BRNO UNIVERSITY OF TECHNOLOGY

Faculty of Electrical Engineering and Communication

BACHELOR'S THESIS

Brno, 2023   Anna Mojžišová

BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

MOBILE APPLICATION FOR ELECTROPHORETIC GEL IMAGE PROCESSING

MOBILNÍ APLIKACE PRO ZPRACOVÁNÍ OBRAZU ELEKTROFORETICKÉHO GELU

BACHELOR'S THESIS
BAKALÁŘSKÁ PRÁCE

AUTHOR   Anna Mojžišová
AUTOR PRÁCE

SUPERVISOR   Ing. Helena Škutková, Ph.D.
VEDOUCÍ PRÁCE

BRNO 2023

Bachelor's Thesis

Bachelor's study program **Biomedical Technology and Bioinformatics**

Department of Biomedical Engineering

Student: Anna Mojžišová   **ID:** 223378

Year of study: 3   **Academic year:** 2022/23

TITLE OF THESIS:

Mobile application for electrophoretic gel image processing

INSTRUCTION:

1) Learn the basic methods of electrophoretic separation of biological macromolecules using 1D gel electrophoresis. 2) Conduct a literature review of electrophoretic methods and focus on digitalization of the gel electrophoresis result and methods of image analysis of the gel image. 3) Design a mobile phone application to process the gel image and implement the sub-steps: acquisition of the gel image, simple segmentation of the sample lines and conversion to 1D signal representations of the samples. 4) Create a mobile application that allows, in addition to the previous steps, to retrieve images from external sources, evaluate the molecular weights of separated molecules in the gel, create templates for mass markers, easily adjust the brightness and contrast of the image, save, export and share the results. 5) Test the program on real images of gel electrophoresis results. 6) Discuss the results obtained.

RECOMMENDED LITERATURE:

ŠKUTKOVÁ, H.; VÍTEK, M.; KŘÍŽKOVÁ, S.; KIZEK, R.; PROVAZNÍK, I. Preprocessing and Classification of Electrophoresis Gel Images Using Dynamic Time Warping. INTERNATIONAL JOURNAL OF ELECTROCHEMICAL SCIENCE, 2013, roč. 2013(8), č. 2, s. 1609-1620. ISSN: 1452- 3981.

SIM, J.-Z.; P.-V. NGUYEN; Y. ZANG; et al. DNA2App: Mobile sequence analyser. Scientific Phone Apps and Mobile Devices. 2016, 2(1), 2. ISSN 2364-4958.

Date of project specification: 6.2.2023   **Deadline for submission:** 29.5.2023

Supervisor: Ing. Helena Škutková, Ph.D.

doc. Ing. Jana Kolářová, Ph.D.
Chair of study program board

Abstrakt

Analýza obrazů 1D gelové elektroforézy je klíčovým krokem ve výzkumu molekulární biologie a umožňuje identifikaci a kvantifikaci proteinů, fragmentů DNA a dalších molekul. Tradiční proces analýzy však může být časově náročný a vyžaduje specializovaný software a vybavení. Tato bakalářská práce představuje nové řešení tohoto problému zavedením mobilní aplikace, která pomáhá při analýze obrazů 1D gelové elektroforézy. Aplikace poskytuje uživatelsky přívětivé rozhraní a sadu výkonných nástrojů pro zpracování obrazu, segmentaci drah, detekci proužků a aproximaci molekulové hmotnosti. Navrhované metody byly podrobně testovány na sadě různých gelových snímků a výsledky ukazují, že aplikace dokáže přesně a efektivně analyzovat elektroforetické snímky. Přenosnost aplikace umožňuje výzkumným pracovníkům provádět analýzy na cestách, čímž se snižuje čas a náklady spojené s tradičními metodami analýzy. Tato mobilní aplikace představuje pohodlný a praktický nástroj pro výzkumné pracovníky i lékaře, díky němuž je analýza elektroforetických obrazů dostupnější a efektivnější.

Klíčová Slova

1D gelová elektroforéza, analýza obrazu, kvantitativní analýza, mobilní aplikace

Abstract

The analysis of 1D gel electrophoresis images is a crucial step in molecular biology research, as it allows the identification and quantification of proteins, DNA fragments, and other molecules of interest. However, the traditional analysis process can be time-consuming and requires specialized software and equipment. This paper presents a novel solution to this problem by introducing a mobile application that assists in the analysis of 1D gel electrophoresis images. The application provides a user-friendly interface and a set of powerful tools for image processing, band detection, lane segmentation, and molecular weight approximation. The designed methods were extensively tested using a dataset of diverse gel images, and the results show that the application can accurately and efficiently analyze electrophoresis images. The portability of the application allows researchers to perform analyses on the go, reducing the time and cost associated with traditional analysis methods. This mobile application provides a convenient and practical tool for researchers and clinicians alike, making the analysis of electrophoresis images more accessible and efficient.

Keywords

1D gel electrophoresis, image analysis, quantitative analysis, mobile application

Rozšířený Abstrakt

Tato bakalářská práce popisuje vývoj mobilní aplikace, která pomáhá při analýze snímků 1D gelové elektroforézy. Jednorozměrná gelová elektroforéza je široce používaná technika pro separaci proteinů nebo nukleových kyselin ve vzorku na základě jejich velikosti a náboje. Vzorek se vloží na polyakrylamidový gel, který funguje jako molekulární síto. Gel se poté umístí do pufrovacího roztoku a na gel se působí elektrickým polem. Molekuly ve vzorku migrují gelem různou rychlostí v závislosti na svém náboji a velikosti, což vede k vytvoření pásových vzorů, které lze vizualizovat pomocí barvicích technik. Porovnáním vzorů pásů vzorku se známými markery molekulových hmotností nanesenými na gel je možné provést kvantitativní analýzu molekulových hmotností separovaných molekul. Vzdálenost, kterou urazí každý pás, je úměrná logaritmu molekulové hmotnosti molekuly, což umožňuje odhadnout molekulovou hmotnost separovaných molekul.

Oblast gelové elektroforézy hraje zásadní roli v molekulární analýze tím, že usnadňuje separaci a zkoumání molekul na základě jejich velikosti a náboje. S rostoucí poptávkou po účinnějších a uživatelsky přívětivějších metodách zpracování a analýzy obrazů z gelové elektroforézy je však třeba překonat omezení stávajících nástrojů. I když bylo vyvinuto několik softwarových nástrojů pro stolní počítače, často jim chybí univerzálnost kvůli jejich závislosti na specifických elektroforetických nebo optických systémech. Mobilní aplikace nabízejí potenciál pro rychlou a uživatelsky přívětivou analýzu gelové elektroforézy a eliminují složitost spojenou s obsluhou specializovaných systémů. Využitím všudypřítomnosti a pohodlí mobilních zařízení mohou výzkumní pracovníci dosáhnout průběžných a dostupných výsledků, což v konečném důsledku šetří cenný čas a zdroje.

Praktická část této bakalářské práce zahrnuje několik klíčových bodů. Za prvé byl vyvinut algoritmus pro analýzu obrazu v jazyce Python, který provádí různé úlohy, včetně základního zlepšení kvality obrazu, segmentace drah, detekce proužků a kvantitativní analýzy separovaných molekul. Tento algoritmus slouží jako základ pro další fáze projektu. Kromě toho byla pomocí aplikace Android Studio vyvinuta aplikace pro Android 33, která poskytuje uživatelsky přívětivé rozhraní pro snímání a zpracování snímků gelové elektroforézy na mobilních zařízeních. Funkce analýzy obrazu byla implementována na serveru a poskytuje komunikační koncové body pro interakci mezi klientem, serverem a databázovými entitami. V neposlední řadě praktická část práce zahrnuje rozsáhlé testování jednotlivých kroků algoritmu s cílem zajistit přesnost, spolehlivost a výkonnost. Prostřednictvím důkladného testování byla ověřena účinnost vyvinutého algoritmu analýzy obrazu a celková funkčnost aplikace pro systém Android.

Vyvinutá metoda analýzy obrazu zahrnuje několik klíčových kroků. Nejprve se vstupní obraz převede do stupňů šedi a vyhodnotí se průměrná intenzita pixelů. Byly testovány různé techniky, včetně ekvalizace histogramu, gama korekce, kusové lineární úpravy kontrastu a prahování pozadí. Nejúčinnější zlepšení kvality gelového obrazu

přinesla gama korekce. Pro segmentaci drah algoritmus vypočítává střední hodnotu intenzity každého sloupce v horní třetině obrazu gelu, aby identifikoval vrcholy a označil polohu hranice každé dráhy. Algoritmus pak sleduje pixely každé hraniční čáry až na dno gelu. Pro získání 1D reprezentace intenzity signálu se profily ve stupních šedi jednotlivých pruhů vzorku zploští a použije se mediánová filtrace. To pomáhá při analýze a porovnávání polohy pásů. Korekce stínování pozadí se provádí použitím metody lokálních maxim pro odhad obálky. K výpočtu maximální hodnoty v rámci okna se použije posuvné okno vhodné velikosti a výsledek se interpoluje, aby se získala křivka pokrývající celý signál. Obálka se odečte od původního signálu, aby se snížilo zastínění pozadí. Detekce pásma zahrnuje analýzu jednopásmového signálu invertováním signálu a výpočtem maximální amplitudy. Píky se detekují pomocí funkce find_peaks, která poskytuje umístění pásem v signálu. Nakonec se vytvoří kalibrační křivka pro odhad molekulových hmotností neznámých vzorků proteinů nebo nukleových kyselin. Zjistí se migrační vzdálenosti známých proužků markerů molekulových hmotností a interpolací se vytvoří kalibrační křivka. Pomocí této kalibrační křivky lze odhadnout molekulové hmotnosti neznámých pásů v pruzích vzorků.

Infrastruktura mobilní aplikace byla vyvinuta pomocí Android Studia v jazyce Kotlin pro platformu Android. Komponenta uživatelského rozhraní (UI) slouží jako hlavní prostředek pro interakci uživatelů s aplikací. Aplikace umožňuje uživatelům pořizovat snímky pomocí fotoaparátu zařízení a poskytuje funkce pro ořezávání snímků a zadávání potřebných informací pro analýzu.

Komunikace mezi aplikací na straně klienta a serverem probíhá prostřednictvím požadavků HTTP. Při nahrávání obrázku k analýze je na server odeslán požadavek HTTP POST, který obsahuje obrazová data a parametry analýzy. Server požadavek zpracuje, uloží obrázek a zahájí analýzu. Klient pak může provádět požadavky HTTP GET pro získání stavu a výsledků analýzy.

Server komunikuje s databází SQLite pomocí příkazů SQL. Po nahrání obrázku server vytvoří záznam v databázi a odpovídajícím způsobem aktualizuje stav analýzy. Server pravidelně kontroluje databázi, zda se obrazy nacházejí ve stavu "TO_ANALYZE", aktualizuje jejich stav na "IN_PROGRESS", provede analýzu, a nakonec nastaví stav na "FINISHED. Databáze je také dotazována za účelem načtení dat obrázků nebo tabulek a poskytování odpovědí na požadavky klientů.

Testovací soubor dat zahrnuje laboratorní měření a snímky gelové elektroforézy z databáze, které poskytují reprezentativní vzorek reálných snímků. K některým snímkům jsou přiložena základní pravdivá data, která slouží jako reference pro hodnocení.

Pro posouzení robustnosti aplikace a algoritmu byly vybrány různé kategorie kvality snímků. Mezi tyto kategorie patří standardní, šum, nerovnoměrné hranice, rozmazání a bubliny. Představují běžné problémy při zpracování obrazů gelové elektroforézy. Kategorie standardních snímků představuje snímky s dobrou kvalitou a slouží jako měřítko. Kategorie šumu obsahuje odchylky a artefakty, které jsou výzvou pro schopnosti

redukce šumu. Kategorie nerovnoměrných hranic hodnotí přesnou segmentaci pruhů. Kategorie rozmazání testuje schopnost algoritmu poradit si s rozmazáním a provést přesnou detekci pásů. Kategorie bublin hodnotí robustnost algoritmu při identifikaci a kompenzaci artefaktů.

Při testování segmentace algoritmus automaticky detekoval dráhy a označil jejich hranice. Výsledek bol vyhodnocován ručně. Algoritmus prokázal vynikající výkon ve všech kategoriích kvality obrazu, s výjimkou nejednotných hranic, kde dosáhl mírně nižší míry přesnosti. Pro zlepšení výkonu při řešení takových případů může být zapotřebí dalšího zdokonalení.

Testování detekce pásů zahrnovalo porovnávání polohy detekovaných pásů s informacemi z databáze. Pro posouzení přesnosti algoritmu při identifikaci pásů byla vypočtena přesnost a citlivost. Algoritmus měl tendenci vykazovat vyšší míru nadměrné detekce, což mělo za následek více falešně pozitivních než falešně negativních výsledků. Vyřešení tohoto problému a zlepšení korekce pozadí u rozmazaných snímků by mohlo zvýšit výkonnost.

Testování odhadu molekulové hmotnosti zahrnoval porovnání odhadnutých hmotností detekovaných pásů s hodnotami v databáze. Jako hodnotící metrika byla použita střední kvadratická chyba (RMSE). Hodnoty RMSE poskytly náhled na výkonnost algoritmu, ale interpretace byla náročná kvůli chybám při přípravě gelu a vizualizaci. Implementace funkce pro úpravu vzájemné polohy proužků by mohla pomoci zlepšit přesnost. Navzdory těmto problémům zůstala výkonnost algoritmu při odhadu molekulových hmotností robustní bez ohledu na kvalitu obrazu, pokud nebyl přítomný smile efekt.

