University of Brasília
Institute of Exact Sciences
Department of Statistics

Master’s Dissertation

Essays on fingerprint data statistical analysis

by

Gabriel Ângelo da Silva Gomes

Brasília, December 13, 2023

Essays on fingerprint data statistical analysis

by

Gabriel Ângelo da Silva Gomes

Dissertation submitted to the Department of Statistics at the University of Brasília, as part of the requirements required to obtain the Master Degree in Statistics.

Advisor: Prof. Dr. Raul Yukihiro Matsushita

Brasília, December 13, 2023

Dissertation submitted to the Graduate Program in Statistics of the Department of Statistics at the University of Brasília, as part of the requirements required to obtain the Master Degree in Statistics.

Approved by:

Prof. Dr. Raul Yukihiro Matsushita

Advisor, EST/UnB

Prof. Dra. Regina Célia Bueno da Fonseca

MAT/IFG

Prof. Dr. Roberto Vila Gabriel

EST/UNB

Prof. Dr. Gladston Luiz da Silva

EST/UnB

In the fear and alarm, you did not desert me, my brothers in arms.

(Dire Straits, "Brothers In Arms")

To my dedicated wife Lorena, whose attributes of love strengthen me in the face of life's hardships. Along with her, my pets, with all their gratitude and unconditional devotion, have achieved some success in the difficult task of convincing me to follow the different paths of happiness that life has offered me.

Acknowledgments

I would like to thank everyone who makes up the select Postgraduate Program in Statistics at UnB - PPGEst/UnB, in particular, my advisor Prof. Raul Yukihiro Matsushita. Due to notable personality characteristics, combined with solid theoretical and practical knowledge, I had the honor of having the best guidance possible.

My most sincere gratitude to my friends Federal Police Fingerprint Experts Lílian Pedrosa Marouelli de Oliveira, Daniel da Silva Carvalho, and Felipe Carlos de Araújo Brito, whose vigor of technical-scientific knowledge and willingness to work hard were fundamental for the studies present here to materialize on a level that, only I, would limit myself to the field of ideas only. Quoting these three friends, I extend my deference to the dedicated forensic identification experts from all over Brazil.

Gratitude to the great statistics companion Pedro Carvalho Brom, whose empathy can only be measured by his intelligence. The partnership in testing neural networks was a pleasant and surprising opportunity.

To the exponents of forensic sciences Christophe Champod, Glenn Langenburg and Clemil José de araújo for their generosity in giving me valuable suggestions in the essays presented here.

Finally, I would like to thank the institutional support of the Brazilian Federal Police.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

This dissertation is organized as a collection of five articles regarding applying statistical tools in fingerprint studies. The first applies convolutional neural networks to fingerprint data for predicting human attributes such as sex, hand types (left or right), and position of fingers (right index finger, for example). The second presents a bibliometric review from 2018 to 2023 of automated minutiae counting initiatives, we noted that most involve convolutional neural networks. The third deals with a statistical analysis of the distribution of Level 2 details concerning levels 1 and 3, in addition to considering sex and type of finger. The fourth suggests an initiative to disseminate 1,000 fingerprints sampled from Brazilians (50 males and 50 females) for ethical, non-profit academic and scientific research. This initiative aims to promote fingerprint identification studies. Finally, the fifth essay suggests Rényi's divergence as an alternative to the traditional chi-square test to evaluate goodness-of-fit, homogeneity, and independence in contingency tables involving rare events. We illustrate this method using fingerprint minutiae data sampled from the Brazilian Federal Police records.

Keywords: Convolutional Neural Networks, statistical analysis, fingerprints minutiae, Brazilian population.

Resumo Expandido

Ensaios sobre análise estatística de dados de impressões digitais

Resumo

O presente resumo expandido é uma síntese dos cinco artigos que compõem esta dissertação, os quais são decorrência de alguma necessidade de ordem prática em que a estatística pôde colaborar. O primeiro ensaio é referente à predição de atributos humanos a partir de Redes Neurais Convolucionais aplicadas às impressões digitais. O segundo trabalho trata de uma revisão bibliométrica que abrangeu o período de 2018 a 2023 em que foram propostos métodos automatizados de contagem de minúcias em impressões digitais. O terceiro artigo é resultado de um estudo estatístico referente à distibuição de fequências das *minutiae* e suas relações com os detalhes de níveis 1 e 3, e também seu comportamento diante do tipo de sexo e dedo. O quarto paper resulta de uma inicativa inédita de disponibilizar uma amostra de impressões digitais representativa da população brasileira e, com isso, espera-se fomentar pesquisas acadêmicas e científicas com propósito ético, não comercial e sem fins lucrativos. Por fim, o quinto estudo trata da aplicação da divergência de Rényi, como uma opção ao teste qui-quadrado, ao realizar testes de hipótese envolvendo contagens menores que cinco de minúcias em impressões digitais.

Palavras-chave: Redes neurais convolucionais, divergência de Rényi, distribuições de frequências das minutiae em impressões digitais, população brasileira.

Introdução

Neste primeiro tóppico destinou-se um parágrafo para introduzir cada um dos cinco ensaios que compõe desta dissertação.

O primeiro texto explora duas categorias de impressões digitais: batidas e roladas. Enquanto as impressões digitais batidas são obtidas ao pressionar o dedo contra uma superfície, as impressões digitais roladas são adquiridas ao rolar o dedo lateralmente, com o objetivo de capturar áreas cruciais como o núcleo e o delta. Ambos os métodos de coleta são realizados de maneira controlada, resultando em impressões digitais de alta qualidade. O enfoque principal do estudo está na previsão do gênero, mão e dedo de origem das impressões digitais, empregando Redes Neurais Convolucionais. A pesquisa envolveu a análise de dois conjuntos de dados: o *Sokoto Coventry Fingerprint Dataset - SOCOFing*, contendo 6.000 impressões digitais batidas de africanos, e uma amostra de 1.401 impressões digitais roladas de brasileiras.

O segundo estudo enfatiza a avaliação das impressões digitais em três níveis distintos dentro do campo das ciências forenses. O primeiro nível aborda as características de fluxo das linhas que formam o campo digital, o segundo nível trata das alterações ao longo dessas linhas, especialmente as minúcias, e o terceiro nível considera atributos dimensionais e morfológicos referentes às cristas de fricção. A atenção principal está voltada para o segundo nível, que se revela crucial para a comparação de impressões digitais. O texto ressalta a importância do emprego de métodos de inteligência artificial na marcação automatizada de minúcias, destacando avanços significativos nos últimos cinco anos. Essa constatação é respaldada por uma revisão bibliométrica conduzida por meio do pacote R bibliometrix, seguindo a abordagem teórica *Consolidated Meta-Analytic Approach Theory - CMAAT*.

No terceiro paper, a pesquisa concentra-se primariamente nos *Level 2 Details - L2D* (minúcias), mas também considara as características de *Level 1 Details - L1D* (tipo fundamental) e *Level 3 Details - L3D* (presença de poros e linhas incipientes), devido à possível correlação estatisticamente significativa entre os três níveis. A distribuição de *L1D* é detalhada, considerando a amostra total ($n = 600$) e os sexos masculino ($n = 300$) e feminino ($n = 300$). A análise estatística indica uma associação entre tipos fundamentais e sexos (p-valor = 0,028) e entre tipos fundamentais e tipos de dedos ($p < 0.001$). Embora não haja evidências contrárias à hipótese de homogeneidade na ocorrência de cicatrizes por sexo (p-valor = 0,074), observa-se falta de homogeneidade em relação ao tipo fundamental (p-valor = 0,032) e tipo de dedo (p-valor = 0,012). A presença de linhas incipientes mostra não homogeneidade por sexo (p-valor ¡ 0.001) e tipo fundamental (p-valor = 0,001), mas foi possível descartar a homogeneidade em relação ao tipo de dedo (p-valor = 0,187). A ocorrência de poros não é homogênea por sexo (p-valor = 0,001) e tipo fundamental (p-valor = 0,040), mas é não houve elementos contrários à homogêneidade em relação ao tipo de dedo (p-valor = 0,621). A análise incluiu o cálculo da média de pontos característicos por impressão digital em relação ao sexo, tipo fundamental e tipo de dedo. Os resultados indicam que o sexo masculino apresenta uma média de pontos característicos maior. A ordem decrescente das médias para tipos fundamentais é verticilo, loops (direito e esquerdo) e arco. Em relação aos dedos, a ordem de médias, do maior para o menor, é polegares, anulares, médios, indicadores e mínimos, com a observação interessante de que os dedos indicadores de ambas as mãos têm a mesma média.

No quarto ensaio, destaca-se a importância de disponibilizar uma amostra de impressões digitais brasileiras para incentivar estudos acadêmicos sobre a população do Brasil. Acredita-se que essa iniciativa trará benefícios significativos para a identificação humana, especialmente em casos críticos, como a identificação de cadáveres desconhecidos e locais de crimes. O objetivo é disponibilizar o BRAFing em um repositório científico renomado, preferencialmente associado a periódicos de alta classificação. A base de dados proposta é considerada de alto valor para pesquisas acadêmicas, especialmente nas ciências forenses, devido aos critérios de seleção, à dimensão continental do Brasil e à riqueza antropológica de sua população. O BRAFing é visto como uma contribuição significativa que preencherá uma lacuna na comunidade científica. Importante notar que o Comitê de Ética em Pesquisa da Faculdade de Medicina da UnB autorizou, em 30/11/2023, o avanço da pesquisa que resultará no BRAFing (projeto CAAE nº 73627423.3.0000.5558).

No quinto e último trabalho, é abordado o teste qui-quadrado (χ^2) como uma ferramenta estatística convencional para avaliar a concordância entre frequências observadas e esperadas em dados categóricos, com ênfase nos desafios associados a frequências esperadas reduzidas. Para contornar esse problema, são examinadas abordagens como simulações de Monte Carlo e bootstrapping, além da exploração de alternativas assintóticas ao χ^2. Uma alternativa considerada é a divergência de Kullback-Leibler, que está relacionada ao teste de razão de verossimilhança generalizada. O texto também introduz a (h, ϕ)-divergência, abarcando diversas métricas teóricas da informação. O ponto central da pesquisa é a aplicação da divergência de Rényi em testes de hipóteses para tabelas de contingência. Destaca-se a característica distintiva da divergência de Rényi de evitar divisões por frequências, especialmente em situações de frequências baixas. O estudo concentra-se no cenário em que $0 < \alpha < 1$, sendo α o índice da divergência de Rényi. A pesquisa explora valores específicos de α para assegurar que a estatística proposta siga uma distribuição qui-quadrado ao lidar com frequências esperadas reduzidas.

Metodologia

Na sequência, serão descritas uma breve síntese da metodologia utilizada em cada um dos cinco ensaios.

No primeiro texto, as imagens de impressões digitais, inicialmente convertidas para tons de cinza e com dimensões de 64×64 pixels, foram submetidas à normalização e divididas em conjuntos de treinamento (70%) e teste (30%). A arquitetura da *Convolutional Neural Network - CNN* é composta por duas camadas convolucionais seguidas por camadas de pooling máximo e dropout, sendo as saídas são processadas por camadas densas. A camada de saída utiliza a função Softmax para classificação, com duas unidades para sexo e mão, e cinco unidades para dedos. Os modelos são compilados usando o otimizador Adam *(Adaptive Moment Estimation)*, com uma taxa de aprendizado de 0,001 e entropia cruzada binária como função de perda. A métrica de avaliação escolhida é a AUC-ROC *(Area Under the ROC Curve)* média, visando reduzir falsos positivos. A validação do modelo é conduzida utilizando 50% dos dados de treinamento, e a qualidade preditiva é avaliada no conjunto de teste. Durante o treinamento, a função de perda é monitorada, interrompendo o processo em caso de falta de melhora após um número pré-determinado de épocas e restaurando os pesos para garantir o melhor resultado.

No segundo estudo, foram empregados recursos do pacote Bibliometrix no ambiente R, juntamente com a metodologia *Consolidated Meta-Analytic Approach Theory - CMAAT*, em uma pesquisa exploratória. O Bibliometrix é um pacote de código aberto no R destinado a análises bibliométricas, enquanto o *CMAAT* é um método integrativo de revisão sistemática. A fonte de metadados escolhida para a pesquisa foi a Scopus, e a busca por artigos foi realizada em 27 de abril de 2023, abrangendo o período de 2018 a 2023, utilizando as palavras-chave *"fingerprint AND minutiae"*.

O terceiro artigo trata de uma amostra composta por 600 impressões digitais de 60 indivíduos (30 homens e 30 mulheres, maiores de 18 anos) igualmente distribuídos entre as cinco regiões geográficas do Brasil. Essas impressões foram coletadas pela Polícia Federal empregando o método rolado e a seleção seguiu critérios elevados de qualidade. Foram analisados os três níveis de características de uma impressão digital, o nível 2 foi o foco principal do trabalho e considerou o mais abrangente de rol de minúcias (54 minúcias e suas variações) já considerado em um estudo até então.

No quarto ensaio, está prevista a criação de um banco de dados contendo 1.000 impressões digitais de 100 brasileiros, coletadas entre 1990 e 2004 por agentes de segurança pública, utilizando o método rolado. A amostra é composta por 50 homens e 50 mulheres, distribuídos igualmente entre as cinco regiões do Brasil. As impressões foram selecionadas a partir de registros decadactilares, priorizando aquelas sem cicatrizes capazes de comprometer a qualidade dos datilogramas. Após digitalização a 500 dpi e avaliação de qualidade pelo NIST, apenas aquelas com índice máximo (NFIQ 1) foram mantidas. Cada impressão foi renomeada para preservar sua origem anônima, e sua qualidade foi certificada por pelo menos dois especialistas com 10 anos de experiência.

Por fim, o quinto trabalho explora a relação entre as divergências de Rényi e a estatística qui-quadrado, utilizando o método delta para auxiliar nas derivações de grandes amostras. Uma seção específica do estudo emprega uma abordagem de Monte Carlo para fundamentar as descobertas teóricas. O estudo examina dois exemplos nos quais as frequências absolutas esperadas são inferiores a cinco, avaliando a conformidade com as propriedades do qui-quadrado para diferentes tamanhos de amostra em uma variedade de valores α. Além disso, compara as funções de potência empíricas para alguns valores de α usando valores críticos assintóticos. O texto também faz menção a uma aplicação prática na análise de uma amostra de dados de impressões digitais da polícia brasileira, onde são testadas diferenças na distribuição de minúcias entre os gêneros, com foco em tipos de minúcias raramente observados na população.

Resultados e discussões

A seguir serão resumidos e discutidos os principais resultados dos 5 ensaios.

No primeiro texto, o modelo proposto exibiu um desempenho de previsão geralmente sólido e, em alguns casos, excepcional ao estimar o gênero, mão e dedo com base em imagens de impressões digitais. Os resultados superaram um estudo anterior realizado no conjunto de dados SOCOFing, que utilizava uma arquitetura de Convolutional Neural Network (CNN) baseada em ResNet. Ambos os conjuntos de dados permitiram uma previsão de gênero com uma precisão superior a 98%, e a previsão da mão foi bem-sucedida em mais de 90% dos casos. Embora a previsão dos dedos tenha alcançado precisões acima de 92%, surgiram desafios no conjunto de dados brasileiro, especialmente para os dedos médio e anelar, que foram confundidos entre si em pouco mais de 55% dos casos. A discrepância notável nas dimensões das impressões digitais roladas entre os conjuntos de dados pode explicar o desempenho notável na previsão do polegar e dedo mínimo. Ao mesmo tempo, dimensões semelhantes para os dedos médio e anelar influenciaram a classificação, destacando a importância do fluxo das linhas que definem os detalhes da impressão digital no Nível 1. A análise da matriz de confusão revelou a maioria dos erros na classificação foram entre os dedos médio e anelar.

No segundo estudo, foram identificados 696 documentos provenientes de 382 fontes no período de 2018 a 2023. A idade média desses documentos foi de 3,11 anos, com uma média de 4.963 citações por artigo. Dos 1.692 autores, 14,7% colaboraram internacionalmente, e 42 artigos foram de autoria única. Os cinco principais periódicos, que representam 56,9% dos artigos, incluem *Advances in Intelligent Systems and Computing* (AISC), *Lecture Notes in Computer Science* (LNCS), *Multimedia Tools and Applications* (MTA), *ACM International Conference Proceeding Series* (ACM ICPS), e *IEEE Access*. Houve um notável aumento nas publicações sobre o tema nas cinco principais revistas científicas de 2018 a 2023.

