned. This band is identified by creating a 1D signal of the control sample by calculating the mean value for each row of the image. HbF band can be identified from this signal by locating the second peak over a certain threshold, since the order of the bands in a control sample is known to be HbA, HbF, HbS, and HbC. This corresponds to step three of the process summarized in Fig. 4.

After this, the fourth step of the process is done, where the whole sample area is moved so that the HbF band of the first control sample is aligned with a preselected y-coordinate. This ensures that the y-coordinate of this band is definite, and it can be used as a reference location. It also establishes that all the images straightened will have the sample area at the same position.

Then, the 1D signals are created the same way for rest of the samples, and the highest peaks of all these signals are stored. The highest peak in each samples' signal is used to align the sample with the first control samples HbF band. In normal newborn samples and in most other newborn samples, the band with the greatest quantity, and therefore the band with the highest peak in the signal, is HbF. These two steps are steps five and six in Fig. 4.

Control samples have generally HbA with the greatest quantity. So, the alignment of control samples is fixed, by identifying HbA band in each control sample, the same way as identifying HbF band in the third step of the process, and then aligning each control samples' HbA band with the HbA band of the first control sample, shown as the seventh step in Fig. 4.

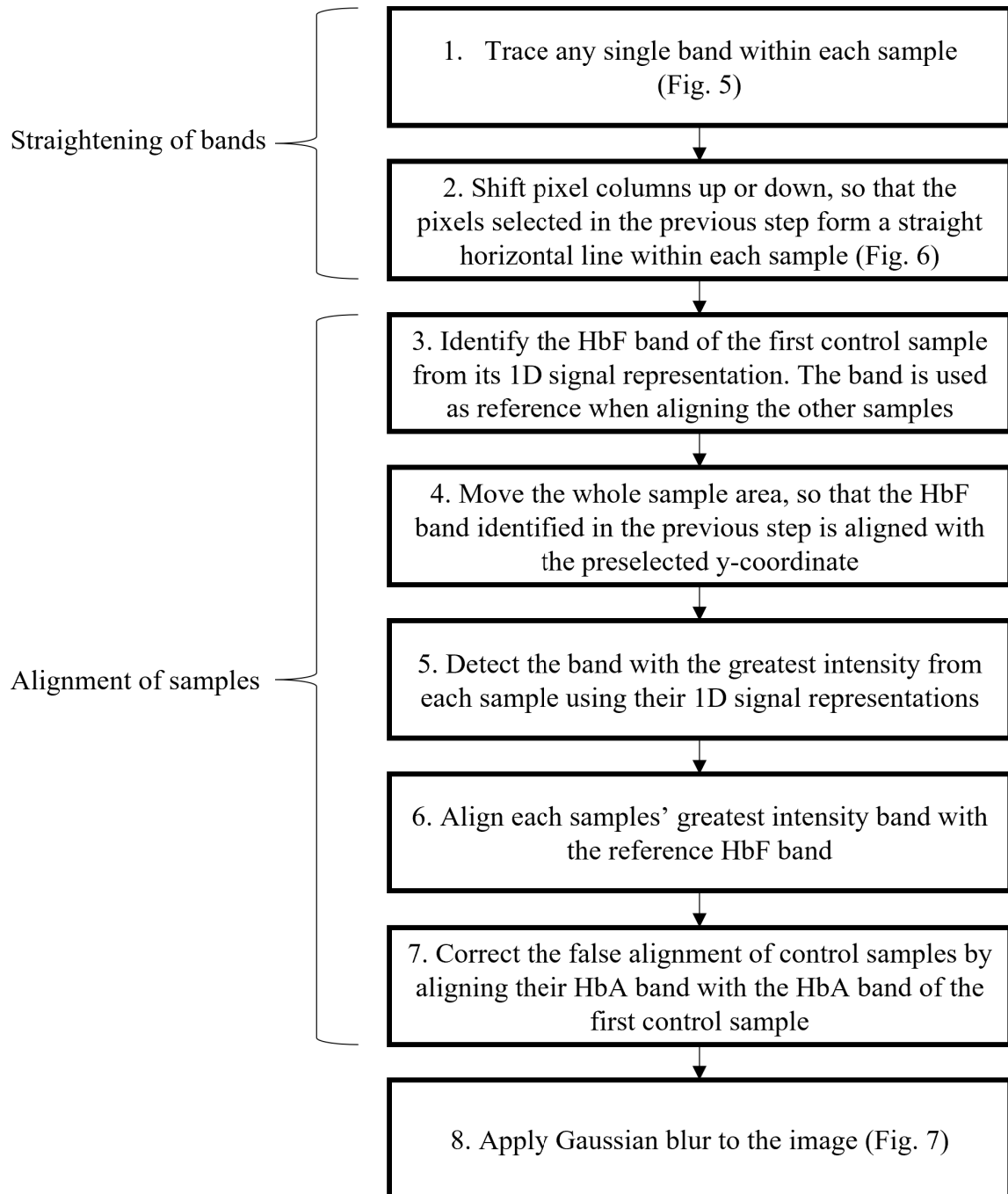


FIGURE 4. Summary of the steps in the geometric correction process.