Závěrem lze říct, že tato bakalářská práce úspěšně vyvinula mobilní aplikaci pro zpracování obrazu gelové elektroforézy, která nabízí uživatelsky přívětivý a pohodlný nástroj pro analýzu snímků gelu. Testování a vyhodnocení aplikace a implementovaných algoritmů prokázalo jejich účinnost a spolehlivost. Práce poukázala na silné stránky a omezení metod a zdůraznila potenciál mobilní aplikace jako časově úsporného a nákladově efektivního řešení. Budoucí směry zahrnují zlepšení kvality obrazu, implementaci automatického ořezu, umožnění manuálního nastavení a začlenění kompenzace zkreslení obrazu a klasifikace vzorků.

Bibliographic Citation

MOJŽIŠOVÁ, Anna. Mobile application for electrophoretic gel image processing. Brno, 2023. Also available from: https://www.vut.cz/studenti/zav-prace/detail/150819. Bachelor's Thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Biomedical Engineering. Advised by Ing. Helena Škutková, Ph.D.

Author's Declaration

Author:	Anna Mojžišová
Author's ID:	223378
Paper type:	Bachelor's Thesis
Academic year:	2022/23
Topic:	Mobile application for electrophoretic gel image processing

I declare that I have written this paper independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the paper and listed in the comprehensive bibliography at the end of the paper.

As the author, I furthermore declare that, with respect to the creation of this paper, I have not infringed any copyright or violated anyone's personal and/or ownership rights.
In this context, I am fully aware of the consequences of breaking Regulation S 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll. of the Czech Republic, Section 2, Head VI, Part 4.
.

.
in Brno: 29.05.2023

----
author's signature

Acknowledgment

I would like to express my sincere gratitude to my advisor, Ing. Helena Škutková, Ph.D., for her invaluable help, advice, and guidance throughout this bachelor's thesis.

In Brno: 29.05.2023

----

author's signature

Contents

LIST OF SYMBOLS AND ABBREVIATIONS

Abbreviations:

FEKT	Fakulta elektrotechniky a komunikačních technologií
VUT	Vysoké učení technické v Brně
2DGE	two dimensional gel electrophoresis
CNN	convolutional neural network
UI	user interface
RMSE	relative mean squared error

Symbols:

γ	gamma value	[ ]

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Gel electrophoresis is a widely used technique for separating and analyzing molecules based on their size and charge. In recent years, there has been an increased demand for more efficient and user-friendly gel electrophoresis image processing and analysis methods. This has led to the development of numerous software tools for desktop computers that allow researchers to analyze their gel electrophoresis images.

However, many of these desktop tools are tied to specific electrophoretic or optical systems, limiting their usefulness in other settings. Mobile applications have the potential to provide a quick and easy-to-use solution for continuous results without the need for operating complex systems, which could save time and money.

The main goal of this thesis is to develop a mobile application for gel electrophoresis image processing and analysis. The thesis will focus on creating an image analysis algorithm with basic image quality enhancement and contrast adjustment, lane segmentation, band detection, and quantitative analysis of separated molecules. The mobile application will include functions for image capture, definition of new molecular weight markers, image analysis, export, and sharing of results.

To achieve these goals, the thesis will begin with a literature review of electrophoresis basics and principles and methods used for gel image processing and analysis. This will be followed by the development of the mobile application and the image analysis algorithm. Testing of the algorithm and mobile application will be performed on real electrophoresis gel images with different image quality.

The thesis is divided into three parts. Chapters 1 and 2 introduce theoretical knowledge concerning electrophoresis as a technique for molecule separation, focusing on gel electrophoresis and methods for image processing and analysis of gel images, respectively. Chapter 3 describes the practical realization of the mobile application. Finally, chapter 4 focuses on the testing methodology, results, and discussion, while the Conclusion assesses these results and introduces future directions.

1. ELECTROPHORESIS

Electrophoresis is a widely used laboratory method for the separation of charged molecules under the influence of an electric field [54]. This electrokinetic phenomenon was first observed by Peter Ivanovic Strakhov and Ferdinand Frederic Reuss in 1807 when they applied an electrical field to a clay particle solution, which caused the particles to migrate non-randomly [35]. Early forms of electrophoresis were crude and lacked the necessary resolution to separate complex mixtures of molecules [9].

The modern method of electrophoresis was first introduced by the Swedish biochemist Arne Tiselius in 1930 [49]. Moving-boundary electrophoresis was a significant advancement over earlier forms of electrophoresis as it allowed for the separation of different protein fractions based on their charge and molecular weight. This method laid the foundation for the development of many electrophoresis techniques used for the separation of a wide range of molecules, including DNA, RNA, and proteins [47].

1.1 Principle of Electrophoresis

Electrophoresis is the movement of charged particles under the influence of an electric field between two electrodes in an electrolyte medium [18]. Various biological molecules such as amino acids, peptides, proteins, nucleic acids, and nucleotides have ionizable groups that exist in a solution as a charged species (cation or anion) at a particular pH [12]. When a potential difference is applied between the electrodes, molecules with different charges migrate toward the anode or cathode. Separation of molecules with different net charges and different molecular sizes is possible through electrophoresis. The speed of particle movement depends on the size of the total surface charge, the size, and shape of the molecule, and its concentration in the solution [54]. Electrophoresis is an incomplete form of electrolysis. Staining techniques and molecular weight markers are used to visualize and analyze the separated molecules.

The charge of particles influences the direction of migration under the influence of an electric field: particles with a positive charge move toward the anode, while particles with a negative charge move toward the cathode [24]. The magnitude of the charge increases the rate of migration. According to the Henderson-Hasselbalch equation, the charge's size depends on pH [54]. The speed of migration is affected by an increase in molecular size and differences in the shape of the sample. There is also a frictional force that slows down the movement of charged molecules [52].

The electric field is a property that describes the space surrounding electrically charged particles. The movement of ions depends on the voltage, current, and resistance of the electric field. The distance ions travel is directly proportional to the applied current and time. The electrical resistance of the electrophoretic system depends on the type of system, the size and thickness of the supporting medium, the amount and conductivity of

the buffer, and the temperature [55]. One of the main problems with most forms of electrophoresis is the generation of heat [51].

1.2 Types of Electrophoresis

Electrophoresis is used in various modalities. Depending on the presence or absence of a supporting medium, the technique can be classified as free-flow electrophoresis or zone electrophoresis. Some authors also include capillary electrophoresis, isoelectric focusing, and isotachophoresis as part of free-flow electrophoresis, while others classify them as separate electrophoresis groups [18].

Capillary electrophoresis is a group of electrokinetic separation methods performed in capillaries with a submillimeter diameter and in micro- and nanofluidic channels [16]. Isoelectric focusing is a technique for separating amphoteric protein molecules based on their isoelectric point [8]. Isotachophoresis enables the separation of analytes consisting exclusively of anions or cations. The method separates smaller ionized substances into adjacent zones that do not overlap and all migrate at the same speed [9].

1.2.1 Gel Electrophoresis

In one-dimensional gel electrophoresis, the sample is typically loaded onto a polyacrylamide gel or an agarose gel, which acts as molecular sieves. The gel is made porous by adding a cross-linking agent to the polymer solution, which forms a three-dimensional network of polymer chains when exposed to an initiator under UV light [5]. The gel is then placed in a buffer-filled electrophoresis chamber, and an electric current is applied to the gel, as illustrated in Figure 1.1. The current causes the negatively charged molecules in the sample to migrate through the gel toward the positive electrode. The size and shape of the pores in the gel determine the rate of migration, resulting in the separation of the molecules based on their size and charge. Several different staining techniques can be used to visualize the separated molecules in 1D electrophoresis, including Coomassie blue staining, silver staining, and fluorescent staining [28][42]. Once the gel is stained, the separated molecules can be visualized as distinct bands or spots. The gel image can then be captured by a digital camera or scanner and stored as a digital image file.
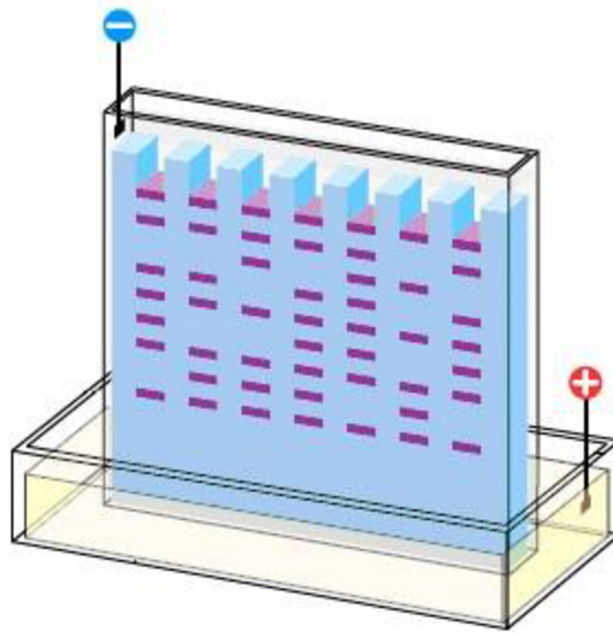
Figure   1.1   Gel electrophoresis apparatus.

Processing and analyzing these digital gel images is a crucial step in protein research. Researchers use image analysis software to quantify the amount of protein in each band or spot, compare protein expression levels across different samples, and identify differences between normal and diseased tissues. The software can also be used to analyze the molecular weight and isoelectric point of proteins, providing valuable insights into their structure and function [38]. The information obtained from electrophoresis can be used to construct phylogenetic trees and to study the evolutionary relationships between different species [9].

Molecular weight markers, also known as protein ladders, are often used in quantitative analysis of electrophoresis. These markers are composed of known proteins with well-defined molecular weights and are loaded onto the gel alongside the sample [21]. By comparing the banding pattern of the sample to that of the molecular weight markers, it is possible to estimate the molecular weight of the separated molecules. The distance traveled by each band is proportional to the logarithm of the molecule's molecular weight [28].

There are two types of conditions used in protein electrophoresis: native and denaturing. Native conditions preserve the protein's natural structure and charge, while denaturing conditions cause the protein to unfold and lose its native charge. Denaturing conditions are often used for the separation of complex protein mixtures [27].

Two-dimensional gel electrophoresis (2DGE) is a powerful technique that separates complex protein mixtures based on their isoelectric point and molecular weight. It is

commonly used for the analysis of protein expression patterns and the identification of differentially expressed proteins [38].

1.3 Detection and Analysis

After molecules have been separated by electrophoresis, various methods can be used to detect and analyze them. Staining techniques such as Coomassie blue staining, silver staining, and fluorescent staining can be used to visualize the separated molecules [28][42]. Autoradiography can be used to detect radiolabeled molecules, and immunoblotting can be used to detect specific proteins by using antibodies [24].

Each method has its advantages and disadvantages. Staining techniques are relatively simple and inexpensive but may not be sensitive enough for low-abundance molecules. Autoradiography is highly sensitive but requires the use of radioactive isotopes. Immunoblotting is specific and sensitive but can be time-consuming and expensive. The choice of method depends on the goals of the experiment and the nature of the molecules being analyzed.

1.3.1   Image Distortions in Gel Electrophoresis

In gel electrophoresis, we can expect various types of interferences caused by external influences, errors in the methodology during gel preparation, or the selection of incorrect method parameters [55]. Distortions can also occur when acquiring the image itself due to incorrect focusing, faults in the optical system, or movement of the sensor relative to the object. These distortions reduce image quality and need to be sufficiently reduced or compensated for before analysis.
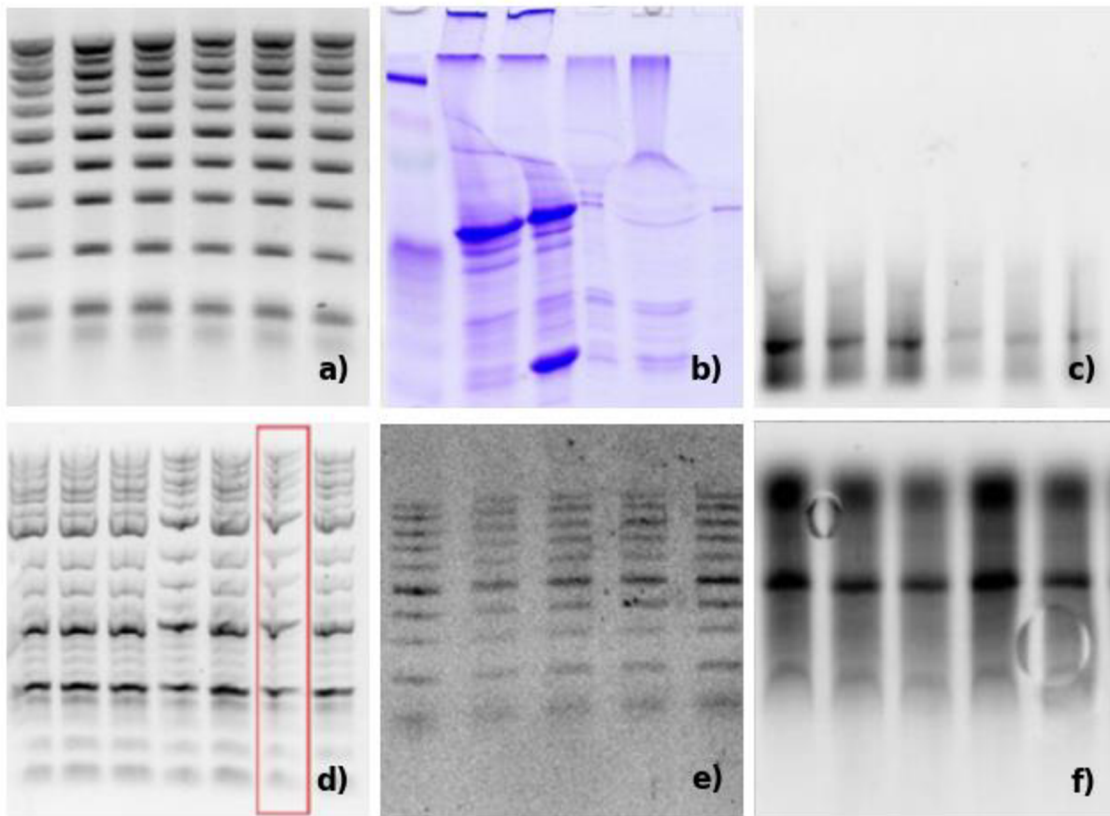
Figure 1.2 Image distortions in gel electrophoresis: gel images with a) concave curvature of sample lanes; b) non-uniform lane boundaries; c) blurred bands; d) damaged wells; e) background noise; f) air bubbles [19] [39].