No terceiro paper, o estudo se concentra principalmente no L2D (*minutiae*), mas também considera as características de L1D (tipo fundamental) e L3D (presença de poros e linhas incipientes), devido à possível relação estatisticamente significativa entre os três níveis. A distribuição de L1D é descrita na Tabela 1, considerando a amostra total ($n = 600$) e os sexos masculino ($n = 300$) e feminino ($n = 300$). A análise estatística indica dependência entre tipos fundamentais e sexos (p-valor = 0,028) e entre tipos fundamentais e tipos de dedos (p-valor = 0,001). Não há evidências contrárias à hipótese de homogeneidade na ocorrência de cicatrizes por sexo (p-valor = 0,074), mas a não homogeneidade é observada diante do tipo fundamental (p-valor = 0,032) e tipo de dedo (p-valor = 0,012). A incidência de linhas incipientes não é homogênea por sexo (p-valor = 0,001) e tipo fundamental (p-valor = 0,001), mas é homogênea quanto ao tipo de dedo (p-valor = 0,187). A ocorrência de poros não é homogênea por sexo (p-valor = 0,001) e tipo fundamental (p-valor = 0,040), mas é homogênea quanto ao tipo de dedo (p-valor = 0,621). A análise incluiu o cálculo do número médio de pontos característicos por impressão digital em relação ao sexo, tipo fundamental e tipo de dedo. Os resultados mostram que o sexo masculino tem uma média de pontos característicos maior, enquanto a ordem decrescente de médias para tipos fundamentais é verticilo, presilhas (direito e esquerdo) e arco. Em relação aos dedos, a ordem de médias, do maior para o menor, é polegares, anelares, médios, indicadores e mínimos, com uma observação interessante de que os dedos indicadores de ambas as mãos apresentaram a mesma média.

No quarto ensaio destaca-se a importância de disponibilizar uma amostra de impressões digitais de brasileiros, denominada BRAFing, para incentivar estudos acadêmicos sobre a população do Brasil. Acredita-se que essa iniciativa trará benefícios significativos para a identificação humana, principalmente em casos críticos, a exemplo de identificação de cadáveres desconhecidos e fragmentos de datilogramas revelados em locais de crimes. O objetivo é tornar o BRAFing disponível em um repositório científico renomado, preferencialmente associado a periódicos de alta classificação. A base de dados proposta é considerada de alto valor para pesquisas acadêmicas, especialmente nas ciências forenses, devido aos critérios de seleção, à dimensão continental do Brasil e à riqueza antropológica da sua população. O BRAFing é visto como uma contribuição significativa que preencherá uma lacuna na comunidade científica. Importante frisar que o Comitê de Ética em pesquisa da Faculdade de Medicina da UnB autorizou, no dia 30/11/2023, o progresso da pesquisa que terá como resulado a BRAFing.

Por fim, o quinto trabalho conduziu experimentos de Monte Carlo para validar a estatística proposta pela equação (6.9) e para avaliar o desempenho de potência, considerando dois exemplos com frequências absolutas esperadas abaixo de cinco. As simulações abrangeram tamanhos de amostra aleatórios e testes de ajuste e homogeneidade, totalizando 500 simulações para cada cenário, explorando valores de α de $0,1$ a $0,9$. As distribuições empíricas foram comparadas com a distribuição teórica do χ^2 por meio de gráficos QQ e testes de Kolmogorov-Smirnov.

A escolha da estatística de Rényi foi motivada pela possibilidade de frequências absolutas esperadas abaixo de cinco para H_0, o que tornaria o resultado assintótico da estatística do qui-quadrado potencialmente não confiável. Realizaram-se replicacões de Monte Carlo com diferentes tamanhos de amostra, variando α de $0,1$ a $0,9$, e testando π_0 em diferentes valores. Os testes de Kolmogorov-Smirnov indicaram um ajuste razoável ao qui-quadrado esperado para a estatística sugerida em alguns cenários, mesmo quando as frequências absolutas esperadas eram menores que cinco. As funções de potência empíricas para diferentes valores de α não apresentaram diferenças substanciais no poder do teste em um cenário específico com $n = 2.000$.

Este ensaio final também aborda a identificação de impressões digitais, uma área essencial na ciência forense, focando nas minúcias. Utilizando uma amostra de 7.031 minúcias de um registro policial brasileiro (3.702 de homens e 3.329 de mulheres), a análise revelou frequências esperadas inferiores a cinco em células da tabela de contingência. O teste de homogeneidade, baseado na estatística $R_{\alpha,n,m}$, obteve resultados significativos para $\alpha = 0,1$ ($50,47$, p-valor = $0,0019$) e $\alpha = 0,2$ ($50,69$, p-valor = $0,0018$). Uma simulação de Monte Carlo validou esses resultados, não rejeitando a hipótese nula devido à falta de robustez devido às contagens pequenas. O estudo confirma diferenças sutis nos padrões minuciosos entre homens e mulheres, sugerindo um potencial aplicação na determinação de gênero em investigações forenses.

Conclusões

As conclusões decorrentes dos cinco papers são sumarizadas nos parágrafos seguintes.

O primeiro texto aborda a aplicação recente de Redes Neurais Convolucionais (CNNs) no estudo de impressões digitais, que são tradicionalmente reconhecidas como um método eficaz de identificação humana. O foco do trabalho está na previsão de atributos, como sexo, posição da mão e dedos, a partir de datilogramas. Foram utilizados dois conjuntos de dados, um contendo impressões digitais batidas e outro com impressões digitais roladas. Os resultados indicam taxas de precisão notáveis na determinação de sexo, mão e dedo, alcançando médias de 100%, 98,28% e 96,34%, respectivamente, para as impressões digitais batidas do conjunto SOCOFing. No conjunto de dados brasileiro (datilogramas rolados), as médias de precisão foram de 99,10%, 94,32% e 77,49%, para a determinação de sexo, mão e dedo. O estudo conclui que as CNNs podem ser eficazes na previsão de atributos a partir de impressões digitais, proporcionando utilidade na filtragem de grandes conjuntos de dados e otimizando o trabalho de especialistas ou sistemas automatizados.

No segundo estudo, há uma abordagem ao crescente interesse na análise automatizada de características biométricas, com ênfase no uso de redes neurais profundas para a contagem de minúcias em impressões digitais. Embora experimentos nesse domínio tenham apresentado resultados promissores para classificações precisas, existe uma lacuna na literatura em relação à identificação e totalização de todas as minúcias por tipo usando inteligência artificial. O texto sugere que esta é uma área promissora para futuras pesquisas, destacando desafios como a necessidade de dados mais abertos e marcados, a ausência de um padrão internacional para a contagem de minúcias e a falta de um sistema codificado para esse processo. A aplicação do aprendizado profundo para identificar padrões complexos em impressões digitais é ressaltada como uma contribuição significativa para o avanço da ciência forense.

O terceiro artigo conduziu a primeira análise estatística das distribuições de minúcias na população brasileira, explorando as interconexões entre diferentes níveis de detalhes nas impressões digitais (L2D, L1D, L3D) e variáveis como sexo e tipo de dedo. Apesar da falta de padronização oficial dos tipos de minúcias, o estudo identificou 54 tipos com base em observações anteriores. A amostra diversificada, representando todas as regiões do Brasil, possibilita inferências sobre o comportamento das minúcias na população brasileira. Os resultados detalhados das distribuições por sexo, tipo de dedo e padrão fundamental contribuem para uma compreensão holística da análise de impressões digitais no Brasil. O estudo sugere que pesquisas futuras podem explorar a relação entre a orientação de minúcias, o tipo fundamental e o tipo de dedo, além de abordar a probabilidade de incidência em áreas específicas das impressões digitais.

O quarto ensaio representou o passo inicial para a materialização da BRAFing. As etapas subsequentes dependiam da autorização do Comitê de Ética em pesquisa da Faculdade de Medicina da UnB, o que foi obtido em 30/11/2023 (projeto CAAE nº 73627423.3.0000.5558). Portanto, com o término das pesquisas, a expectativa é disponibilizar a BRAFing em um renomado repositório científico, preferencialmente associado a periódicos de alta classificação. Isso possibilitará o acesso a uma base de impressões digitais inédita e selecionada de brasileiros para pesquisas científicas éticas e sem fins lucrativos.

Finalmente, o quinto estudo investigou o uso da divergência de Rényi como uma alternativa ao teste qui-quadrado convencional para avaliar a qualidade do ajuste, homogeneidade e independência em tabelas de contingência. Ficou evidente que a estatística da divergência de Rényi se aproxima de maneira eficaz de uma distribuição qui-quadrado, mesmo em cenários com contagens de frequência esperada reduzidas. O estudo empregou análise de Monte Carlo em dois exemplos com baixas frequências absolutas esperadas, destacando que um índice de Rényi α reduzido aprimora a concordância da estatística proposta com a distribuição qui-quadrado, especialmente em situações de baixas frequências esperadas. Além disso, a metodologia foi aplicada a dados reais de impressões digitais da polícia brasileira para identificar disparidades de gênero na distribuição de minúcias, incluindo aquelas raramente observadas na população em geral. Uma simulação de Monte Carlo confirmou os resultados, indicando que as percentagens de realizações simuladas que excedem o valor crítico assintótico estão alinhadas com o nível nominal de 5%, especialmente com um α reduzido (inferior a $0,2$).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In Brazil, the role of forensic experts is still rare and, consequently, there are few scientific productions that study Brazilians' fingerprints for identification purposes. The four articles that make up this dissertation are the result of research that sought to fill gaps in knowledge related to the behavior and characteristics of Brazilian fingerprints.

The first paper delves into applying Convolutional Neural Networks (CNNs) to predict specific human attributes from fingerprints, namely gender, hand, and finger position. For the first time, a study applies CNNs to plain and rolled fingerprints, utilizing two distinct datasets, the Sokoto Coventry Fingerprint Dataset (SOCOFing) and a sample of 1401 Brazilian fingerprints, this research demonstrates the efficacy of CNNs in this domain. We tested our proposed CNN architecture, yielding impressive accuracies: 100% for gender, 98.28% for hand, and 96.34% for finger position using the SOCOFing dataset. The Brazilian dataset also showed promising results with 99.10% accuracy for gender. However, differences in prediction accuracies were observed between plain (SOCOFing) and rolled Brazilian fingerprints, highlighting the nuances in data representation. This study underscores the potential of deep learning in enhancing the efficiency of fingerprint-based identification systems, offering a robust method to filter and reduce database sizes, thereby optimizing processing efforts.

The second paper reviews recent literature on fingerprints and minutiae, focusing on the principal methodologies used, most of which are convolutional neural networks. The objective is to analyze the evolution of traditional minutiae counting approaches and identify possible gaps in current research. The methodology used was bibliometric analysis, considering the scientific production between 2018 and 2023. The results indicate increased use of deep neural networks for automatic minutiae detection.

The third paper offers a comprehensive analysis conducted on a sample of 600 fingerprints from 60 individuals (30 males and 30 females), equally distributed across Brazil's five geographic regions. This research is the first of its kind whose research object was the fingerprints of Brazilians. The study primarily focused on L2D minutiae but also considered relationships with L1D and L3D variables and the influence of sex and finger type. Statistical tests revealed dependencies between fundamental fingerprint types and sex and between fundamental types and finger types. Notably, males presented a higher average of characteristic points than females. When considering fundamental types, the order of average minutiae was whorl, loops (right and left), and arch. Thumbs had the highest average minutiae for individual fingers, followed by ring fingers, middle fingers, index fingers, and little fingers. The study's results contribute significantly to understanding the distribution and occurrence of minutiae in the fingerprints of the Brazilian population, highlighting the influence of various factors on these distributions.

The fourth article innovates by proposing the creation of BRAFing (Brazilian Fingerprints Dataset), a biometric fingerprint database for academic and scientific research. BRAFing will consist of 1,000 fingerprint images from 100 Brazilians distributed equally across the five geographic regions of Brazil. The anonymized sample will only contain characteristics of gender, hands, and fingers, without informing the names of biometric donors. The database will be made available free of charge with targeted access to non-profit research and a renowned scientific repository. The project initiative was the result of research within the scope of the Postgraduate Program in Statistics at UnB.

Finally, the discussion centers on the chi-square test (χ^2) as a traditional statistical tool employed to assess the concordance between observed and expected frequencies in categorical data. The emphasis is placed on the challenges arising from diminished expected frequencies. Various strategies, including the examination of Monte Carlo simulations and bootstrapping, are investigated to address this issue. Additionally, the study explores asymptotic alternatives to χ^2. Another considered alternative is the Kullback-Leibler divergence, which is linked to the generalized likelihood ratio test. The (h, ϕ)-divergence is also introduced, encompassing multiple information-theoretical metrics.

The focal point of the investigation is the utilization of Rényi divergence in the context of hypothesis testing for contingency tables. A noteworthy characteristic of Rényi divergence is its ability to circumvent frequency divisions, particularly evident in scenarios with low frequencies. The research hones in on situations where $0 < \alpha < 1$, denoting the Rényi divergence index. Specific values of α are explored to ensure that the proposed statistic adheres to a chi-square distribution, especially when confronting reduced expected frequencies.

Chapter 2

Predicting sex, hand and finger position by applying convolutional neural networks to plain and rolled fingerprints

2.1 Introduction

The conscious use of the fingerprint as a unique and personal nature of the raised patterns on the fingers dates back to ancient times. Archaeologists have found records of fingerprints as seals affixed to legal contracts in Babylonia (1855–1913 BC). Contemporary references to the code of Hammurabi (1792-1750 BC) point out that law enforcement officers were authorized to collect the fingerprints of prisoners. Hypotheses for using typescripts by ancient civilizations are for identification, decoration, symbolism, and commerce Ashbaugh, 1999a. The Chinese civilization in the Qin Dynasty (221–206 BC) was the first to use fingerprints with the proven purpose of human identification, describing how to employ epidermal prints as forensic identification evidence (Barnes, 2011a).

The scientific foundations of human identification using fingerprints date from the end of the modern age to the beginning of the contemporary one. The uniqueness of fingerprints was clearly expressed in 1788 by J. C. Mayer, a German physician, and anatomist, in his illustrated atlas of anatomy. The invariability of fingerprints is generally a discovery attributed to Sir William Herschel in 1877 (Cummins and Midlo, 1943), and Francis Galton established the basic principles of uniqueness and permanence (Galton, 1892). In 1877, Thomas Taylor and three years later, Henry Faulds and William Herschel proposed fingerprints as a crime-solving tool (Ashbaugh, 1999a; Cummins and Midlo, 1943).

As an identification resource consecrated for over 140 years, dactylograms relate with specific attributes of those who produced them, as dermatoglyphic features statistically vary between the sexes, ethnic groups, and age categories (Cummins, Waits, and McQuitty, 1941).

This work considers two types of fingerprints: plain and rolled. The plain type is collected by pressing the finger down on a card or placing the finger flat on a scanner (Figure 2.1, left). The rolled fingerprint is taken by rolling the finger from one side of the nail to the other (namely, nail-to-nail or N2N) on a card or inside a platen scanner (Figure 2.1, right), this technique aims to capture as many important areas of a fingerprint as possible, such as the core and delta. As both types are taken in a controlled mode, they typically show good-quality prints with many detailed and particular features.

In particular, the present work focuses on predicting the gender, hand, and finger that originated the fingerprints using deep Convolutional Neural Networks (CNNs). Two datasets were analyzed: the Sokoto Coventry Fingerprint Dataset (SOCOFing) (Shehu et al., 2018d; Shehu et al., 2018c) containing 6000 plain fingerprints and a sample of 1,401 Brazilian nail-to-nail fingerprints. The bases will be described in due course.

2.2 Related Works

There are some ways to infer the gender, hand, and even finger of the person who originated a fingerprint from its characteristics. Pioneering studies of fingerprints focused on manually counting their morphological characteristics, which include the number (or density) of minutiae, the ratio of ridge thickness to trough thickness, ridge width, and fingerprint pattern. Most research has inferred that female fingerprints contain more minutiae in a given area (Ramanjit and Garg, 2011). Analyzes show that females tend to have significantly higher ridge density than males in the distal (radial and ulnar) region of all ten digits in diverse populations across the planet, including Spanish Caucasians (Gutiérrez-Redomero et al., 2008), Argentinians from the regions of Puna-Quebrada and Ramal (Gutiérrez-Redomero et al., 2013), northern Malaysians (Abdullah, Rahman, and Abas, 2015) and Chinese Han youth (Xie and Lin, 2020). Another type of work showed that the digital ridge count test and its relation to fingertip size (Gnanasivam and Vijayarajan, 2019) achieved significant results for gender classification with a maximum success rate of 88.41% and a success rate of 90.11% for the right-hand ring finger.