The eighth and final step of the process, after straightening the bands and aligning all the samples, is applying a Gaussian blur to the image, to smoothen any bands that might still have rough edges caused by the pixel columns shifting. This is done to transform the corrected samples to appear more lifelike. The Gaussian blur is applied with a Gaussian kernel that has a width and height of 7 and a standard deviation of 1 both in X and Y directions. Fig. 7 shows the effect of the blur to a single sample.

E. HARDWARE AND SOFTWARE

All software was written in a JupyterLab 0.35.4 environment with Python 3.6.7 programming language. The following Python libraries were used for data analysis and algorithm development: pandas 0.23.4, NumPy 1.15.4, OpenCV 4.5.4, and Matplotlib 3.0.2. Keras and Tensorflow 2.2.0 were used to fit the Sig2Img model, and to generate synthetic variations of sample images. R Open 4.0.2 and RStudio 1.4.1106 were used to calculate the statistics in Tables 1 and 2.

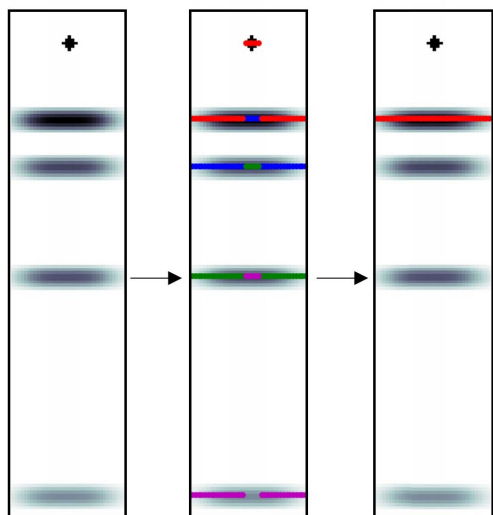


FIGURE 5. Example of the tracing of a single band. The added dust particle causes the highest peak determination logic to break. From the first image the four highest peaks (intensities from highest to lowest are red, blue, green, and purple) are calculated for each pixel column. The particle causes highest peak to shift for five columns in the second image. The third image shows the successfully selected points after the distance comparison process.

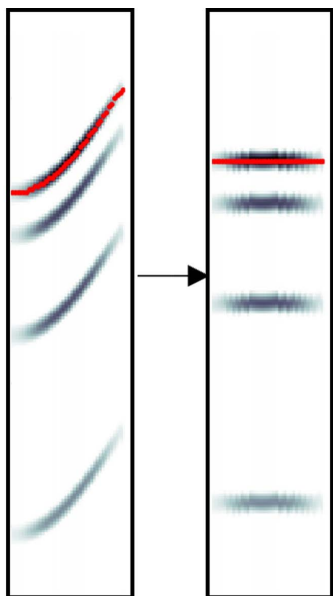


FIGURE 6. Example of a single band, before and after shifting each pixel column up or down, so that the selected points form a straight horizontal line.

Intel® Core™ i9-9940X processor was used for running the software.

III. RESULTS AND DISCUSSION

Before and after images for experiments 1.1 to 1.4, 2.1 to 2.4, 3.1 to 3.4, and 4.1 are presented in Figs. 8, 9, 10, and 11, respectively. Also, Figs. 10 and 11 both have one

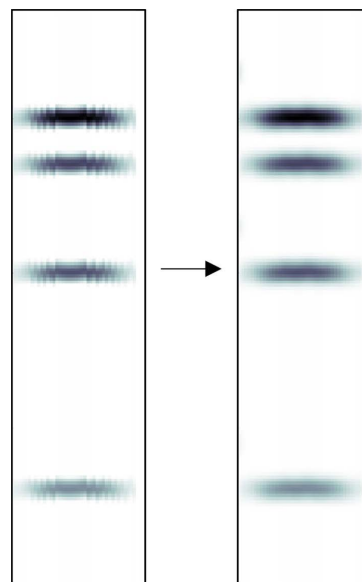


FIGURE 7. Before and after applying a Gaussian blur to a straightened sample.

subfigure that show an example image from the 50 runs of the experiments 3.5 and 4.2.

Experiments 1.1, 2.1, 3.1, and 4.1, shown in subfigures (a) of these figures, demonstrate that if the gel side to be corrected does not contain any distortions, the correction method does not visually hinder the image quality. The relative quantitation differences in Table 1 also illustrates this, as the biggest mean difference in relative quantitation for these experiments was 0.46% in HbF for experiment 1.1 and maximum absolute difference was 0.67 in HbF for experiment 3.1. These also show the minimal effect of the Gaussian blur to the quantitation, since there is no other image processing done in these experiments.

Experiments 1.2, 2.2, and 3.2, show results for correcting images that have sample-to-sample distortions in subfigures (b) of Figs. 8, 9, and 10. The outcome for these experiments look faultless visually inspected, with no clear errors in the straightening or alignment. Table 1 also supports this argument, with again only slight differences in the relative quantitation.

Experiments 1.3, 2.3, and 3.3 have randomly applied within-sample distortions. Fig. 8 (c), shows experiment 1.3, that has six bands affected with this kind of distortion. Similarly, Figs. 9 (c) and 10 (c) show experiments 2.3 and 3.3, with within-sample distortions affecting three and two bands, respectively. Again, the outcomes after the correction in these experiments are flawless.