The "smile effect" is a popular term for the upward curve caused by the slow migration of molecules in the outer or inner lanes of the gel due to uneven gel placement during electrophoresis or uneven heat distribution in the gel [5]. It manifests as a convex or concave curvature of all the lines in the gel, as seen in Figure 1.2 a). The most common cause of this distortion is the application of too high voltage, which can be prevented by suitable gel cooling during electrophoresis.

The "high salt effect" refers to the expanded and curved lines on the electrophoretic gel due to a high concentration of salt in the solution. It leads to the formation of non-uniform borders between the sample lanes, as seen in Figure 1.2 b) [39].

Blurring of the bands can occur due to several mechanisms, such as incorrectly selected gel concentration or its uneven distribution, low voltage, subsequent long electrophoresis running time, or excessive sample volume applied to the wells [5], as seen in Figure 1.2 c). The wells may be damaged during careless removal of the comb or removal of the comb before the gel fully solidifies, which will subsequently affect all the bands in that lane, as seen in Figure 1.2**Error! Reference source not found.** d). The well

can also be punctured during sample pipetting, and the puncture will again be visible in all bands in that lane.

The noise in the background of the electropherogram may be caused by gel contamination [37], as seen in Figure 1.2 e). Low contrast can result from an incorrect visualization procedure. Air bubbles can form during gel casting if not removed before the gel solidifies, causing optical deformation where they occur, as seen in Figure 1.2 f). Physical damage to the gel, such as cutting, can also cause distortion.

1.4 Applications

Electrophoresis has numerous applications in molecular biology, biochemistry, and clinical diagnostics. In molecular biology, electrophoresis is widely used for DNA fingerprinting, gene mapping, and studying gene expression [25]. In biochemistry, it is used to analyze proteins, enzymes, and other biomolecules [38]. In clinical diagnostics, electrophoresis is used for detecting and quantifying various disease markers, such as hemoglobin variants and serum proteins [37]. Overall, electrophoresis is an essential technique for various fields of research and plays a crucial role in advancing our understanding of biological systems.

2. GEL IMAGE PROCESSING AND ANALYSIS METHODS

The main objective of this chapter is to provide an overview of various image processing techniques and algorithms that can be applied to enhance the quality of gel images and extract useful information from them. These methods include pre-processing techniques such as noise reduction, gamma correction, and histogram equalization, and segmentation techniques such as lane detection and band detection. Additionally, this chapter will cover post-processing techniques such as molecular weight marker calibration and quantification of separated molecules. The information provided in this chapter will serve as a foundation for the development of the mobile application discussed in the next chapter.

2.1 Extraction of the Area of Interest

We need to perform border detection to extract the area of interest (gel with sample lanes) from the photo captured by the user's phone. This involves identifying the edges of the gel in the image. Apart from manual cropping of the image in the mobile application user interface, several automatic methods for border detection are available, including the Canny detector, Sobel operator, Prewitt operator, and Gaussian filter.

The Canny detector is an edge detection algorithm that uses a multi-stage process to detect edges in an image. It first applies a Gaussian filter to the image to reduce noise, then calculates the gradient magnitude and direction, and finally applies non-maximum suppression and hysteresis thresholding to produce the final edge map [22].

The Sobel operator is a simple edge detection algorithm that uses a 3x3 kernel to calculate the gradient of the image in the x and y directions. The magnitude of the gradient is then used to detect edges [15].

The Prewitt operator is similar to the Sobel operator but uses a slightly different kernel to calculate the gradient in the x and y directions [6].

The Gaussian filter is not a border detection algorithm per se, but rather a smoothing filter that can be used to reduce noise in an image before applying a border detection algorithm. It works by convolving the image with a Gaussian kernel to blur the image and reduce high-frequency noise [10].

2.2 Image Quality Enhancement

To perform analysis on the gel electrophoresis image captured by the user's phone, it is essential to enhance the image quality to mitigate distortions and to increase the accuracy of subsequent analysis steps. Several methods for image quality enhancement can be employed, including gamma correction, histogram equalization, noise reduction filters,

unsharp masking, color balance adjustment, and image registration.

Gamma correction is a powerful image enhancement technique that can significantly improve the quality of an image by adjusting its brightness and contrast. This non-linear method works by modifying the gamma value, which is a measure of the relationship between the brightness of an image and the brightness of the pixels that make up the image. By changing the gamma value, gamma correction can improve the visibility of details in the shadows and highlights of an image, resulting in a more visually pleasing and accurate representation of the original scene [34]. The effects of gamma correction with different gamma values is shown in Figure 2.1.
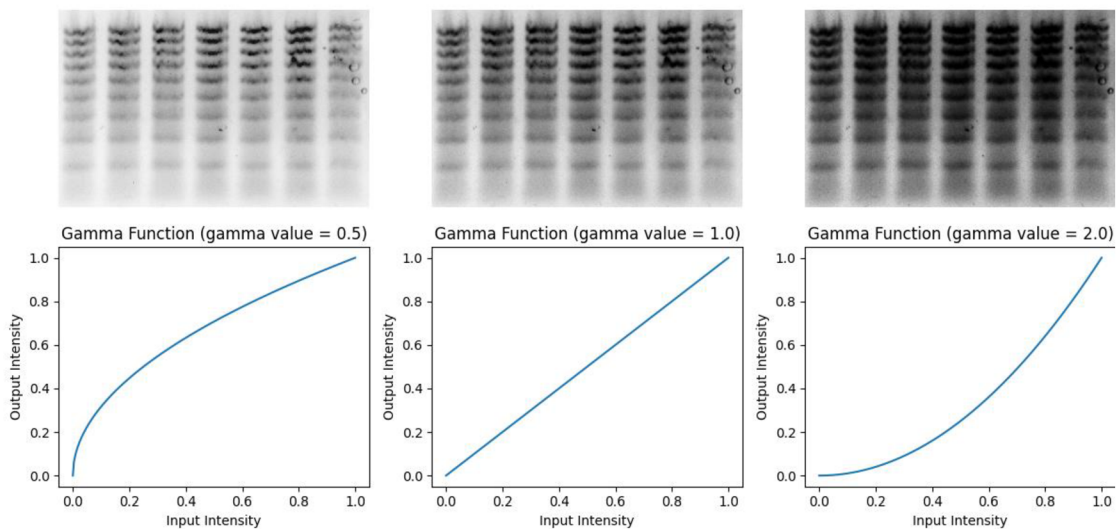
Figure   2.1   The effect of the gamma value on the transformation of the mean
              values of image brightness and the shape of the transformation
              function.

Histogram equalization is a well-known image processing technique that enhances the contrast of an image by redistributing the pixel intensities in the histogram, as shown in Figure 2.2.. The method works by computing the image histogram, which represents the frequency of occurrence of each intensity level. The histogram is then used to generate a cumulative distribution function, which maps the original pixel intensities to new, more evenly distributed values. This technique is particularly useful for enhancing images with poor contrast or for emphasizing features in images with low contrast [48].
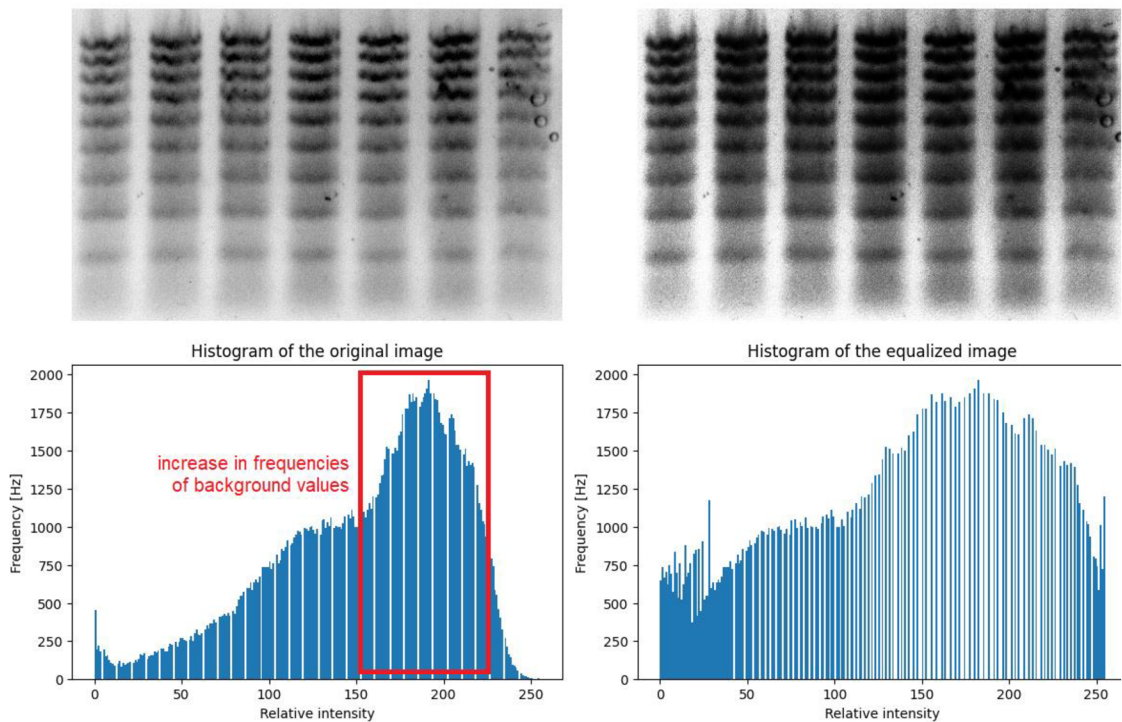
Figure   2.2   Principle of histogram equalization: original gel image and
corresponding histogram (left) and gel image after histogram
equalization and corresponding (right).

Noise reduction filters are a type of image processing technique used to remove unwanted noise or artifacts from an image caused by the imaging process. One of the most common types of noise reduction filters are median and Gaussian filters. Median filters work by replacing the value of each pixel with the median value of its neighboring pixels, while Gaussian filters use a weighted average of neighboring pixels to smooth the image [2][10].

Unsharp masking is a widely used image sharpening technique that emphasizes the edges in an image. This technique involves creating a blurred copy of the original image, which is then subtracted from the original image to enhance the edges. The result is an image with higher contrast and sharper edges. Unsharp masking is commonly used in image processing applications to improve the visual appearance of images without significantly altering their content. It is particularly effective for enhancing the edges of low-contrast images [41].

Image registration is a process used to align two or more images of the same object. In the context of gel electrophoresis, image registration is often used to align multiple images of the same gel taken at different times or under different lighting conditions. This technique can improve the overall quality and clarity of the image by reducing the impact of artifacts caused by changes in experimental conditions. Image registration can be

achieved using various techniques such as feature-based methods or intensity-based methods [53].

2.3 Lane Segmentation

To perform analysis on individual sample lanes, the gel image needs to be segmented. This involves separating the lanes from the background and/or from each other.

One approach to lane segmentation is to use machine learning. This method involves training a machine learning model to segment the lanes. The model is trained on a large dataset of images with ground truth segmentation labels. A deep learning model like a convolutional neural network (CNN) can be trained to perform the segmentation [11]. CNNs are well suited for this task as they can learn to recognize features of the lanes and accurately separate them from the background. Once the CNN is trained, it can be used to segment lanes in new images.

Until recently, the majority of presented approaches to lane segmentation were based on a principle called "lane tracking", which lies in tracking the center of each lane. The disadvantage of such an approach lies in ignoring the information from the area outside the central line [43]. One such method of lane segmentation involves projecting the image intensity information onto the horizontal direction [46]. This method utilizes the sum of intensities along the column for better discrimination between large bands in neighboring lanes, rather than the maximum value in each row. To smooth out the irregular profile of the projection, a smoothing operation is applied. The resulting curve's extreme points indicate the presence and central positions of each lane, with each lane limited by two maximal points of the curve. The points where the projection is minimal are associated with the lane central positions, as seen in Figure 2.3. For each lane, only the area that surrounds its central position is used for calculation purposes (10% for each side of the lane's central position).
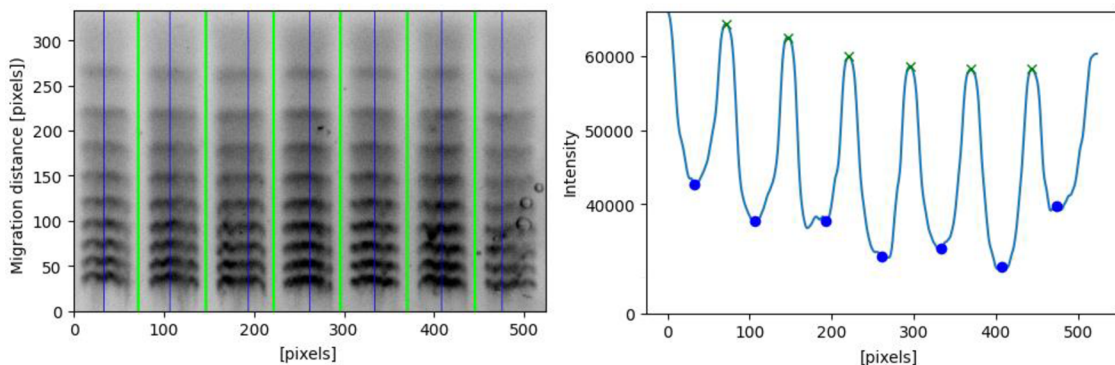

Figure   2.3   Principle of automatic lane detection based on the mean intensity profile: original image with detected lane central positions (left) and intensity profile of the image projection onto the horizontal direction (right).