It is possible to relate the hand based on the characteristic outline of the fingerprints of each finger, the direction of the ridges that form the marginal system of the fingerprints, the direction of the ridges close to the interphalangeal folds, the finger position when placing sequential prints on objects, the position of deltas, and the position of the nuclei, especially for whorls in which spiral nuclei to the right/left are more common in the left/right hand (Ayala, 1981). Core in "S" are more common in the left hand and "Z" in the right. Empirical studies along the same lines are present by Caballero (2012) and detailed analysis of whorl nuclei angles were done by (Kapoor and Badiye, 2015; Singh, Chattopadhyay, and Garg, 2005).

Combining Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD) for gender classification (Gnanasivam and Muttan, 2012) analyzed 3,570 fingerprints (1980 from males and 1590 from females). They obtained an overall classification rate of 84,69%, with a rate of 94.32% for the left-hand little fingers of female persons and 95.46% for the left-hand index finger of male persons. Still using DWT and SVD, Shinde and Annadate, 2015 considered a database of 1000 fingerprints (500 from males and 500 from females). They obtained a success rate of 82.60% for the left-hand little fingers of females and 82.90% for the left-hand index finger of males. For any other finger of male/female in their study, they get rates of 80.40% and 76.84%, respectively. In the study using DWT and Multilayer Perceptron (MLP) networks techniques for analyzing the NIST (National Institute of Standart and Technology) dataset with 750 fingerprints (375 males and 375 females), Suwarno (2023) observed an overall gender classification rate of 80.1%. The Adaptive Neuro-Fuzzy Inference System (ANFIS) is also a possible model to classify the gender from a given fingerprint image, providing better accuracy than using NN and Fuzzy separately (Sahu, Rao, and Mishra, 2015).

More recently, the CNNs have shown promising results for classifying gender and finger position based on fingerprint characteristics (Hsiao et al., 2022). In order to better investigate the performance of the CNN method, we applied it to two different data sets covering two different populations in this work. The first is the Sokoto Coventry Fingerprint Dataset (SOCOFing) (Shehu et al., 2018a; Iloanusi and Ejiogu, 2020; Shehu et al., 2018d; Shehu et al., 2018c), and the second is a sample of Brazilian fingerprints from different geographic regions of Brazil (Silva Carvalho et al., 2022).

The determination of gender, hands and fingers is the scope of this work, and for this two sets of data were used. The first is the Sokoto Coventry Fingerprint Dataset (SOCO-Fing) and our results are related to two other studies that also used SOCOFing (Shehu et al., 2018a; Iloanusi and Ejiogu, 2020; Shehu et al., 2018d; Shehu et al., 2018c). The second data set refers to a sample of fingerprints collected from Brazilians from all geographic regions (Silva Carvalho et al., 2022).

2.3 Dataset

The SOCOFing dataset (Shehu et al., 2018d; Shehu et al., 2018c) is freely available at `https://www.kaggle.com/datasets/ruizgara/socofing`. The resolution of all image files is $1 \times 96 \times 103$ (gray $\times$ width $\times$ height), as illustrated in Fig. 1 (left). The SOCOFing comes from the ten fingers of 600 adult Africans (477 males and 123 females). Thus, this database contains 6000 plain fingerprints: 600 for each finger position, 300 for each hand.

The second database is from a study where the gender is predicted from the ridge density

Figure 2.1: Plain (left) and Rolled (right) fingerpints n.d.(b).

of Brazilian N2N fingerprints (Silva Carvalho et al., 2022). The original database consists of 2250 datagrams collected from 10 fingers of 225 adult Brazilians (109 men and 116 women) from all geographic regions of Brazil. As these images are from digitized paper cards, some background noise hindered the application of CNNs (Fig. 1, right). Thus, we screened these images, resulting in a subsample of 1401 suitable fingerprints, 605 male and 796 female (Tab. 4.6). The biometry images of Brazilians are in grayscale with 500 DPI resolution and NIST Fingerprint Image Quality - NFIQ level 1.

Table 2.1: Brazilian dataset: Summary of finger positions by gender

Finger Position	Male (n=109)	Female (n= 116)
Right Thumb	84	102
Right Index	57	79
Right Middle	68	85
Right Ring	63	82
Right Little	14	34
Left Thumb	95	109
Left Index	65	88
Left Middle	69	95
Left Ring	68	88
Left Little	22	34
Total	605	796

2.4 Metodology

We loaded the images in grayscale and converted them to 64×64 arrays. After normalization, we split the database into training (70%) and testing (30%) sets. The proposed architecture for gender, finger, and hand classification based on fingerprints comprises a Convolutional Neural Network (CNN) algorithm. The input to the network is an image of 64×64 pixels. Initially, the network implements a convolutional layer with 32 filters of size 5×5, using the ReLU activation function and padding the same to maintain the image's original dimensions. This layer is followed by a max pooling layer with pool size 2×2 and then a dropout layer with a rate of 50%. Then, the process is repeated with another convolutional layer of 32 filters of size 5×5 and ReLU activation, followed by a Max Pooling layer and a Dropout Layer. The outputs from these layers are flattened and passed through three consecutive dense layers with 128, 64, and 32 units, respectively, all using the ReLU activation function. In the end, a dense output layer uses the softmax activation function, with two units for classifying gender and hand and five for finger classification (Fig. 2.2).

We compiled the gender and hand classifier models using the Adam optimizer with a learning rate of 0.001 and the binary cross entropy as the loss function. We applied the average AUC (Area Under the ROC Curve) as the evaluation metric to reduce the false positives.

We validated the model using 50% of the training data and assessed the model's predictive quality over the (shuffled) testing data set. The algorithm monitored the loss function, stopping the training if there was no improvement after a predetermined amount of epochs and restoring the weights that ensured the best result.

2.5 Results and Discussion

The test results of our model are in Tables 2.2 – 2.4 (SOCOFing dataset) and Tables 2.5 – 2.7 (Brazilian dataset). In general, they show that our proposed scheme provides a good, and eventually excellent, predictive performance to predict gender, hand and finger from a fingerprint

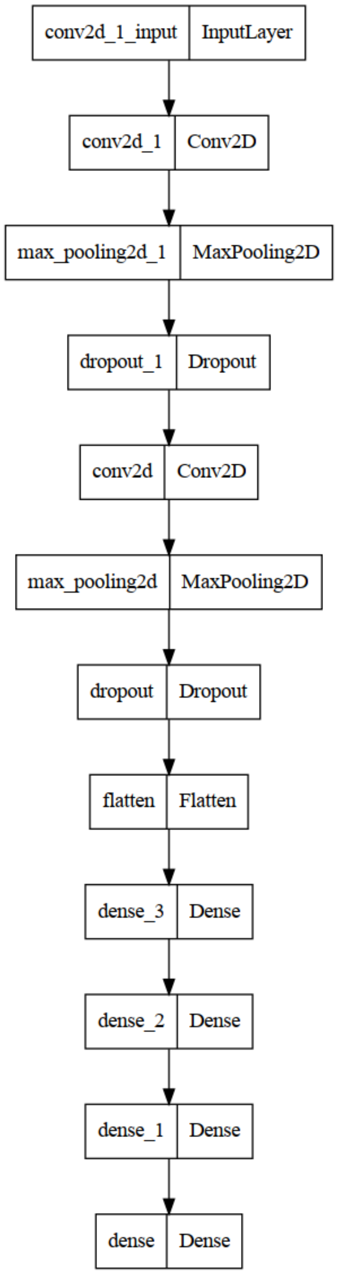

Figure 2.2: Our convolutional neural network architecture.

image. Our approach overperformed the previous study on the SOCOFing dataset that used a ResNet-based CNN architecture (Shehu et al., 2018a).

For both datasets, the gender can be predicted from an image almost perfectly, with accuracies superior to 98% (Tables 2.2 and 2.5).

Presenting lower accuracies but still expressive, the hand that generated a fingerprint can be predicted based on its image in at least 90% of the cases.

Regarding the finger prediction, although our model overperformed the previous study Shehu et al., 2018a with accuracies over 92%, the results for the Brazilian dataset were expressive for the thumb, index, and little. But for the middle and ring were only suitable. Table 2.5 presents accuracies superior to 77% for these fingers.

The middle and ring of the Brazilian dataset were almost randomly confused with each other, with accuracies of just over 55%. To seek an answer to this misclassification, recall that the SOCOFing and Brazilian datasets have essential differences. While the SOCOFing dataset consists of 6000 plain fingerprints of the African population, equally distributed between genders and fingers, the Brazilian dataset comprises 1401 rolled fingerprints of Brazilians distributed according to Table 4.6.

The notable extremes of dimensions between their rolled fingerprints may explain the exceptional performance of 100% in predicting an image as a thumb or little finger. Thus, the size differences may have facilitated their classifications.

In contrast, rolled fingerprints from the middle and ring exhibit similar dimensions Araújo and Pasquali, 2012, which may influence the classification based on the flow of the lines that define the Level 1 fingerprint details.

This hypothesis is corroborated by observing the confusion matrix in Table 2.5, where 76.92% of the times that CNN misclassified the middle finger was for classifying it as the ring finger. Similarly, 77.78% of CNN's ring finger classification errors were in its classification as a middle.

Table 2.2: SOCOFing dataset: Confusion matrix for the gender prediction

Gender	Male	Female	rate (%)
Female	369	0	100
Male	0	1431	100

Note: Shehu et al. (2018a) showed 73.98% (female) and 76.42% (male)

Table 2.3: SOCOFing dataset: Confusion matrix for the hand prediction

Hand	Left	Right	rate (%)
Left	881	12	98.66
Right	19	888	97.90

Note: Shehu et al. (2018a) showed 94.44% (left) and 92.56% (right)

Table 2.4: SOCOFing dataset: Confusion matrix for the finger prediction

Finger	Thumb	Index	Middle	Ring	Little	rate (%)
Thumb	355	0	0	0	0	100
Index	0	340	1	1	4	98.27
Middle	1	8	344	5	14	92.47
Ring	0	1	5	354	1	98.06
Little	4	11	10	0	341	93.17

Note: Shehu et al. (2018a) showed 95.84% (thumb), 77.78% (index), 72.50% (middle), 66.12% (ring), 71.39% (little)

Table 2.5: Brazilian Dataset: Confusion matrix for the gender prediction

Gender	Male	Female	rate (%)
Female	223	0	100
Male	3	163	98.19

Table 2.6: Brazilian Dataset: Confusion matrix for the hand prediction

Hand	Left	Right	rate (%)
Left	153	3	98.07
Right	22	211	90.56

2.6 Conclusions

Established over a hundred years ago as one of the most efficient human identification resources (Galton, 1892) and consolidated as robust forensic evidence (Neumann, Evett, and Skerrett, 2012; Neumann et al., 2015), fingerprints have recently been studied from the perspective of

Table 2.7: Brazilian Dataset: Confusion matrix for the finger prediction

Finger	Thumb	Index	Middle	Ring	Little	rate (%)
Thumb	108	0	0	0	0	100
Index	2	68	10	8	0	77.27
Middle	3	6	48	30	0	55.17
Ring	2	4	28	44	2	55.00
Little	0	0	0	0	25	100

Convolutional Neural Networks. The CNNs applied to Fingerprints have shown that this biometrics can help predict other human attributes in addition to providing proper identification (Hsiao et al., 2022).

In the fingerprint science literature, several works relate the gender/hand to a fingerprint. Nevertheless, few works deal with predicting the finger position from its fingerprint. The present study dealt with the application of CNNs in determining the gender, hand, and finger position of who originated a fingerprint. We used two databases with different types of fingerprints, plain and rolled. With expressive results from SOCOFing plain fingerprints, we get highly accurate predictive performance in determining sex, hand, and finger (100%, 98.28%, and 96.34% on average, respectively). Using the database composed of a sample of Brazilian rolled fingerprints, the average accuracies found to determine gender, hand, and finger are 99.10%, 94.32%, and 77.49%.

Hence, differences in prediction accuracies were observed between plain (SOCOFing) and rolled Brazilian fingerprints, highlighting the nuances in data representation. This distinction underscores the importance of understanding not only the type and quality but the purpose of fingerprint data when applying deep learning models, as the nature of the fingerprint can influence the model's performance.

Our study shows that it is possible to apply CNNs to predict gender, hand, and finger from plain and rolled fingerprint images. This approach can be applied as an efficient filter to reduce the enormous amount of fingerprints available in a database, offering a reduction in work time from an expert or processing from an automated system.

Bibliography

[1] D. R. Ashbaugh, *Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology*. Boca Raton, FL, USA, CRC Press, 1999.

[2] J. G. Barnes, *Chapter 1: History, in Fingerprint sourcebook*, National Institute of Justice. Washington, DC, USA, A. McRoberts, 2011.

[3] H. Cummins and C. Midlo, *Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics*. New York, USA, Dover, 1943.

[4] F. Galton, *Fingerprints*. London, U.K.: Macmillan Co, 1892.

[5] H. Cummins, W. J. Waits, J. T. McQuitty, *The breadths of epidermal ridges on the finger tips and palms: A study of variation.*. American Journal of Anatomy, 68(1), 127–150. doi: 10.1002/aja.1000680106.

[6] K. Ramanjit, R. K. Garg, *Determination of Gender Differences from Fingerprint Ridge Density in Two Northern Indian Populations*. Probl. Forensic Sci. 2011, 85, 5–10.

[7] E. Gutiérrez-Redomero, C. Alonso, E. Romero and V. Galera, *Variability of fingerprint ridge density in a sample of Spanish Caucasians and its application to sex determination*. Forensic Science International, 2008 180(1), pp.17-22.

[8] E. Gutiérrez-Redomero, Á. Sánchez-Andrés, N. Rivaldería, Noemí, C. Alonso-Rodríguez, J. E. Dipierri and L. Martín, *A Comparative Study of Topological and Sex Differences in Fingerprint Ridge Density in Argentinian and Spanish Population Samples*.Journal of Forensic and Legal Medicine, 2013, 20.5, pp. 419-29. Web.

[9] S. F. Abdullah, A. F. N. A. Rahman and Z. A. Abas, *Classification of gender by using fingerprint ridge density in northern part of Malaysia*. ARPN Journal of Engineering and Applied Sciences, 2015, 10(22), 10722–10726.

[10] H. Xie and Z. Lin, *Gender Difference Analysis of Chinese Han Youth based on Fingerprint Ridge Density*. E3S Web Conf., 198 (2020) 04014.DOI: https://doi.org/10.1051/e3sconf/202019804014.

[11] P. Gnanasivam, R. Vijayarajan, *Gender classification from fingerprint ridge count and fingertip size using optimal score assignment*. Complex Intell. Syst. 2019, 5, 343–352.

[12] J. A. G. Ayala, *Determinación de mano y dedo: técnicas para la formación de hipótesis en los procesos de identificación de huellas digitales anónimas*. Madrid,ES, Dirección de la Seguridad del Estado, 1981. ISBN 84-500-4860-5.

[13] S. A. D. Caballero, *Papiloscopia: Certeza ou dúvida? Apologia à micropapiloscopia*. Campinas, SP, Millennium, 2012.

[14] N. Kapoor and A. Badiye. *An Analysis of Whorl Patterns for Determination of Hand*. Journal of Forensic and Legal Medicine, 2015, 42-46. Web.

[15] I. Singh, P.K. Chattopadhyay, R.K. Garg, *Determination of the hand from single digit fingerprint: a study of whorls*. Forensic Science International, Volume 152, Issues 2–3, 2005.

[16] C. Huynh, E. Brunelle, L. Halamkova, J. Agudelo, and J. Halamek, *Forensic Identification of Gender from Fingerprints*. Analytical Chemistry, 87.22 (2015).

[17] I. M. Alaoui and J. Halamek, *Fluorescence of 1,2-Indanedione with Amino Acids Present in the Fingerprint Residue: Application in Gender Determination*. Journal of Forensic Sciences, 64.5 (2019): 1495-499. Web.

