

Article

Enhanced neonatal screening for sickle cell disease: Human-guided deep learning with CNN on isoelectric focusing images

Kpangni Alex Jérémie Koua^{1,2,*}, Cheikh Talibouya Diop¹, Lamine Diop³, Mamadou Diop¹¹ University Gaston Berger of Saint-Louis, BP 234, Nationale 2, route de Ngallèle, Saint-Louis 32000, Senegal² University Jean Lorougnon Guédé of Daloa, Daloa BP 150, Côte d'Ivoire³ EPITA Research Laboratory (LRE), 14–16 rue Voltaire, 94270 Le Kremlin-Bicêtre, France* **Corresponding author:** Kpangni Alex Jérémie Koua, jeremickoua@ujlg.edu.ci

CITATION

Koua KAJ, Diop CT, Diop L, Diop M. (2024). Enhanced neonatal screening for sickle cell disease: Human-guided deep learning with CNN on isoelectric focusing images. *Journal of Infrastructure, Policy and Development*. 8(9): 6121. <https://doi.org/10.24294/jipd.v8i9.6121>

ARTICLE INFO

Received: 30 April 2024

Accepted: 28 May 2024

Available online: 5 September 2024

COPYRIGHT



Copyright © 2024 by author(s).

Journal of Infrastructure, Policy and Development is published by EnPress Publisher, LLC. This work is licensed under the Creative Commons Attribution (CC BY) license. <https://creativecommons.org/licenses/by/4.0/>

Abstract: Accurate detection of abnormal hemoglobin variations is paramount for early diagnosis of sickle cell disease (SCD) in newborns. Traditional methods using isoelectric focusing (IEF) with agarose gels are technician-dependent and face limitations like inconsistent image quality and interpretation challenges. This study proposes a groundbreaking solution using deep learning (DL) and artificial intelligence (AI) while ensuring human guidance throughout the process. The system analyzes IEF gel images with convolutional neural networks (CNNs), achieving over 98% accuracy in identifying various SCD profiles, far surpassing the limitations of traditional methods. Furthermore, the system addresses ambiguities by incorporating an “Unconfirmed” category for unclear cases and assigns probability values to each classification, empowering clinicians with crucial information for informed decisions. This AI-powered tool, named SCScreen, seamlessly integrates machine learning with medical expertise, offering a robust, efficient, and accurate solution for SCD screening. Notably, SCScreen tackles the previously challenging diagnosis of major sickle cell syndromes (SDM) in newborns. This research has the potential to revolutionize SCD management. By strengthening screening platforms and potentially reducing costs, SCScreen paves the way for improved healthcare outcomes for newborns with SCD, potentially saving lives and improving the quality of life for affected individuals.

Keywords: SCD; IEF; agarose gel; CNN; visualization; healthcare data analysis

1. Introduction

SCD is a common genetic disorder affecting hemoglobin, recognized as a major global health issue by the WHO (World Health Organization, 2006) and the UN (United Nations, 2009). With over 0.3 million annual cases, SCD poses a significant burden, particularly in Africa, where its prevalence exceeds 10% in the general population, with nearly 2000 children (0.5%) born with a severe form known as “SDM”. Early and accurate detection of SCD is crucial for newborn care, as over half of affected children do not survive beyond the age of five (Gueye et al., 2020). However, SCD screening remains rare in low-income countries, often leading to late diagnosis based solely on clinical signs.

Various screening techniques exist (Frömmel, 2018). One of the most common methods for neonatal SCD screening is IEF (Daniel et al., 2019; El-Haj and Hoppe, 2018) with agarose gels, although this method faces challenges such as variability in image staining, interpretation errors, and certain ambiguities. To improve screening efficiency, we propose an AI-based solution using deep learning, which analyzes agarose gel images for enhanced classification. This approach automates SCD patient

detection and helps identify different sickle cell profiles, overcoming challenges associated with manual interpretation of images. Our main goal is to develop a system based on deep learning classifiers for early SCD detection by analyzing IEF patterns from neonatal agarose gel images to achieve fast and accurate results.

By integrating new data science tools, we aim to improve productivity, collaboration, and the overall efficiency of data-driven research (Bäuerle et al., 2022). Our approach aims to provide a precise and efficient resource for SCD screening and follow-up, with the potential to enhance patient management and reduce the economic burden, especially in resource-limited settings and countries with limited neonatal screening capacity. By integrating biological expertise into our tool, we aim to provide a robust solution for image interpretation and SCD profile detection, including SDM, for which diagnosis represents a major challenge in neonatal laboratories. Our key contributions are:

- Development of an SCD detection system in newborns through the analysis of neonatal agarose gel images, identifying the specific type of SCD in case of abnormal hemoglobin detection.
- Integration of the NC (Unconfirmed) category to address image clarity issues, with a predictive model guiding biologists decisions on the profile type. Reevaluation of these images with the assignment of a profile and associated probability using our second-intention model.
- Experimental demonstration of strong predictive accuracy for complex IEF patterns using our tool, surpassing the baseline condition of 67% set by biologists.

The subsequent sections are structured as follows: Section 2 situates our work in the field of SCD and image classification. Section 3 introduces essential definitions for the problem. Section 4 delves into our main contributions. Section 5 presents experimental results. Finally, Section 6 offers a discussion and conclusion of the document.

2. Related works

This section presents neonatal screening techniques for SCD, along with related works on CNN-based image classification and data augmentation techniques for medical images.

2.1. Neonatal screening techniques for SCD

SCD, also known as sickle cell anemia, is a genetic disorder characterized by an anomaly of hemoglobin linked to a genetic mutation causing hemolysis of red blood cells, resulting in anemia and other symptoms. This disease is caused by a genetic mutation involving the substitution of glutamic acid with valine at position 6 of the beta (β) globin chain (Elendu et al., 2023). While present from birth, the manifestation of the disease is less obvious in infants due to the predominance of fetal hemoglobin in their red blood cells, providing protection, especially before the age of three months. This period is crucial for early screening (Adekile, 2021). Hemoglobin electrophoresis at acidic or alkaline pH is a technique that allows for the separation of different hemoglobin fractions based on their electrical charge and molecular weight, facilitating diagnosis by highlighting the presence of hemoglobin S, associated with

other normal or abnormal hemoglobin. The principle of Rapid Diagnostic Tests (RDTs) using the Sickie SCAN™ device is a technique for detecting AS, AC, SS/Sβ⁰thal, SC, CC/Cβ⁰ that phenotypes (Nguyen-Khoa et al., 2018). It is fairly effective in adults but is currently under evaluation for neonatal screening. Capillary electrophoresis is a rapid electrokinetic separation technique with good resolution. High Performance Liquid Chromatography (HPLC) provides precise measurement of different hemoglobin fractions simultaneously and is used as a first-line screening method in certain specialized laboratories. Polymerase Chain Reaction (PCR) allows for the identification of haplotypes and is used for prenatal diagnosis (Arishi et al., 2021). Mass spectrometry MALDI-TOF is a method used to identify proteins such as hemoglobin in biological samples (Li et al., 2022). This method involves sample preparation and the use of a matrix to facilitate ionization. A laser ionizes the sample, producing ions that are then accelerated in a flight tube. By measuring their flight times, their mass can be determined and a mass spectrum generated. This spectrum is compared to a database to identify hemoglobin. However, mass spectrometry equipment, including MALDI-TOF instruments, can be expensive to purchase and maintain. In addition, issues such as sample degradation or contamination can affect the accuracy of the analysis, potentially leading to false-positive or false-negative results.

These different techniques either have technical limitations or result in very expensive costs for neonatal screening of SCD in developing countries. IEF appears as a method of choice for neonatal screening (Frömmel, 2018). The storage and transportation conditions and the significant number of samples per manipulation series in IEF are particularly important in high-prevalence contexts. Moreover, this technique provides good resolution of hemoglobin fractions (Arishi et al., 2021). However, challenges persist in making IEF more comprehensive. IEF images present reading challenges, with multiple images on a gel representing different situations. Identification and interpretation errors lead to repeated analyses, requiring more time and resources.

Our main goal is to enhance the manual approach currently utilized for SCD screening in some African countries by leveraging automation and advanced technologies. We aim to streamline and optimize the screening process, ensuring timely and accurate detection of SCD in neonates. By integrating cutting-edge AI technologies, we seek to enhance SCD screening, making it more efficient, accessible, and cost-effective. Through this initiative, we aspire to improve healthcare outcomes for newborns affected by SCD, ultimately contributing to better overall health and well-being in communities.

2.2. Image classification based on deep learning

Machine learning, particularly deep learning with CNNs, has shown significant effectiveness in disease detection using medical images (Adeniyi et al., 2024). This section reviews current research on image classification for blood diseases, focusing on SCD, and explores the potential of combining image analysis of agarose gel images with IEF using CNNs. In the context of blood diseases, a study by Loey et al. (2020), compared traditional methods with deep learning, highlighting the efficiency of CNNs

in feature extraction and leukemia classification. However, this study mainly focuses on leukemia classification, and its results may not be directly applicable to SCD diagnosis. Another approach that combines the InceptionV3 model with a wrapper-based feature selection and SVM classification is developed for sickle cell detection. Chen et al. (2023) and Alagu et al. (2022) propose a method based on holographic cytometry and deep learning for accurate biophysical profiling of sickle cell disease. Although promising, this study is based on the analysis of single red blood cell images and does not include the sickle cell trait. Jennifer et al. (2023) compares six deep learning models, including ResNet-50, MobileNet, and VGG-16, using the Erythrocytes IDB dataset for SCD classification. These aforementioned previous studies do not provide any information on patient age, and broader validation on more diverse populations is required. Several studies have been conducted using artificial intelligence for sickle cell disease and IEF. For example, Aliyu et al. (2019) proposed a two-phase method to detect red blood cell abnormalities in sickle cell anemia from blood smear images. Xu et al. (2017) proposed an automated framework using deep learning to classify red blood cell shapes in sickle cell patients, potentially aiding in disease prognosis. However, this study focused on sickle cell shapes and included only 8 patients, limiting the model generalizability. Das et al. (2020) reviewed existing automated methods for sickle cell detection, focusing on segmentation, classification, and challenges related to blood cell image analysis. They emphasized the importance of precise segmentation for reliable detection but did not present original research or propose new sickle cell detection methods. De Haan et al. (2020) proposed a deep learning framework using a smartphone microscope to automatically analyze blood smears and classify sickle cell disease, potentially providing a cost-effective screening tool for resource-limited settings. However, this study focused on blood smears and classified patients as sick or non-sick. Alzubaidi et al. (2020) used deep learning to classify red blood cells into three categories: circular (normal), elongated (sickle cell), and other blood constituents, achieving high performance. However, this study focused on blood smear images, which may not provide the same level of detail as IEF or agarose gel images for sickle cell diagnosis. Salman Khan et al. (2022) presented a method to automate the analysis of hemoglobin electrophoresis images to detect thalassemia, using a convolutional neural network named U-NetComet for automatic comet segmentation. While this study is closely related to SCD, it focuses on the presence or absence of thalassemia and does not reference subtype classification, which is important for patient management. Additionally, Tian et al. (2023) introduced an unsupervised marker less IEF method based on a CNN for effectively identifying meat species and meat cuts from 105 IEF patterns. This approach opens up possibilities for large-scale IEF analyses of complex protein samples for various applications. These advancements underscore the significance of AI in disease detection and classification. However, few studies have specifically addressed sickle cell profiles, agarose gels, and their combination with IEF.