Random Gaussian noise was applied for experiments 1.4, 2.4, and 3.4, shown in subfigures (d) of Figs. 8, 9, and 10. These figures show that the method works as intended, even after applying Gaussian noise, that has a magnitude estimated to be five times greater than the background noise.

It is clear from Table 1 that in each singleton experiment the mean relative quantitation difference is very

TABLE 1. Relative quantitation difference results for each singleton experiment and each Hb band. Mean difference and absolute maximum of the differences (%) were calculated for each experiment. Band counts reflect the number of samples used per experiment.

Exp	Hb	Count	Mean diff	Abs max	Exp	Hb	Count	Mean diff	Abs max	Exp	Hb	Count	Mean diff	Abs max	Exp	Hb	Count	Mean diff	Abs max	
1.1	F ^{ac}	40	-0.31	0.38	2.1	F ^{ac}	18	-0.31	0.35	3.1	F ^{ac}	40	-0.26	0.37	4.1	F ^{ac}	18	-0.20	0.34	
1.1	A	44	-0.18	0.28	2.1	A	22	-0.17	0.28	3.1	A	38	-0.19	0.32	4.1	A	16	-0.16	0.30	
1.1	F	44	0.46	0.63	2.1	F	22	0.43	0.62	3.1	F	44	0.41	0.67	4.1	F	22	0.28	0.64	
1.1	S	4	-0.02	0.03	2.1	S	4	-0.02	0.03	3.1	S	10	-0.03	0.07	4.1	S	10	0.02	0.09	
1.1	C	4	-0.01	0.02	2.1	C	4	-0.01	0.02	3.1	C	4	-0.01	0.02	4.1	C	4	-0.01	0.02	
1.2	F ^{ac}	40	-0.48	0.72	2.2	F ^{ac}	18	-0.48	0.53	3.2	F ^{ac}	40	-0.41	0.55						
1.2	A	44	-0.30	0.44	2.2	A	22	-0.28	0.43	3.2	A	38	-0.31	0.47						
1.2	F	44	0.74	1.16	2.2	F	22	0.68	0.94	3.2	F	44	0.65	1.01						
1.2	S	4	-0.02	0.03	2.2	S	4	-0.02	0.03	3.2	S	10	-0.03	0.08						
1.2	C	4	-0.01	0.02	2.2	C	4	-0.01	0.02	3.2	C	4	-0.01	0.01						
1.3	F ^{ac}	40	-0.34	0.55	2.3	F ^{ac}	18	-0.34	0.51	3.3	F ^{ac}	40	-0.27	0.57						
1.3	A	44	-0.20	0.42	2.3	A	22	-0.19	0.39	3.3	A	38	-0.20	0.40						
1.3	F	44	0.51	0.97	2.3	F	22	0.47	0.89	3.3	F	44	0.43	0.95						
1.3	S	4	-0.02	0.03	2.3	S	4	-0.02	0.03	3.3	S	10	-0.03	0.07						
1.3	C	4	-0.01	0.02	2.3	C	4	-0.01	0.02	3.3	C	4	-0.01	0.02						
1.4	F ^{ac}	40	-0.45	1.09	2.4	F ^{ac}	18	-0.38	0.70	3.4	F ^{ac}	40	-0.48	1.13						
1.4	A	44	-0.28	0.69	2.4	A	22	-0.24	0.67	3.4	A	38	-0.33	0.77						
1.4	F	44	0.70	1.25	2.4	F	22	0.56	1.10	3.4	F	44	0.76	1.54						
1.4	S	4	-0.10	0.33	2.4	S	4	-0.09	0.18	3.4	S	10	-0.17	0.40						
1.4	C	4	-0.06	0.16	2.4	C	4	0.01	0.13	3.4	C	4	0.01	0.14						

TABLE 2. Relative quantitation difference results for each repetition experiment and each Hb band. Mean difference, absolute maximum and standard deviation of the differences (%) were calculated from 50 repeated tests per experiment.

Exp	Hb	Count	Mean diff	Abs max	ISD diff	Exp	Hb	Count	Mean diff	Abs max	ISD diff
3.5	F ^{ac}	2000	-0.55	1.46	0.29	4.2	F ^{ac}	900	-0.45	1.29	0.33
3.5	A	1900	-0.41	1.20	0.27	4.2	A	800	-0.34	1.06	0.31
3.5	F	2200	0.89	2.19	0.47	4.2	F	1100	0.68	1.89	0.53
3.5	S	500	-0.15	0.84	0.21	4.2	S	500	-0.12	0.98	0.21
3.5	C	200	-0.07	0.34	0.11	4.2	C	200	-0.07	0.36	0.12