An alternative method introduced by Škutková et al. [43] is based on tracking lane borders. The result of such an approach is full segmentation of the lanes, and therefore, information from all pixels of each lane can be used for subsequent analysis. The algorithm requires the number of lanes in the image as an input, and the rest of the segmentation process is fully automatic. In the first step, the intensity mean value is calculated for each column of pixels from the upper third of the image, and the peaks in the obtained 1D signal representation mark the position of the first pixel of each border, as seen in Figure 2.4 (upper). The next step involves the tracking of other pixels of each border line using an algorithm that compares the intensity values of the three pixels below the current pixel and selects the pixel with the highest intensity, as seen in Figure 2.4 (lower).
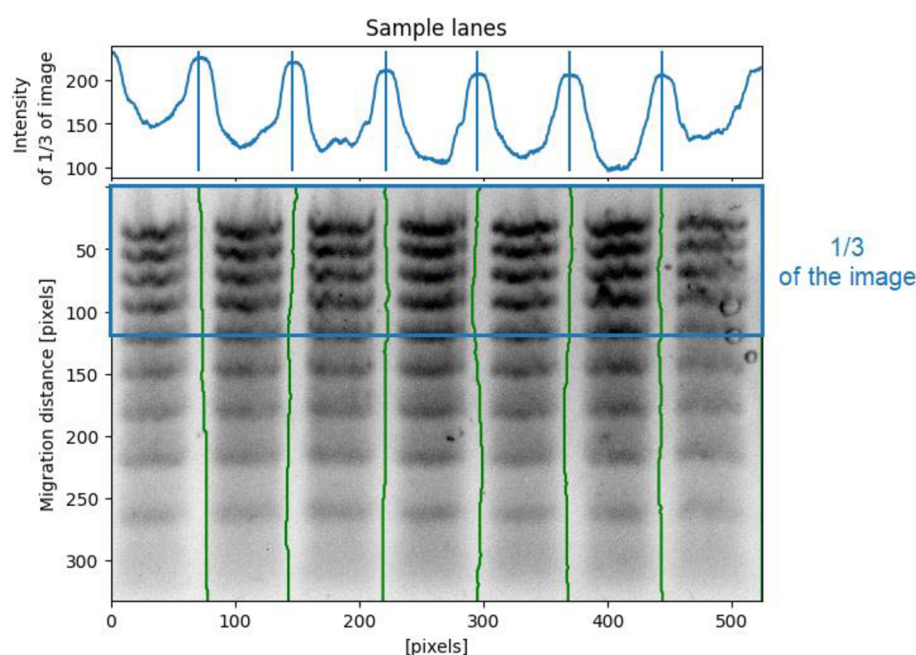
Figure 2.4 Principle of gel image segmentation using tracking of lane boundaries: detected positions of intensity peaks (upper) and segmented image with marked boundaries of each lane (lower).

2.4 Band Detection

Band detection is a critical step in gel electrophoresis image analysis as it allows us to identify and quantify the different bands in the gel lane. There are various methods for band detection, including peak detection, wavelet transform, and Fourier transform.

Peak detection involves detecting peaks in the 1D signal of the gel lane. The peaks correspond to the bands in the lane. This method can be performed using algorithms such as the Savitzky-Golay filter or the local maxima approach. The Savitzky-Golay filter applies a moving polynomial function to the signal to smooth it and reduce noise [17],

while the local maxima approach identifies local peaks based on their relative height and prominence [27].

Wavelet transform is another method for band detection that involves transforming the signal into the frequency domain. This method is useful for detecting overlapping or closely spaced bands. The wavelet transform uses a family of functions called wavelets to decompose the signal into different frequency components. The position and magnitude of the peaks in the transformed signal correspond to the bands in the gel lane [3].

Fourier transform is a method that also involves transforming the signal into the frequency domain. The Fourier transform uses sine and cosine waves of different frequencies to represent the signal. The peaks in the transformed signal correspond to the frequency components and their magnitudes correspond to the amplitudes of the bands in the gel lane [4].

2.5 Molecular Weight Estimation

A common method for molecular weight estimation is interpolation of weight values and migration distances of molecular weight marker to create a calibration curve [23][29]. In this method, a set of molecular weight markers of known sizes are run on the gel alongside the samples. The migration distances of the molecular weight markers are measured, and their sizes are known. A calibration curve is then created by plotting the migration distance of each marker against its known size. This curve is used to estimate the sizes of the sample bands based on their migration distances.

This method assumes that the migration distance of each band is directly proportional to its molecular weight and that the relationship between migration distance and the molecular weight is linear. However, the relationship may not be perfectly linear, especially for very large or very small molecules. Additionally, this method requires a set of molecular weight markers of known sizes, which may not always be available or may not include markers of the appropriate sizes for the samples being analyzed.

There are alternative methods for molecular weight estimation, such as using ladderless gel electrophoresis or DNA sequencing for DNA samples or using mass spectrometry for protein samples [54]. However, these methods may require specialized equipment and expertise and are not suitable for routine analysis in a mobile application.

3. Development of the Mobile Application and Image Processing Algorithm

The practical part of this thesis aims to develop a mobile application for gel electrophoresis image processing and analysis. The scope of this practical endeavor encompasses the implementation of image processing algorithms, band detection techniques, lane segmentation methods, and molecular weight approximation algorithms. The primary objective is to create a user-friendly and efficient mobile application that streamlines the analysis process of 1D gel electrophoresis images. The practical part of this thesis also involves testing the application using a diverse dataset of gel images to evaluate its performance, accuracy, and usability. The goal is to provide researchers and clinicians with a powerful tool that simplifies and enhances the analysis of electrophoresis images, thereby facilitating scientific discoveries and advancements in various fields such as molecular biology, genetics, and biochemistry.

3.1 User Interface

The User Interface (UI) component of the application was developed using Android Studio in Kotlin for the Android 33 platform. It serves as the primary means for users to interact with the application. Through the UI, users input all the necessary information required for the analysis and subsequently receive and review the results. Individual fragment screenshots are shown in Figure 3.1. The whole application run can be found in Attachment A.
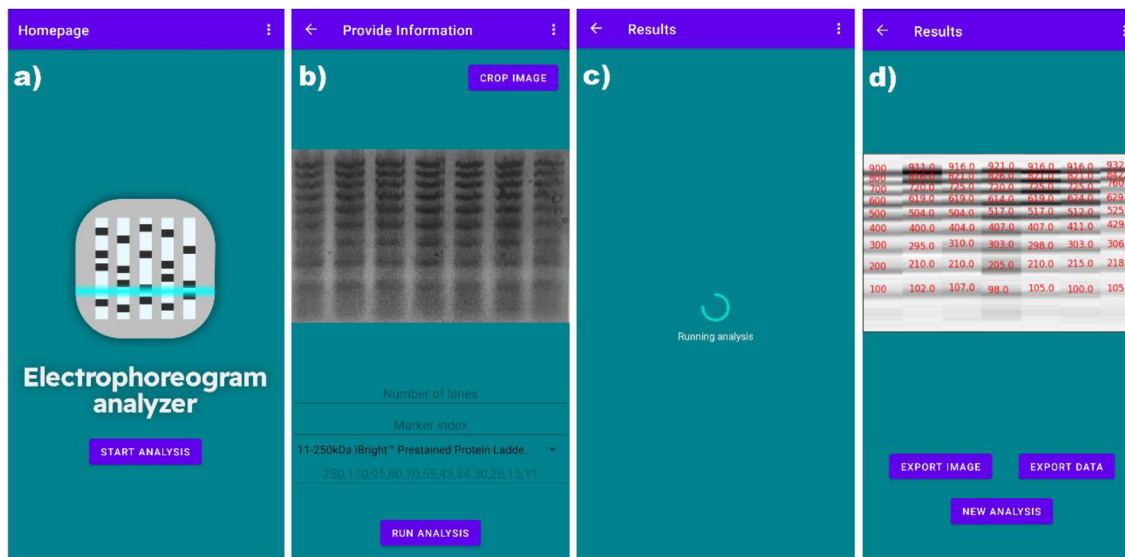

Figure 3.1 Screenshots of individual fragments: a) start analysis; b) provide information; c) loading; d) end analysis.

3.1.1 Capture Image

The CaptureImageFragment is a crucial part of the application's user interface that allows users to capture images using the device's camera. It requires camera permissions, which need to be granted by the user. Once permissions are granted, the camera starts, providing a live preview for users to compose their shots.

Capture of an image is carried out by tapping the "Capture" button. If the capture is successful, the image is saved to the MediaStore, and its URI and Base64-encoded data are stored in the SharedViewModel for sharing between fragments. The user is then navigated to the next fragment.

3.1.2 Provide Information

When the ProvideInformationFragment is opened, the user can see the captured image. They also have the option to crop the image by clicking the "Crop Image" button, which opens the image cropping activity.

To proceed with the analysis, the user needs to enter the number of lanes and the index of the lane containing the marker. Spinner allows the user to select a predefined molecular weight marker or choose the option "New" to enter a custom marker value. If the "New" option is chosen, the values of marker EditText becomes editable, allowing the user to input their desired marker values. If a predefined marker is selected, the EditText is disabled and displays the marker values of the selected marker. All user-inputted values are stored in the SharedViewModel for sharing between fragments. The input is validated, and if any field is missing or not entered correctly, an error message is displayed. Once the user has filled in all the required fields correctly, they can click the "Run Analysis" button. This action triggers the navigation to the next fragment. The screenshot of this fragment is shown in Figure 3.1 b).

3.1.3 Loading

The LoadingFragment is a crucial stage in the user interface of the application, despite not having a multitude of UI elements. It serves as the component responsible for sending the captured image data to the server for analysis.

It utilizes the sharedViewModel to access the image data and other relevant information provided by the user. By utilizing the lifecycleScope and Dispatchers.IO, the analysis process is performed asynchronously to prevent blocking the main thread. The image data, number of lanes, marker lane, and marker values are passed as a request to the server.

While waiting for the analysis to be completed, the LoadingFragment continuously checks the status of the analysis. The function then delays execution for five seconds before checking the status again. This process continues until the status is "finished," indicating that the analysis is complete.

Once the analysis is finished, the LoadingFragment retrieves the analyzed image and table containing result values from the. The files are then saved locally in the cache directory for future use. Finally, it utilizes the sharedViewModel to store the image URI, table URI, and updated image data. The application then navigates to the next fragment. The screenshot of this fragment is shown in Figure 3.1 c).

3.1.4 End Analysis

The EndAnalysisFragment is the final stage of the user interface in the application. In this fragment, the user is presented with the results of the image analysis and is provided with options to export the analyzed image and the corresponding CSV file.

Upon reaching the EndAnalysisFragment, the application retrieves the analyzed image URI and the CSV file URI from the sharedViewModel. Results can be exported using buttons labeled "Export Image" and "Export CSV". The screenshot of this fragment is shown in Figure 3.1 d).

3.2 Image Processing and Analysis Algorithm

The algorithm is designed to automate the process of analyzing electrophoretic gel images. It starts by importing and preprocessing the image, including conversion to grayscale and gamma correction. Then it performs lane segmentation to identify individual sample lanes and processes these lanes into one-dimensional representations by applying a median filter and background correction. Afterward, it detects the bands within each lane, and using a marker (a lane with known molecular weights), it interpolates a calibration curve to estimate the molecular weights of bands in other lanes. Finally, the algorithm visualizes the reconstructed image with the detected bands and their estimated molecular weights and saves values into a CSV file. The flow of the algorithm is shown in Figure 3.2.
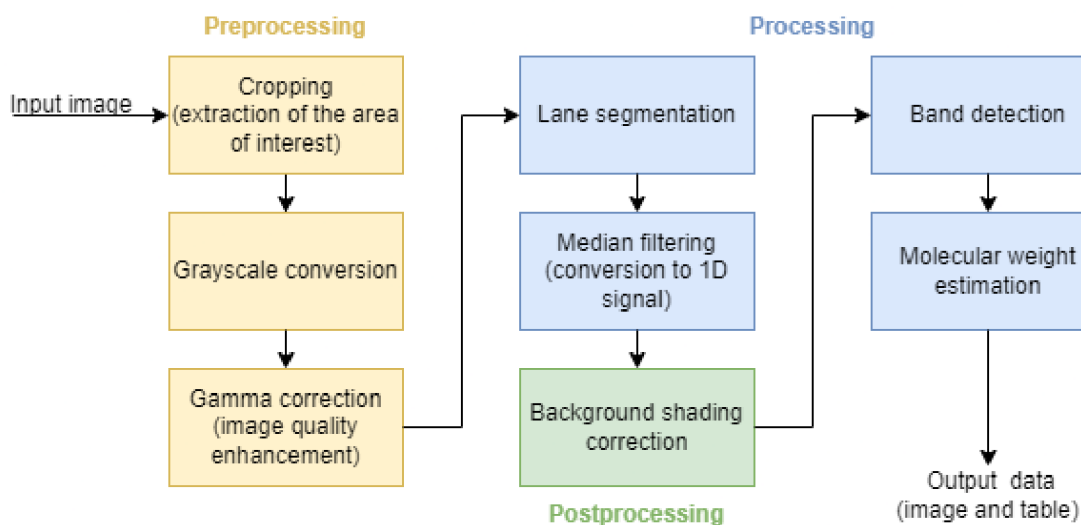
Figure   3.2   Block diagram of image processing and analysis algorithm.