Our goal is to introduce a method for classifying IEF using CNN models with agarose gel images, aiming to identify normal and abnormal profiles essential for accurate patient monitoring.

2.3. Image data augmentation

Data augmentation in machine learning is a crucial technique for expanding and diversifying datasets by generating new data from existing samples. It involves applying various transformations to create new, yet similar, versions of the original images. We explore current approaches and will use the one that is most suitable for our use case. Shorten and Khoshgoftaar (2019) present some commonly used transformations. This survey explores data augmentation techniques, a strategy aimed at artificially increasing the size and diversity of datasets to combat overfitting in deep learning models, especially for tasks with limited data availability. It should be noted that augmented data may not always reflect real-world scenarios, which can introduce artifacts that may mislead the model. Careful design and evaluation of augmentation strategies are crucial. Kandel et al. (2022) examine the effectiveness of brightness adjustments as a data augmentation technique for image classification using CNN and reveal limited benefits, with geometric augmentations proving more effective. The study demonstrates that brightness adjustments can, in some cases, degrade model performance compared to no augmentation or geometric augmentations. However, this study focuses on histopathological datasets and may not be generalizable to other types of images where brightness variations play a more significant role. Khachnaoui et al. (2022) propose a method for diagnosing pulmonary embolism (PE). The authors applied operations such as rotation, resizing, cropping, and zooming to the images to augment the dataset for improved performance. Geng et al. (2023) propose a study highlighting the innovative use of contrast to effectively augment synthetic radar imagery in the SAMPLE dataset, thereby improving the quality of training data. Contrast-based data augmentation methods contribute to improving the accuracy of target classification and OOD detection performance in SAR-ATR systems based on deep learning. In comparing different data augmentation approaches, Nanni et al. (2021) demonstrate the effectiveness of Gaussian blur and other approaches. This study uses pre-trained ResNet50 networks. Reza et al. (2019) propose a deep learning approach using transfer learning and data augmentation to classify crop pests from images. However, the specific data augmentation techniques used are not explicitly mentioned, limiting the reproducibility of the results. The study does not address potential variations in image quality or lighting conditions that could affect actual performance. Goceri (2023) proposes a study on data augmentation of medical images, comparing different techniques. The author examines data augmentation methods to improve the performance of medical diagnostics based on CNNs. He emphasizes the importance of carefully choosing augmentation techniques based on image types to achieve effective results. Gaussian blur is often used to reduce noise and details in an image, while brightness adjustment can simulate changes in lighting conditions. Min-max scaling helps normalize pixel values, and contrast adjustment modifies the contrast in the image. By combining these techniques, researchers can create diverse sets of images that simulate various conditions, such as low-light environments or enhanced contrast while maintaining normalized pixel values.

Our choice of augmentation techniques involves a combination of these transformations to enhance the quality and quantity of images for our machine learning tasks. Machine learning approaches, including CNN models and data augmentation,

help overcome the challenges associated with reading patient results in medicine (Yan et al., 2019).

3. Preliminaries and problem statements

In this section, we establish the foundational concepts and notations necessary for formulating our problem on IEF pattern classification for SCD profile detection. This includes defining IEF and the associated pattern language.

3.1. Basic definitions

Definition 1. (IEF). *Isoelectric Focusing is a molecular dissociation technique based on differences in isoelectric point (pI) (Righetti, 1983). The IEF pattern is analyzed to identify the types and quantities of hemoglobin in the sample, typically HbS (sickle hemoglobin, indicating the disease) for SCD and HbA (normal hemoglobin). Various types of SCD, such as SS/Sβ⁰, SC, ... can be distinguished based on the specific IEF pattern. To obtain IEF data, a blood sample from the patient is subjected to IEF using specialized equipment. This results in a set of patient and control outcomes in each column (see **Figure 1**). In a given column, we have results of patient and control images. The controls allow identifying, on a given gel, the positions of the various fractions of hemoglobin and technically validating the gel results. The resulting IEF is then analyzed by a healthcare professional to diagnose and classify the specific type of SCD. The following definition describes the type of IEF-based data obtained in a biology laboratory for neonatal screening using the “RESOLVE™ Hemoglobin kit FR-9436.”*

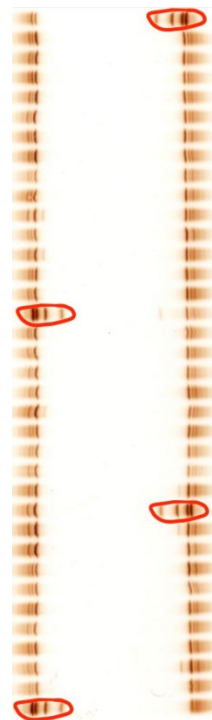


Figure 1. IEF Gel data.

Definition 2. (Gel IEF data and Gel IEF pattern). *An IEF gel data, denoted by I , is an agarose gel containing $2 \times (\alpha + \beta)$ images grouped in two columns of results patients*

and controls images in each column, with α and β respectively the number of patient results and the number of controls. A gel IEF pattern, denoted by φ , is one of the $2 \times (\alpha + \beta)$ images of an IEF gel data I , $\varphi \in I$. The set of IEF gel data forms an IEF database denoted by $G = \{I_1, \dots, I_n\}$, and the set of all gel IEF patterns is the pattern language denoted by $L(G) = \{\varphi \in I: I \in G\}$.

In our case, the biologists use gels with maximum 34 patients per column ($\alpha \in [1, \dots, 34]$) and $\beta = 2$.

Example 1. *Figure 1* presents an example of an IEF gel data with $\alpha = 34$. The β reference images (control image) that we call reference IEF patterns encircled in red characterize the absence or presence of SCD and its type, if any.

Definition 3. (Reference IEF pattern). A reference IEF pattern, denoted as φ_r , consists of k distinct features that characterize the presence or absence of SCD and, if applicable, its type. The letters in this pattern have the following meanings:

- 1) A: normal hemoglobin
- 2) F: fetal hemoglobin
- 3) S: hemoglobin S (sickle cell)
- 4) C: hemoglobin C.

Example 2. In our case, the biologists use the reference IEF pattern with $k = 4$, $\varphi_r = AFSC$, and its features distribution is shown in **Figure 2**.

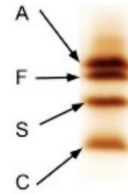


Figure 2. Reference IEF (control image).

For instance, a pattern $\varphi = AFC$ or $\varphi = FC$ means that the patient has hemoglobin C while $\varphi = AF$ indicates a normal health status.

It is interesting to note that each non-reference IEF pattern corresponds to one patient. In addition, our approach is general in the sense that α , β , and k can take larger numbers. But for simplicity and specific use cases with the Senegalese biologists, we take $\alpha = 34$, $\beta = 2$ and $k = 4$ in the rest of this paper. **Figure 3** shows the different profiles sought in this study.

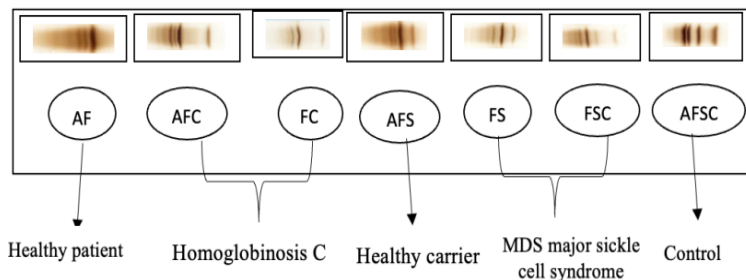


Figure 3. Various profiles.

This figure also presents the normal form, hemoglobinopathies, and our reference

image. Among hemoglobinopathies, there are heterozygous forms (composed of two different hemoglobin fractions in addition to F) and homozygous forms (composed of a single hemoglobin fraction in addition to F). However, in this study, our categories correspond to the most common hemoglobin profiles. The AF class corresponds to the AA profile. The AFS class is, therefore, associated with the AS profile. Similarly, AFC corresponds to AC, FS to SS, FC to CC, FSC to SC. These profiles can also be associated with thalassemia, which is not identified by isoelectric focusing (e.g., SB^+ and SB^0 will be in the AFS and FS classes, respectively).

3.2. Problem formulation

Deploying AI models successfully in this context presents challenges. The accuracy of AI-based classification relies on data quality. Theoretically, it is possible to generate $2^k - 1$ different IEF patterns from our reference pattern. However, our biologists focus only on specific patterns. This set would be $C_b = \{AF, AFC, FC, AFS, FS, FSC\}$ where FS and FSC represent SDM (severe forms) which is a crucial element in this study. These patterns constitute the distinct labels or classes for prediction in our classification framework.

Yet, factors like inconsistent staining, image sharpness, diverse feature arrangements within the same class, and interpretation variations in agarose gel images introduce inaccuracies, leading to incorrect results. To illustrate these ambiguities, a few images are shown in **Figure 4** and explained below.

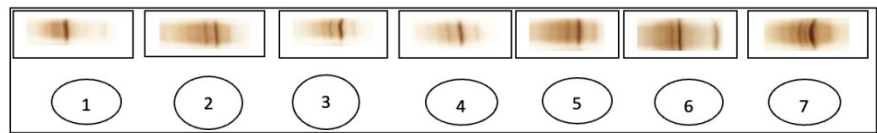


Figure 4. Examples of ambiguous IEF pattern images.

Here are some definitions and explanations of the images classified as ambiguous.

- 1) For the first image, we suspect the presence of hemoglobin C, but we cannot confirm whether hemoglobin A is also present. This result could correspond to AF, AFC, or FC.
- 2) In the second image, we can verify the presence of hemoglobin A, but hemoglobin S cannot be confirmed. If hemoglobin S is present, it is relatively close to hemoglobin F. This result could be AF, AFS, or AFD.
- 3) Images 3 to 5 show definite presence of hemoglobins F and S, but uncertainties about hemoglobin A lead to their classification as ambiguous images. This result could correspond to FS or AFS.
- 4) For the sixth image, hemoglobins F and C are clearly identified, but the presence of hemoglobin A is uncertain. This result should have been classified as either FC or AFC based on the available information.
- 5) In the last image, hemoglobin A and F are evident. However, uncertainty arises about the spot on the right. Is it another hemoglobin or a staining/fixing issue? This phenomenon typically occurs near the position of hemoglobin C or S, located towards the front of the migration. This result could correspond to AF, AFC or AFS.