[18] P. Gnanasivam, and S. Muttan, *Fingerprint Gender Classification Using Wavelet Transform and Singular Value Decomposition*.IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012. https://doi.org/10.48550/arXiv.1205.674.

[19] M. K. Shinde and S. A. Annadate, *Analysis of Fingerprint Image for Gender Classification or Identification: Using Wavelet Transform and Singular Value Decomposition*. IEEE - International Conference on Computing Communication Control and Automation, Pune, India, 2015, pp. 650-654, doi: 10.1109/ICCUBEA.2015.133.

[20] S. Suwarno, *Gender Classification Based on Fingerprint Using Wavelet and Multilayer Perceptron*. Jurnal Dan Penelitian Teknik Informatika, 2023, 8(1), 139-144. https://doi.org/10.33395/sinkron.v8i1.11925.

[21] S. Sahu, A. Prabhakar Rao, and S. T. Mishra. *Fingerprints Based Gender Classification Using Adaptive Neuro Fuzzy Inference System*. IEEE - 2015 International Conference on Communications and Signal Processing (ICCSP) (2015): 1218-222.

[22] Chung-Ting Hsiao, Chun-Yi Lin, Po-Shan Wang, and Yu-Te Wu. *Application of Convolutional Neural Network for Fingerprint-Based Prediction of Gender, Finger Position, and Height*. Entropy, Basel, Switzerland, 24.4 (2022): 475. Web.

[23] Y. I. Shehu, A. Ruiz-Garcia, V. Palade, and A. James. *Detailed Identification of Fingerprints Using Convolutional Neural Networks*. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (2018): 1161-165. Web.

[24] O. N. Iloanusi, and U. C. Ejiogu. *Gender Classification from Fused Multi-fingerprint Types*. Information Security Journal A Global Perspective,(2020): 209-19. DOI: 10.1080/19393555.2020.1741742 29.5.

[25] Y.I. Shehu, A. Ruiz-Garcia, V. Palade and A. James, *Sokoto Coventry Fingerprint Dataset*. 2018. ArXiv preprint arXiv:1807.10609.

[26] Y.I. Shehu, A. Ruiz-Garcia, V. Palade and A. James, *Detection of Fingerprint Alterations Using Deep Convolutional Neural Networks*. In Proceedings of the International Conference on Artificial Neural Networks (ICANN 2018), Rhodes – Greece, 5th - 7th October 2018. Springer-Verlag Lecture Notes in Computer Science.

[27] Daniel da Silva Carvalho, Marcos Martinho Viana de Alecrim, Rafael Timóteo de Sousa Júnior, Luiz Antônio Ribeiro Júnior, Outcome of sex determination from ulnar and radial ridge densities of Brazilians' fingerprints: Applying an existing method to a new population, Science & Justice, Volume 62, Issue 2, 2022, Pages 181-192, ISSN 1355-0306, https://doi.org/10.1016/j.scijus.2022.01.005.

[28] M. E. C. Araújo and L. Pasquali, *Datiloscopia - A determinação dos dedos*. Brasília, BR, LABPAM, 2012.

[29] C. Neumann, I. W. Evett, and J. Skerrett, *Quantifying the Weight of Evidence from a Forensic Fingerprint Comparison: A New Paradigm*. Journal of the Royal Statistical Society. Series A, Statistics in Society 175.2 (2012): 371-415. Web.

[30] C. Neumann, C. Champod, M. Yoo, T. Genessay, and G. Langenburg. *Quantifying the Weight of Fingerprint Evidence through the Spatial Relationship, Directions and Types of Minutiae Observed on Fingermarks*. Forensic Science International, 248 (2015): 154-71. Web.

[31] Avalabe at https://www.gambitid.com/international-fingerprinting-faq.html. Acessed in 08/28/2023.

Chapter 3

A bibliometric review of automated methods for counting fingerprint minutiae

3.1 Introduction

Sir Arthur Conan Doyle's Sherlock Holmes detective solves his cases using reason, logic, deduction, and science. In his tales, he gets fingerprints as valuable evidence since "The Sign of Four" case, published in 1890, while the real-world Scotland Yard began to use fingerprints after 1901. Indeed, on the fingertips are prominent skin features that distinguish anyone from everyone else (Barnes, 2011b). For modern Sherlocks, how is the state of the art of automated methods for identifying minutiae in fingerprints?

Often done manually, fingerprint specialists use a peer-review-based methodology called ACE-V (Analysis, Comparison, Evaluation, and Verification) for identification expertise (SWG-FAST, 2002). This acronym describes the sequence of steps performed during identification (Champod et al., 2016; Ashbaugh, 1999b).

From a forensic science point of view, fingerprints are studied at three levels(Ashbaugh, 1999b). Level 1 is the set of characteristics related to the general morphology of the lines and their flows that form the typescript, defining its primary and secondary classification (subclas-

Figure 3.1: Level 2 details, called minutiae by Galton Galton, 1892.

sification). Level 2 is the association of characteristics related to changes observed along a line and which forms a minutia. The dactylogram's angle, position, and minutiae arrangements are part of the bulge of level 2. The term *minutiae* was first defined by Sir Francis Galton in 1892 (Galton, 1892), when he introduced six types of them (Figure 1): (a) bifurcation, (b) confluence, (c) enclosure or lake, (d) ridge beginning, (e) ridge ending, and island (f). Level 3 is the composition of dimensional and morphological attributes along each line, such as thickness, the shape of its edge, and quantity and relative position of pores. Figure 2 illustrates the three levels in a fingerprint.

Among these levels, the second is the most important for comparing fingerprints, as the other two criteria individualize the fingerprint (Champod et al., 2016). Thus, this work aims to present a database literature review on the automated marking of fingerprints' minutiae performed by artificial intelligence methods over the last five years. In Section 3.2, we present our bibliometric review with the help of the R package bibliometrixTeam, 2023; Aria and Cuccurullo, 2017 and the steps based on the Consolidated Meta-Analytic Approach Theory (CMAAT) Mariano and Rocha, 2017. Results and discussion are in Section 3.3, and Section 3.4 concludes.

3.2 Methodology

We used Bibliometrix and the Consolidated Meta-Analytic Approach Theory (CMAAT) method to conduct our exploratory survey. The first is an open-source R-package for performing bib-

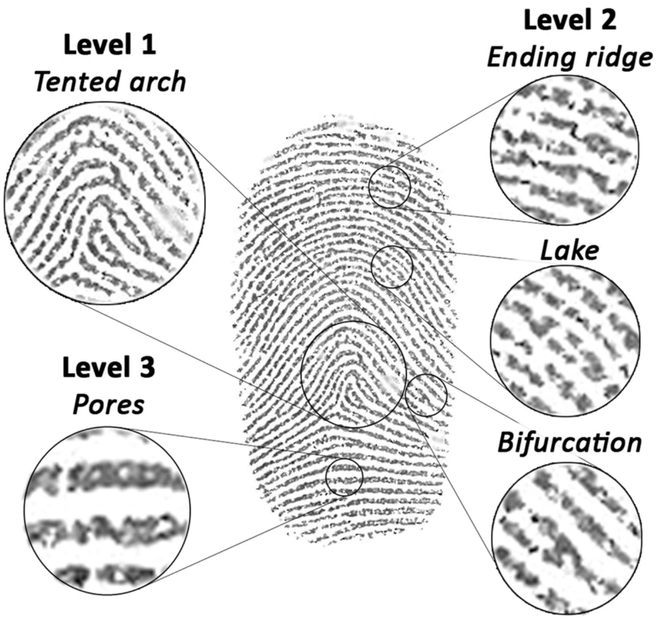

Figure 3.2: The 3 levels of analysis of a fingerprint. Image replicated from Scotcher and Bradshaw (2018).

liometric analyses and studies (Aria and Cuccurullo, 2017), and the latter is an integrative systematic review method (Mariano and Rocha, 2017). We defined Scopus as our metadata source, and on April 27, 2023, we performed the search for papers published in the 2018-2023 period using the keywords *fingerprint AND minutiae*.

3.3 Results and Discussion

3.3.1 Metadata

We found 696 documents published from 382 sources in 2018-2023. The average age of articles is 3.11 years, with an average of 4,963 citations per paper. There are 42 single-authored papers; among 1,692 authors, 14.7% collaborated internationally. The top 5 journals that concentrate 396 articles (56.9%) are *Advances in Intelligent Systems and Computing* (AISC), *Lecture Notes in Computer Science* (LNCS), *Multimedia Tools and Applications* (MTA), *ACM International Conference Proceeding Series* (ACM ICPS), and *IEEE Access*. Table 3.1 shows a significant growth from 2018 to 2023 in the evolution of publications on the subject in the top 5 scientific journals.

Table 3.1: Number of articles on *fingerprint AND minutiae*

Year	AISC	LNCS	MTA	ACM ICPS	IEEE Access
2018	6	6	2	6	1
2019	13	15	4	8	6
2020	24	19	4	10	11
2021	26	21	9	12	13
2022	26	24	12	13	13
2023	26	24	16	13	13

3.3.2 Main articles and collaboration networks

Yang et al. (2018) in *Pattern Recognition* have the highest total citations (142) and an annual citation rate of 23.7, resulting in a normalized citation index of 17.2. This paper proposes a cancelable multibiometric system that combines information from fingerprints and finger veins, with a feature fusion strategy presenting matching performance due to partial discrete Fourier transform. The system performance is evaluated using three performance indices: genuine acceptance rate, false acceptance rate, and global error rate. As for the information construction process, which starts in the dataset until the Fourier transforms, it requires a number of steps that go from the coordinate systems of the image to the search for features.

The second most cited article is by Chugh, Cao, and Jain (2018) in *IEEE Transactions on Information Forensics and Security*, with 119 total citations and a normalized citation index of 14.4. This work proposes an approach based on deep convolutional neural networks that utilize centralized and aligned local patches using fingerprint minutiae. Experimental results on three datasets show 99.03% of accuracy.

Among the top 3, Cao and Jain (2019) in *IEEE Transactions on Pattern Analysis and Machine Intelligence* have 103 citations and a normalized index of 14.7. They suggest an automated latent fingerprint recognition method using convolutional neural networks to extract features such as crest flow and minutiae. Performance is evaluated using the NIST SD27 and WVU databases, with an identification accuracy of 64.7% and 75.3%, respectively.

Recent work by Gao et al. (2023) stands out for performing automatic minutiae detection and classification using the YOLOv5*Ultralytics.YOLOv5* n.d. model, a deep convolutional neural network model. The experimental results showed that the model precision reached an accuracy of 97.22% in detecting six types of minutiae. Based on 619,297 fingerprint images, the results show differences in the distribution frequency of the six types of minutiae in the ten-finger positions and the finger patterns (arch, left loop, right loop, and whorl).

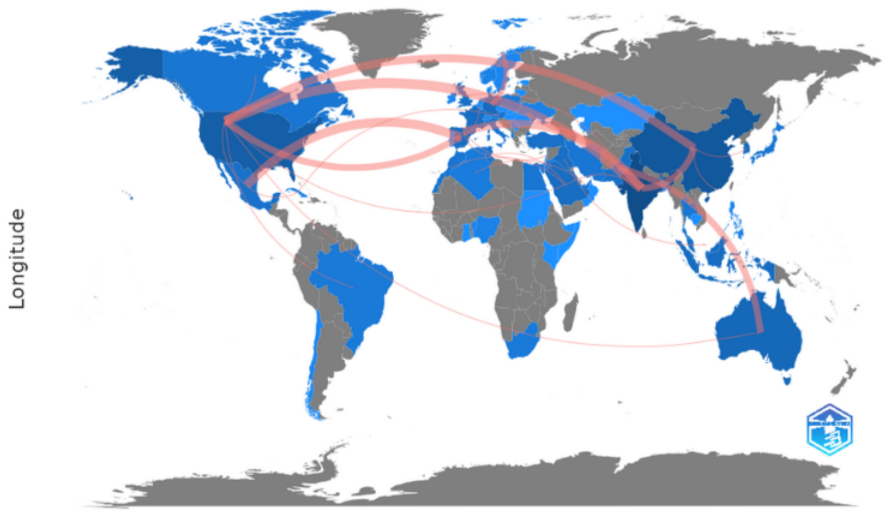

Figure 3.3: Collaboration network.

Finally, regarding the collaboration network (Figure 3.3), the countries with the most collaborations include India/United States, China/United States, India/France, Germany/Norway, and Spain/Mexico. We also observe collaborations between countries in the Middle East (such as Saudi Arabia, United Arab Emirates, Lebanon, Jordan, and Iran), and European countries (such as France, the United Kingdom, Germany, Spain, and Sweden). Analyzing these collaborations between countries allows for identifying trends in scientific and technological cooperation and understanding how research networks develop globally.

3.3.3 Cluster analysis

Taking as reference the most frequent words in the abstracts, groups by thematic areas emerge using Network Analysis. Figure 3.4 plots Callon's centrality measure (Callon, 1999) against the network density. The higher these indicators are, the greater the theme's degree of importance (centrality) and development (density). The fingerprint templates cluster has high centrality (9.17) and density (21.28), with a frequency of 599. Palmprint recognition has a centrality of 8.53 and a density of 26.09 (frequency of 1833). The algorithm cluster has the lowest centrality (1.54) and density (19.50), with a frequency of 43.

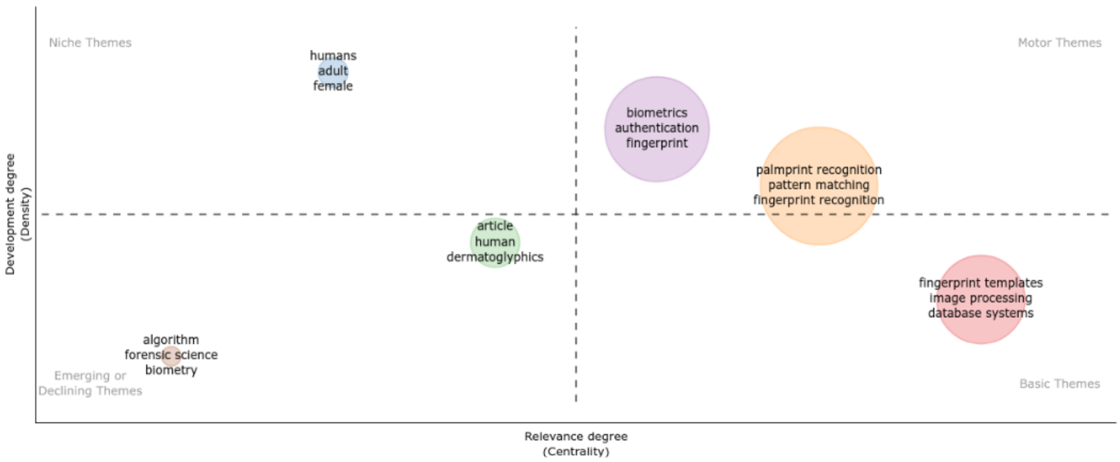

Figure 3.4: Trend of search topics.

From the factorial analysis, we find groups of similar terms given their contexts in the articles (Figure 4.3). *Cluster* 1 refers to keywords associated with biometric recognition techniques, such as *biometrics*, *fingerprint recognition*, *minutiae*, *image processing*, *extraction of features* , and *pattern matching*. These words are linked to extraction processes, authentication, and identification of biometric characteristics, such as fingerprints and palmprints. Cluster 2 indicates using neural networks and deep learning techniques applied to biometric image processing and analysis, where we find the keywords *deep learning*, *neural networks*, *convolutional networks*, and *deep neural networks*. *Cluster* 3 encompasses data protection, security, cryptography, and sensitive information in biometric systems (*cryptography*, *template protection*, and *fingerprint* *authentication*). Finally, *cluster* 4 refers to human aspects with keywords such as *human* and *dermatoglyphics*. These terms highlight the focus on analyzing human skin's unique physical characteristics and patterns.