3.2.1   Library Dependencies and Their Roles

In the development of the image processing and analysis algorithm, several libraries were used to leverage prebuilt functions and facilitate the implementation process. The following libraries played a crucial role in the algorithm:

OpenCV is a popular computer vision library that provides a wide range of functions for image processing and analysis. It offers extensive support for image manipulation, including color conversion, normalization, and various filtering techniques. The cv2 module from OpenCV was used to perform these operations in the algorithm.

NumPy is a fundamental library for numerical computing in Python. It provides efficient array operations and mathematical functions that are essential for image-processing tasks. The algorithm utilized NumPy arrays extensively for storing and manipulating image data.

SciPy is a scientific computing library that builds on top of NumPy. It offers additional functionality for scientific and technical computing, including signal processing and interpolation. The scipy.signal module was specifically used in the algorithm for peak detection and width measurement in the signals.

Matplotlib is a plotting library that provides a convenient interface for creating visualizations in Python. It was employed in the algorithm to generate plots and visualize the results of the image analysis.

Pandas is a powerful data analysis library that provides data structures and functions for efficient data manipulation and analysis. In the algorithm, Pandas was used to organize the marker values and lane positions into a structured table format, which was then saved as a CSV file.

27

The integration of these libraries enabled the algorithm to leverage prebuilt functions and achieve efficient image processing and analysis. The algorithm utilized various prebuilt functions from the libraries to implement specific image processing and analysis tasks:

cv2.normalize(img, None, 0, 255, cv2.NORM_MINMAX): This function is part of the OpenCV library and is used for image normalization. It takes an input image img and scales its pixel values to a specified range (0 to 255 in this case) using min-max normalization.

scipy.signal.find_peaks: This function, provided by the SciPy library, identifies peaks in a one-dimensional signal. It takes a signal as input and detects prominent peaks based on specified prominence and distance criteria.

scipy.signal.peak_widths: Another function from SciPy's signal module, peak_widths is used to compute the widths of peaks in a signal. It returns an array of widths corresponding to the identified peaks.

scipy.interpolate.interp1d: This function enables interpolation of a one-dimensional function. It constructs a callable interpolation function based on given data points. In the algorithm, it is used to create a calibration curve mapping band positions to marker values.

3.2.2   Image Preprocessing

Before applying any preprocessing techniques, it is essential to identify and select the area of interest in the image. In the context of the algorithm, the selection of the area of interest is performed on the client side of the application. This allows users to define the specific region within the image that contains the relevant information for further analysis. The selected area of interest is then passed to the preprocessing algorithm for subsequent steps.

The first preprocessing step applied to the input image is grayscale conversion. Grayscale conversion simplifies the image representation by transforming the original colored image into a grayscale image. This conversion eliminates the color information while retaining the intensity or brightness information of each pixel. The grayscale image is easier to process and analyze since it contains a single channel instead of the three channels (red, green, and blue) present in the colored image. The method used for grayscale conversion in this algorithm is implemented using the OpenCV library. The algorithm converts the input image from the BGR color space to the grayscale color space using the cv2.cvtColor() function. This function applies a predefined formula to calculate the grayscale intensity value for each pixel based on its color channels. The resulting grayscale image represents the original image with darker bands and a lighter background, which aids in subsequent analysis steps.

Numerous methods for image quality enhancement described in the previous chapter have been tested and evaluated. Out of these techniques, only gamma correction provided sufficient improvement in gel image quality. The proposed function provides a simple

and effective way to apply gamma correction to gel electrophoresis images, with the gamma value automatically calculated based on the mean intensity of the input image. The input to the function is an image represented as a NumPy array, and the output is the gamma-corrected image as a NumPy array. The function starts by calculating the mean intensity of the input image using the np.mean function. This mean intensity is then used to calculate the gamma value using the formula:

$$\gamma = \frac{\log 0.5}{\log \frac{mean\ intensity}{255}},$$ (3.1)

which calculates the gamma value γ that will result in an image with a mid-tone gray level of 0.5, which is typically considered to be visually pleasing.

Next, the function checks if the calculated gamma value is greater than 3, which is an arbitrary threshold. If the gamma value is greater than 3, it is set to a fixed value of 1.5. This is done to avoid excessive gamma correction, which can lead to over-saturation and loss of detail in the image.

Finally, the gamma correction is applied to the input image. The result of this calculation is then scaled back to the range [0, 255] and converted to an unsigned 8-bit integer using the np.uint8 function. The gamma-corrected image is then returned by the function.

3.2.3 Lane Segmentation

The algorithm of lane detection was slightly modified from the method proposed by Škutková et al. and requires an input value of the number of lanes of the analyzed electrophoretic image [43].

The method involves two main steps. Firstly, the algorithm calculates the intensity mean value of each column in the upper third of the gel image and identifies peaks using the find_peaks method to mark the position of the first pixel of each border. The algorithm automatically fills in the expected lane boundary if none is found. Secondly, the algorithm starts tracking the boundary of each lane by iterating over the y-axis of the image. For each row, the algorithm takes a window of three pixels centered at the x-coordinate of the previous row's pixel and selects the pixel with the highest intensity. If the highest intensity is not in the center of the window, it is shifted by 1.04 times the width of the lane in the direction of the highest intensity. The x-coordinate of the selected pixel is then added to the lane.

The process continues until the bottom of the image is reached, and a list of lanes is returned. Each lane is represented as a list of x-coordinates, where the y-coordinate is given by the index of the list. The algorithm also adds left and right border lanes to the list, represented by a list of 0's and the width of the image, respectively.

3.2.4 Conversion to 1D Signal

To obtain a 1D signal representation of the intensity of pixels from 2D grayscale profiles

representing individual sample lanes in horizontal orientation, the function takes as input the gel image lane, as well as the left and right borders of the current lane, and returns a 1D signal that represents the intensity of the lane across its length.

The algorithm first iterates over the rows of the image, and for each row, it selects the pixels that belong to the lane, based on the left and right borders. Then, the algorithm calculates the median intensity value of the selected pixels, which represents the intensity of the lane at that row. Finally, the algorithm appends the median intensity value to the output signal and repeats the process for all the rows in the image.

In this way, the 2D image representation of an individual lane is transformed into a 1D signal that represents the intensity of the lane across its length, as seen in Figure 3.3, which can be further analyzed and processed by the gel electrophoresis image processing algorithm.
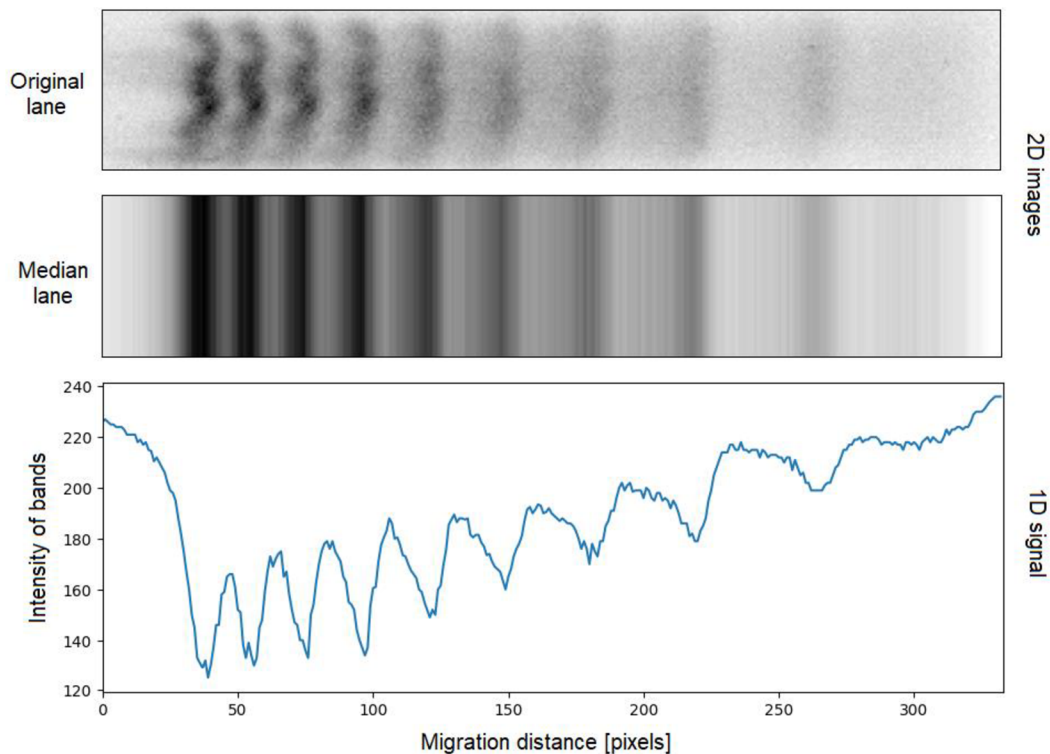
Figure   3.3   Principle of conversion of a lane image to signal representation using median filtering: original lane image (upper), lane image after median filtering (middle), and its intensity profile (bottom).

3.2.5   Background Shading Correction

In electrophoretic gels, the background can have a non-uniform shading due to factors such as uneven lighting or variations in the thickness of the gel. By correcting the shading of the background, the positions of bands can be more accurately measured and compared between different samples.

To achieve this, the algorithm first identifies the peaks of the lane signal using the find_peaks method and calculates the width of each peak using the peak_widths function. The largest peak width is then used to define a sliding window size, which is moved along the lane signal with a step size of one-quarter of the window size.

Before the sliding window operation, buffer values are added at the beginning and end of the lane signal to ensure that the window operation can be performed throughout the entire signal. The maximum value of each window is then recorded, and a piecewise linear interpolation function is created using these values. The function is then used to construct the envelope of the signal by evaluating the function at each point that is not a maximum.

Finally, the lane signal is corrected by subtracting the envelope from the original signal. The corrected lane signal is obtained by cropping the buffer values at the beginning and end of the signal and is returned by the function. The whole process can be seen in Figure 3.4.
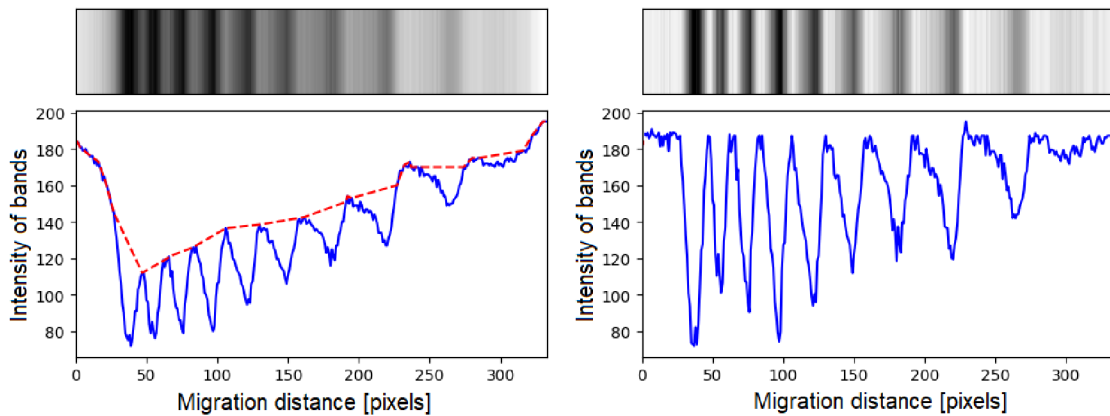
Figure   3.4   Principle of background shading correction: single lane image after median filtering with its intensity profile and estimated envelope shape (left) and single image lane after shading correction by subtraction of estimated signal envelope from single lane signal with its intensity profile (right).

3.2.6   Band Detection

The band_detection function is used for detecting bands or peaks in a single lane signal. It follows several steps. The input signal is multiplied by -1 to invert the signal, making the peaks positive and easier to detect. The maximum amplitude of the inverted signal is calculated using the np.amax(signal) function. A prominence threshold p is set to 0.2 times the maximum amplitude. This threshold determines the minimum height of a peak to be considered significant. A distance threshold d is set to 0.02 times the length of the signal. This threshold determines the minimum distance between peaks. The find_peaks function from the scipy.signal module is called with the inverted signal, prominence

threshold p, and distance threshold d. It returns the indices of the detected peaks in the signal. The function returns the indices of the detected peaks as an array. A lane signal with detected peaks is shown in Figure 3.5.

On the other hand, the marker_detection function is similar to the band_detection function but is modified to find a specific number of peaks corresponding to the provided marker values. It enables future interpolation of these values. The input marker_signal is multiplied by -1 to invert the signal. The maximum amplitude of the inverted marker signal is calculated using the np.amax(marker_signal) function. A prominence threshold p is set to 0.2 times the maximum amplitude. This threshold determines the minimum height of a peak to be considered significant. A distance threshold d is set to 0.02 times the length of the marker signal. This threshold determines the minimum distance between peaks. The find_peaks function is called with the inverted marker signal, prominence threshold p, and distance threshold d. It returns the indices of the detected peaks in the marker signal. The detected peaks are then sorted in descending order based on their corresponding values in the marker signal using the sorted function with a custom sorting key. The highest_peaks variable is assigned the first marker_length number of peaks from the sorted list. Finally, the highest_peaks are sorted in ascending order, ensuring that the peaks are in the correct order relative to the marker values. The function returns the sorted highest_peaks as an array.
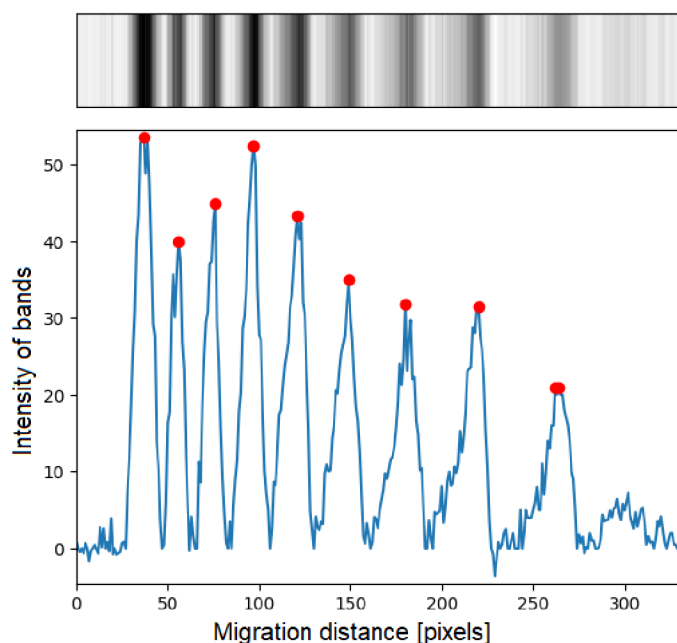

Figure   3.5   Principle of band detection: single lane 2D image representation (upper) and its intensity profile with detected band peaks (lower).