To tackle this problem, our approach introduces a novel class: the “Unconfirmed” (NC) class for images not verified by biologists. This class addresses the complexity of ambiguous patterns, acknowledging classification uncertainty. Images labeled as “NC” serve as early alerts, indicating potential ambiguity. Such images undergo a retesting process facilitated by a well-designed second-intention model. Thus, we added an additional class called NC. We consider that if a pattern is classified as NC, then we retest it with a second-intention model. Therefore, the set of target classes C of our method that we seek to predict consists of the base classes C_b , the reference IEF pattern $\varphi_r = \text{AFSC}$, and the $\varphi_0 = \text{NC}$ class for recommending new experiments to biologists. **Figure 3** presents the different labels of C_b and φ_r that we need to predict.

The problem that we want to solve can be finally reformulated as follows. Given an IEF database G , we aim to build a system trained on G such that:

- 1) detect or predict the label (or class) of any IEF pattern in a new IEF data gel, where the image belongs to C_b , including SDMs, whose diagnosis at the neonatal age poses a real challenge for the laboratory;
- 2) calculate the validity of the gel data;
- 3) recommend new biological experiments for patients whose IEF patterns pose diagnostic challenges (those belonging to NC whose probability threshold may not be suitable for biologists);
- 4) develop an application capable of detecting SCD in newborns by analyzing neonatal agarose gel images using pre-registered analysis models.

In the remainder of this paper, we will provide a detailed presentation of our proposed solution to address this issue. We will outline our problem modeling, offer theoretical background, and present extensive experimental results that substantiate our theoretical approach.

4. IEF pattern classification

This section explores the creation of IEF pattern classifiers using IEF data, then demonstrates the use of their classifiers to assess the accuracy of the data and recommend potential additional biological experiments, where appropriate.

4.1. Classifier building

The approach to building the classifier for the IEF database involves two main steps: (a) creating a training database by segmenting the IEF data and assigning labels to each IEF pattern, and (b) training the model using a CNN with the training database as input.

Segmentation and labeling: The individual IEF data is divided into images of size 130×46 , resulting in a total of $2 \times (\alpha + \beta)$ images. Each image corresponds to an IEF pattern and is annotated with its class and patient identifier. The labeling process for each IEF pattern is labor-intensive but has been eased through the utilization of Excel files generated by users over time and a collaborative contribution using a WebApp (depicted in **Figure 5**) will be described in the experimental phase. These Excel files contain details like labels, patient identifiers, and the associated IEF patterns.

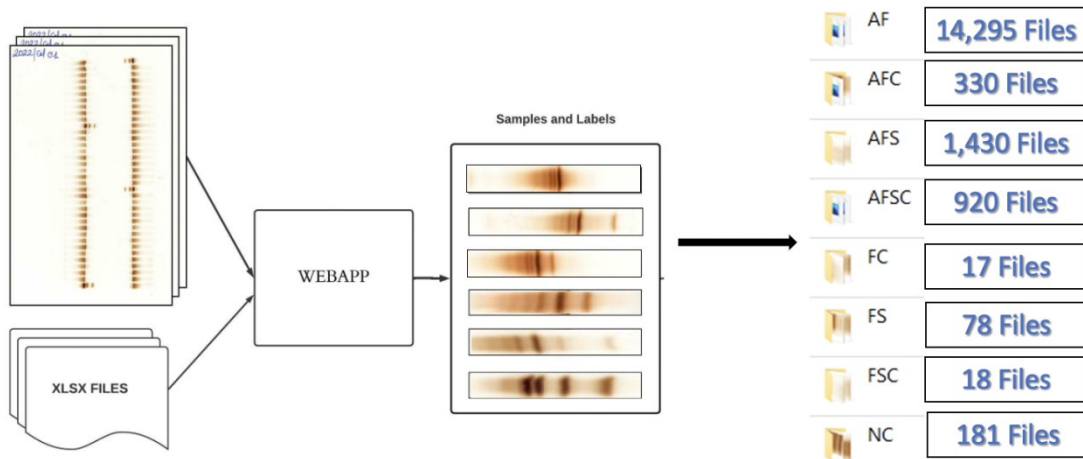


Figure 5. Structure of agarose gel image pre-processing.

Model training: After the preprocessing phase, different image classification methods could be applied to the labeled data. In this study, we chose to use CNNs, which have already proven to be efficient in image classification. The proposed model takes as input the pattern language obtained from preprocessing and learns to predict its label as output. The CNN output provides the different classes that we aim to predict, including AF, AFC, FC, AFS, FS, FSC, and AFSC.

4.1.1. Naive approach

We provide an overview of the naive approach that we first propose in this work in **Figure 6**.

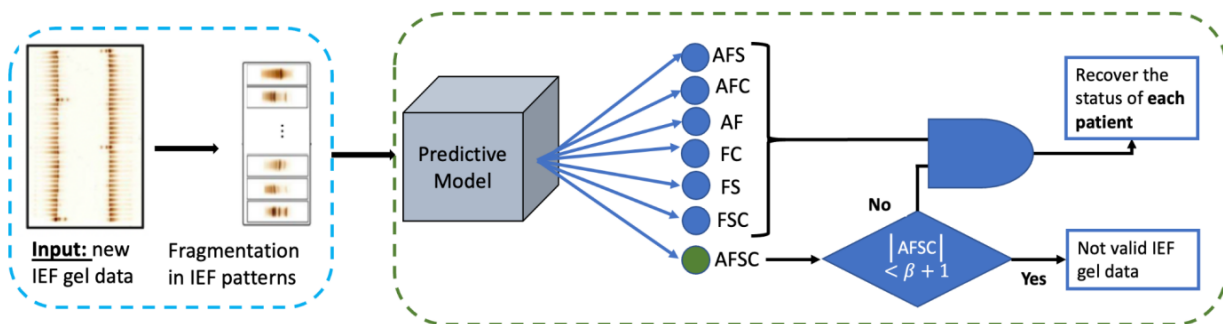


Figure 6. Naive approach.

Feedback about the naive approach: The naive approach presented in this workflow follows a typical pattern of deep learning, with input data directly feeding into the predictive model. However, this approach carries significant risks as it does not take into account human intervention, particularly from the technician, in the decision-making process, even in the presence of unconfirmed results (NC).

Indeed, not all images are of optimal or perfect quality. Some images raise doubts about their classification, making it difficult to place them in a specific class. These images are often set aside for training, limiting the effectiveness of the naive approach. Furthermore, this approach does not offer any solution for handling imperfect images.

These limits underscore the need to adopt a more robust approach, such as a two-stage model. This model could involve a first stage where a threshold is set by a human before the prediction is confirmed. Additionally, the non-optimal images will be

redirected to a class for special processing in the second stage. Human intervention will be crucial in the decision-making process for uncertain images. This process aligns more with the approach of screening laboratories, which generally have a method of first and second intentions.

4.1.2. Our proposed solution: SCSScreen

Figure 7 serves as an illustrative guide to the workflow (SCSScreen) embedded within our proposed solution, encompassing both the first and second intention models elucidated in section 4.2. Our approach stands out for its commitment to addressing the concerns of technicians, offering a tailored solution for handling imperfect images while maintaining the involvement of specialists in the decision-making process. Through the CNN output (First-intention), a range of classes are predicted, including AF, AFC, FC, AFS, FS, FSC, AFSC, and the newly introduced NC class. The incorporation of this NC class is a pivotal step, providing a mechanism to categorize IEF patterns that fall below a user-defined acceptance probability threshold for truth or are otherwise unclear. This integration ensures a holistic evaluation, preventing the oversight of ambiguous patterns during analysis. By integrating the user into the decision process, our approach fosters collaboration between technicians and specialists, facilitating a comprehensive assessment of the data. This collaborative framework empowers users to contribute their expertise, enhancing the accuracy and relevance of the decision-making process. Additionally, our approach promotes transparency and accountability by allowing users to actively participate in the interpretation and validation of results. Through this user-centric approach, we aim to bridge the gap between automated analysis and human expertise, ultimately fostering more informed and effective decision-making in clinical settings.

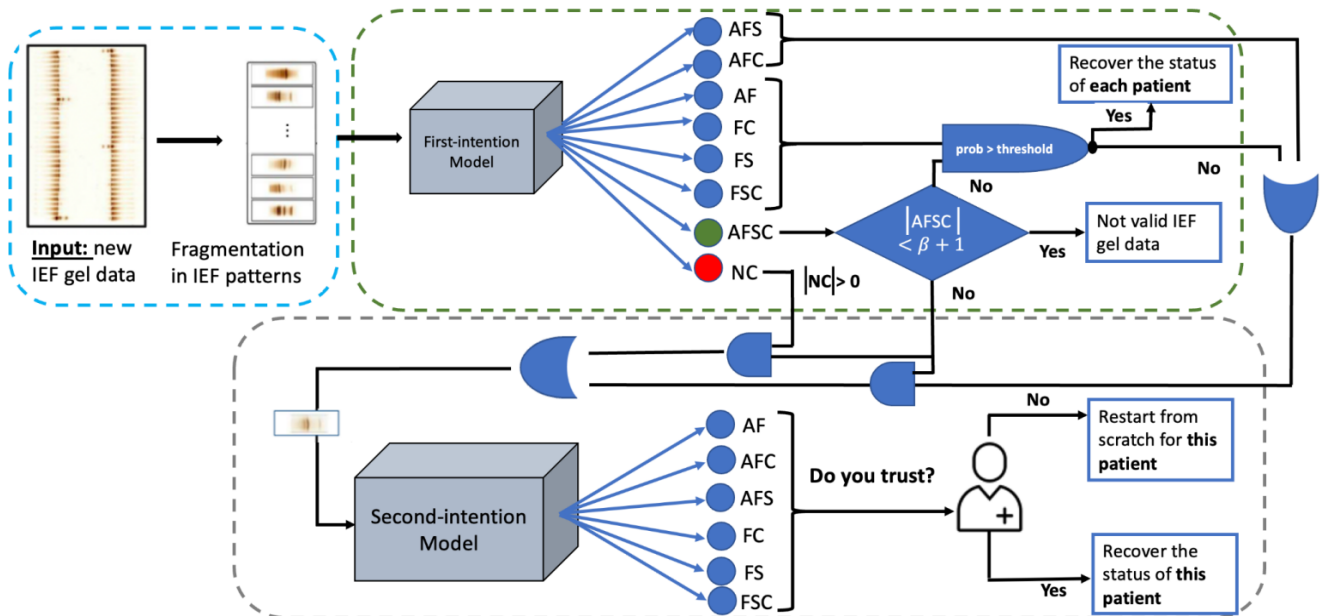


Figure 7. SCSScreen workflow.

4.2. Theoretical analysis of our models

First-intention model: In this context of interpretability, we perform a discriminant analysis on the C_b classes, the reference IEF pattern φ_r and the

unconfirmed class $\varphi_0 = \text{NC}$. We recall that C_b is the set of classes that characterize the state of health of the patients, so it is used to compute the prediction rate, while φ_r and φ_0 are used to judge the validity of the biological tests.