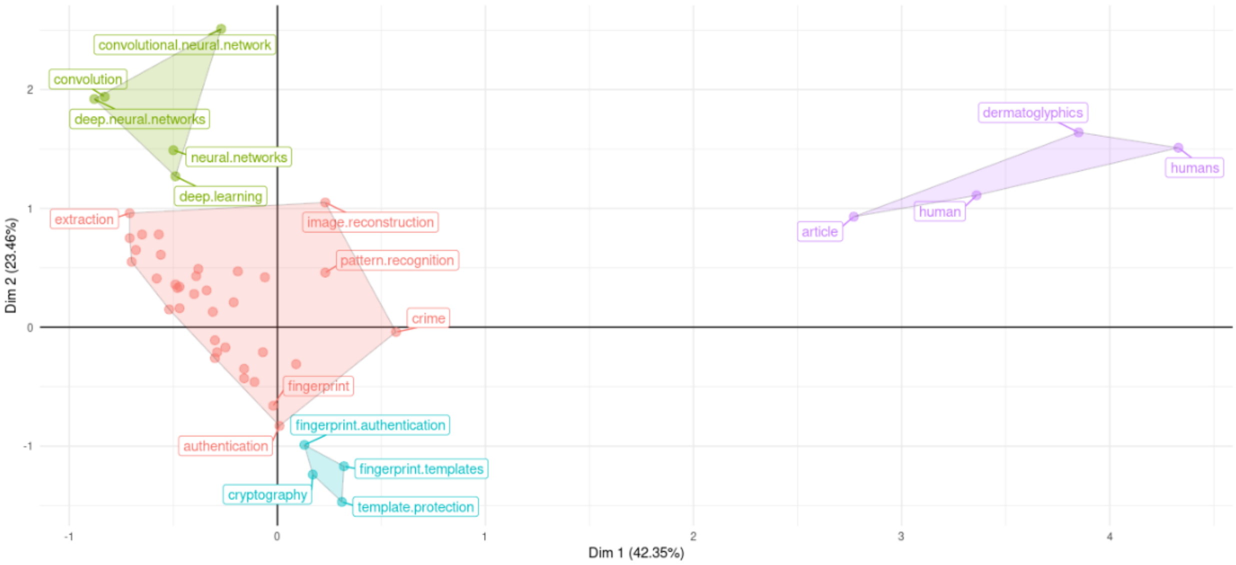

Figure 3.5: Cluster analysis.

3.4 Conclusions

The growing interest in the automated analysis of biometric characteristics mobilizes international research efforts. Deep neural network approaches stand out because of their high image analysis and processing accuracy. Experiments based on this technique show promising results for automating minutiae counting with precise fingerprint classifications. However, no paper has proposed a study using artificial intelligence to identify and total all minutiae in a fingerprint by type. It indicates a niche for future research in this area.

It is challenging because more open fingerprint data with minutiae markings are needed. Furthermore, we emphasize the need for an internationally adopted counting standard, which demonstrates a gap in the standardization of this process. Equally important is the lack of a coded system for counting minutiae, whose implementation could enhance and standardize such activity. These points reveal gaps in the literature regarding current procedures.

Deep learning to identify complex and distinct patterns in fingerprints that require time-consuming analysis by a fingerprint expert represents a significant contribution to the advancement of forensic science. It is a tool worth exploring.

Bibliography

[1] J. G. Barnes, *Chapter 1: History, in Fingerprint sourcebook*, National Institute of Justice. Washington, DC, USA, A. McRoberts, 2011.

[2] SWGFAST, *Friction Ridge Examination Methodology for Latent Print Examiners*, Scientific Working Group on Friction Ridge Analysis Study and Technology. USA, 2002, version 1.01.

[3] Champod, C., Lennard, C.J., Margot, P., and Stoilovic, M. *Fingerprints and Other Ridge Skin Impressions*, CRC Press, Boca Raton, 2nd ed. 2016.

[4] D. R. Ashbaugh, *Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology*. Boca Raton, FL, USA, CRC Press, 1999.

[5] F. Galton, *Fingerprints*. London, U.K.: Macmillan Co, 1892.

[6] Scotcher, K., Bradshaw, R.,*The analysis of latent fingermarks on polymer banknotes using MALDI-MS*, Sci Rep 8, 8765 (2018). Nature Potfolio. https://doi.org/10.1038/s41598-018-27004-0. Acesso em 10/09/2022.

[7] M. C. Puttini Fernandes, *Tecnicas de processamento de impressoes digitais para criação de banco de minúcias de alta qualidade*. Dissertação de Mestrado, Universidade de Brasília, Brasília, 2008, 90 páginas. [Online] Disponível em: `https://repositorio.unb.br/handle/10482/4123`. Acesso em: Abril 30, 2023.

[8] A. M. Mariano and M. S. Rocha, *Revisão da Literatura: Apresentação de uma Abordagem Integradora*, in AEDM International Conference–Economy, Business and Uncertainty: Ideas for a European and Mediterranean industrial policy, Reggio Calabria, Italy, 2017.

[9] Aria, M. and Cuccurullo, C., *Bibliometrix: An R-tool for comprehensive science mapping analysis*, Journal of Informetrics, vol. 11, no. 4, pp. 959-975, 2017.

[10] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023. [Online]. Disponível em: `https://www.R-project.org/`. Acesso em: Abril 27, 2023.

[11] CAPES, *Portal de Periodicos da Capes*. [Online]. Disponível em: `https://www.periodicos.capes.gov.br`. Acesso em: Abril 27, 2023.

[12] RNP - Rede Nacional de Pesquisa. (2023). [Online]. Disponível em: `https://www.rnp.br/`. Acesso em: Abril 27, 2023.

[13] Elsevier B.V., *Scopus*, [Online]. Disponível em: `https://www-scopus.ez110.periodicos.capes.gov.br/search/form.uri?display=basic#basic`. Acesso em: Abril 27, 2023.

[14] W. Yang, S. Wang, J. Hu, G. Zheng e C. Valli, *A fingerprint and finger-vein based cancelable multi-biometric system*, Pattern Recognition, vol. 78, pp. 242-251, 2018, doi: 10.1016/j.patcog.2018.01.026.

[15] T. Chugh, K. Cao and A. K. Jain, *Fingerprint Spoof Buster: Use of Minutiae-Centered Patches* in IEEE Transactions on Information Forensics and Security, vol. 13, no. 9, pp. 2190-2202, Sept. 2018, doi: 10.1109/TIFS.2018.2812193.

[16] K. Cao and A. K. Jain, *Automated Latent Fingerprint Recognition* in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 4, pp. 788-800, 1 April 2019, doi: 10.1109/TPAMI.2018.2818162.

[17] M. Gao, Y. Tang, H. Liu and R. Ma, *Statistics of fingerprint minutiae frequency and distribution based on automatic minutiae detection method*, Forensic Science International, vol. 344, p. 111572, 2023. doi: 10.1016/j.forsciint.2023.111572.

[18] Ultralytics.YOLOv5. [Online]. Disponível em: `https://github.com/ultralytics/yolov5`. Acesso em: Abril 27, 2023.

[19] M. Callon, *Actor-Network Theory: the Market Test* in J. Law and J. Hassard (Eds.), Actor Network and After, Oxford and Keele, Blackwell and the Sociological Review, 1999, pp. 181-195.

Chapter 4

Analyzing Minutiae Distributions in Brazilian Fingerprints: A Statistical Perspective on Sex, Finger Type, and Fundamental Patterns

4.1 Introduction

The use of fingerprints as a human identification resource dates to antiquity (4000 BC – 476 BC). It is understood that the Chinese civilization in the Qin Dynasty (221- 206 BC) was the first to use fingerprints for human identification, describing how to employ skin impressions as forensic identification evidence (Barnes, 2011b). The scientific basis for using fingerprints in human identification is expressed in mainly two principles: uniqueness and invariability. The uniqueness was defended in 1788 by the German physician and anatomist J. C. Mayer, and the permanence by the Englishman Sir William James Herschel in 1877 (Cummins and Midlo, 1943; Langenburg, 2011). It was up to Francis Galton to establish the two principles (Galton, 1892).

The first probabilistic models, which are those of Galton (1892), Henry (1900), and Balthazard (1911), made it possible to infer that a complete fingerprint was not necessary for secure human identification. Thus, the next question was to establish a minimum limit so that a partial fingerprint could identify a person, the first criterion for this being Locard's tripartite rule (Locard, 1914; Champod et al., 2016). There are two main approaches to identifying fingerprints: the quantitative method and the holistic, mixed, or qualitative-quantitative method. The number of marked features is considered in the quantitative method, while the holistic approach also considers the rarity of the characteristic points involved (Locard, 1914). Galton (1892) referred to these fingerprint features as Minutiae, which Vucetich (1904) called characteristic points. Ashbaugh (1999b) classified the fingerprint features into three distinctive levels of detail: I, II, and III (L1D, L2D, and L3D). Level I refers to the overall pattern defined by the ridge flow on the papillary surface. Second-level detail refers to friction ridge path deviations such as ridge ending, bifurcation, and island. The third level details are small shapes on the ridge, the relative location of pores, and the minor features in accidental damage to the friction ridges Ashbaugh, 1999b. Since the 1973 IAI Resolution (*International Association for Identification* n.d.), endorsed by the IAI and Ne'Urim Declaration (Ne'urim Declaration, 1995), several countries have moved towards the holistic approach (Champod et al., 2016). No valid basis exists to require a predetermined number of characteristics to exist between two fingerprint impressions to establish a positive identity (IAI, 1973-5 Resolution). In Brazil, initiatives have already been in this direction (*Ordem de serviço, da Polícia Civil do DF* n.d.; *The Federal District and the 12-Point Rule in Brazil, Dias da Costa* n.d.; *Manual Técnico de Datiloscopia* 2002). However, a scientific study is still needed to demonstrate the frequency distribution of minutiae in Brazilian fingerprints. Therefore, the objective of the present paper has been the qualitative and quantitative evaluation of the statistical frequency of minutiae in a sample of the Brazilian population.

4.2 Materials and methods

The sample consists of 600 fingerprints from 60 individuals, 30 males, and 30 females, over 18 years old, equally distributed among the five geographic regions of Brazil. To compose this database, the Brazilian Federal Police Fingerprint Experts selected the files whose ten prints were collected, with ink and paper, by the rolled method, applying the finger and rolling it from one end of the nail to the other to obtain the entire digital field. Furthermore, fingerprints without scars or any mark that would significantly compromise the analysis of these biometrics' levels I, II, and III details were chosen (Ashbaugh, 1999b). All fingerprints were digitized at 500 dpi and then subjected to NIST fingerprint image quality classification. Only those that obtained the maximum index, NFIQ 1, remained. Finally, each fingerprint was renamed to anonymize its specific origin, and its quality was certified by at least two experts with ten years of experience. From each of the 600 fingerprints, the characteristics of the three levels were analyzed. From L1D, the primary classification was considered (arch, right loop, left loop, and whorl) according to Vucetich (Figure 1) (Vucetich, 1904).

Figure 4.1: Fingerprints classified in the Vucetich system. From left to right, as arch, right loop, left loop, and whorl Vucetich, 1904.

Regarding L2D, all the minutiae have been accounted for. Finally, pores and incipient ridges for L3D have been observed. The analysis process was carried out and reviewed by two experienced fingerprint specialists with at least a master's scientific background. The markings were calculated using a spreadsheet editor, and a polygon with characteristic colors and shapes was inserted for each minutia, as shown in Figure 2.

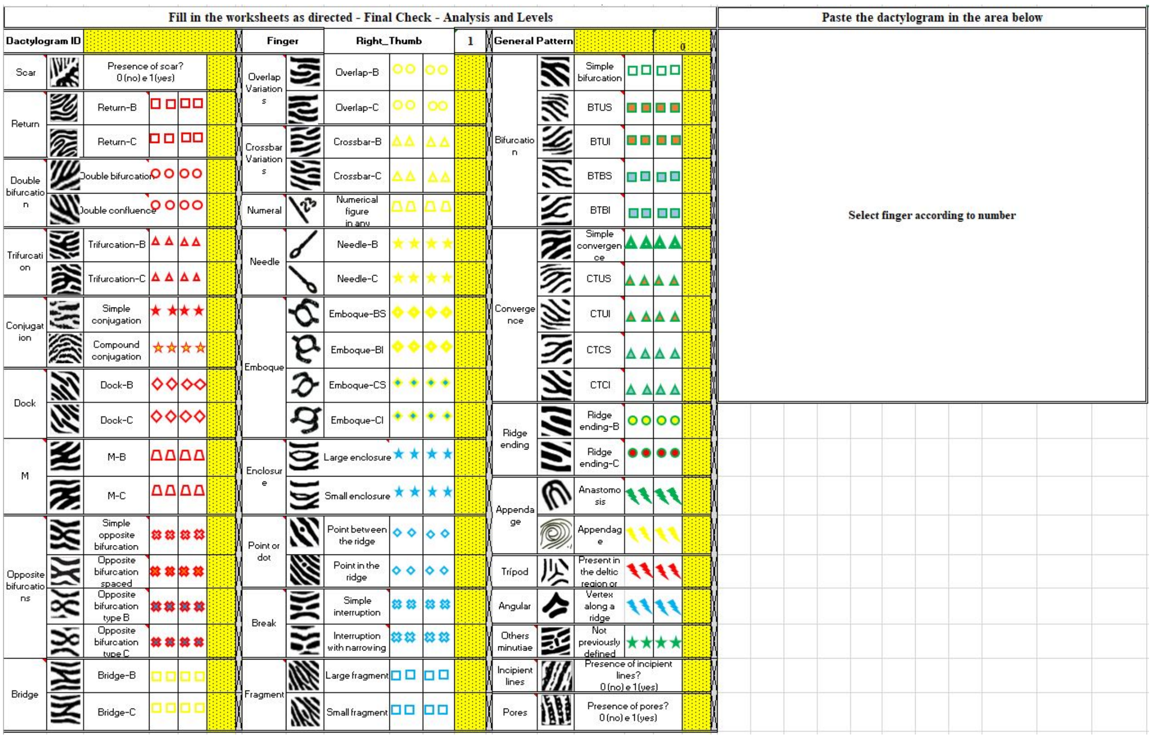

Figure 4.2: Table used to help count and concatenate information from levels I, II, and III. The minutiae representation was taken from fingerprint police manuals and previous studies *Manual Técnico de Datiloscopia* 2002; *Manual Técnico de Datiloscopia* 2002; Investigation, 1984; Champod, 1996; Gutiérrez et al., 2007; Gutiérrez-Redomero et al., 2011; Gutiérrez-Redomero et al., 2012; Rivaldería et al., 2017.

The minutiae counted initially were those considered in Brazil (*Manual Técnico de Datiloscopia* 2002; *Manual Técnico de Datiloscopia* 2002). But, throughout the work, occurrences predicted in other studies were observed (Champod, 1996; Gutiérrez et al., 2007; Gutiérrez-Redomero et al., 2011; Gutiérrez-Redomero et al., 2012; Rivaldería et al., 2017; Santamaría, 1955; Kingston, 1955; Steffens, 1965; Gupta, 1968; Okajima, 1970; Osterburg et al., 1977; Sclove, 1979; Lin, 1981; Stoney and Thornton, 1987; Sarkar, 2004; Fournier and Ross, 2016; Gao et al., 2023). Therefore, the markings were all redone considering a more extensive list of characteristic points, totaling 54 minutiae and their variations, which can be seen in Figure 2 and are defined below. The definitions were created to reduce possible subjectivity in determin-

Figure 4.3: Digital print with polygons of different colors and shapes indicating the L2D markings. As for L1D, it is a left loop type, and the L3D considered (pores and incipient ridges) was not found.

ing each minutia, following the inspiration of Gutiérrez et al. (2007) and Gutiérrez-Redomero et al. (2011).

- Ridge ending-B (E-B): is the minutiae referring to the beginning of the line.

- Ridge ending-C (E-C): is the minutiae referring to the end of the line.

- Bifurcation (B): a point where a line splits into two. In general, the ridges show no interruptions.

Bifurcation Variations (Spur-B):

*BTUS: bifurcation with a tendency to unite in the superior or external branch, whose length is five to ten times the width of the line. Reminds the formation of a large enclosure.

*BTUI: bifurcation with a tendency to unite in the inferior or internal branch, whose length is five to ten times the width of the line. Reminds the formation of a large enclosure.

*BTBS: bifurcation forming a Spur on the superior or external branch whose length is up to four times the width of the ridge. Well-characterized minutia with a proper spur.

*BTBI: bifurcation forming a Spur on the inferior or internal branch whose length is up to four times the width of the ridge. Well-characterized minutia with a proper spur.