3.2.7   Molecular Weight Estimation

The create_calibration_curve function is used to create a calibration curve for molecular

weight estimation. It takes three parameters: bands_detected, marker_values, and marker_signal. The function utilizes the interp1d function from the scipy.interpolate module. It creates a callable object f that performs linear interpolation using the bands_detected and marker_values. The interpolation is set to be linear, and it allows extrapolation beyond the detected bands.

To generate the calibration curve, an x-range array is created using np.arange from 0 to the length of the marker_signal. This array represents the positions at which the interpolation will be evaluated. The x_range array is then passed to the f object, which returns the interpolated values as the calibration curve. The resulting curve is returned from the function and is shown in Figure 3.6.
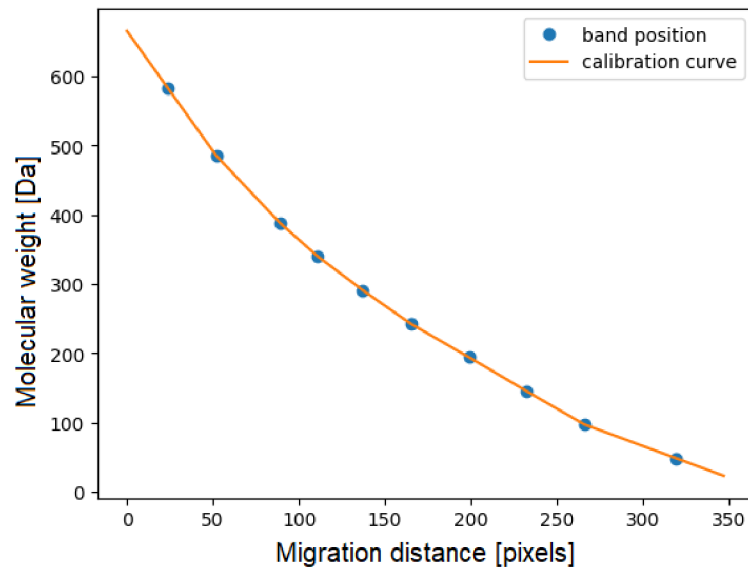
Figure   3.6   Calibration curve for molecular weight estimation.

The molecular_weights_evaluation function is responsible for evaluating molecular weights based on an input image. It takes five parameters: img, num_of_lanes, lanes, marker_values, and marker_lane. The function begins by calling the marker_detection function to identify the positions of the marker bands in the lane specified by marker_lane. The width of each lane is calculated by dividing the width of the input image (img.shape[1]) by the total number of lanes (num_of_lanes). This width will be used for positioning the text annotations in the output plot.

The create_calibration_curve function is then called with the marker band positions, marker values, and the signal of the marker lane. This generates the calibration curve for subsequent weight estimation.

The function initializes three lists: x_list, y_list, and text_list. These lists will store the x-coordinates, y-coordinates, and text values for annotating the plot. The maximum band count is determined by finding the maximum length among the detected marker band positions and the band positions in all the lanes.

The function iterates over the range from 0 to num_of_lanes to process each lane. Within the loop, it determines the band positions and weights for each lane. If the current lane index matches the marker lane index, the detected marker band positions and marker values are used directly. Otherwise, the band positions for the current lane are obtained using the band_detection function, and the corresponding weights are calculated using the calibration curve.

The band positions, weights, and rounded weights are added to the respective lists (lane_band_positions, band_weights, and rounded_list). These values will be used for generating the plot annotations. Arrays are padded with empty strings to ensure consistent length for generating the table data. The number of empty strings added depends on the difference between the maximum band count and the length of the lane-specific band positions.

Based on the lane index, the table data is constructed in a dictionary format, where keys represent lane names and values represent migration distances and molecular weights. The table data is used to create a DataFrame using the pd.DataFrame function. This DataFrame will be converted to a CSV file for data storage and further analysis.

A plot is created using matplotlib.pyplot.subplots(). The input image (img) is displayed on the plot, with the grayscale colormap applied. The axes are turned off to provide a clean visualization. Text annotations are added to the plot using the ax.text() function. The x-coordinates, y-coordinates, and text values are retrieved from the corresponding lists (x_list, y_list, and text_list). The annotations are positioned based on the x-coordinates and adjusted vertically to align with the bands.

Finally, the function returns the CSV data from the StringIO object and the plot figure object. These can be used for further analysis or visualization and are shown in Figure 3.7 and 3.8, respectively.

Marker lane		Sample lane 1		Sample lane 2		Sample lane 3	
Migration distances [pixels]	Molecular weights [Da]	Migration distances [pixels]	Molecular weights [Da]	Migration distances [pixels]	Molecular weights [Da]	Migration distances [pixels]	Molecular weights [Da]
37	900	35	911.0	34	916.0	33	921.0
56	800	53	816.0	52	821.0	51	826.0
76	700	72	720.0	71	725.0	72	720.0
97	600	93	619.0	93	619.0	94	614.0
121	500	120	504.0	120	504.0	117	517.0
149	400	149	400.0	148	404.0	147	407.0
180	300	182	295.0	177	310.0	179	303.0
220	200	216	210.0	216	210.0	218	205.0
262	100	261	102.0	259	107.0	263	98.0

Figure   3.7   Table with results for electrophoretic gel image analysis.

Figure 3.8 Gel image with molecular weight values of detected bands.

3.3 Client-Server-Database Communication

The core of the client-server-database model is composed of three primary entities: the client initiating service requests, the server accommodating these requests, and the database storing and managing data.

In the context of the ElectrophoreogramAnalyzer, the Android application is identified as the client, enabling user interaction for image upload and analysis result viewing. The server, built with FastAPI, is responsible for image processing and database communication, managing the state of image analysis. Finally, the SQLite database is utilized as a persistent data store, keeping information about each image intact. The interactions between the three entities are shown in Figure 3.9.

This model's implementation was driven by its capacity to distinctly compartmentalize components, facilitating modularity and scalability in the design, and supporting complex network operations.
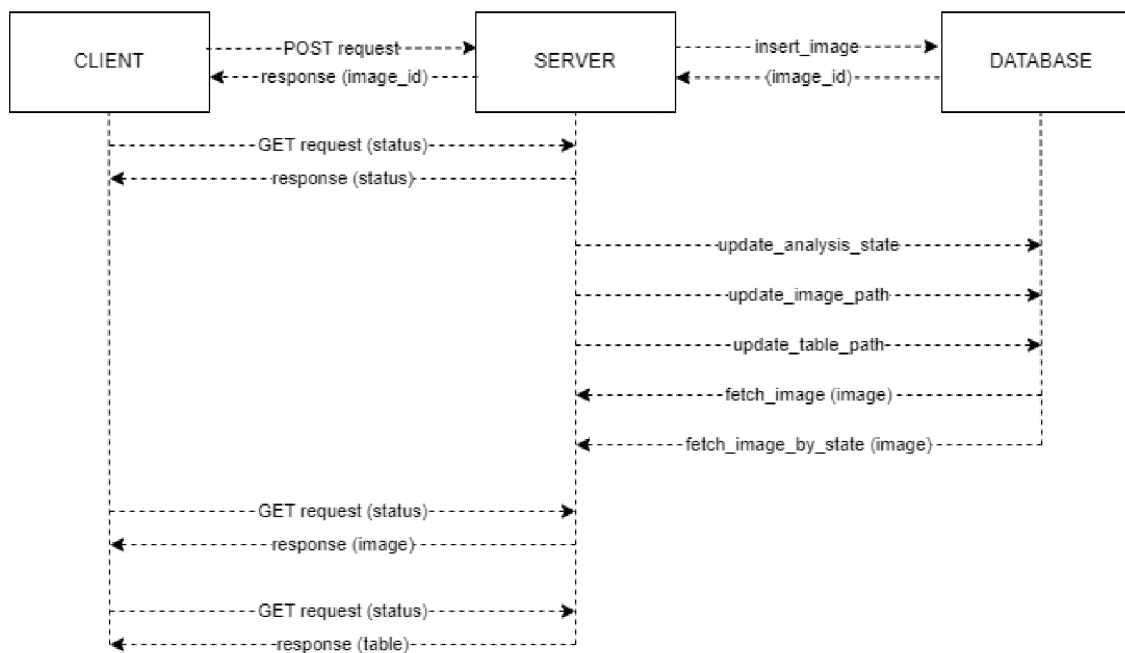
Figure   3.9   Client-server-database interaction flow for electrophoretic gel
image analysis.

3.3.1   Communication between the Client and the Server

When an image for analysis is to be uploaded, an HTTP POST request is sent by the client. This request carries a JSON payload consisting of image data, the number of lanes, marker lane, and marker values. Upon receipt, this request is processed by the server, leading to the image's storage and subsequent initiation of the analysis.

For the status or result of the analysis, HTTP GET requests are made by the client. The server then processes these requests and replies with the current status or final result. This can take the form of an image or table of data. JSON is chosen as the primary data format for these communications due to its lightweight, flexible, and human-readable nature, allowing for seamless data exchange.

Endpoints are clearly defined for communication. For example, "/upload" is used for image upload, while "/status/{image_id}" and "/result/{image_id}" are used to fetch the status and result of the analysis respectively.

3.3.2   Interactions between the Server and the Database

The server communicates with the SQLite database through SQL commands. When an image upload occurs, an entry in the database is created by the server. At this stage, the analysis state is marked as 'TO_ANALYZE'. As the analysis progresses, the image's state in the database is updated by the server, changing to 'IN_PROGRESS', and ultimately to 'FINISHED' when the analysis is complete.

Fetching operations are performed to retrieve the status of an image, obtain the image or table data, or acquire a list of images awaiting analysis. These interactions are central to the system's functionality and enable the server to provide accurate and timely responses to client requests.

3.3.3 Data Flow

The journey of an image from being uploaded to the analysis result involves several stages. Initially, the client uploads an image using a POST request. This request is received and stored by the server while creating a new database entry. The analysis state is marked as 'TO_ANALYZE' at this point.

The server, during its periodic database checks, identifies images with the 'TO_ANALYZE' state, updates their state to 'IN_PROGRESS', and initiates the analysis.

Throughout this period, the client can request the status of the analysis with a GET request. In response, the server provides the current state retrieved from the database.

On completion of the analysis, the state is updated to 'FINISHED' by the server, and the result is stored. A GET request can now be made by the client to retrieve the result. This result is fetched from the database and relayed back to the client by the server.

3.3.4 Error Handling and Exceptions in Client-Server-Database Communication

Errors and exceptions are unavoidable in any system. These are managed effectively in client-server-database communication with the use of HTTP status codes. For instance, if a request is made for the result of an image that is not yet analyzed or does not exist, the server responds with an HTTP status code to indicate an error.

4. RESULTS AND DISCUSSION

This chapter provides a comprehensive analysis of the performance and accuracy of the implemented algorithms for lane segmentation, band detection, and molecular weight estimation. The obtained results are presented and discussed in detail, highlighting the strengths, limitations, and implications of the developed methods. Additionally, this chapter serves as a platform for a critical examination of the achieved outcomes, comparison with existing software tools, and exploration of potential areas for further improvement and future research.

4.1 Dataset Creation and Image Quality Categories

In this section, we discuss the creation of the testing dataset for evaluating the mobile application and the quantitative analysis algorithm. The dataset was compiled using a combination of laboratory measurements from previous years' students and images obtained from a database of gel electrophoresis images [19][36]. This diverse dataset provides a representative sample of real-world gel images encountered in practice. Furthermore, a subset of the images in the dataset was accompanied by ground truth information, which included the molecular weights of the individual detected bands. This ground truth data serves as a reference for evaluating the accuracy and performance of the quantitative analysis algorithm.

To simulate various image quality scenarios and assess the robustness of the mobile application and algorithm, different image quality categories were chosen. These categories include standard, noise, non-uniform, blurred, and bubbles. While additional categories could exist, these were selected to encompass a range of common challenges encountered in gel electrophoresis image processing.

The standard category represents images with good overall quality, minimal noise, and well-defined lane boundaries. These images serve as a benchmark for evaluating the baseline performance of the mobile application and algorithm.

The noise category comprises images that contain random variations or artifacts caused by factors such as sensor noise, lighting fluctuations, or other sources of interference. These images challenge the algorithm's ability to distinguish true bands from noise and assess its noise reduction capabilities.

The non-uniform boundaries category includes images where the lane boundaries are irregular or discontinuous. This category aims to evaluate the algorithm's ability to accurately segment lanes under challenging conditions.

The blurred category consists of images with reduced sharpness or blurriness due to factors such as motion blur, focus issues, or image acquisition limitations. This category assesses the algorithm's ability to handle blurry images and perform accurate band detection.

The bubbles category represents images with the presence of air bubbles or artifacts that can obscure or distort the bands. This category tests the algorithm's robustness in identifying and compensating for such artifacts during the analysis process.