First, it is important to note that IEF data is not always valid, hence the use of reference IEF patterns like positive controls (samples that are known to produce specific bands on the gel based on their pI). These are important aspects of agarose gel IEF experiments, as they serve as a reference or standard for interpreting the results. Controls can help to ensure that the IEF gel has been properly prepared, the protein sample is suitable for separation, and the staining and detection methods are functioning correctly. By including appropriate controls, we can ensure that the experimental conditions are consistent and that any detected bands are not due to artifacts or other factors.

Thus, for the IEF data to be valid, we need β reference IEF patterns in each column of the IEF data. In other words, we need all $2 \times \beta$ reference IEF patterns to be classified in the class if the gel is perfect. Otherwise, we relax these constraints as technicians do. For biologists, the gel is probably correct if there is at least one clear reference in each column. Therefore, the number of IEF checks that our AI model must perform to ensure the correctness of a given gel data must be greater than or equal to $\beta + 1$. We therefore state the following property for a theoretical guarantee:

Property 1. (Validity of an IEF data). An IEF data I is valid if the number of IEF patterns φ of I classified in φ_r is greater or equal than to $(\beta + 1)$.

Proof of Property 1. This property is somewhat trivial. Let β be the number of IEF references set by the technician in each column. Suppose the number of IEF references found by the AI model, denoted $|\text{AFSC}|$, is less than $\beta + 1$: $|\text{AFSC}| \leq \beta$. In this case, our AI model cannot immediately determine if there is at least one IEF reference per column. In other words, the IEF references could all be from a single column. However, if $|\text{AFSC}| \geq \beta + 1$, then at least each column provides one IEF reference since at most a column can contain β IEF references based on the technician settings. Hence the result. \square

Then, some IEF patterns are not necessarily clear due to biological effects or technical problems. In both cases, we expect these IEF patterns to be classified in the class φ_0 . As a result, recommendations for the re-examination of the involved patients are made to the biologists. Therefore, they can resubmit new tests for these patients.

How to build the second-intention model? In the context of sickle cell disease classification, our second-intention model is trained using a dataset of IEF patterns from patients without the NC class. Then, the model uses this information to predict the SCD status of new patients whose IEF patterns are unclear or ambiguous (classified to $\varphi_0 = \text{NC}$). Furthermore, we have included the AFS and AFC profiles to reduce false negatives based on feedback from biologists and technicians. Indeed, specialists prefer to include a person suspected of being ill in the management process, awaiting confirmation of their Hb profile at a slightly older age, rather than risk excluding a person with an SDM-type illness. To evaluate the performance of this model, a validation dataset is used. The model predictions for these patients are compared to their known SCD status to determine the accuracy of the model.

Additionally, different types of machine learning models are compared to determine which one performs best in this classification task.

We retain the following formalism. Let I be the IEF database containing IEF patterns, and let C be the set of IEF pattern classes. We can define C_b as the base set of patient health classes, including SDM, and φ_r as the reference IEF pattern.

The prediction or detection of the label y for a given IEF pattern x in a new IEF data gel can be formulated as a function $f: X \rightarrow C$, where X is the space of IEF patterns. This can be represented as:

$$y = f(x, I, C_b, \varphi_r) \quad (1)$$

where y is the predicted label, x is the given IEF pattern.

In our context with CNNs, the function f can be modeled by the convolutional neural network, and the specific prediction of SDMs can be incorporated into the function. For example, if y is the class prediction, we can define an auxiliary function $g: C \rightarrow \{0,1\}$ such that $g(y) = 1$ if y corresponds to an SDM class and $g(y) = 0$ otherwise. Thus, the complete formalization can be given by:

$$y = f(x; I, C_b, \varphi_r) \times g(y) \quad (2)$$

Thus, our classification approach can both validate the input agarose gel. Moreover, it can identify the type of hemoglobinopathy, specifically if the patient has the severe form of SCD, thereby requiring follow-up. Based on a defined probability threshold, it suggests a redo of biological tests if needed.

Evaluation based on recommended activation function: Various activation functions are available for neural networks, such as sigmoid, ReLU (Rectified Linear Unit), and softmax. The choice of activation function significantly affects the model performance. In the case of the second-intention model (as well as the first-intention model) for SCD classification based on IEF patterns, the suitable activation function is softmax. This function is commonly used for multiclass classification tasks and generates a probability distribution across output classes. This makes it ideal for predicting the probability of a patient belonging to different SCD classes. The softmax function formula is:

$$\text{softmax}(z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}} \quad (3)$$

where, z represents the input to the activation function, i is the index of the output neuron (or φ_i), and K is the total number of output neurons. The softmax function exponentiates the input and normalizes it by the sum of exponentials across all output neurons.

This function ensures a probability distribution since $\text{softmax}(z)_i \in [0, 1]$ and $\sum_i \text{softmax}(z)_i = 1$. Consequently, we recommend the class φ_i with the highest probability $P(\varphi_i = \text{argmax}(\text{softmax}(z)_i) = \text{softmax}(z)_i$.

5. Experiments

This section describes the experimental protocol, the analysis of the results obtained, and the interpretation of the conclusions drawn from our experiments.

5.1. Experimental protocol for classifier building

Data collection and processing: **Figure 5** illustrates the framework used to process the agarose gel images received from the laboratory. Each agarose gel image contains seventy-two (72) images, grouped into two (2) columns, with thirty-four (34)

patient results and two (2) control results (used to validate the gel). The WebApp component of the framework was created using the algorithms showcased in **Figures 8** and **9**, which will be detailed in the following sections. Our goal was to obtain a dataset of individual patient images to solve an initial identification problem. Additionally, this segmentation allowed us to set up our learning database. Cutting multiple layers of agarose gel produced a dataset containing approximately 17,269 images, including 14,295 images of healthy patients and 1873 images of patients with hemoglobinopathies. The set also includes 920 reference images and 181 NC images. Among patients with hemoglobinopathies, we obtained 1430 SCD images and 347 hemoglobin C images, including 330 AFC patient images and 17 FC images. Furthermore, there were 96 images of major forms of SCD, including 78 FS images and 18 FSC images. Consequently, our dataset is imbalanced, which is common but unsuitable for medical imaging studies. To enhance medical image data for training machine learning models, we used data augmentation techniques such as Gaussian blur, brightness adjustment, min-max scaling, and contrast adjustment. These methods aimed to address the issue of imbalanced data and resulted in a diverse collection of augmented images. By combining these techniques, we were able to simulate various lighting and contrast conditions. After data augmentation, we retained 500 images per class for training. Therefore, this study used 4000 images, including 500 unaffected patient images (AF) and 2500 diseased patient images, as well as 1000 images of controls (AFSC) and unconfirmed cases (NC). For diseased patients, images of healthy carriers (AFS), images of hemoglobins C (AFC and FC), and images of SDM patients (FS and FSC) were used.

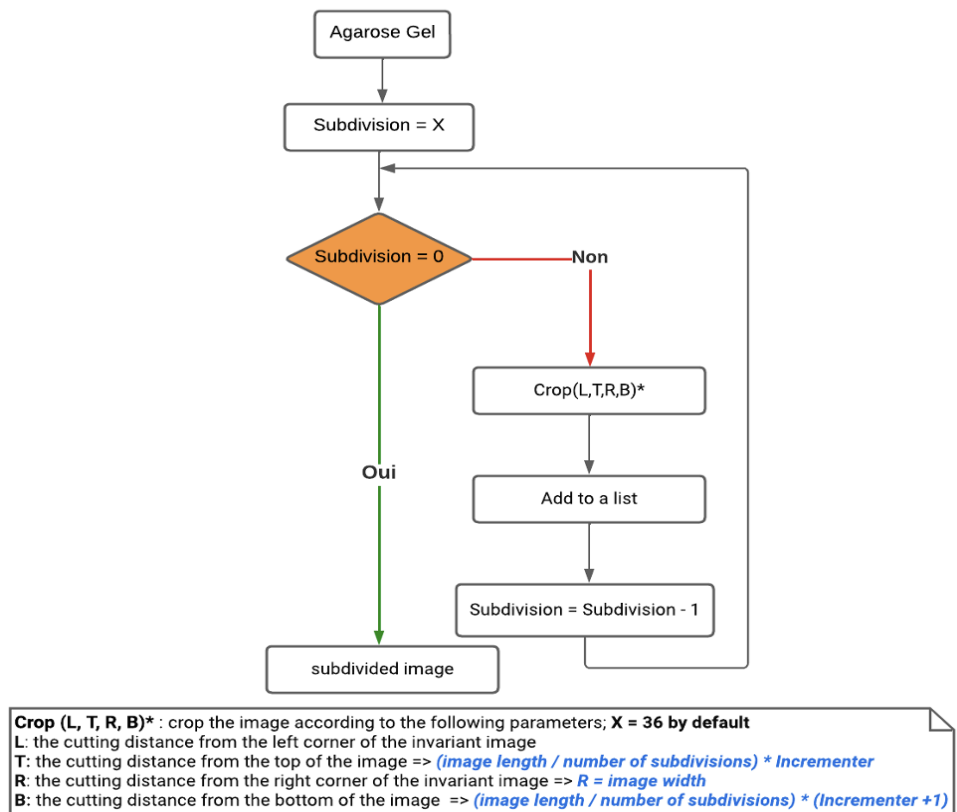


Figure 8. Segmentation of the agarose gel layer into individual images.