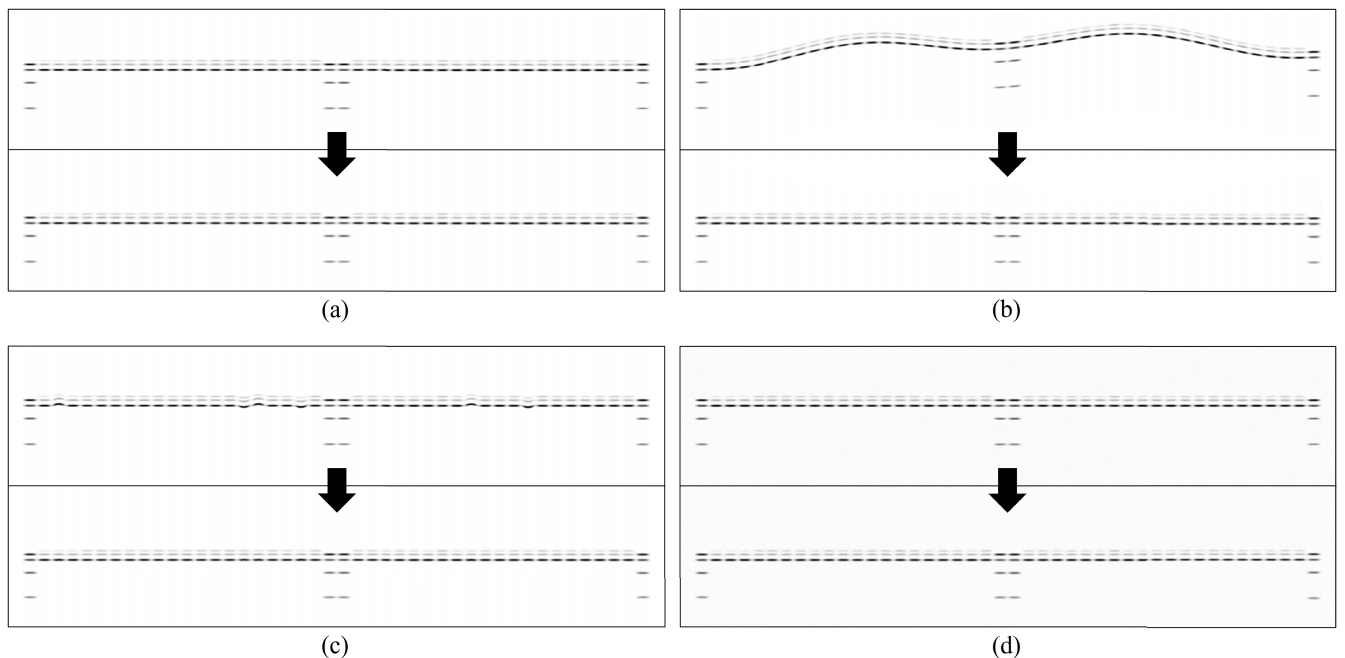


FIGURE 8. (a) Experiment 1.1 (b) Experiment 1.2 (c) Experiment 1.3 (d) Experiment 1.4.

low and therefore would not affect the identification of hemoglobinopathies with naked eye. The natural variation for the mean quantity of both HbA and HbF in a study by Chandrakasan and Kamat was 2.5% [6] and 2.28% for

HbF^{ac} in a study by Shiao and Ou [7], which indicates that the differences shown in Table 1 are not significantly greater, and thus not enough to influence the interpretation of the gel. Also, in each experiment the largest difference is