The characterization of the images in each category provides valuable insights into the specific challenges and complexities encountered in gel electrophoresis image processing. It allows for a systematic evaluation and comparison of the algorithm's performance across different image quality scenarios, leading to a comprehensive assessment of the mobile application's effectiveness in real-world scenarios.

4.2 Testing Methods and Results

Testing and evaluation of the mobile application and the quantitative analysis algorithm are of paramount importance in ensuring their reliability, accuracy, and effectiveness. Through comprehensive testing, the performance of the application can be assessed in various scenarios and under different conditions, validating its functionality and usability. The evaluation process allows for the identification and resolution of any potential issues, such as bugs or errors, ensuring a smooth and seamless user experience. Additionally, testing provides an opportunity to compare the results generated by the algorithm with ground truth data or established methods, establishing the algorithm's accuracy and effectiveness in quantitative analysis. Ultimately, the thorough testing and evaluation of the mobile application and the quantitative analysis algorithm are crucial in guaranteeing their robustness, performance, and suitability for real-world applications in gel electrophoresis image processing.

4.2.1 Lane Segmentation

The lane segmentation testing methodology involved the segmentation of all images in the dataset using the developed algorithm. The algorithm automatically detected and marked the lane boundaries for each image. The segmented images with the marked lane boundaries were then exported for further evaluation.

To assess the accuracy of the lane segmentation algorithm, a manual evaluation was performed. Each segmented lane in the exported images was visually inspected and compared against the ground truth or expected lane boundaries. The manual evaluation involved checking the alignment and accuracy of the segmented lanes concerning the actual lanes in the gel images.

The accuracy of the lane segmentation algorithm for individual image quality categories was calculated by determining the number of correctly segmented lanes and dividing it by the total number of lanes in the dataset, as seen in Table 4.1. This provided a quantitative measure of the algorithm's performance in accurately identifying and delineating the lanes in the gel electrophoresis images.

By employing this testing methodology, the effectiveness, and reliability of the lane segmentation algorithm could be assessed. The manual evaluation of the segmented lanes

allowed for a thorough examination of the algorithm's ability to accurately reproduce the true lane boundaries, providing insights into its performance and potential areas for improvement.

Table 4.1   Lane Segmentation Results

Category	No. of lanes tested	Accuracy [%]
standard	112	100.0
noise	84	100.0
non-uniform	34	92.7
blurred	135	100.0
bubbles	47	100.0

The lane segmentation algorithm demonstrated excellent performance across all image quality categories. In the standard, noise, blurred, and bubbles categories, the algorithm achieved a perfect accuracy rate of 100.0%, accurately segmenting all lanes in the corresponding images. This indicates that the algorithm is robust and reliable, capable of accurately detecting and delineating lanes even in the presence of noise, blurring, or the presence of air bubbles. Other articles report lane segmentation accuracy ranging from 95% to 100% [23][30][46].

However, in the non-uniform category, where the gel lanes had irregular boundaries or variations in intensity, the algorithm achieved a slightly lower accuracy rate of 92.7%. Although the algorithm was successful in segmenting most lanes in these images, it encountered some challenges when dealing with non-uniform lane boundaries. This suggests that further refinements or adjustments to the algorithm may be necessary to improve its performance in handling such cases. Other lane segmentation methods may be more fitted for these images.

4.2.2   Band Detection

The band detection testing methodology involved the following steps. First, bands were detected on all correctly segmented images from the dataset, and their positions were recorded into an array. Next, this array of detected band positions was compared algorithmically with the ground truth information, which provided the number and/or position of the actual bands in the image (depending on the specific dataset). A small allowance of a 5% difference in band positions was considered acceptable, as this metric was evaluated separately from band detection.

To assess the performance of the band detection algorithm, two key metrics were calculated: precision and sensitivity. Precision was determined by dividing the number of true positive detections (correctly detected bands) by the total number of detected bands. This metric indicates the accuracy of the algorithm in correctly identifying and localizing the bands within the image. Sensitivity, on the other hand, was calculated by dividing the

number of true positive detections by the total number of ground truth bands. This metric measures the algorithm's ability to capture and detect the actual bands present in the image. The results can be seen in Table 4.2.

By evaluating both precision and sensitivity, a comprehensive assessment of the band detection algorithm's performance can be obtained. These metrics provide insights into the algorithm's accuracy in detecting bands and its ability to minimize false positives and false negatives.

Table   4.2   Band Detection Results

Category	No. of bands tested	Precision [%]	Sensitivity [%]
standard	1038	99.6	99.8
noise	862	98.9	99.4
non-uniform	295	92.7	92.4
blurred	1264	94.0	94.8
bubbles	486	96.2	96.5

It can be observed from the results that the sensitivity values were higher than the precision values across all categories. This suggests that the algorithm tends to exhibit a higher rate of overdetection, leading to more false positives than false negatives. To address this issue, one potential improvement could be to increase the prominence parameter in the methods utilizing find_peaks. However, it should be noted that increasing this parameter can result in underdetection. Consequently, a balance must be struck to achieve accurate band detection.

The presence of decreased image quality, particularly blur, contributes to the higher rate of overdetection. In blurred images, the uneven background shading becomes more prominent, leading to a higher number of false positive detections. To mitigate this effect, improving the background correction method to perform more precise envelope estimation could be beneficial.

One notable finding is that band detection on non-uniform images yielded the lowest success rate in terms of both precision and sensitivity. This can be attributed to the significant distortion of bands in these images, where median filtration may not be sufficient for reconstructing the lane image accurately. A potential solution could involve using a narrower range of x-values for median filtration, enabling the restoration of bands more effectively.

4.2.3   Molecular Weight Estimation

The testing methodology for molecular weight estimation involved the following steps. First, bands were detected on all correctly segmented images from the dataset, and their positions were recorded in an array. Only images with provided ground truth information about the position of true bands were selected for this evaluation. Due to the limited

number of such images, the evaluation was conducted on two categories of image quality: standard and blurred.

Next, the array of detected band positions was algorithmically compared with the ground truth information of the true band positions in the image. The aim was to assess the accuracy of the molecular weight estimation method. In this evaluation, the root mean squared error (RMSE) was used as the evaluation metric.

RMSE calculates the square root of the average squared difference between the predicted values (detected band positions) and the true values (ground truth band positions). This metric provides a comprehensive measure of the estimation accuracy, as it gives more weight to larger errors due to the squaring operation.

By evaluating the RMSE, the algorithmic accuracy of the molecular weight estimation method was assessed. Lower RMSE values indicate better agreement between the estimated and true molecular weights, while higher RMSE values indicate larger discrepancies. The results are shown in Table 4.3.

Table 4.3   Molecular Weight Estimation Results

Category	No. of bands tested	Relative RMSE [%]
standard	148	5.4
blurred	184	6.8

The results obtained from the molecular weight estimation evaluation using the RMSE metric provide valuable insights into the performance of the algorithm. However, it is important to note that interpreting this metric can be challenging due to various factors that can contribute to errors in the estimated values.

One notable factor is the presence of errors in gel preparation and visualization, which can result in a phenomenon known as the "smile effect." The smile effect causes distortions in the gel image, leading to differences between the estimated molecular weights and the expected values. Consequently, the RMSE values may be influenced by these errors, making it difficult to solely attribute the discrepancies to the algorithm's performance. To address this issue, a potential improvement could be the implementation of a function for adjusting the mutual positions of bands, such as utilizing dynamic time warping [43]. This approach could help mitigate the impact of gel preparation and visualization errors on the molecular weight estimation accuracy.

It is worth mentioning that another article has reported differences in the range of 0.73% to 21.43% between the predicted and expected values of molecular weights [29]. These variations can be attributed to the different types of commercial pre-stained standards and separated proteins used in that study. The differences observed in molecular weight estimation between various commercial pre-stained standards and proteins indicate the inherent challenges in accurately estimating molecular weights and the influence of the specific samples being analyzed.

Interestingly, when considering images without the smile effect, the RMSE values were significantly lower, approximately 1.2% for both standard and blurred images. This suggests that the quality of the images had minimal impact on the accuracy of molecular weight estimation. Therefore, it can be inferred that the algorithm's performance in estimating molecular weights remains robust, regardless of the image quality, as long as the smile effect is effectively addressed.

5. CONCLUSION

This bachelor's thesis successfully achieved its goal of developing a mobile application for gel electrophoresis image processing. The application was developed for Android 10 using the Kotlin programming language, providing a user-friendly and convenient tool for on-the-go analysis of electrophoresis images.

The results obtained from the testing and evaluation of the mobile application and the implemented algorithms demonstrate their effectiveness and reliability. The lane segmentation algorithm achieved high accuracy across different image quality categories, ensuring accurate identification of lane boundaries. The band detection algorithm showcased excellent precision and sensitivity, effectively detecting bands in the gel images. The molecular weight estimation algorithm exhibited low relative RMSE, providing accurate estimates of molecular weights.

A critical analysis of the achieved results revealed the strengths and limitations of the developed methods. The mobile application showcased its potential as a time-saving and cost-effective solution compared to existing desktop software tools. The contribution of this thesis lies in the development and implementation of robust algorithms for lane segmentation, band detection, and molecular weight estimation within a user-friendly mobile application framework.

Looking ahead, several ideas for future directions can be explored. These include incorporating manual contrast tuning to enhance image quality, implementing automatic cropping of the area of interest to focus on relevant regions, allowing manual adjustment of automatic segmentation and detection for increased flexibility, and providing options to select different segmentation and detection methods based on specific image distortions. Furthermore, image distortion compensation, sample classification, and the export of additional data such as calibration curves and graphs can further enhance the functionality and versatility of the mobile application.

In conclusion, this thesis has made significant contributions to the field of gel electrophoresis image processing by developing a mobile application with robust algorithms and providing insights for future enhancements. The application serves as a valuable tool for researchers and practitioners in the molecular biology and biochemistry domains, offering a portable and efficient solution for analyzing gel electrophoresis images.

REFERENCES

[1]   ALBERTS, Bruce. Molecular biology of the cell. 5th ed. New York: Garland Science, c2008. ISBN 9780815341055.

[2]   ANAM, C. Noise Reduction in CT Images Using a Selective Mean Filter. Journal of Biomedical Physics and Engineering [online]. 2020, 10(5) [cit. 2023-04-27]. ISSN 22517200. Dostupné z: doi:10.31661/jbpe.v0i0.2002-1072

[3]   ARTS, Lukas P. A. a Egon. L. VAN DEN BROEK. The fast continuous wavelet transformation (fCWT) for real-time, high-quality, noise-resistant time–frequency analysis. Nature Computational Science [online]. 2022, 2(1), 47-58 [cit. 2023-04-30]. ISSN 2662-8457. Dostupné z: doi:10.1038/s43588-021-00183-z

[4]   AZIZA, H., K. COULIE, W. RAHAJANDRAIBE a R. VAUCHE. Using short-term fourier transform for particle detection and recognition in a CMOS oscillator-based chain. In: 2018 IEEE 19th Latin-American Test Symposium (LATS) [online]. IEEE, 2018, 2018, s. 1-5 [cit. 2023-04-30]. ISBN 978-1-5386-1472-3. Dostupné z: doi:10.1109/LATW.2018.8349684

[5]   AZQUETA, Amaya, Kristine B. GUTZKOW, Gunnar BRUNBORG a Andrew R COLLINS. Towards a more reliable comet assay: Optimising agarose concentration, unwinding time and electrophoresis conditions. Mutation Research/Genetic Toxicology and Environmental Mutagenesis [online]. 2011, 724(1-2), 41-45 [cit. 2023-01-02]. ISSN 13835718. Dostupné z: doi:10.1016/j.mrgentox.2011.05.010

[6]   BALOCHIAN, Saeed a Hossein BALOOCHIAN. Edge detection on noisy images using Prewitt operator and fractional order differentiation. Multimedia Tools and Applications [online]. 2022, 81(7), 9759-9770 [cit. 2023-04-27]. ISSN 1380-7501. Dostupné z: doi:10.1007/s11042-022-12011-1

[7]   BHATTACHARYYA, Siddhartha, Indrajit PAN, Anirban MUKHERJEE a Paramartha DUTTA, ed. Hybrid Intelligence for Image Analysis and Understanding [online]. Wiley, 2017 [cit. 2023-01-02]. ISBN 9781119242925. Dostupné z: doi:10.1002/9781119242963

[8]   BJELLQVIST, Bengt, Kristina EK, Pier GIORGIO RIGHETTI, Elisabetta GIANAZZA, Angelika GÖRG, Reiner WESTERMEIER a Wilhelm POSTEL. Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. Journal of Biochemical and Biophysical Methods [online]. 1982, 6(4), 317-339 [cit. 2023-01-02]. ISSN 0165022X. Dostupné z: doi:10.1016/0165-022X(82)90013-6.

[9]   BOYER, Rodney. Modern experimental biochemistry (3rd edition). 2000, Pearson Publishing. ISBN 9780805331110.

[10]   DENG, G. a L.W. CAHILL. An adaptive Gaussian filter for noise reduction and edge detection. In: 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference [online]. IEEE, 1993, s. 1615-1619 [cit. 2023-04-27]. ISBN 0-7803-1487-5. Dostupné z: doi:10.1109/NSSMIC.1993.373563

[11]   DIN, Nizam Ud a Ji YU. Training a deep learning model for single-cell segmentation without manual annotation. Scientific Reports [online]. 2021, 11(1) [cit. 2023-04-27]. ISSN 2045-2322. Dostupné z: doi:10.1038/s41598-021-03299-4

[12]   HANAOR, Dorian, Marco MICHELAZZI, Cristina LEONELLI a Charles C. SORRELL. The effects of carboxylic acids on the aqueous dispersion and electrophoretic deposition of ZrO2. Journal of the European Ceramic Society [online]. 2012, 32(1), 235-244 [cit. 2023-01-02]. ISSN 09552219. Dostupné z: doi:10.1016/j.jeurceramsoc.2011.08.015

[13]   Handbook of medical imaging: processing and analysis. Editor I. N. BANKMAN. San Diego: Academic Press, c2000. ISBN 9780120777907.