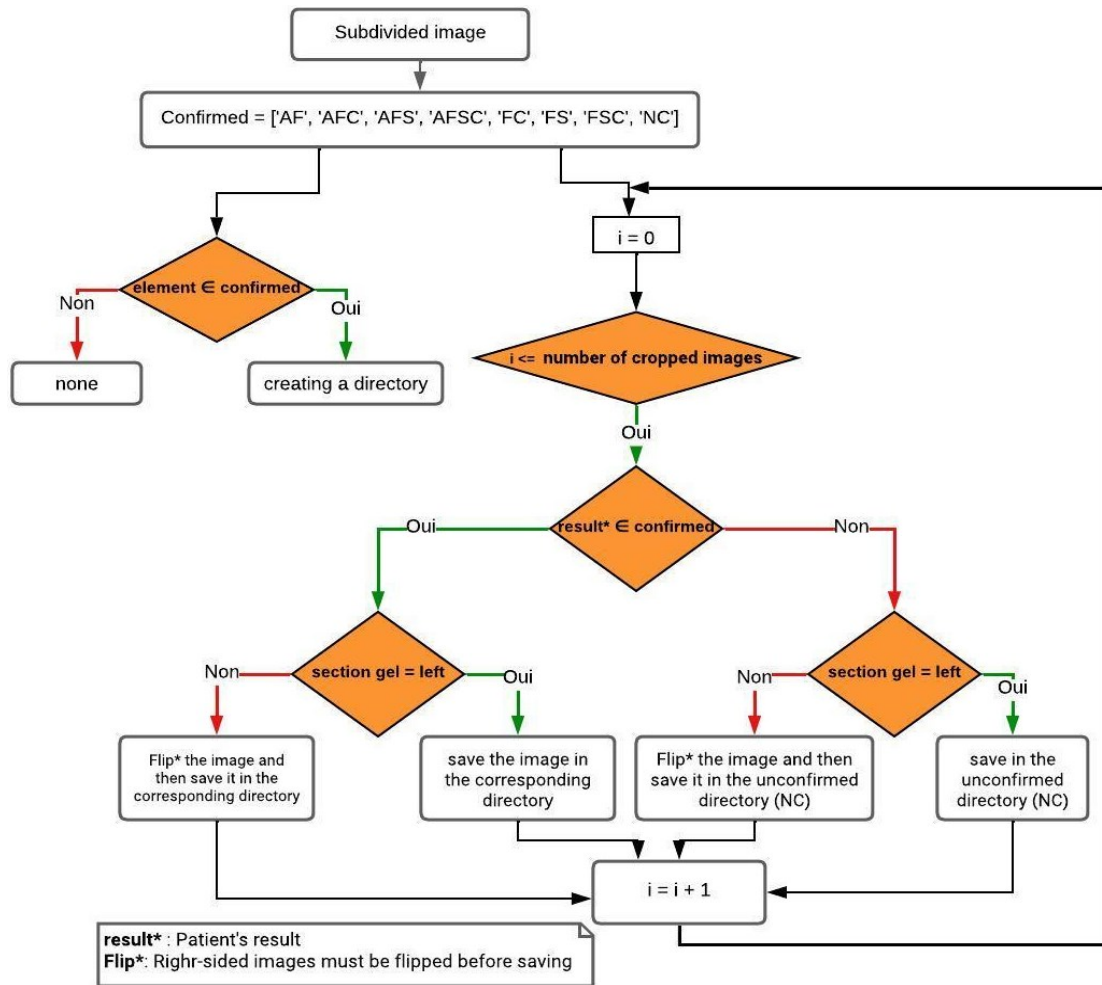


Figure 9. Distribution of individual images into different classes.

WebApp explanation: This phase aims to prepare data for training machine learning models from images to ensure accurate classification. The first algorithm (Figure 8) segments the image. The initial algorithm divides the image layer into two separate data lists based on the image left, top, right, bottom and the number of subdivisions, producing individual patient images. Additionally, it is the same used for deployment in the real-world testing phase of implementing our tool (Appendix, Figure A1). Next, a second algorithm (Figure 9) organizes these images into different directories based on the labels (confirmed), after verifying the orientation of each image. These two steps result in a dataset ready for learning. To facilitate the task for future users with upcoming gels, this process has been integrated into a framework shown in (Appendix, Figure A1) with screenshots of the web application provided in (Appendix Figures A2 and A3).

Presentation of the selected models: VGG (Visual Geometry Group) and ResNet (Residual Neural Network) are two widely used pre-trained models in computer vision and image processing due to their performance and unique characteristics. VGG is known for its simple and uniform architecture, mainly composed of small-sized convolutional layers (3 × 3), followed by pooling layers and a few fully connected layers. This modular structure makes it easy to understand and manipulate the model, while being effective at learning complex visual representations.

On the other hand, ResNet introduced the concept of residual connections, allowing information to skip layers, thereby facilitating the learning of very deep networks (over 100 layers) while avoiding gradient vanishing problems such as the gradient disappearance problem. This architecture has led to superior performance over previous models on many computer vision tasks.

Compared to other existing models, VGG and ResNet offer several advantages:

- 1) Performance: They have demonstrated superior performance on standard computer vision datasets.
- 2) Ease of use: Their modular architecture and implementation available in popular frameworks like TensorFlow and PyTorch make them easy to use and adapt to different problems.
- 3) Transfer learning: is particularly relevant to our use case, as these models have been trained on extensive datasets like ImageNet. This allows them to serve as foundational models for transfer learning, a technique where pre-trained models are adapted and fine-tuned for specific tasks using smaller datasets.

In summary, VGG and ResNet are popular choices for computer vision tasks due to their proven performance, modular architecture, and ability to learn complex visual representations.

For this study, we will be using the ResNet50 version, which is a 50-layer deep model from the ResNet family. Additionally, implementations of the VGG19 and VGG16 models will also be employed to compare their performance in the specific classification task at hand.

Model performance indicators: Several evaluation measures are considered to show the effectiveness of a CNN model. Indicators are used to measure performance. Here, TP, TN, FP, and FN are, respectively, the number of true positives, true negatives, false positives, and false negatives of classified images. We use the loss which represents the sum of false predictions divided by the sum of total observations. We also use the well known following metrics to evaluate the selected models with our approach:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}} \quad (4)$$

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (6)$$

$$\text{F1 - score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

$$\text{MCC}_{\text{multiclass}} = \frac{1}{C(C-1)} \sum_{i=1}^C \sum_{j=i+1}^C \text{MCC}_{ij} \quad (8)$$

$$\text{AUC}_{\text{multiclass}} = \frac{1}{C} \sum_{i=1}^C \text{AUC}_{ij} \quad (9)$$

$$\text{MCC}_{ij} = \frac{(A \times B) - (C \times D)}{\sqrt{(A + C)(A + D)(B + C)(B + D)}} \quad (10)$$

$$\text{AUC}_{ij} = \int_0^1 \text{TPR}(\text{FPR}_i^{-1}(t)) dt \quad (11)$$

where C , MCC_{ij} , AUC_i are respectively the number of classes, the Matthews correlation coefficient for each pair of classes i and j (with $i \neq j$) and the formula AUC for each class i . Also, $\text{TPR}(\text{FPR}_i^{-1}(t))$ is the true positive rate (sensitivity) for the class

i for a specific threshold value t , and $FPR_i^{-1}(t)$ is the inverse function of the false positive rate ($1 - \text{specificity}$) for class i for a specific threshold value t .

Finally, A is the number of true positives for classes i and j (observations of classes i and j correctly classified as i and j). B is the number of true negatives for the class i and class j (observations from classes other than i and j correctly classified as no- i and no- j). C is the number of false positives for the class i and class j (observations from classes other than i and j incorrectly classified as i and j). D is the number of false negatives for class i and class j (observations from classes i and j incorrectly classified as respectively the true negatives for class i and the true negatives for class j).

5.2. Empirical study

In this section, we present the parameters used for the training and evaluation of these different models.

5.2.1. Model parameters

This section presents the outcomes of the diverse models employed within our framework for predicting sickle cell profiles. To train and assess each model, the image dataset was divided into two portions: 80% for the training dataset and 20% for the testing dataset. For the required computing power during the training of our models, we opted for a professional Colab environment, equipped with an NVIDIA A100 GPU with 40 GB of RAM, 500 units (which we can supplement as needed by renewing payment), RAM up to 83 GB, and a 166 GB disk. We started with a grid search and then continued with manual research to determine our hyperparameters. The grid search proved challenging due to the large number of combinations for our machine. We conducted several tests to obtain the optimal values selected. After several experiments, we observed that 12 epochs were adequate to train the models to achieve good accuracies, while ResNet required 15 epochs. Among the models, ResNet had the highest number of parameters at 26,813,319 (with 53,120 non-trainable parameters), followed by VGG19 with 20,827,208 parameters and VGG16 with 15,517,512 parameters. **Table 1** provides detailed training parameter information used for classification.

Table 1. Details of training parameters used for classification.

Training Parameters	
Batch size	32
Epochs	12
Optimizer	SGD
Learning rate	0.0001
Momentum	0.9

5.2.2. Model evaluation

This section presents the results of our experiments and the different evaluations of our models.

First-intention model: In the first-intention model analysis, our focus is on newborn sickle cell profile prediction. The results presented in **Table 2** and **Figure 10** provide detailed information on the performance of the predictive models for sickle

cell profile prediction. They highlight the following findings:

- VGG19 model excels with 0.02% training and 0.99% testing loss, achieving 99% accuracy for both. Precision, recall, and F1-score also reach 99%. AUC and MCC exceed 99%.
- VGG16 model performs well, with 0.08% training and 0.15% testing loss. It achieves 97% accuracy in both phases, with 98% precision, recall, and F1-score. MCC and AUC values are 97% and 99%.
- ResNet50 model (15 epochs) yields 0.20% training loss, 97% accuracy in training and testing. Precision, recall, and F1-score are 97%, while MCC and AUC values are 96% and 99%.

Table 2. Performance measures with NC training with 12 epochs.

Model	Accuracy	Loss	Precision	Recall	F1-score	MCC	AUC
VGG19	0.9925	0.0263	0.99	0.99	0.99	0.9914	0.9999
VGG16	0.9787	0.0878	0.98	0.98	0.98	0.9758	0.9990
ResNet50 (15ep)	0.9725	0.2093	0.97	0.97	0.97	0.9685	0.9988

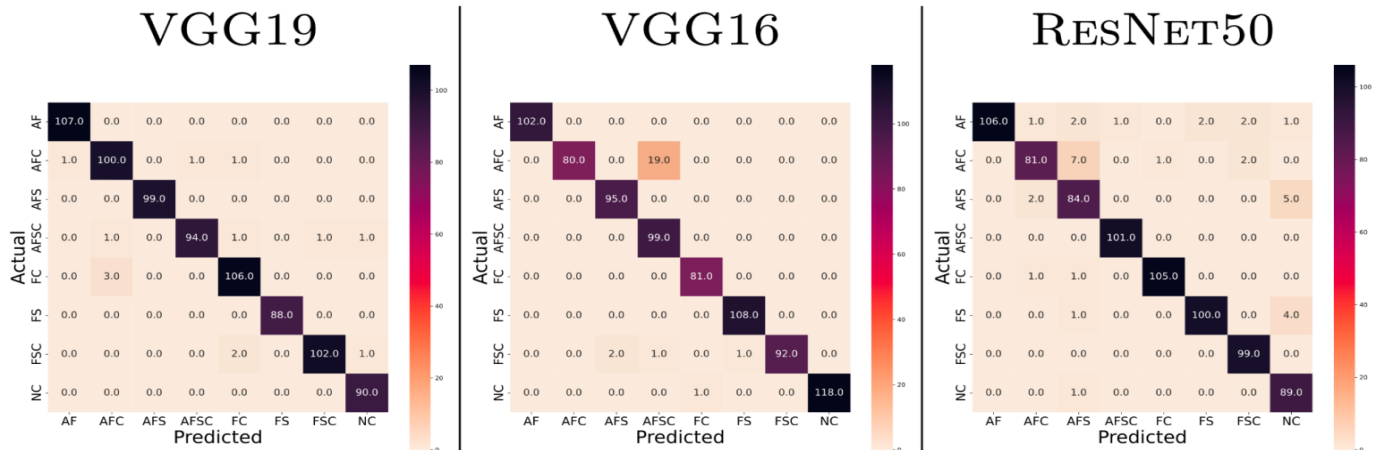


Figure 10. Shows the confusion matrices for each case of the tests of the different models.