- Convergence (C): a point where two ridges merge into one. In general, the ridges show no interruptions. Convergence variations (Spur-C):

*CTUS: convergence with a tendency to unite in the superior or external branch, whose length

is five to ten times the width of the line. Reminds the formation of a large enclosure.
*CTUI: convergence with a tendency to unite in the inferior or internal branch, whose length is five to ten times the width of the line. Reminds the formation of a large enclosure.
*CTCS: convergence forming a Spur on the superior or external branch whose length is up to four times the width of the ridge. Well-characterized minutia with a proper spur.
*CTCI: convergence forming a Spur on the inferior or internal branch whose length is up to four times the width of the ridge. Well-characterized minutia with a proper spur.
- Ridge fragments (F): are line fractions whose length varies from one to ten times the width of the ridge.
Ridge fragment variations:
*Big fragment (F-BG): it is considered significant when its length is five to ten times the width of the ridge.
* Small fragment or island (F-SM): it is considered small when its length is from one to four times the width of the ridge.
* Point between the ridges (P-BW): fragment of ridge whose length and width are of a ridge. If the dimension is less than a ridge, it is classified as an incipient ridge.
* Point in the ridge (P-IN): point at the continuity of a ridge.

Observation: Ridge ending-B and Ridge ending-C are types of Ridge end. Bifurcations and convergences are types of forks. A point is a ridge fragment whose length and width are of a ridge. If the dimension is less than a ridge, it is classified as an incipient ridge. Ridge endings, forks, and points are the fundamental minutiae of this study. Combining these three, called complex minutiae, can form any other one.

- Enclosure (EN): occurs when a ridge forks and merges again. Enclosure variations:
*Big enclosure (EN-BG): when its length is five to ten times the width of a ridge.
*Small enclosure (EN-SM): when its length is up to four times the width of the ridge.

- Break (BR): occurs where the course of a line is interrupted, provided that this discontinuity is one to four times the width of the ridge.

Break variations:

*Simple break (BR-SP) occurs when the two adjacent ridges are not modified.

*Break with narrowing (BR-NW) occurs when the two adjacent ridges narrow.

- Overlap (O): formation composed of two overlapping ridges. The distance between the ends of the ridge must be up to an island, that is, up to four times the width of the line.

Overlap variations:

*Overlap-B (O-B): When the ridge end of type B is above the ridge end of type C.

*Overlap-C (O-C): When the ridge end of type C is above the ridge end of type B.

- Crossbar (CR): line that changes direction transversely, crossing between two others. The ends of ridges can be far apart or intersect, as in the deviation, for up to the length of an island (up to four times the width of the ridge). Crossbar variations:

*Crossbar-B (CR-B): When the ridge end of type B is above the ridge end of type C.

*Crossbar-C (CR-C): When the ridge end of type C is above the ridge end of type B.

- Bridge (BD): minutiae in which a ridge fragment (up to ten times the width of the ridge) has its ends connected to two adjacent ridges, forming a bridge between them. Bridge variations:

*Bridge-B (BD-B): When the bifurcation is above the convergence. o Bridge-C (BD-C): When the convergence is above the bifurcation.

*Opposite bifurcations (OB): two ridges that join in one (single) or two points (double). The latter can be type spaced, B or C.

Opposite bifurcations variations:

*Opposite bifurcations spaced (OB-SP): when the convergence and bifurcation are separated by up to one island (up to four times the width of the line).

*Opposite bifurcations secant type B (OB-SB): the respective formation of an enclosure with conjugate bifurcation (simple or varied) is observed.

*Opposite bifurcations secant type C (OB-SC): the respective formation of a convergence (simple or varied) with conjugate enclosure is observed.

*Dock (D): ridge ending that enters between two others, with the distance between the ends of up to an island (four times the width of the ridge). Small misalignments can be accepted up to the width of a line.

Dock variations:

*Dock-B (D-B): when, from the left side, one ridge ending enters between two others.

*Dock-C (D-C): when, from the right side, one ridge ending enters between two others.

- Double bifurcation (DB-B): occurs when a ridge splits into three others. The distance between the two bifurcations is up to an island (four times the width of the ridge).
- Double Convergence (DB-C): occurs when three ridge merges into one. The distance between the two confluence points is up to an island (four times the width of the ridge).
- Trifurcation (TF): it is a particular case of Double bifurcation (TF-B) or Double Convergence (TF-C). In these cases, the three ridges split or merge regularly, forming a trident.
- M (M): minutia characterized by three fusion points of two bifurcations and a convergence (M-B) or vice versa (M-C), which presents a pattern like the letter M. The fusion points can be up to an island apart (up to four times the width of the line). This minutia can be characterized as a Y between two lines.
- Return (R): ridge, which does not belong to a core and which returns in its flow direction.

Return variations:

*Return type B (R-B) when directed to the right.

*Return type C (R-C) when directed to the left.

- Appendage (A): Minutia formed by joining a ridge or ridge fragment into a loop. Fusion can take place internally or externally.

Appendage variation:

*Anastomosis (An): Small ridge fragment (island or up to four times the width of a thread) that connects one loop to another more internal one by the apex.

- Needle (N): Ridge that begins (N-B) or ends (N-C) with an enclosure.

- Numeral (NM): Formations, in any orientation, of numerical figures.
- Tripod (TP): Three ridges (or fragments), in the deltic region or not, with openings practically at the same angle to each other (120°). Each ridge (or fragment) length doesn't need to be the same.
-Emboque (EB): incidence of the ridge in part of an enclosure. Emboque variations:
*Emboque type BS: Ridge ending in an enclosure's superior (EB-BS) part.
*Emboque type BI: Ridge ending in an enclosure's inferior (EB-BI) part.
*Emboque type CS: Incidence of the ridge ending C in the superior (EB-CS) part of an enclosure.
*Emboque type CI: Incidence of the ridge ending C in the inferior (EB-CI) part of an enclosure.
- Simple conjugation (CJ-S): Association of two complex minutiae distant up to an island or that share adjacent parts.
- Compound Conjugation (CJ-S): Association of three or more complex minutiae distant up to an island or that share adjacent parts.
- Angular (AG): vertex along a ridge.
- Others minutiae (OM): a formation that symbolizes a minutia but which has not been previously defined.

4.3 Results and discussion

The focus of the present study is on L2D. However, as noted in the introduction, the characteristics of L1D (fundamental type) and L3D (presence of pores and incipient lines) were also considered, as the three levels may present a statistically significant relationship. The distribution of L1D, according to Vucetich's classification (Vucetich, 1904), is as described in Table 1 and was carried out considering both the total fingerprint sample ($n = 600$) and the male ($n = 300$) and female sexes ($n = 300$). It is also essential that arch, as expected in the Brazilian population [35], had the lowest incidence and, even so, occurred 44 times. Thus, there is an adequate quantity to conduct statistical tests by fundamental type, sex, and type of fingers.

Table 4.1: Distribution of total L1D (general pattern) by sex.

Empty cell	Total (%)	Male (%)	Female (%)
Arch	7.33	4.33	10.33
Rigth Loop	30	30.33	29.67
Left Loop	31.33	34.33	28.33
Whorl	31.33	31	31.67

According to the chi-square test, there is dependence between fundamental types and sexes ($p = 0.028$), as well as between fundamental types and finger types ($p < 0.001$). Still, according to the aforementioned statistical test, there is no statistical evidence against the hypothesis of homogeneity in the occurrence of scars by sex ($p = 0.074$), but the non-homogeneity in the face of fundamental type ($p = 0.032$) and the type of finger ($p = 0.012$). The incidence of incipient lines is not homogeneous by sex ($p < 0.001$) and fundamental type ($p < 0.001$), but they are homogeneous in terms of finger type ($p = 0.187$). The occurrence of pores is not homogeneous by sex ($p < 0.001$) and the fundamental type ($p = 0.040$), but cannot reject the hypothesis of homogeneity when considering the type of finger ($p = 0.621$). Table 2 illustrates the percentage of occurrence of scars, incipient ridges, and pores by sex and, based on the above, comparisons that showed dependence and/or non-homogeneity have statistically significant differences. Thus, the difference between the incidence of incipient lines in men and women represented in Table 2 is relevant, as the chi-square test reported the non-homogeneity of incipient lines about sex.

Table 4.2: Percentage of occurrence of scars, incipient ridges, and pores concerning males and females.

Empty cell	Total (%)	Male (%)	Female (%)
Scars	35.83	32.33	39.33
Incipient ridges	20.5	15	26
Pores	65.83	74	57.67

The average number of characteristic points per fingerprint was also calculated in relation to sex, fundamental type, and finger type. The results are compiled in Tables 3, 4, and 5. The sum of L2D by fundamental type is in Table 6. It is observed that males presented a higher average of characteristic points than females, a result compatible with (Gutiérrez et al., 2007). Regarding the fundamental types, the decreasing order of average is whorl, loops (right and left), and arch, consistent with (Gutiérrez et al., 2007; Gao et al., 2023; Dankmeijer, Waltman, and De Wilde, 1980). As for the fingers, the corresponding fingers on both hands are in the same order of averages, from largest to smallest: thumbs, ring fingers, middle fingers, index fingers, and little fingers. It is interesting to note that the index fingers of both hands had the same average.

Table 4.3: Mean of minutiae in the total sample and divided by sex.

Empty cell	Total (n)	Male (n)	Female (n)
Mean of minutiae	82.28	86.62	77.94

Table 4.4: Mean minutiae by fundamental type.

Empty cell	Whorl	Right Loop	Left Loop	Arch
Mean of minutiae	87.87	81.06	79.68	74.5

Table 4.5: Mean minutiae by finger type.

Empty cell	RT	RR	RM	RI	Rm	LT	LR	LM	LI	Lm
Mean of minutiae	104.43	84.42	81.4	76.58	68.52	101.7	83.28	79.3	76.58	66.6

Note: RT=Right Thumb, RR=Right Ring, RM=Right Middle, RI=Right Index, Rm=Right Minimum, LT=Left Thumb, LR=Left Ring, LM=Left Middle, LI=Left Index, Lm=Left Minimum

Table 4.6: Total of minutiae by fundamental type.

Empty cell	Whorl	Right Loop	Left Loop	Arch
Total of minutiae	16520	14591	14980	3278

In Table 12, the results of the current study were compiled with those who considered a similar list of minutiae, which are those of Gutiérrez et al. (2007), E. Gutiérrez-Redomero et

Table 4.7: Dwass-Steel-Critchlow-Fligner multiple comparison test.

L1D	L1D	W	p
Arch	Rigth Loop	3.02	0.142
Arch	Left Loop	2.04	0.474
Arch	Whorl	5.77	¡0.001
Rigth Loop	Left Loop	-1.51	0.708
Rigth Loop	Whorl	4.93	0.003
Left Loop	Whorl	5.98	¡0.001

Table 4.8: Descriptive statistics of minutiae and their variations.

Minutia	Mean	Median	S.D.	Occ. (n)	Occ. (%)
Ridge ending-B	18.5	17.0	7.74	600	100
Ridge ending-C	17.7	17.0	8.28	599	99.83
Bifurcation	14.5	14.0	6.67	600	100
BTUS	0.390	0.00	0.667	182	30.33
BTUI	0.342	0.00	0.613	166	27.67
BTBS	0.362	0.00	0.620	177	29.5
BTBI	0.330	0.00	0.604	162	27
Convergence	11.3	10.0	6.19	598	99.67
CTUS	0.242	0.00	0.500	127	21.17
CTUI	0.263	0.00	0.508	140	23.33
CTCS	0.273	0.00	0.531	141	23.5
CTCI	0.318	0.00	0.592	156	26
Small fragment	1.97	1.00	2.01	460	76.67
Large fragment	1.22	1.00	1.28	396	66
Point between	2.17	1.00	3.25	372	62
Point in	0.280	0.00	0.675	114	19
Simple break	2.02	1.00	3.00	375	62.5
Break with narrowing	0.238	0.00	0.525	120	20
Small enclosure	0.900	1.00	1.11	328	54.67
Large enclosure	0.757	1.00	0.953	302	50.33
Overlap C	0.733	0.00	0.987	282	47
Overlap B	0.713	0.00	0.948	279	46.5
Simple conjugation	0.900	1.00	1.16	318	53
Compound Conjugation	0.340	0.00	0.738	151	25.17
Crossbar-C	0.602	0.00	0.853	249	41.5
Crossbar-B	0.562	0.00	0.900	230	38.33
Bridge-C	0.528	0.00	0.762	243	40.5
Bridge-B	0.457	0.00	0.690	213	35.5
Tripod	0.687	1.00	0.692	333	55.5
Simple Opp. Bif.	0.198	0.00	0.493	98	16.33
Spaced Opp. Bif.	0.138	0.00	0.399	72	12
Opp. Bif. Secant B	0.0700	0.00	0.268	40	6.67
Opp. Bif. Secant C	0.0600	0.00	0.251	34	5.67
Others minutiae	0.420	0.00	0.834	170	28.33
Dock-B	0.245	0.00	0.640	102	17
Dock-C	0.155	0.00	0.501	71	11.83
Appendage	0.240	0.00	0.435	142	23.67
Anastomosis	0.0833	0.00	0.300	46	7.67
Angular	0.267	0.00	0.503	143	23.83
Double bifurcation	0.187	0.00	0.482	91	15.17
M-B	0.102	0.00	0.376	48	8
M-C	0.0767	0.00	0.307	40	6.67
Emboque-BS	0.0383	0.00	0.192	23	3.83
Emboque-BI	0.0267	0.00	0.161	16	2.67
Emboque-CS	0.0617	0.00	0.241	37	6.17
Emboque-CI	0.0150	0.00	0.122	9	1.5
Double convergence	0.115	0.00	0.349	63	10.5
Return - C	0.0700	0.00	0.280	39	6.5
Return - B	0.0417	0.00	0.200	25	4.17
Trifurcation-B	0.0400	0.00	0.196	24	4
Trifurcation-C	0.0150	0.00	0.122	9	1.5
Needle-C	0.0233	0.00	0.151	14	2.33
Needle-B	0.0117	0.00	0.107	7	1.17
Numerical	0.0183	0.00	0.146	10	1.67

Table 4.9: Frequency distributions of minutiae, ordered from highest to lowest, considering the total sample and divisions by sex.

Minutiae	Total (%)	Male (%)	Female(%)
Ridge ending-B	22.508	22.996	21.966
Ridge ending-C	21.497	21.272	21.747
Bifurcation	19.308	18.817	19.853
Convergence	15.0641	13.683	16.598
Fragments	3.869	4.148	3.558
Points	2.976	3.663	2.211
Breaks	2.751	3.325	2.113
Enclosures	2.013	1.912	2.126
Overlaps	1.758	1.947	1.548
Conjugations	1.507	1.674	1.322
Crossbars	1.414	1.628	1.176
Bridges	1.197	1.07	1.339
Tripods	0.835	0.773	0.902
Opp. Bifurcs.	0.567	0.512	0.629
Others minutiae	0.510	0.435	0.594
Docks	0.486	0.473	0.5
Appendage	0.393	0.4	0.385
Angular	0.324	0.339	0.308
Double bifurc.	0.227	0.2	0.257
Ms	0.217	0.158	0.282
Emboques	0.172	0.165	0.18
Double converg.	0.14	0.115	0.167
Returns	0.136	0.173	0.094
Trifurcations	0.067	0.077	0.056
Needles	0.043	0.027	0.06
Numericals	0.022	0.015	0.03

Table 4.10: Frequency distributions of minutiae, ordered from highest to lowest, considering the general pattern.

Empty cell	Whorl %	Arch %	Right Loop %	Left Loop %
Ridge ending-B	22.815	22.239	16.113	28.458
Ridge ending-C	21.059	23.185	28.175	15.107
Bifurcation	19.08	20.012	13.652	24.913
Convergence	13.82	16.199	19.978	11.402
Fragments	4.558	3.386	3.619	3.458
Breaks	2.9	2.471	3.262	2.15
Points	2.742	1.983	3.584	2.857
Enclosures	2.185	1.952	1.926	1.923
Overlaps	1.834	1.739	1.68	1.756
Conjugations	1.538	1.617	1.474	1.482
Crossbars	1.429	1.19	1.48	1.382
Tripods	1.253	0.244	0.679	0.654
Bridges	1.041	1.342	1.295	1.242
Others minutiae	0.672	0.153	0.466	0.454
Opp. Bifurcs.	0.545	0.61	0.569	0.581
Docks	0.539	0.427	0.398	0.527
Appendage	0.46	0.214	0.404	0.347
Angular	0.393	0.092	0.329	0.294
Emboques	0.206	0.244	0.137	0.154
Returns	0.2	0.153	0.096	0.1
Double bifurcs.	0.2	0.244	0.192	0.287
Double converg.	0.182	0.122	0.144	0.093
Ms	0.182	0.061	0.226	0.28
Needles	0.061	0.061	0.027	0.033
Trifurcations	0.054	0.061	0.089	0.06
Numerals	0.054	0	0.007	0.007

al. [20, 21] and N. Rivaldería et al. [22]. Thus, the samples are from populations in Brazil (current study), Spain [19, 20] and Argentina [21, 22]. Comparing the results, we can see unity between the frequency distributions of the minutiae considered. In all samples, Ridge ending, Bifurcation, and Convergence had an incidence greater than 10%, and for the remaining L2D, it was less than 5%.