[14]   KAABOUCH, Naima, Richard R. SCHULTZ, Barry MILAVETZ a Lata BALAKRISHNAN. An analysis system for DNA gel electrophoresis images based on automatic thresholding an enhancement. In: 2007 IEEE International Conference on Electro/Information Technology [online].

IEEE, 2007, 2007, s. 26-31 [cit. 2023-01-02]. ISBN 978-1-4244-0940-2. Dostupné z: doi:10.1109/EIT.2007.4374496

[15] KANOPOULOS, N., N. VASANTHAVADA a R.L. BAKER. Design of an image edge detection filter using the Sobel operator. IEEE Journal of Solid-State Circuits [online]. 23(2), 358-367 [cit. 2023-04-27]. ISSN 00189200. Dostupné z: doi:10.1109/4.996

[16] KEMP, Graham. Capillary Electrophoresis. Biotechnology and Applied Biochemistry [online]. 1998, 27(1), 9-17 [cit. 2023-01-02]. ISSN 08854513. Dostupné z: doi:10.1111/j.1470-8744.1998.tb01369.x

[17] KENNEDY, Hugh L. Improving the frequency response of Savitzky-Golay filters via colored-noise models. Digital Signal Processing [online]. 2020, 102 [cit. 2023-04-30]. ISSN 10512004. Dostupné z: doi:10.1016/j.dsp.2020.102743

[18] KRÁLOVÁ, Blanka. Bioanalytické metody. 3., přeprac. vyd. Praha: Vysoká škola chemicko-technologická, 2001. ISBN 9788070804490.

[19] KRUPKA, O. Klasifikace vzorku 1D gelové elektroforézy. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Ústav biomedicínského inženýrství, 2015. 98 s. Diplomová práce. Vedoucí práce: Ing. Martin Vítek, Ph.D.

[20] LABYED YASSIN, Naima KAABOUCH, Richard R. SCHULTZ a Brij B. SINGH. Automatic segmentation and band detection of protein images based on the standard deviation profile and its derivative. In: 2007 IEEE International Conference on Electro/Information Technology [online]. IEEE, 2007, 2007, s. 577-582 [cit. 2023-01-02]. ISBN 978-1-4244-0940-2. Dostupné z: doi:10.1109/EIT.2007.4374497

[21] LAN, Vo Thi Thuong, Pham Thi Thanh LOAN, Pham Anh Thuy DUONG, Le Thi THANH, Ngo Thi HA a Ta Bich THUAN. Straightforward Procedure for Laboratory Production of DNA Ladder. Journal of Nucleic Acids [online]. 2012, 2012, 1-4 [cit. 2023-01-02]. ISSN 2090-0201. Dostupné z: doi:10.1155/2012/254630

[22] MA, Xiaoju, Bo LI, Ying ZHANG a Ming YAN. The Canny Edge Detection and Its Improvement. In: LEI, Jingsheng, Fu Lee WANG, Hepu DENG a Duoqian MIAO, ed. Artificial Intelligence and Computational Intelligence [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, 2012, s. 50-58 [cit. 2023-04-27]. Lecture Notes in Computer Science. ISBN 978-3-642-33477-1. Dostupné z: doi:10.1007/978-3-642-33478-8_7

[23] MACHADO, A.M.C., M.F.M. CAMPOS, A.M. SIQUEIRA a O.S.F. DE CARVALHO. An iterative algorithm for segmenting lanes in gel electrophoresis images. In: Proceedings X Brazilian Symposium on Computer Graphics and Image Processing [online]. IEEE Comput. Soc, 1997, s. 140-146 [cit. 2023-05-27]. ISBN 0-8186-8102-0. Dostupné z: doi:10.1109/SIGRA.1997.625161

[24] MARKWELL, John. Fundamental laboratory approaches for biochemistry and biotechnology, 2nd edition. Biochemistry and Molecular Biology Education [online]. 2009, 37(5), 317-318 [cit. 2023-01-02]. ISSN 14708175. Dostupné z: doi:10.1002/bmb.20321

[25] MCCORD, B.R. a E. BUEL. Capillary Electrophoresis in Forensic Genetics. In: Encyclopedia of Forensic Sciences [online]. Elsevier, 2013, 2013, s. 394-401 [cit. 2023-01-02]. ISBN 9780123821669. Dostupné z: doi:10.1016/B978-0-12-382165-2.00050-7

[26] MCREYNOLDS, Tom a David BLYTHE. Advanced graphics programming using openGL. San Francisco, CA: Elsevier Morgan Kaufmann Publishers, 2005. ISBN 9781558606593.

[27] MILLS, R. J. Book Review: Electrophoresis of Proteins in Polyacrylamide and Starch Gels. Annals of Clinical Biochemistry: International Journal of Laboratory Medicine [online]. 1972, 9(1-6), 30-31 [cit. 2023-01-02]. ISSN 0004-5632. Dostupné z: doi:10.1177/000456327200900108

[28] MUHARRAM, Magdy M. a Maged S. ABDEL-KADER. Utilization of gel electrophoreses for the quantitative estimation of digestive enzyme papain. Saudi Pharmaceutical Journal [online]. 2017, 25(3), 359-364 [cit. 2023-04-30]. ISSN 13190164. Dostupné z: doi:10.1016/j.jsps.2016.09.002

[29] NERIS, Rômulo Leão Silva, KAUR, Ajuni, GOMES, Aldrin V. Incorrect Molecular Weights due to inaccurate Prestained Protein Molecular Weight Markers that are used for Gel Electrophoresis and Western Blotting. bioRxiv 2020.04.03.023465. Dostupné z: doi: 10.1101/2020.04.03.023465

[30] PARK, Sang Cheol, In Seop NA, Tae Ho HAN, Soo Hyung KIM a Guee Sang LEE. Lane detection and tracking in PCR gel electrophoresis images. Computers and Electronics in Agriculture [online]. 2012, 83, 85-91 [cit. 2023-05-29]. ISSN 01681699. Dostupné z: doi:10.1016/j.compag.2012.01.016

[31] PCR 100 bp Low Ladder. Sigma-Aldrich Co. [online]. 2014 [cit. 2023-01-02]. Dostupné z: http://www.sigmaaldrich.com/catalog/product/sigma/p1473?lang=en&region=CZ

[32] POYNTON, Charles. Digital video and HDTV: algorithms and interfaces. San Francisco: Morgan Kaufmann Publishers, c2003. ISBN 1558607927.

[33] Proceedings X Brazilian Symposium on Computer Graphics and Image Processing. In: Proceedings X Brazilian Symposium on Computer Graphics and Image Processing SIBGRA-97 [online]. IEEE, 1997, 1997, iii- [cit. 2023-01-02]. ISBN 0-8186-8102-0. Dostupné z: doi:10.1109/SIGRA.1997.625134

[34] RAHMAN, Shanto, Md Mostafijur RAHMAN, M. ABDULLAH-AL-WADUD, Golam Dastegir AL-QUADERI a Mohammad SHOYAIB. An adaptive gamma correction for image enhancement. EURASIP Journal on Image and Video Processing [online]. 2016, 2016(1) [cit. 2023-04-30]. ISSN 1687-5281. Dostupné z: doi:10.1186/s13640-016-0138-1

[35] REUSS, F.F. Sur un nouvel effet de l'électricité galvanique. Mémoires de la Société Impériale des Naturalistes de Moscou, 1809, 2(327–37).

[36] RGP: 332 PCR-based genetic markers on rice chromosomes, National Institute of Agrobiological Sciences licensed under CC Attribution-Share Alike 2.1 Japan

[37] SAMBROOK, Joseph a David W. RUSSELL. Molecular cloning: a laboratory manual. 3rd. ed. New York: Cold spring harbor laboratory press, 2001. ISBN 9780879695767.

[38] SARASWATHY, N., RAMALINGAM, P. Concepts and Techniques in Genomics and Proteomics. 2011, Woodhead Publishing, (147-158). ISBN 9781907568107.

[39] SEE, Yew Phew, Peter M. OLLEY a George JACKOWSKI. The effects of high salt concentrations in the samples on molecular weight determination in sodium dodecyl sulfate polyacrylamide gel electrophoresis. Electrophoresis [online]. 1985, 6(8), 382-387 [cit. 2023-01-02]. ISSN 0173-0835. Dostupné z: doi:10.1002/elps.1150060806

[40] SHAPIRO LINDA, G. a STOCKMAN GEORGE C. Computer vision. New Jersey: Prentice-Hall, 2001. ISBN 9780130307965.

[41] SHUKLA, Akshita, Vikrant BHATEJA, Alka SINGH a Akansha Singh RATHORE. An Improved Unsharp Masking (UM) Filter with GL Mask. In: CHOWDARY, P. Satish Rama, Jaume ANGUERA, Suresh Chandra SATAPATHY a Vikrant BHATEJA, ed. Evolution in Signal Processing and Telecommunication Networks [online]. Singapore: Springer Singapore, 2022, 2022-03-24, s. 225-231 [cit. 2023-04-27]. Lecture Notes in Electrical Engineering. ISBN 978-981-16-8553-8. Dostupné z: doi:10.1007/978-981-16-8554-5_22

[42] SCHÄGGER, Hermann. Tricine–SDS-PAGE. Nature Protocols [online]. 2006, 1(1), 16-22 [cit. 2023-01-02]. ISSN 1754-2189. Dostupné z: doi:10.1038/nprot.2006.4

[43] SKUTKOVA, Helena, Martin VITEK, Sona KRIZKOVA, Rene KIZEK a Ivo PROVAZNIK. Preprocessing and classification of electrophoresis gel images using dynamic time warping. International Journal of Electrochemical Science. 2013.

[44] SMITH, I. Zone Electrophoresis : General Principles of Zone Electrophoresis (4th Edition). Butterworth-Heinemann, 1976, (1-15). ISBN 9780815178408.

[45] SMITH, Steven W. Digital signal processing: a practical guide for engineers and scientists. Amsterdam: Newnes, 2003. ISBN 9780750674447.

[46] SOUSA, Antonio V., Rui AGUIAR, Ana Maria MENDONCA a Aurelio CAMPILHO. Automatic Lane and Band Detection in Images of Thin Layer Chromatography. Image Analysis and Recognition. Springer-Verlag Berlin Heidelberg 2004. 2004, s. 158-165.

[47] STRINGER, R. Encyclopedia of Analytical Science: Electrophoresis | Overview (2nd edition). Elsevier, 2005, (356-363). ISBN 9780123693976.

[48] SVOZILOVÁ, V. Předzpracování obrazu 1D gelové elektroforézy. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2013. 50 s. Vedoucí bakalářské práce Ing. Martin Vítek, Ph.D.

[49] TISELIUS, Arne. Reflections from Both Sides of the Counter. Annual Review of Biochemistry [online]. 1968, 37(1), 1-23 [cit. 2023-01-02]. ISSN 0066-4154. Dostupné z: doi:10.1146/annurev.bi.37.070168.000245

[50] TSENG, Din-Chang a You-Ching LEE. Automatic band detection on pulsed-field gel electrophoresis images. Pattern Analysis and Applications [online]. 2015, 18(1), 145-155 [cit. 2023-04-30]. ISSN 1433-7541. Dostupné z: doi:10.1007/s10044-014-0424-4

[51] TSENG, Shiojenn, Jeng-Yang LIN a Jyh-Ping HSU. Theoretical study of temperature influence on the electrophoresis of a pH-regulated polyelectrolyte. Analytica Chimica Acta [online]. 2014, 847, 80-89 [cit. 2023-01-02]. ISSN 00032670. Dostupné z: doi:10.1016/j.aca.2014.08.045

[52] VIOVY, Jean-Louis. Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms. Reviews of Modern Physics [online]. 2000, 72(3), 813-872 [cit. 2023-01-02]. ISSN 0034-6861. Dostupné z: doi:10.1103/RevModPhys.72.813

[53] WANG, Hao-Jen, Chia-Yen LEE, Jhih-Hao LAI, Yeun-Chung CHANG a Chung-Ming CHEN. Image registration method using representative feature detection and iterative coherent spatial mapping for infrared medical images with flat regions. Scientific Reports [online]. 2022, 12(1) [cit. 2023-04-27]. ISSN 2045-2322. Dostupné z: doi:10.1038/s41598-022-11379-2

[54] WILSON, Keith a John WALKER, ed. Principles and Techniques of Biochemistry and Molecular Biology [online]. Cambridge University Press, 2012 [cit. 2023-04-30]. ISBN 9780521535816. Dostupné z: doi:10.1017/CBO9780511813412VIOVY, J.L. Viovy. "Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms," in Reviews of Modern Physics, vol. 3, pp.813-872. 2000. ISSN 0034-6861

[55] ZHU, Mingde, David L. HANSEN, Sam BURD a Frank GANNON. Factors affecting free zone electrophoresis and isoelectric focusing in capillary electrophoresis. Journal of Chromatography A [online]. 1989, 480, 311-319 [cit. 2023-01-02]. ISSN 00219673. Dostupné z: doi:10.1016/S0021-9673(01)84300-4

LIST OF ATTACHMENTS

Attachment A　　List of Electronic Attachments

Zip folder "Mojzisova_Anna_BP_electronic_attachments.zip" contains the following files:

- Video of the application run
  (Mojzisova_Anna_BP_electronic_attachments\application_run.mp4)
- Android project ElectrophoreogramAnalyzer – client side of the application
  (Mojzisova_Anna_BP_electronic_attachments\ElectrophoreogramAnalyzer)
- Python project Analyzer – server side of the application
  (Mojzisova_Anna_BP_electronic_attachments\Analyzer)
- Script for testing image analysis
  (Mojzisova_Anna_BP_electronic_attachments\image_analysis.ipynb)