It also shows the learning and loss evolution curves. We can say that from the beginning, we observe for the models with VGG19 and VGG16, a good evolution of the learning, which is not the case for ResNet50. In other words, we see models that stabilize rather quickly compared to ResNet50. In general, the results of the first-intention models for sickle cell profile prediction are highly promising. The VGG19 model outperforms the other models, achieving an accuracy score of 99%. This indicates its superior ability to accurately predict sickle cell profiles. Additionally, the VGG19 model demonstrates balanced performance, with high precision, recall, and F1-score values of 99%. These results suggest that the VGG19 model is a reliable and effective tool for sickle cell profile prediction.

Second-intention model: In the second-intention phase, we suggest sickle cell profiles for challenging images (NC class images) from the previous experiment. The model not only proposes profiles but also assigns a probability value, aiding clinicians in validation. As seen in **Table 3**, the second-intention model excels in suggesting profiles for these images, affirming its effectiveness.

- VGG19 model excels with 0.01% training and 0.03% testing loss. Its 99% accuracy, precision, recall, and F1-score highlight consistent and balanced performance. AUC and MCC values are over 99%, providing clinicians with reliable probability values for informed decisions.
- VGG16 model achieves 97% accuracy in both training and testing. Its precision, recall, and F1-score remain at 97%, demonstrating its effectiveness in suggesting sickle cell profiles for challenging images.
- ResNet50 model (15 epochs) exhibits impressive results, with an almost negligible 0.16% training loss. Its 98% precision, recall, and F1-score, along with AUC and MCC measures exceeding 99%, enhance its credibility in providing valuable recommendations.

Table 3. Performance measures with NC training without 12 epochs.

Model	Accuracy	Loss	Precision	Recall	F1-score	MCC	AUC
VGG19	0.9928	0.0335	0.99	0.99	0.99	0.9917	0.9998
VGG16	0.9714	0.0890	0.97	0.97	0.97	0.9666	0.9985
ResNet50 (15ep)	0.9785	0.1645	0.98	0.98	0.98	0.9750	0.9992

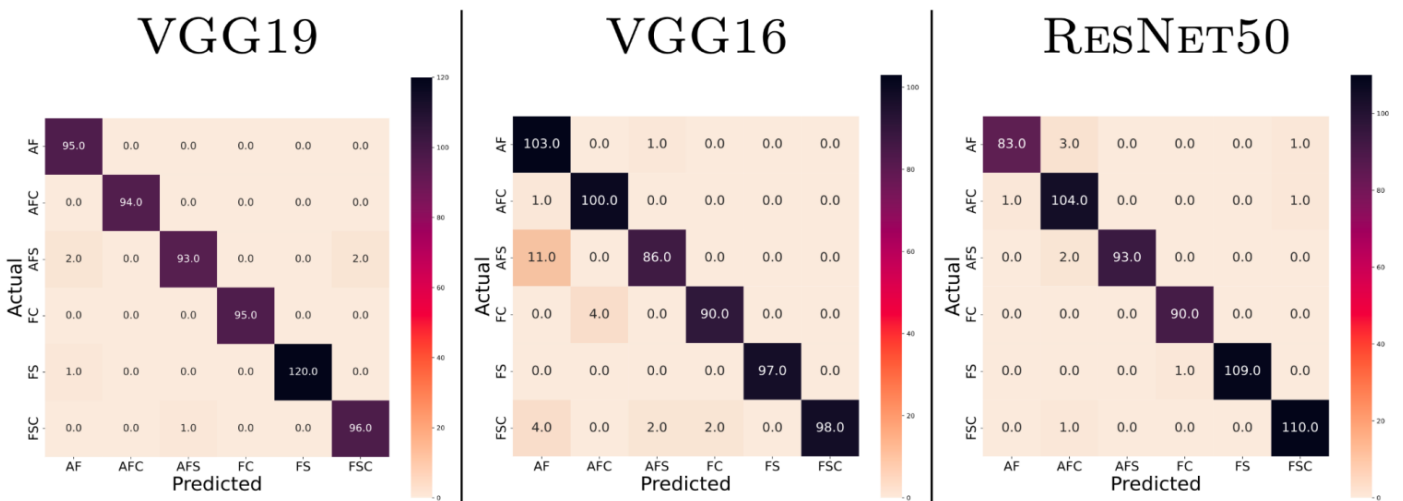


Figure 11. Shows the confusion matrices for each case of the tests of the different models.

Overall, the second-intention model, as reflected in **Table 3** and **Figure 11**, proves to be a robust and effective tool in suggesting sickle cell profiles for the hard-to-classify images. Our second-intention tool was developed based on the results obtained with the VGG19 model because of its superior performance and its ability to provide highly accurate predictions. The integration of probability values enhances the second-intention tool by offering a quantitative measure of certainty. Biologists can leverage these probabilities to validate their choices or determine if further experimentation or analysis is necessary. This feature significantly aids in reducing repetitive analyses caused by reading and profile identification challenges, allowing for a more efficient workflow in sickle cell profile assessment. The VGG19 model, with its accurate predictions and probability-based recommendations, provides biologists with a valuable resource to support their decision-making process. It helps them navigate the complexities of sickle cell profile identification and ultimately

contributes to more effective and informed healthcare practices.

5.2.3. Interpretation and discussion of the results

Our models were trained over 12 and 15 epochs using a dataset of 4000 and 3500 images, respectively, with or without NC. It is observed that the losses on the VGG19 and VGG16 models decrease steadily from the initial epochs, while the loss on the ResNet50 model exhibits a different pattern, initially evolving before eventually dropping to low values similar to the other models. The accuracy of the models shows consistent improvement with epochs, achieving over 99% accuracy in training and over 98% accuracy in testing, which is a highly satisfactory rate. The minimal difference between the accuracies of the models in training and testing indicates good generalization, except for the ResNet50 model, which requires more epochs to converge. Both VGG19 and VGG16 models demonstrate superior performance in terms of training, testing, and validation results. The slight difference between the two models can be attributed to the number of additional layers in VGG19. The analysis of confusion matrices for our models reveals the overall accuracy of predictions. The models exhibit almost correct predictions for image classes, with only a few instances in the first-intention model where images are misclassified as NC. The second intention model further enhances the accuracy of predictions and provides valuable insights for specialists to consider. The integration of the second-intention model, particularly for the NC class, enables the provision of probability values that indicate the level of trust associated with the proposed classifications. Considering these metrics and the analysis of confusion matrices, we can conclude that our classification model demonstrates satisfactory performance. The VGG19 model, in particular, holds promise for predicting IEF gel image results and identifying different types of sickle cell patients. The development of the second-intention model reinforces the usefulness of our approach by providing an answer for NC class images. It offers valuable support to specialists by suggesting potential sickle cell profiles and providing probabilities that indicate the level of trust associated with each recommendation. This additional layer of information assists clinicians in making informed decisions and reduces the need for repetitive analyses caused by challenges in reading and identifying sickle cell profiles.

5.3. Biologists and technicians feedback on SCScreen

After the deployment, the tool has been used with new gels data. **Table 4** presents feedback from specialists on the tool usage. The different percentages represent the correct predictions for each class for the three recently processed gels, totaling 216 results including 204 patients and 12 controls. $2 \times \beta$ represents the controls used for the gels. Our framework outperforms the naive approach with 1.85% fewer errors compared to over 8%. We recall that this was possible because a little over 15.27% of the results (33 results) went through the second-intention model with the NC class. It should also be noted that our framework corrected two patient results identified by the biologist after comparing the results. Indeed, these two patients had been declared AF while our system predicted them as AFS.

Reading the table: Example of Gel 1 of our framework, the 72 IEF patterns including 4 references are distributed as explained below. The tool predictions are

outside parentheses. Thus, the tool predicted 3 IEF references correctly, which is greater than or equal to $(\beta + 1)$ since $\beta = 2$, so Gel 1 remains valid. It appears that for the AF class, there are 55 correct predictions out of 56 and 03 correct predictions out of 03 for the AFC class. Respectively, 08 out of 08 and 01 out of 01 for the AFS and FS classes that we denote. 02 false predictions were identified, resulting in a success rate of 97.22%. 07 results initially predicted as NC with the first-intention model were eventually identified by the second-intention model. The reading of the lines is done similarly. For the AF line of our framework, we have 55 correct predictions out of 56 for Gel 1, 65 correct predictions out of 66 for Gel 2, 62 out of 63 for Gel 3, and a success rate of 98.38% for this class. Thus, respectively, 52 out of 56, 60 out of 66, 59 out of 63 for Gel 1, Gel 2, and Gel 3 for the naive approach. We note a success rate of 92.43%. As for the NC class line, it only concerns our framework, which is one of the contributions of this paper.

Table 4. Tool usage feedback.

	Our framework				Naïve approach			
	Gel 1	Gel 2	Gel 3	%	Gel 1	Gel 2	Gel 3	%
$2 \times \beta$ (technician)	2×2				2×2			
AFSC	3(4)	4(4)	4(4)	75%	4(4)	4(4)	4(4)	100%
AF	55 (56)	65 (66)	62 (63)	98.38%	52 (56)	60 (66)	59 (63)	92.43%
AFC	03 (03)			100%	03 (03)			100%
AFS	08 (08)	02 (02)	05 (05)	100%	08 (08)	02 (02)	03(05)	86.66%
FC								
FS	01 (01)			100%	01 (01)			100%
FSC								
NC	07	13	13	15.27%				
%	97.22%	98.61%	98.61%	98.14%	94.44%	91.66%	91.66%	92.59%
Number of false predictions	02	01	01	01.85%	04	06	06	08.33%

6. Conclusion

This study focuses on addressing the significant challenge of SCD in West Africa, particularly the lack of reliable screening platforms. It highlights the effectiveness of IEF and AI, especially Deep Learning, in precise disease diagnosis. Through a classification study using agarose gel images and three distinct CNN architectures, the research achieved outstanding results, surpassing a 98% accuracy rate in predicting the disease and identifying sickle cell profiles. The integration of a “Non-Confirmed” (NC) category further enhances comprehensiveness and fairness in the approach. Additionally, introducing probability values provides valuable insights for clinicians to make informed decisions. This fairness aspect ensures that the recommendation model (Second-intention Model) not only suggests potential sickle cell profiles but also conveys the level of trust associated with each prediction. The proposed approach has the potential to enhance screening platforms, reduce costs associated with repetitive analyses, and ultimately contribute to reducing newborn mortality rates.

Despite these achievements, limitations exist, notably the reliance of the second-

intention model solely on agarose gel image information, neglecting additional clinical data or patient-specific factors. Healthcare professionals should incorporate these factors into their decision-making process alongside the model's recommendations to ensure well-informed clinical decisions. Furthermore, this holistic approach can lead to more personalized and effective patient care.