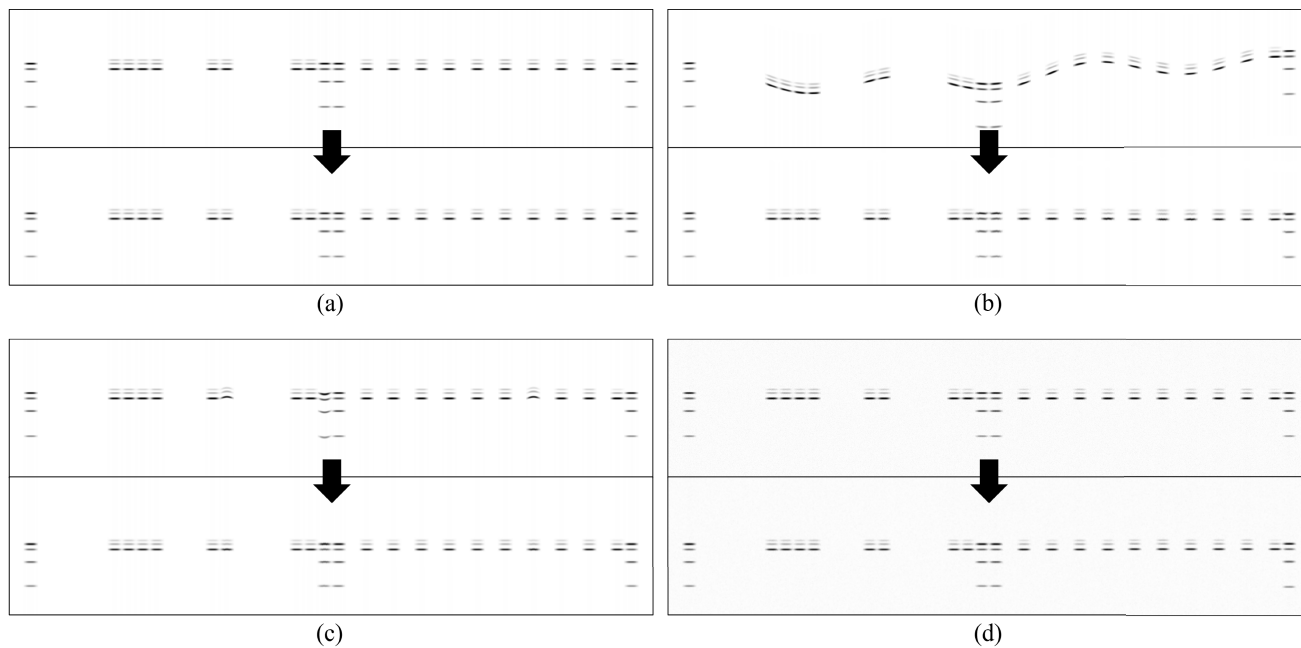


FIGURE 9. (a) Experiment 2.1 (b) Experiment 2.2 (c) Experiment 2.3 (d) Experiment 2.4.

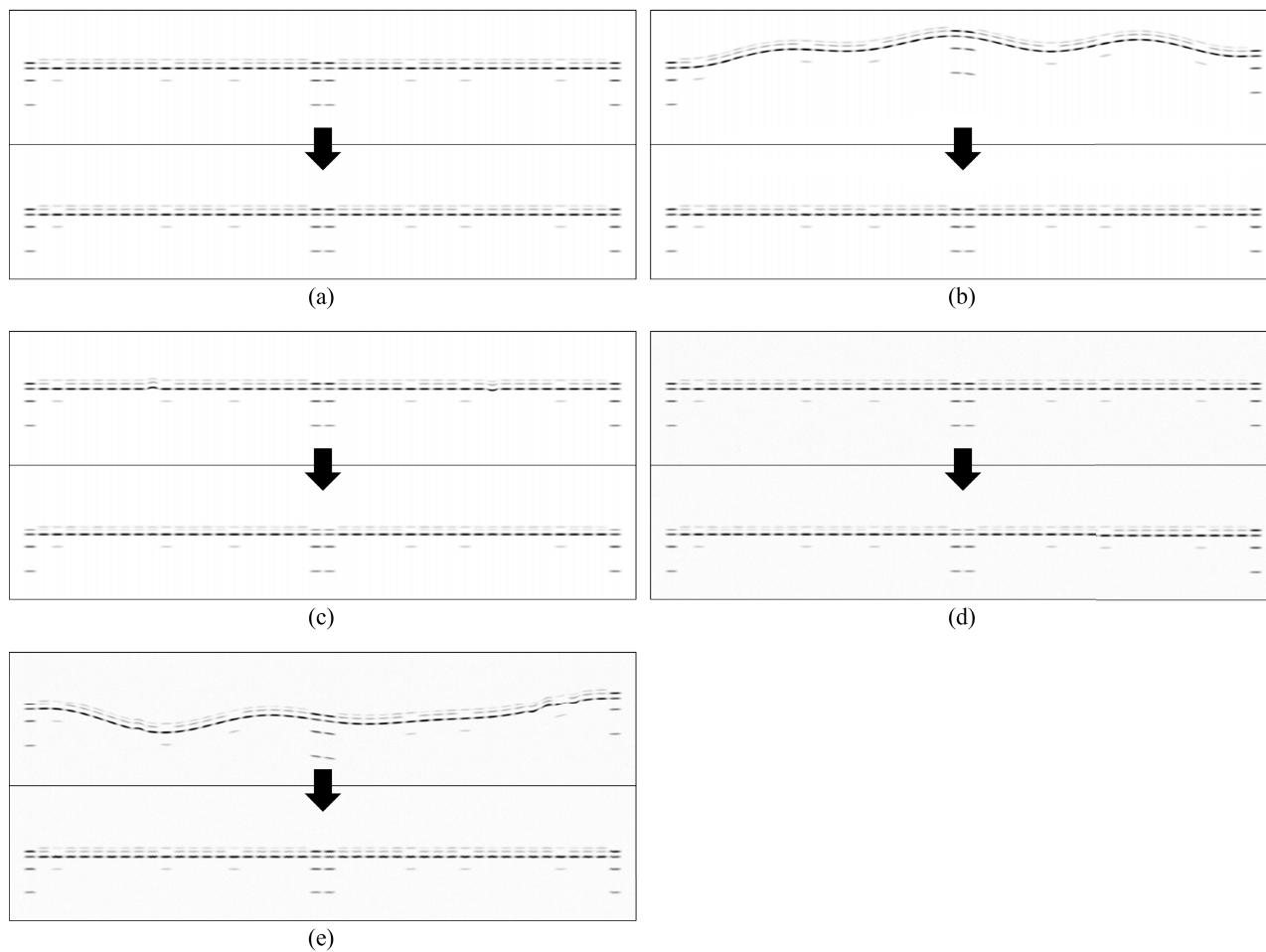


FIGURE 10. (a) Experiment 3.1 (b) Experiment 3.2 (c) Experiment 3.3 (d) Experiment 3.4 (e) Example image of 50 runs from experiment 3.5.