Table 4.11: Frequency distributions of minutiae, ordered from highest to lowest, considering the finger type.

Empty cell	RT%	RI %	RM %	RR %	Rm %	LT %	LI %	LM %	LR %	Lm %
Ridge ending-C	29.014	25.092	28.706	24.383	24.155	15.995	18.672	16.309	16.09	14.965
Ridge ending-B	17.236	21.610	17.363	18.164	18.292	29.482	25.767	27.806	24.355	24.925
Convergence	16.965	16.17	18.817	18.0	19.874	10.046	11.251	12.568	13.868	14.064
Bifurcation	11.81	15.125	14.21	18.026	16.784	21.485	22.459	24.842	25.395	25.175
Fragments	4.692	3.874	4.259	3.751	3.381	4.195	3.460	3.363	3.502	3.779
Points	4.469	3.156	2.969	2.804	2.238	3.163	3.0	2.501	2.461	2.302
Breaks	3.894	3.46	2.785	2.488	2.408	2.9	3.591	1.828	1.541	2.202
Enclosures	2.123	1.654	2.027	2.093	2.043	2.163	1.915	1.702	2.041	2.327
Overlaps	1.931	1.697	1.392	1.658	2.068	1.967	1.436	1.639	2.081	1.602
Conjugations	1.835	1.371	1.269	1.224	1.289	2.0	1.415	1.492	1.361	1.577
Crossbars	1.724	1.24	1.372	1.58	1.44	1.66	1.04	1.03	1.681	1.126
Bridges	0.91	1.11	1.126	1.2	1.557	0.967	1.328	1.156	1.341	1.527
Tripods	0.606	1.044	0.839	0.869	0.973	0.721	0.871	0.82	0.861	0.876
Docks	0.511	0.479	0.328	0.375	0.292	0.574	0.805	0.378	0.6	0.475
Opp. Bifurcs.	0.431	0.718	0.471	0.632	0.657	0.492	0.588	0.462	0.66	0.651
Angular	0.431	0.37	0.307	0.237	0.365	0.361	0.326	0.231	0.24	0.35
Others minutiae	0.335	0.588	0.594	0.77	0.486	0.475	0.61	0.441	0.48	0.35
Appendage	0.319	0.392	0.45	0.434	0.535	0.344	0.37	0.357	0.3	0.5
Returns	0.176	0.196	0.102	0.0395	0.097	0.147	0.196	0.231	0.1	0.05
Ms	0.176	0.152	0.041	0.336	0.292	0.180	0.196	0.210	0.260	0.375
Double converg.	0.112	0.261	0.123	0.257	0.122	0.066	0.109	0.168	0.1	0.1
Double bifurc.	0.08	0.065	0.184	0.217	0.292	0.344	0.24	0.25	0.32	0.3
Emboques	0.08	0.109	0.143	0.276	0.122	0.131	0.239	0.126	0.26	0.275
Trifurcations	0.064	0.022	0.061	0.059	0.17	0.066	0.087	0.063	0.04	0.05
Needles	0.064	0.044	0.041	0	0.073	0.066	0.022	0.021	0.04	0.05
Numericals	0.016	0	0.02	0.118	0	0.016	0	0	0.02	0.025

Still scrutinizing Tables 9, 10 and 11, the minutiae can be categorized into three groups: the first group is made up of common characteristic points that are the beginning of a line, end of a line, bifurcation, and convergence, which, in isolation, have an average incidence greater than or equal to 10% of the total observed, and together they represent more than 60% of the characteristic points recorded. The second group comprises rare minutiae, with an average occurrence of less than or equal to 5% and greater than 1%. This group includes, among others, enclosures, line fragments, and overlaps. The third group comprises rare characteristic points with an average occurrence of less than or equal to 1%. This set includes, among others, returns and trifurcations, in addition to conjunctions and other details. These last two types of minutiae were included in the sporadic group because their calculation had all the possibilities the previous definitions did not cover. It is important to note that common and rare minutiae variations, such as CTUS and interruption with narrowing, can be sporadic, as seen in Table 12. Table 13 compiled the frequencies of the minutiae variations considered in the present study. In the one performed by Gutiérrez et al. (2007), compatible behaviors can be seen between the two results. Although bifurcation and confluence are common characteristic points, their variations are rare.

Break with narrowing presents a sharp drop about a simple break. The incidence of big and small enclosures is practically the same. A noticeable difference is suggested in the opposite bifurcation, but for this detail, there was a difference in the variations considered, which could compromise the comparison.

Table 4.12: Frequency distributions of minutiae and their variations, ordered from highest to lowest, considering the total sample and divisions by sex.

Minutia	Total (%)	Male (%)	Female(%)
Ridge ending-B	22.508	22.996	21.966
Ridge ending-C	21.497	21.272	21.747
Bifurcation	17.578	17.043	18.172
BTUS	0.474	0.477	0.470
BTUI	0.415	0.477	0.346
BTBS	0.440	0.427	0.453
BTBI	0.401	0.393	0.411
Convergence	13.731	12.333	15.285
CTUS	0.294	0.281	0.308
CTUI	0.320	0.327	0.312
CTCS	0.332	0.354	0.308
CTCI	0.387	0.389	0.385
Small fragment	2.392	2.628	2.130
Large fragment	1.477	1.520	1.428
Point between	2.635	3.225	1.980
Point in	0.340	0.439	0.231
Simple break	2.461	2.978	1.886
Break with narrowing	0.290	0.346	0.227
Small enclosure	1.094	1.116	1.069
Large enclosure	0.920	0.797	1.056
Overlap C	0.891	1.027	0.740
Overlap B	0.867	0.920	0.808
Simple conjugation	1.094	1.212	0.962
Compound Conjugation	0.413	0.462	0.359
Crossbar-C	0.731	0.850	0.599
Crossbar-B	0.683	0.777	0.577
Bridge-C	0.642	0.604	0.684
Bridge-B	0.555	0.466	0.654
Tripod	0.835	0.773	0.902
Simple Opp. Bif.	0.241	0.192	0.295
Spaced Opp. Bif.	0.168	0.162	0.175
Opp. Bif. Secant B	0.085	0.100	0.068
Opp. Bif. Secant C	0.073	0.058	0.090
Others minutiae	0.510	0.435	0.594
Dock-B	0.298	0.300	0.295
Dock-C	0.188	0.173	0.205
Appendage	0.292	0.289	0.295
Anastomosis	0.101	0.112	0.090
Angular	0.324	0.339	0.308
Double bifurcation	0.227	0.200	0.257
M-B	0.124	0.085	0.167
M-C	0.093	0.073	0.115
Emboque-BS	0.047	0.042	0.051
Emboque-BI	0.032	0.027	0.038
Emboque-CS	0.075	0.077	0.073
Emboque-CI	0.018	0.019	0.017
Double convergence	0.140	0.115	0.167
Return - C	0.085	0.108	0.060
Return - B	0.051	0.065	0.034
Trifurcation-B	0.049	0.058	0.038
Trifurcation-C	0.018	0.019	0.017
Needle-C	0.028	0.015	0.043
Needle-B	0.014	0.012	0.017
Numerical	0.022	0.015	0.030

Table 4.13: Compilation of minutiae variations, in common with the current study and the one by E. Gutiérrez et al. Gutiérrez et al., 2007.

Empty cell	Total (%)	Males (%)	Females (%)		Empty cell	Total (%)	Males (%)	Females (%)
B	17.578	17.043	18.172		B	12.91	12.76	13.09
SB	15.848	15.269	16.491		SB	11.27	11.07	11.50
BTUS	0.474	0.477	0.470		BTUS	0.59	0.58	0.60
BTUI	0.415	0.477	0.346		BTUI	0.56	0.62	0.49
BTBS	0.440	0.427	0.453		BTBS	0.26	0.24	0.29
BTBI	0.401	0.393	0.411		BTBI	0.23	0.24	0.21
C	13.731	12.333	15.285		C	15.32	15.12	15.56
SC	12.398	10.982	13.972		SC	13.70	13.44	14.00
CTUS	0.294	0.281	0.308		CTUS	0.60	0.65	0.54
CTUI	0.320	0.327	0.312		CTUI	0.58	0.61	0.54
CTCS	0.332	0.354	0.308		CTCS	0.22	0.20	0.24
CTCI	0.387	0.389	0.385		CTCI	0.22	0.22	0.22
F-SM	2.392	2.628	2.130		F-SM	3.12	3.07	3.19
F-BG	1.477	1.520	1.428		F-BG	2.18	2.30	2.03
BR-SP	2.461	2.978	1.886		BR-SP	2.53	3.13	1.85
BR-NW	0.290	0.346	0.227		BR-NW	0.79	0.94	0.62
EN-SM	1.094	1.116	1.069		EN-SM	1.26	1.13	1.42
EN-BG	0.920	0.797	1.056		EN-BG	1.07	0.97	1.21
OB-SP	0.241	0.192	0.295		OB-SP	0.16	0.20	0.12
OB-SC	0.158	0.158	0.158		OB-SC	0.13	0.15	0.11

4.4 Conclusion

This study presents the inaugural statistical analysis of minutiae frequency distributions within the Brazilian populace. The research delves into the connections between L2D and factors associated with L1D and L3D and examines the impact of sex and finger type. Given Brazil's vast territorial expanse and the diverse anthropological makeup of its inhabitants, this research holds significant importance. The results were statistically scrutinized and presented a scientific basis to infer the occurrences of minutiae in the fingerprints of Brazilians. In Brazil, essential steps towards adopting the holistic criterion have already been taken (*Ordem de serviço, da Polícia Civil do DF* n.d.; *The Federal District and the 12-Point Rule in Brazil, Dias da Costa* n.d.; *Manual Técnico de Datiloscopia* 2002). The present work is yet another in this direction by showing that, statistically, a scarce characteristic point such as a Trident-C occurs, on average, 1,234 times less than a typical point like a Ridge ending-C. Thus, the presence of rare (and even more scarce) corresponding points confers a relevant probability of identity between a fragment and a fingerprint, even with a limited number of characteristic points (¡12), as long as there are no relevant and unexplained divergences. Thus, faced with cases involving latent prints with few details, the expert, even if he cannot categorically affirm the identification, can indicate an extreme probability depending on how rare the L2D involved are and their relationship with the

L1D and L3D in each specific case. The quantitative weight of each minutia presented in this study must be weighed in each real case by experts in fingerprint examination who used the definitions espoused here to classify each minutia. In this way, valuable traces of fingerprints, even with less than 12 characteristic points, will be able to provide their due collaboration, together with the entire body of evidence, in the elucidation of critical cases. This context has been a reality for decades [8] in countries that, with scientific responsibility, adopt holistic criteria for examining fingerprints. The classification of minutiae into common, rare, and very rare is another contribution of this work, indicating each characteristic point's qualitative weight. The group of fingerprints examined was taken randomly. It included individuals from all geographic regions of Brazil, with the results being discussed in detail and in line with previous studies of the same type. Thus, it is possible to infer that the sample studied is representative of the behavior of occurrence of minutiae in Brazilians. In the present study, all areas of the fingerprints had marked details, so we focused on the qualitative-quantitative relationship of the characteristic points. Specificity, which is the probability of incidence of minutia in a given region, will be addressed in a future study using learning systems, which is why polygons of different colors and shapes were used to mark the characteristic points, as illustrated in Figures 2 and 3. Finally, the presented results can contribute to studies on the relationship between minutiae orientation, fundamental type, and finger type. The knowledge here can also be helpful for mathematical research involving L2D probability distributions.

Bibliography

[1] J. G. Barnes, Chapter 1: History, in Fingerprint sourcebook, National Institute of Justice. Washington, DC, USA, A. McRoberts, 2011.

[2] H. Cummins and C. Midlo, Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics. New York, USA, Dover, 1943.

[3] Langenburg, G. Scientific research supporting the foundations of friction ridge examinations. In The Fingerprint Sourcebook, chapter 1, pages 14-5. National Institute of Justice, Office of Justice Programs, U.S. Department of Justice, 2011.

[4] F. Galton, Fingerprints. London, U.K.: Macmillan Co, 1892.

[5] Henry, E. R. Classification and Uses of Fingerprints; 1st ed.; Routledge & Sons: London, 1900.

[6] Balthazard, V. De l'identification par les empreintes digitales. Comptes Rendus, des Academies des Sciences 1911, 152, 1862–1864.

[7] Locard, E. (1914), La preuve judiciaire par les empreintes digitales, Archives d'anthropologie criminelle, de médecine légale et de psychologie normale et pathologique, 29, 321–348.

[8] Champod, C., Lennard, C.J., Margot, P., and Stoilovic, M. Fingerprints and Other Ridge Skin Impressions. CRC Press, Boca Raton, 2nd ed. 2016.

[9] Vucetich, J. Dactiloscopía comparada, el nuevo sistema argentino: trabajo hecho expresamente para el 2. Congreso médico latino-americano, Buenos Aires, 3-10 de abril de 1904. La Plata. 1904.

[10] Ashbaugh, D. R. Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology; CRC Press: Boca Raton, FL, 1999.

[11] International Association for Identification, 1973-5 Resolution.

[12] Ne'urim Declaration. Symposium Report (1995) J Forensic Identifi 45: 489-578.

[13] Ordem de serviço, da Polícia Civil do DF, publicada em 5 de março de 2008.

[14] The Federal District and the 12-Point Rule in Brazil, Dias da Costa, N. JFI-2008-6-621.

[15] Manual Técnico de Datiloscopia. Instituto Félix Pacheco de Identificação/PCERJ. Rio de Janeiro. 2002.

[16] Manual de Identificação Papiloscopica. Instituto Nacional de Identificaçao/Polícia Federal. Brasil. 1987.

[17] Federal Bureau of Investigation, U.S. Department of Justice. Fingerprint Training Manual, U.S. Department of Justice; U.S. Government Printing Office: Washington, D.C., 1984.

[18] C. Champod, Reconnaissance automatique et analyse statistique des minuties sur les empreintes digitales, PhA Thesis 1, Institut de Police Scientifique et de Criminologie, Université de Lausanne, 1996.

[19] E. Gutiérrez, V. Galera, J.M. Martínez, C. Alonso, Biological variability of the minutiae in the fingerprints of a sample of the Spanish population, Forensic Sci. Int., 172 (2007), pp. 98-105.

[20] E. Gutiérrez-Redomero, C. Alonso-Rodríguez, L.E. Hernández-Hurtado, J.L. Rodríguez-Villalba, Distribution of the minutiae in the fingerprints of a sample of the Spanish population, Forensic Sci. Int., 208 (2011), pp. 79-90.

[21] Esperanza Gutiérrez-Redomero, Noemí Rivaldería, Concepcion Alonso-Rodríguez, Luis M. Martín, José E. Dipierri, Miguel A. Fernández-Peire, Ricardo Morillo, Are there population differences in minutiae frequencies? A comparative study of two Argentinian population samples and one Spanish sample, Forensic Science International, Volume 222, Issues 1–3, 2012, Pages 266-276, ISSN 0379-0738, https://doi.org/10.1016/j.forsciint.2012.07.003.

[22] Rivaldería, Noemí, Esperanza Gutiérrez-Redomero, Concepcion Alonso-Rodríguez, José E Dipierri, and Luis M Martín. "Study of Fingerprints in Argentina Population for Application in Personal Identification." Science & Justice 57.3 (2017): 199-208. Web.

[23] F.A. Santamaría, A new method for evaluating ridge characteristics, Fingerprint Identification Mag., 36 (1955), pp. 3-8, 16–18.

[24] C.R. Kingston, Probabilistic analysis of partial fingerprint patterns, D. Crim. Dissertation, University of California, Berkeley, 1964.

[25] C. Steffens, Vergleichende Untersuchungen der Minutien der Fingerbeerenmuster bei Familien und eineiigen Zwillingspaaren, Anthrop. Anz., 29 (1965), pp. 234-249.