To conclude, we further explore the interpretability and explainability of the AI models used for sickle cell disease screening. It is essential to enable healthcare professionals to clearly understand the model predictions, which will enhance their trust in these results. Following the implementation of continuous learning and adaptation mechanisms for sickle cell disease screening models to ensure their long-term effectiveness and relevance by integrating new data and considering disease developments. The integration of multimodal data into sickle cell disease screening models could significantly improve their accuracy and robustness by combining information from different sources for a more comprehensive and precise evaluation of the disease.

Author contributions: Conceptualization, KAJK, CTD, LD and MD; methodology, KAJK, CTD, LD and MD; software, KAJK; validation, KAJK, CTD, LD and MD; formal analysis, KAJK, CTD, LD and MD; investigation, KAJK, CTD, LD and MD; resources, KAJK and MD; data curation, KAJK and MD; writing—original draft preparation, KAJK, CTD, LD and MD; writing—review and editing, KAJK, CTD, LD and MD; visualization, KAJK and LD; supervision, CTD and LD; project administration, CTD; funding acquisition, KAJK, CTD, LD and MD. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This research was supported by the Partnership for skills in Applied Sciences, Engineering and Technology (PASET)—Regional Scholarship and Innovation Fund (RSIF). We would like to thank the Center for Research and Ambulatory Care of the Sickle Cell Disease (CERPAD) for providing us with the data and explanation necessary for the realization of our study. We express our gratitude to Ibrahima Diagne, Coordinator, Dominique Doupa, Head of the laboratory, Demba Makalou, Hematologist, and the entire CERPAD team for their invaluable guidance, support, and expertise in the field of neonatal sickle cell screening. Our thanks go to the CEA-MITIC (African Center of Excellence in Mathematics, Computer Science, and ICT) of UGB for their guidance and support.

Conflict of interest: The authors declare no conflict of interest.

References

- Adekile, A. (2020). The Genetic and Clinical Significance of Fetal Hemoglobin Expression in Sickle Cell Disease. *Medical Principles and Practice*, 30(3), 201–211. <https://doi.org/10.1159/000511342>
- Adeniyi, A. E., Ayoola, J. B., Farhaoui, Y., et al. (2024). Comparative Study for Predicting Melanoma Skin Cancer Using Linear Discriminant Analysis (LDA) and Classification Algorithms. In: Farhaoui, A. Y., Hussain, T., Saba, H., et al. (editors). *Artificial Intelligence, Data Science and Applications*. Springer Nature Switzerland. pp. 326–338. https://doi.org/10.1007/978-3-031-48465-0_42
- Alagu, S., Ganesan, K., & K., B. B. (2022). A novel deep learning approach for sickle cell anemia detection in human RBCs using an improved wrapper-based feature selection technique in microscopic blood smear images. *Biomedical Engineering*,

- Biomedizinische Technik, 68(2), 175–185. <https://doi.org/10.1515/bmt-2021-0127>
- Aliyu, H. A., Razak, M. A. A., & Sudirman, R. (2019). Segmentation and detection of sickle cell red blood image. AIP Conference Proceedings. <https://doi.org/10.1063/1.5133919>
- Alzubaidi, L., Fadhel, M. A., Al-Shamma, O., et al. (2020). Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis. *Electronics*, 9(3), 427. <https://doi.org/10.3390/electronics9030427>
- Arishi, W. A., Alhadrami, H. A., & Zourob, M. (2021). Techniques for the Detection of Sickle Cell Disease: A Review. *Micromachines*, 12(5), 519. <https://doi.org/10.3390/mi12050519>
- Bäuerle, A., Cabrera, Á. A., Hohman, F., et al. (2022). Symphony: Composing Interactive Interfaces for Machine Learning. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3491102.3502102>
- Chen, C. X., Funkenbusch, G. T., & Wax, A. (2023). Biophysical Profiling of Sickle Cell Disease Using Holographic Cytometry and Deep Learning. *International Journal of Molecular Sciences*, 24(15), 11885. <https://doi.org/10.3390/ijms241511885>
- Daniel, Y., Elion, J., Allaf, B., et al. (2019). Newborn Screening for Sickle Cell Disease in Europe. *International Journal of Neonatal Screening*, 5(1), 15. <https://doi.org/10.3390/ijns5010015>
- Das, P. K., Meher, S., Panda, R., et al. (2020). A Review of Automated Methods for the Detection of Sickle Cell Disease. *IEEE Reviews in Biomedical Engineering*, 13, 309–324. <https://doi.org/10.1109/rbme.2019.2917780>
- De Haan, K., Ceylan Koydemir, H., Rivenson, Y., et al. (2020). Automated screening of sickle cells using a smartphone-based microscope and deep learning. *Npj Digital Medicine*, 3(1), 76. <https://doi.org/10.1038/s41746-020-0282-y>
- Elendu, C., Amaechi, D. C., Alakwe-Ojimba, C. E., et al. (2023). Understanding Sickle cell disease: Causes, symptoms, and treatment options. *Medicine*, 102(38), e35237. <https://doi.org/10.1097/md.00000000000035237>
- El-Haj, N., & Hoppe, C. C. (2018). Newborn Screening for SCD in the USA and Canada. *International Journal of Neonatal Screening*, 4(4), 36. <https://doi.org/10.3390/ijns4040036>
- Frömmel, C. (2018). Newborn Screening for Sickle Cell Disease and Other Hemoglobinopathies: A Short Review on Classical Laboratory Methods—Isoelectric Focusing, HPLC, and Capillary Electrophoresis. *International Journal of Neonatal Screening*, 4(4), 39. <https://doi.org/10.3390/ijns4040039>
- Geng, Z., Xu, Y., Wang, B. N., et al. (2023). Target Recognition in SAR Images by Deep Learning with Training Data Augmentation. *Sensors*, 23(2), 941. <https://doi.org/10.3390/s23020941>
- Goceri, E. (2023). Medical image data augmentation: techniques, comparisons and interpretations. *Artificial Intelligence Review*, 56(11), 12561–12605. <https://doi.org/10.1007/s10462-023-10453-z>
- Gueye, B., Tacko Diop, C., Marième Diagne, N., et al. (2020). Study of Prognostic Factors of Death in Children with Sickle Cell Diseases Followed at the Albert Royer National Children’s Hospital Center, Dakar, Senegal. *American Journal of Pediatrics*, 6(1), 1. <https://doi.org/10.11648/j.ajp.20200601.11>
- Jennifer, S. S., Shamim, M. H., Reza, A. W., et al. (2023). Sickle cell disease classification using deep learning. *Heliyon*, 9(11), e22203. <https://doi.org/10.1016/j.heliyon.2023.e22203>
- Kandel, I., Castelli, M., & Manzoni, L. (2022). Brightness as an Augmentation Technique for Image Classification. *Emerging Science Journal*, 6(4), 881–892. <https://doi.org/10.28991/esj-2022-06-04-015>
- Khachnaoui, H., Agrebi, M., Halouani, S., et al. (2022). Deep Learning for Automatic Pulmonary Embolism Identification Using CTA Images. In: Proceedings of the 2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). <https://doi.org/10.1109/atsip55956.2022.9805929>
- Li, D., Yi, J., Han, G., et al. (2022). MALDI-TOF Mass Spectrometry in Clinical Analysis and Research. *ACS Measurement Science Au*, 2(5), 385–404. <https://doi.org/10.1021/acsmeasuresciau.2c00019>
- Loey, M., Naman, M. R., & Zayed, H. H. (2020). A Survey on Blood Image Diseases Detection Using Deep Learning. *International Journal of Service Science, Management, Engineering, and Technology*, 11(3), 18–32. <https://doi.org/10.4018/ijssmet.2020070102>
- Nanni, L., Paci, M., Brahmam, S., et al. (2021). Comparison of Different Image Data Augmentation Approaches. *Journal of Imaging*, 7(12), 254. <https://doi.org/10.3390/jimaging7120254>
- Nguyen-Khoa, T., Mine, L., Allaf, B., et al. (2018). Sickle SCAN™ (BioMedomics) fulfills analytical conditions for neonatal screening of sickle cell disease. *Annales de Biologie Clinique*, 76(4), 416–420. <https://doi.org/10.1684/abc.2018.1354>
- Reza, M. T., Mehedi, N., Tasneem, N. A., et al. (2019). Identification of Crop Consuming Insect Pest from Visual Imagery Using Transfer Learning and Data Augmentation on Deep Neural Network. In: Proceedings of the 2019 22nd International

- Conference on Computer and Information Technology (ICCIT). <https://doi.org/10.1109/iccit48885.2019.9038450>
- Righetti, P. G. (1983). *Isoelectric focusing: Theory, methodology, and applications*. Elsevier Science.
- Salman Khan, M., Ullah, A., Khan, K. N., et al. (2022). Deep Learning Assisted Automated Assessment of Thalassaemia from Haemoglobin Electrophoresis Images. *Diagnostics*, 12(10), 2405. <https://doi.org/10.3390/diagnostics12102405>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0197-0>
- Tian, Y., Cao, Y., Zha, G., et al. (2023). Marker-Free Isoelectric Focusing Patterns for Identification of Meat Samples via Deep Learning. *Analytical Chemistry*, 95(37), 13941–13948. <https://doi.org/10.1021/acs.analchem.3c02461>
- United Nations. (2009). Resolution adopted by the General Assembly on 22 December 2008. Available online: https://digitallibrary.un.org/record/644334/files/A_RES_63_237-EN.pdf?version=1 (accessed on 21 January 2024).
- World Health Organization. (2006). Fifty-ninth World Health Assembly, Geneva, 22–27 May 2006: Reports of Committees. Available online: https://iris.who.int/bitstream/handle/10665/21483/WHA59_REC3-en.pdf?sequence=1&isAllowed=y (accessed on 21 January 2024).
- Xu, M., Papageorgiou, D. P., Abidi, S. Z., et al. (2017). A deep convolutional neural network for classification of red blood cells in sickle cell anemia. *PLOS Computational Biology*, 13(10), e1005746. <https://doi.org/10.1371/journal.pcbi.1005746>
- Yan, R., Ren, F., Wang, Z., et al. (2020). Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*, 173, 52–60. <https://doi.org/10.1016/j.ymeth.2019.06.014>

Appendix

This section provides some insights into the reproducibility of the work. First, the general framework will be detailed. Then, some screens of the analysis will be presented to you.

A. Reproducibility

This section presents a tool (**Figure A1**) that implements two pipelines to address the stated problem. The first pipeline examines the implementation of IEF motif classifiers from IEF data and then demonstrates their use to evaluate data accuracy and recommend additional biological experiments if necessary. The second pipeline presents the approach to implementing the analysis tool. More precisely, this framework presents a decision support tool for neonatal sickle cell disease screening based on agarose gel images. Our approach can be broken down into four (04) steps by pipelines (**Figure A1**). Loading the agarose gel image and its reference sheet. This step is followed by a subdivision to obtain individual results for identified patients. The first two steps are almost identical. The second step of the model building pipeline already had the classes which helped in setting up the classes in our supervised learning. In the case of SCScreen, all patient images remain in a single directory which will be used for predictions. Steps 3 and 4 of the model building pipeline perform the learning and save our prediction and recommendation models for IEF patterns. Steps 3 and 4 of the SCScreen pipeline import the pre-trained models, perform predictions followed by displaying the sickle cell or non-sickle cell result associated with its prediction probability. In these pipelines, all packages used at each stage are given.