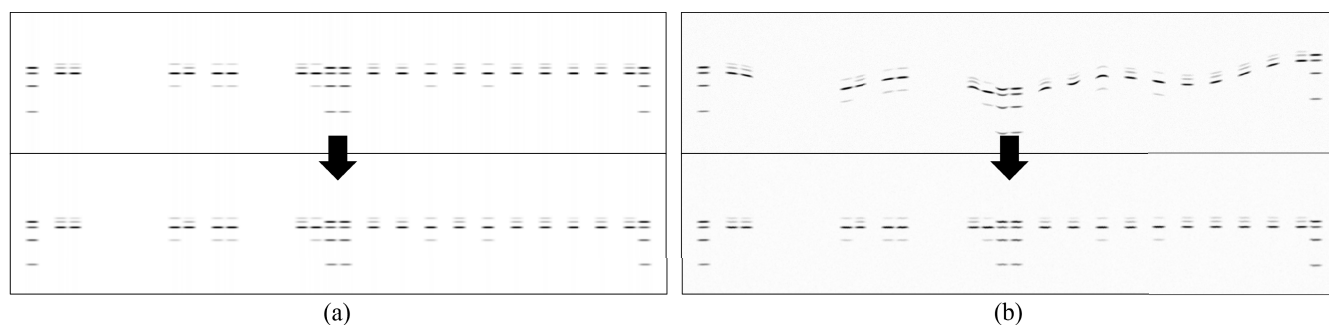


FIGURE 11. (a) Experiment 4.1 (b) Example image of 50 runs from experiment 4.2.

in F and HbF^{ac} bands, with HbF increasing and HbF^{ac} decreasing. Since both bands represent the same hemoglobin, even a large difference would not affect the actual quantitation of that hemoglobin.

Experiments 3.5 and 4.2 were worst-case repetition experiments. Example images from both experiments are shown in Figs. 10 (e) and 11 (b). For both experiments, 50 worst-case gels were created, with SCD samples, within-sample and sample-to-sample distortion and added Gaussian noise. Experiment 3.5 contained full gel sides, therefore the 50 gel sides contained 2200 individual samples. Experiment 4.2 contained partial gel sides, so there were 1100 individual samples. In both experiments, the correction method successfully aligned and straightened every sample in each gel side.

Even though each gel side was straightened and aligned correctly, the worst-case scenario was clearly, as expected, the most difficult for the correction method, which can be seen from the increase in absolute maximum difference in Table 2. The mean difference is in line with Table 1 results, with largest mean difference in HbF, but as mentioned earlier, this amount is still acceptable.

IV. CONCLUSION

In this study, we proposed novel methods for the generation, distortion, and correction of IEF gel results. The novel Sig2Img neural network allows us to create synthetic IEF samples, and therefore gives the opportunity to create gel sides with any kind of samples, with redefinable band intensities and positions. This model combined with the image processing functions to create artificial distortions gives huge opportunities since there are no publicly available datasets containing images of IEF gels.

Thirteen singleton and two repetition experiments with different gel sides were created with the aforesaid method to test the limits of the geometric correction. The experiments showed that the geometric correction method not only managed to straighten and align each gel side successfully, but also retained the relative quantities of the bands accurately. This is crucial for IEF gels, since when screening for hemoglobinopathies, in addition to detecting abnormal bands in a sample, the quantities of the bands are also as meaningful.

Therefore it is important that the straightening does not hinder the image result. As the Tables 1 and 2 show, the mean relative quantitation difference is below the normal variation of Hb bands in each experiment.

The correction also enables relative quantitation of the bands in situations where it is not possible otherwise. Because the relative band quantities are calculated from the peak areas in the 1D signal representation, the signals created from distorted samples do not represent the bands accurately. However, after correctly straightening, the signal representation now contains distinct peaks which correspond to the actual bands, and therefore the quantitation is made possible.

The main limitation of this study was the limited image data of real gel material, which reduces our Seg2Img model's ability to produce more diverse and representative synthetic gel samples. Future work would include additional data gathering and model evaluation.

It should be noted, that the gel correction method was also tested to perform well with a real-life screening gels, where quantitation was made possible for highly distorted bands [5]. The results showcased in this publication provide further evidence of this, by systematically testing for common normal and abnormal situations arising from routine hemoglobinopathy screening.

Enabling quantitation in situations where the gel is too distorted for interpretation is what gives the method clinical significance. NBS laboratories using IEF for screening hemoglobinopathies could include relative quantitation to their clinical workflow, which in turn would enable more sophisticated gel analysis such as hemoglobin migration assessment during a gel run.

ACKNOWLEDGMENT

The authors would like to thank PerkinElmer Wallac Oy for the gel material and samples used in the study.

REFERENCES

- [1] P. Jindatanmanusan, S. Riolueang, W. Glomglao, Y. Sukontharangsri, S. Chamnanvanakij, K. Torcharus, and V. Viprakasit, "Diagnostic applications of newborn screening for α -thalassaemias, haemoglobins e and h disorders using isoelectric focusing on dry blood spots," *Ann. Clin. Biochemistry, Int. J. Lab. Med.*, vol. 51, no. 2, pp. 237–247, Mar. 2014, doi: 10.1177/0004563213491078.

- [2] Centers for Disease Control and Prevention. (2016). *Hemoglobinopathies: Current Practices for Screening, Confirmation and Follow-Up DECEMBER 2015*. Accessed: Feb. 28, 2022. [Online]. Available: https://www.cdc.gov/ncbddd/sickleccll/documents/nbs_hemoglobinopathy-testing_122015.pdf
- [3] W. Clarke and M. Marzinke, *Contemporary Practice in Clinical Chemistry*. New York, NY, USA: Academic, 2020, pp. 413–427.
- [4] N. Volpi, “‘Fast moving’ and ‘slow moving’ heparins, dermatan sulfate, and chondroitin sulfate: qualitative and quantitative analysis by agarose-gel electrophoresis,” *Carbohydrate Res.*, vol. 247, pp. 263–278, Sep. 1993, doi: [10.1016/0008-6215\(93\)84259-9](https://doi.org/10.1016/0008-6215(93)84259-9).
- [5] S. Mäntyniemi, “Geometric sample area correction of an IEF gel in a newborn screening application using image processing methods,” M.S. thesis, Dept. Comput. Health Technol., Univ. Turku, Turku, Finland, 2022. Accessed: Feb. 22, 2022. [Online]. Available: <https://urn.fi/URN:NBN:fi-fe2022020918320>
- [6] S. Chandrakasan and D. Kamat, “An overview of hemoglobinopathies and the interpretation of newborn screening results,” *Pediatric Ann.*, vol. 42, no. 12, pp. 502–508, Dec. 2013, doi: [10.3928/00904481-20131122-11](https://doi.org/10.3928/00904481-20131122-11).
- [7] S.-Y.-P. K. Shiao and C.-N. Ou, “Accurate measurements of fetal hemoglobin for neonates with different gestational ages,” *Hemoglobin*, vol. 30, no. 4, pp. 419–435, Jan. 2006, doi: [10.1080/03630260600867883](https://doi.org/10.1080/03630260600867883).
- [8] A. Kutlar, F. Kutlar, J. B. Wilson, M. G. Headlee, and T. H. J. Huisman, “Quantitation of hemoglobin components by high-performance cation-exchange liquid chromatography: Its use in diagnosis and in the assessment of cellular distribution of hemoglobin variants,” *Amer. J. Hematol.*, vol. 17, no. 1, pp. 39–53, Jul. 1984, doi: [10.1002/ajh.2830170106](https://doi.org/10.1002/ajh.2830170106).
- [9] M. H. Hampe, S. N. Panaskar, R. Mestri, N. A. Kamble, P. S. Chaudhari, A. A. Yadav, P. W. Ingale, and B. Kandasamy, “Screening of hemoglobinopathies in 32000 dried blood spot samples by cation exchange high performance liquid chromatography—An Indian study,” *Int. J. Med. Sci.*, vol. 4, no. 7, pp. 1–8, Jul. 2017, doi: [10.14445/23939117/ijms-v4i7p101](https://doi.org/10.14445/23939117/ijms-v4i7p101).
- [10] A. Dasgupta and A. Wahed, “Hemoglobinopathies and thalassemia,” in *Clinical Chemistry, Immunology and Laboratory Quality Control*, 2nd ed., A. Dasgupta and A. Wahed, Eds. Amsterdam, The Netherlands: Elsevier, 2021, ch. 23, pp. 457–487, doi: doi.org/10.1016/B978-0-12-815960-6.00005-4.
- [11] D. Weatherall and J. Clegg, “Inherited haemoglobin disorders: An increasing global health problem,” *Bulletin World Health Org.*, vol. 79, no. 8, pp. 704–712, Feb. 2001. Accessed: Feb. 28, 2022. [Online]. Available: <https://apps.who.int/iris/handle/10665/268402>
- [12] V. M. Pinto, M. Balocco, S. Quintino, and G. L. Forni, “Sickle cell disease: A review for the internist,” *Internal Emergency Med.*, vol. 14, no. 7, pp. 1051–1064, Aug. 2019, doi: [10.1007/s11739-019-02160-x](https://doi.org/10.1007/s11739-019-02160-x).
- [13] G. R. Serjeant, “The natural history of sickle cell disease,” *Cold Spring Harbor Perspect. Med.*, vol. 3, no. 10, Oct. 2013, Art. no. a011783, doi: [10.1101/cshperspect.a011783](https://doi.org/10.1101/cshperspect.a011783).
- [14] D. P. Kaufman. (Mar. 2021). *Physiology, Fetal Hemoglobin*. Accessed: Feb. 2, 2022. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK500011/>
- [15] C. Frömmel, “Newborn screening for sickle cell disease and other hemoglobinopathies: A short review on classical laboratory methods—Isoelectric focusing, HPLC, and capillary electrophoresis,” *Int. J. Neonatal Screening*, vol. 4, no. 4, p. 39, Dec. 2018, doi: [10.3390/ijns4040039](https://doi.org/10.3390/ijns4040039).
- [16] PerkinElmer. (2022). *Saving Lives with Hemoglobinopathy Detection. MigeleT Gel Electrophoresis Unit*. Accessed: Feb. 28, 2022. [Online]. Available: https://resources.perkinelmer.com/lab-solutions/resources/docs/BRO_Hemoglobinopathies_MigeleTM_Gel_Electrophoresis_Unit_014089_July2018.pdf
- [17] A. Janik, “What is the best measure of the detector signal: The peak area or the peak height?” *J. Chromatograph. Sci.*, vol. 13, no. 2, pp. 93–96, Feb. 1975, doi: [10.1093/chromsci/13.2.93](https://doi.org/10.1093/chromsci/13.2.93).