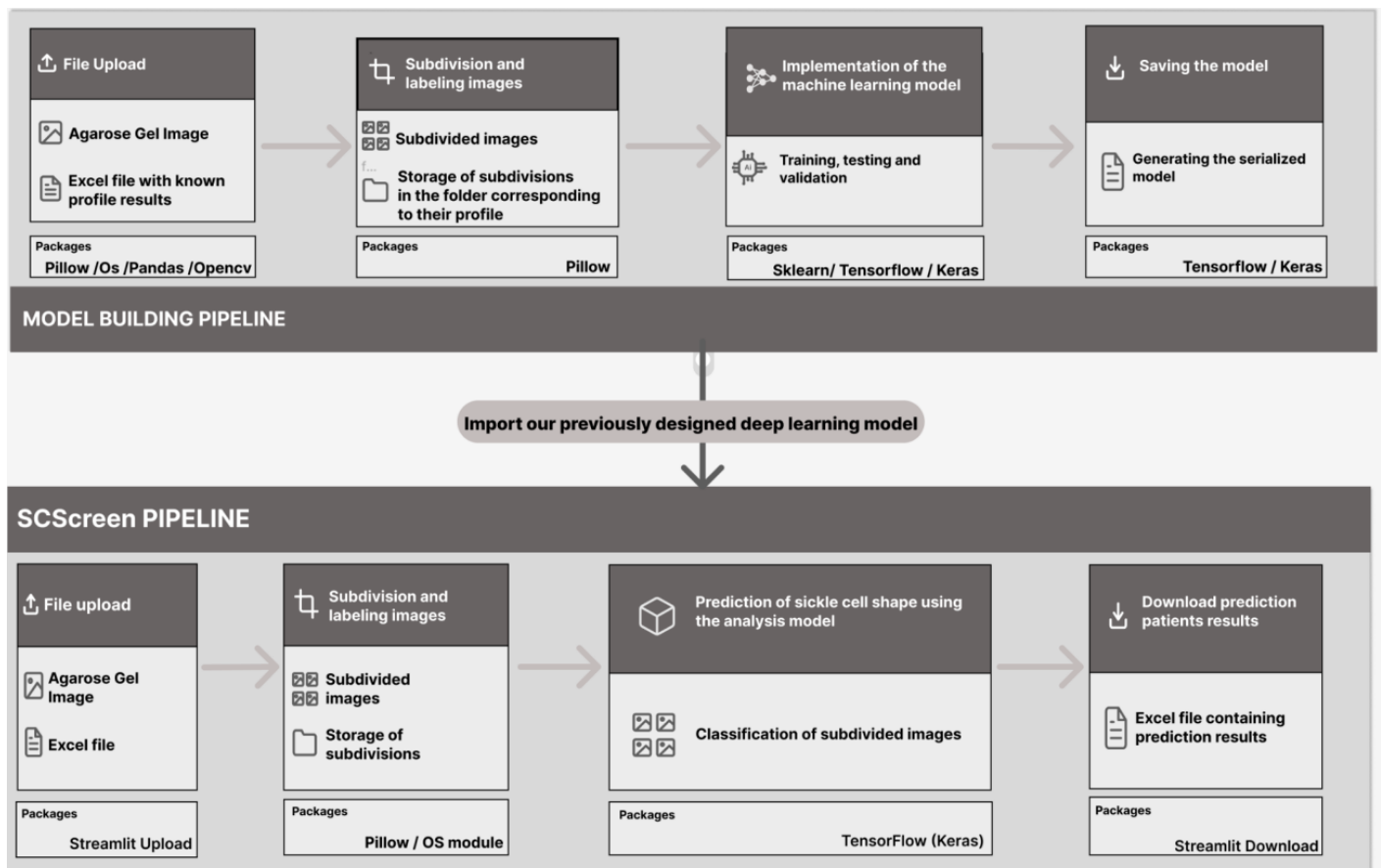


Figure A1. Approach to building the models and the analysis tool.

B. Parameter setting and visualization screen for agarose gel electrophoresis slice

Parameter setting: Here are some screenshots of the deployment and their descriptions. Developed with a focus on applicability and generalizability, our tool is called SCScreen.

Figure A2 shows the screen for cropping agarose gel images into patient results and saving them in a directory.

The screen takes as input the agarose gel image and its bench sheet (sheet with sample identifiers and positions). It then divides the image into several equal parts according to the specified number of subdivisions and returns a list containing these subdivided images. This subdivision is done using the following parameters based on the left (here 180), top (here 0), right (here 260), bottom (688) coordinates and the number of subdivisions, allowing to crop the specified column (here right) correctly and producing individual patient images. A visual check is available on the right to verify the correctness of the cropping. The resolution is also acceptable. Afterward, you need to save using the “save” button to keep the cropping in a directory that will be used for prediction. This implementation is based on **Figure 8** and the first two stages of the pipelines.

Upload Data

The principle of the webcropping interface is to upload a single agarose gel image file for individual patient results. Click on "[How to use ?](#)" to read the instructions.

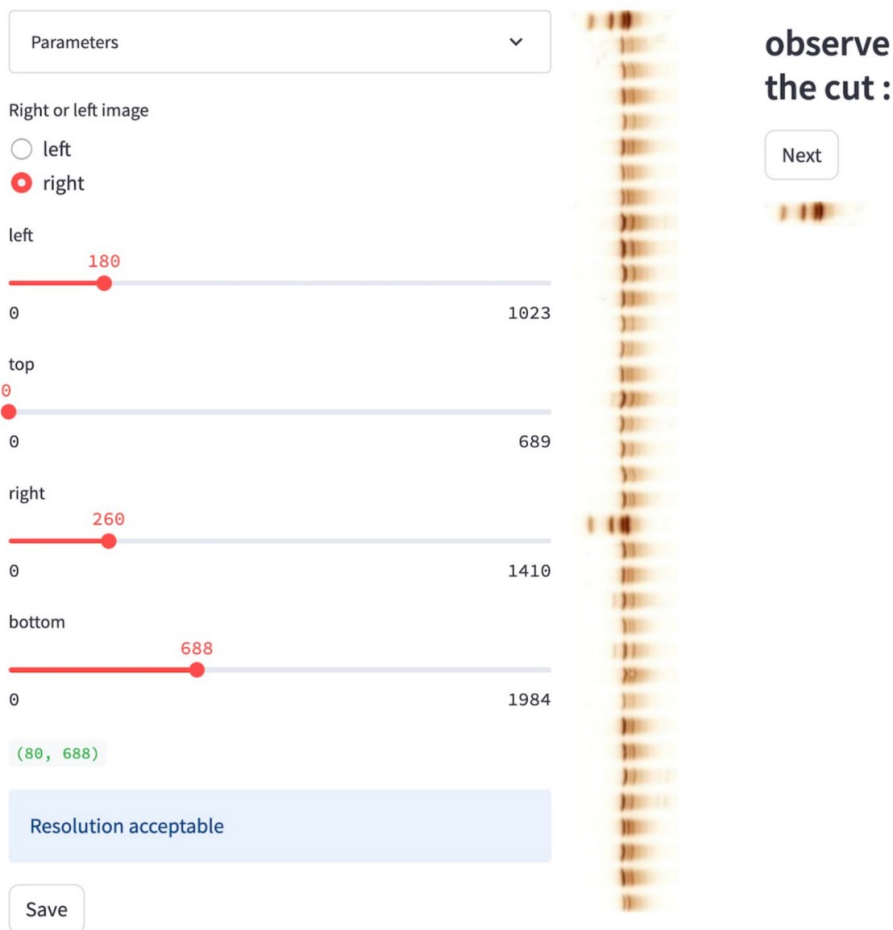


Figure A2. Cutting agarose gel.

C. Visualization screen


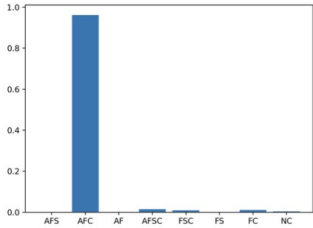

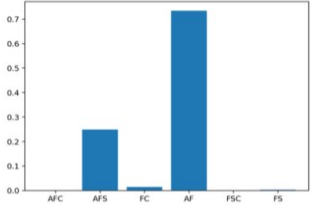

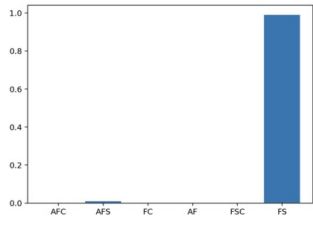
Figure A3 shows the screen for predicting sickle cell profiles. This part of the code takes as input the subdivided images and returns their predictions for the patients, followed by the probability of recommendation to the specialist. Among the profiles in **Figure A3**, the first and last images are predicted by the first-intention model characterized by the set of profiles. Images 3 and 4, whose final predictions were made by the second-intention model after passing through the NC class of the first-intention model. Indeed, the second-intention model does not have the AFSC and NC classes in its prediction set on the histograms. The deployment of the SCScreen tool automates sickle cell screening for

resource-limited environments, integrating agarose gel image analysis and CNNs. The tool reduces reliance on human resources and improves result reproducibility. SCScreen effectively identifies sickle cell profiles, overcomes staining variations, and reduces retesting costs, providing significant benefits to developing countries. It also raises awareness and promotes the adoption of automated technologies for neonatal sickle cell screening, improving patient care.

Predict profiles results

? The predictor interface is used after cutting the agarose gel. In this way, patients results already stored in memory can be analyzed and the prediction restituted. Press the "Predict" button. After prediction, press "remove" to delete the images in memory.

Predict

Image	Predict	Plot the class probability
 IEF2024_001_44	AFC	
 IEF2024_001_19	AF	
 IEF2024_001_25	FS	

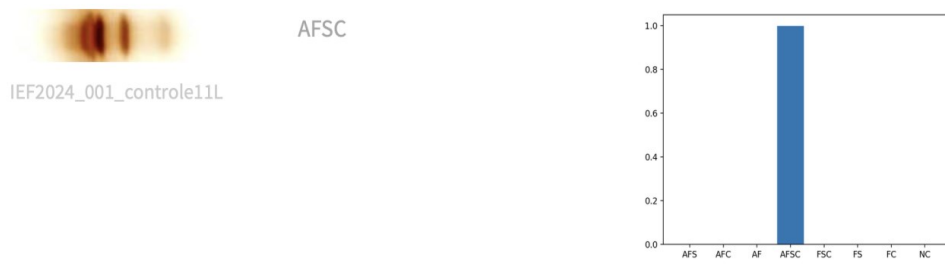


Figure A3. Prediction of profiles.