- [18] D. E. Garfin, “One-dimensional gel electrophoresis,” in *Methods in Enzymology*, vol. 182. New York, NY, USA: Academic, 1990, pp. 425–441.
- [19] L. B. Koutny and E. S. Yeung, “Automated image analysis for distortion compensation in sequencing gel electrophoresis,” *Appl. Spectrosc.*, vol. 46, no. 1, pp. 136–141, Jan. 1992, doi: [10.1366/0003702924444461](https://doi.org/10.1366/0003702924444461).
- [20] I. Bajla, I. Holländer, S. Fluch, K. Burg, and M. Kollár, “An alternative method for electrophoretic gel image analysis in the GelMaster software,” *Comput. Methods Programs Biomed.*, vol. 77, no. 3, pp. 209–231, Mar. 2005.
- [21] I. Bajla, I. Holländer, and K. Burg, “Improvement of electrophoretic Gel image analysis,” *Meas. Sci. Rev.*, vol. 1, no. 1, pp. 5–10, Sep. 2001.
- [22] A. Intarapanich, S. Kaewkamnerd, P. J. Shaw, K. Ukosakit, S. Tragoonrungs, and S. Tongsimma, “Automatic DNA diagnosis for 1D gel electrophoresis images using bio-image processing technique,” *BMC Genomics*, vol. 16, no. S12, p. S15, Dec. 2015, doi: [10.1186/1471-2164-16-s12-s15](https://doi.org/10.1186/1471-2164-16-s12-s15).
- [23] M. M. Goetz, M. C. Torres-Madroño, S. Röthlisberger, and E. Delgado-Trejos, “Preprocessing of 2-dimensional gel electrophoresis images applied to proteomic analysis: A review,” *Genomics, Proteomics Bioinf.*, vol. 16, no. 1, pp. 63–72, Feb. 2018.
- [24] T. Barbu, “Variational image denoising approach with diffusion porous media flow,” *Abstract Appl. Anal.*, vol. 2013, pp. 1–8, Jan. 2013, doi: [10.1155/2013/856876](https://doi.org/10.1155/2013/856876).
- [25] G. Ndezi, C. Kiyaga, A. G. Hernandez, D. Munube, T. A. Howard, I. Ssewanyana, J. Nsungwa, S. Kiguli, C. M. Ndugwa, R. E. Ware, and J. R. Aceng, “Burden of sickle cell trait and disease in the Uganda sickle surveillance study (US3): A cross-sectional study,” *Lancet Global Health*, vol. 4, no. 3, pp. e195–e200, Mar. 2016, doi: [10.1016/s2214-109x\(15\)00288-0](https://doi.org/10.1016/s2214-109x(15)00288-0).
- [26] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Netw.*, vol. 61, pp. 85–117, Oct. 2015, doi: [10.1016/j.neunet.2014.09.003](https://doi.org/10.1016/j.neunet.2014.09.003).
- [27] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. Cham, Switzerland: Springer, 2006, pp. 54–55.
- [28] G. Hinton, N. Srivastava, and K. Swersky. (2012). *Lecture 6A Overview of Mini-Batch Gradient Descent*. Coursera Lecture Slides. Accessed: Feb. 28, 2022. [Online]. Available: <https://class.coursera.org/neuralnets-2012-001/lecture>
- [29] A. Savitzky and M. J. E. Golay, “Smoothing and differentiation of data by simplified least squares procedures,” *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964, doi: [10.1021/ac60214a047](https://doi.org/10.1021/ac60214a047).
- [30] M. Starita-Geribaldi, A. Hourri, and P. Sudaka, “Lane distortions in gel electrophoresis patterns,” *Electrophoresis*, vol. 14, no. 1, pp. 773–781, 1993, doi: [10.1002/elps.11501401121](https://doi.org/10.1002/elps.11501401121).



SANTERI MÄNTYNIEMI received the M.Sc. (Tech.) degree in information and communication technology from the University of Turku, Turku, Finland, in 2022. His current research interests include artificial intelligence, machine learning, and computer vision.



EERO LEHTONEN (Member, IEEE) received the M.Sc. degree in mathematics and the D.Sc. (Tech.) degree in electronics from the University of Turku, Turku, Finland, in 2006 and 2013, respectively. He is currently working as a Senior Researcher with the Digital Health Technology Group, Department of Computing, University of Turku. His research interests include developing computer vision and sensor fusion algorithms for medical technology and imaging.



AKI KOIVU (Member, IEEE) received the Ph.D. degree in computer science from the University of Turku, Turku, Finland, in 2022. He has had a Data Scientist and a Data Analyst roles at PerkinElmer, Turku, from 2016 onwards, where he is currently employed as a Principal Data Scientist. His current research interests include neural networks, health informatics, and risk calculation for health applications. He has acted as the Student Representative Section Officer in IEEE Finland Section for the 2020 to 2021 term.

• • •