[26] S.R. Gupta, Statistical survey of ridge characteristics, Int. Crim. Polic. Rev., 218 (1968), pp. 130-134.

[27] M. Okajima, Frequency of forks in epidermal ridge minutiae in the finger print, Am. J. Phys. Anthrop., 32 (1970), pp. 41-48.

[28] J.W. Osterburg, T. Parthasarathy, T.E.S. Raghavan, S. Sclove, Development of a mathematical formula for the calculation of fingerprint probabilities based on individual characteristics, J. Am. Stat Assoc., 72 (360) (1977), pp. 772-778.

[29] S.L. Sclove, The occurrence of fingerprint characteristics as a two dimensional process, J. Am. Stat. Assoc., 74 (1979), pp. 588-595.

[30] C.H. Lin, The similarity of fingerprints and the development of a single fingerprint filing and searcharching system, Master's Science [PhA Thesis], University of Illinois, Chicago, 1981.

[31] D.A. Stoney, J.L. Thornton, A systematic study of epidermal ridge minutiae, J Forensic Sci., 32 (1987), pp. 1182-1203.

[32] N.Ch. Sarkar, Finger Ridge minutiae: Classification Distribution and Genetics, Anthropological Survey of India, Calcutta (2004).

[33] N. Fournier, A. Ross. Sex, Ancestral, and pattern type variation of fingerprint minutiae: a forensic perspective on anthropological dermatoglyphics. Am. J. Phys. Anthropol., 160 (2016), pp. 625-632, 10.1002/ajpa.22869

[34] Gao, Mengting, Yunqi Tang, Huan Liu, and Rongliang Ma. Statistics of Fingerprint Minutiae Frequency and Distribution Based on Automatic Minutiae Detection Method. Forensic Science International 344 (2023): 111572. Web

[35] M. E. C. Araújo and L. Pasquali, Datiloscopia - A determinação dos dedos. Brasília, BR, LABPAM, 2012.

[36] J. Dankmeijer, J.M. Waltman, A.G. De Wilde. Biological foundations for forensic identifications based on Fingerprints. Acta Morphol. Neerlando Scandinavia, 18 (1980), pp. 67-83.

Chapter 5

BRAFing: Brazilian Fingerprint Dataset

5.1 Introduction

Brazil was one of the first countries to regulate the use of fingerprints for human identification, a fact that occurred as early as 1903 [1] and, at the beginning of the 20th century, it was at the forefront of fingerprint science, whether in the field of application or academic production, with the works of Félix Pacheco [2,3], Galdino Ramos [4] and Manuel Viotti [5]. Currently, Brazil has representatives in important international groups related to fingerprint science, such as the International Fingerprint Research Group *(IFRG – International Fingerprint Research Group)*. However, few studies are available in the literature with data from Brazilians and no fingerprint databases from a sample of the Brazilian population are available for academic research. This study aims to gather, organize, analyze, and disseminate fingerprints already collected for identification purposes and stored in Federal Police – PF databases. The results of this study will provide valuable input for academic research, enabling – among other things – the advancement of human identification processes, especially in which only fragments of typescripts are available, as is common in situations involving the identification of corpses and crime scenes. The data set of Brazilian fingerprints, which is intended to be called BRAFing (Brazilian Fingerprints), was initially used in anonymized studies and resulted from the professional activities of Federal Police (FP) fingerprint experts in master's courses at the University of Brasília-UnB [6,7]. Taking into account that BRAFing has become an important biometric field to be studied, the idea of making it available for research purposes, on a non-profit basis, came up, such as the biometric databases of the National Institute of Standards and Technology of the United States of America. America (NIST - The National Institute of Standards and Technology) the North American Federal Police (Federal Bureau of Investigation - FBI) [8-10] and the Kaggle data science platform - Google LLC. [11].

The management of the National Identification Institute, a central body of the FP well received this initiative. However, as it involves the availability of fingerprint images, focusing on the academic community, ethical examination by the Research Ethics Committee (Conep/CNS/MS) is imperative. On 11/30/2023, the Research Ethics Committee of the UnB Faculty of Medicine authorized the advancement of research that will result in BRAFing (CAAE project nº 73627423.3.0000.5558).

5.2 Hypothesis

A database with Brazilian fingerprints accessible to the scientific community will favor the development of studies in the country's forensic identification area.

5.3 Primary Objective

Make a select sample of Brazilian fingerprints available in a renowned data repository accessible to researchers and postgraduate students.

5.4 Secondary Objective

Encourage studies related to the characteristics of Brazilian fingerprints by providing a database of national fingerprints.

5.5 Proposed Methodology

A database will be created consisting of 1000 fingerprints obtained from Decadactilar Records (medical records with fingerprints of the 10 fingers) of 100 Brazilian individuals, who had their prints collected, in ink and paper, by public security agents, from 1990 to 2004. The sample will consist of 50 male and 50 female people, over 18 years old, equally distributed across the five geographic regions of Brazil. To compose this database, the Federal Police Fingerprint Experts selected the Tenprints Files whose fingerprints were collected by the rolled method, applying the finger and rolling it from one end of the nail to the other, to obtain the entire digital field. Furthermore, fingerprints without scars or any mark that would significantly compromise the analysis of levels I, II, and III of these biometrics were chosen. All fingerprints were digitized at 500 dpi and then subjected to the NIST fingerprint image quality classification (*Nist Fingerprint Image Quality*), only those that obtained the maximum index, that is, NFIQ 1 remained. Finally, each fingerprint was renamed to anonymize its specific origin, and its quality was certified by at least two experts with 10 years of experience.

The three levels of detail of a fingerprint are represented in Figure 2.

5.6 Nomenclature

BRAFing files are .jpg and follow the standard nomenclature represented in Figure 3. Each of the four blocks means:

1- The first three numbers, ranging from 001 to 100, indicate the person. In the example, 001 is person 1;

2- The letter represents the male (M) and female (F) genders;

3- Means the hand, in English, referring to the finger: Right or Left, and;

4- The name of the finger, in English: Thumb, Index, Middle, Ring or Little.

Figure 5.1: Adapted from Map of the Brazilian Institute of Geography and Statistics [12].

5.7 Risks

The selected sample does not relate the fingerprint to any specific index of whoever originated it. Therefore, for the general population, this database is understood to be anonymized. However, it is possible to identify who a fingerprint belongs to, as long as someone with training and access to an automated fingerprint identification system (AFIS or ABIS) performs the search. The people who created the sample's fingerprints are registered in the Federal Police's ABIS, which, like the main systems of its kind, have *logs* access and a small list of public servants who can work You are subject to penalties in case of misuse of the data you have access to. Another

Figure 5.2: Adapted from [13].

$$\underbrace{\text{001}}_{1}\_\underbrace{\text{M}}_{2}\_\underbrace{\text{RIGH}}_{3}\_\underbrace{\text{THUMB}}_{4}$$

Figure 5.3: Caption of the formation of the name of each fingerprint. Inspired by [11].

risk that can be seen is the possibility, albeit remote, of a typescript from this database being used for purposes other than research. However, just as a fingerprint is unique, it is accepted that each fingerprint capture is also unique. This way, it is possible to infer whether two fingerprints from the same finger were produced at different times or whether they are copies of one another.

5.8 Benefits

Have a sample of Brazilian fingerprints available. This will encourage academic studies on the Brazilian population based on their fingerprints. It is believed that the greatest benefits will be in the field of human identification, making it increasingly efficient, especially in critical cases that are those in which only fragments of typescripts are available, as is common in situations involving the identification of unknown corpses and crime locations. The aim is to make BRAFing available in a renowned scientific repository, preferably associated with journals classified as Qualis A1 or Journal Citation Reports-JCR Q1.

5.9 Primary Outcome

The proposed database has a high value for academic research, especially related to forensic sciences, whether due to the criteria used in its selection, the continental dimension of Brazil or the anthropological richness of its population. BRAFing will fill a gap in the scientific community.

5.10 Usage

This dataset and its subsets are for exclusive use for non-commercial and non-profit academic and scientific research purposes.

5.11 Thanks

We want to thank the Brazilian Federal Police, through the National Identification Institute, which encouraged and supported this initiative to provide the academic and scientific world with a high-quality sample of Brazilian fingerprints. This database will encourage research in the field of human identification and fill the gap in accessible material to better understand the behavior of typing in Brazilians. We also extend due recognition to all the valuable fingerprint experts in Brazil, who, with serious and dedicated work, have provided human identification,

which is more efficient and safe every day.

Bibliography

[1] BRASIL. Decreto 4.764, de 5 de fevereiro de 1903.

[2] Pacheco, Felix. O problema da identificação. Rio de Janeiro, 1903.

[3] Pacheco, Felix. A excellencia do systema dactyloscopico Vucetich e a necessidade da creação dos Gabinetes inter-continentaes. Rio de Janeiro, 1905.

[4] Ramos, Galdino. M. S. Da Identificação. Rio de Janeiro, 1906.

[5] Viotti, Manuel. Dactyloscopia. 1908.

[6] Carvalho, D. S. Determinação do sexo a partir da densidade de cristas papilares e da contagem de albodatilares em brasileiros. Dissertação (Mestrado em Ciência de Materiais)–Universidade de Brasília, Brasília, 2019.

[7] Gomes, Gabriel. A. S. Dissertação em elaboração (Mestrado em Estatística - UnB).

[8] Fiumara, G., Flanagan, P., Schwarz, M., Tabassi, E., Boehnen, C. (2018) National Institute of Standards and Technology Special Database 301: Nail to Nail Fingerprint Challenge Dry Run. NIST Technical Note 2002. https://doi.org/10.6028/NIST.TN.2002. Acesso em 28/06/2023.

[9] Fiumara, G., Flanagan, P., Grantham, J., Bandini, B., Ko K, Libert, J. NIST Special Database 300 Uncompressed Plain and Rolled Images from Fingerprint Cards. NIST Technical Note 1993. https://doi.org/10.6028/NIST.TN.1993. Acesso em 28/06/2023.

[10] Fiumara, G., Flanagan, P., Grantham, J., Ko, K., Marshall, K., Schwarz, M., Tabassi, E., Woodgate, B. and Boehnen,C. (2019), NIST Special Database 302: Nail to Nail Fingerprint Challenge, Technical Note (NIST TN), National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.TN.2007. Acesso em 28/06/2023.

[11] Shehu, Y. I., Ruiz-Garcia, A., Palade, V., James, A. (2018). Sokoto Coventry Fingerprint Dataset. arXiv. https://arxiv.org/abs/1807.10609. Acesso em 28/06/2023.

[12] BRASIL. Mapa formulado pelo IBGE em 2010. Disponível em https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/19383-dividir-para-conhecer-as-diversas-divisoes-regionais-do-brasil. Acesso em 05/07/2023.

[13] Scotcher, K., Bradshaw, R. The analysis of latent fingermarks on polymer banknotes using MALDI-MS. Sci Rep 8, 8765 (2018). Nature Potfolio. https://doi.org/10.1038/s41598-018-27004-0. Acesso em 10/09/2022.

Chapter 6

Testing contingency table data with rare categories using Rényi divergence

6.1 Introduction

The chi-squared (χ^2) test is a standard statistical tool used to evaluate how well-observed frequencies match expected frequencies in categorical or discrete data, as well as to test for homogeneity and independence in contingency tables. When the sample size is small, Fisher's exact test is preferred as it calculates an exact p-value. However, the chi-squared test's reliability can diminish when expected frequencies are low, as the approximation it relies on for large samples may not hold. A general guideline is that expected frequencies should be no less than five. Sparse data can challenge the validity of the chi-squared test's asymptotic results, potentially leading to incorrect interpretations of the data's significance (Haberman, 1988; Baglivo, Olivier, and Pagano, 1992; Maiste and Weir, 2004). To circumvent the issue of small expected frequencies, data can be grouped/collapsed; however, this can result in a loss of information (Gautam and Kimeldorf, 1999; Bartfay and Donner, 2000).

This issue can be addressed using computational techniques like Monte Carlo simulations and bootstrapping, which generate an empirical distribution of the chi-squared statistic that reflects the particular characteristics of the dataset (Mudholkar and Hutson, 1997; Vexler et al., 2014). Nonetheless, pursuing asymptotic results for statistics that could serve as alternatives to the χ^2 remains a valuable endeavor. In this quest, divergence-type measures have emerged as promising statistical tools for testing goodness-of-fit (*e.g.* Jager and Wellner, 2007; Vonta, Mattheou, and Karagrigoriou, 2012; Noughabi and Balakrishnan, 2016; Yang and Chen, 2019; Pardo and Martín, 2021; Mirakhmedov, 2023).

An alternative measure is the Kullback-Leibler divergence, which relates to the generalized likelihood-ratio test for count data in multinomial contexts and is asymptotically equivalent to the χ^2 statistic (Lehmann, 1986). Salicru et al., 1994 and Morales et al., 1994 explored the asymptotic distributional characteristics of a broader class of divergence known as (h, ϕ)-divergence, which encompasses various information-theoretic measures like the Kullback-Leibler, J, R, (r, s), and ϕ divergences. Their findings indicate that, under certain conditions, (h, ϕ)-divergences may follow a chi-square or Gaussian distribution. These information-theoretic measures offer a more detailed perspective on the differences between distributions and are particularly useful for complex models or in evaluating the fit of a model in settings involving multiple variables.

In this study, we focus on the Rényi divergence for hypothesis testing within contingency tables. This interest is driven by the characteristic of the Rényi divergence that avoids divisions by frequencies, which can be problematic when frequencies are low. Contrary to other statistics and tests based on Rényi divergence proposed in previous research (Morales, Pardo, and Vajda, 2000; Morales et al., 2004), our investigation is confined to the case where $0 < \alpha < 1$, with α representing the index of the Rényi divergence. Importantly, we investigate a selection of appropriate values of α that ensures our proposed statistic is distributed as a chi-square distribution when dealing with low expected frequencies.

In Section 6.2 we delineate the relationship between the Rényi and chi-squared divergences, utilizing the second-order delta method to assist in large-sample derivations. In Section 6.3, we perform a Monte Carlo study to substantiate our theoretical findings. We examine two examples where the expected absolute frequencies fall below five, evaluating the adherence to chi-square properties across a range of α values for different sample sizes. We also compare the empirical power functions for some values of α using asymptotic critical values. An application of our approach is presented in Section 6.4, where we analyze a fingerprint data sample from a Brazilian police database to test for differences in minutiae distribution across genders, focusing on minutiae types that are infrequently observed in the population.

6.2 Relating the Rényi and χ^2 statistics

Let X be a discrete random variable on a bounded sample space $\mathcal{X}$ with the probability mass function (PMF) denoted as $p_j = P_X(x_j)$ for all $x_j \in \mathcal{X}$, $j \in \{1, 2, \ldots, m\}$. Expressing it as a vector of Bernoulli outcomes $\mathbf{Y} = (Y_1, \ldots, Y_m)'$, we have the PMF in the multinomial form

$$P_{\mathbf{Y}}(y_1, \ldots, y_m) = \prod_{j=1}^{m} p_j^{y_j}, \tag{6.1}$$

where $y_j \in \{0, 1\}$, $y_m = 1 - \sum_{j=1}^{m-1} y_j$ and $p_m = 1 - \sum_{j=1}^{m-1} p_j$. Here, we have a set of $m - 1$ independent trials, each one resulting in any one of m possible outcomes. Thus, denoting a sample of n independent and identical copies of $\mathbf{Y}$ as $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,m})'$, with $i = 1, \ldots, n$, the number of trials resulting in each outcome j is

$$N_j = \sum_{i=1}^{n} Y_{i,j} \sim \text{binomial}(n, p_j),$$

and hence, $(N_1, \ldots, N_m)'$ follows a multinomial distribution. The maximum likelihood estimator of p_j is the relative frequency of outcome j, $\hat{p}_j = N_j/n$. Hence, by taking n sufficiently large, $(\hat{p}_1, \ldots, \hat{p}_m)'$ is multinormal with $\mathrm{E}(\hat{p}_j) = p_j$, $\mathrm{Var}(\hat{p}_j) = p_j(1 - p_j)/n$, and $\mathrm{Cov}(\hat{p}_j, \hat{p}_{j'}) = -p_j p_{j'}/n$, $\forall\, 1 \le j, j' \le m$ for $j \ne j'$. Consequently, we can write

$$\hat{p}_j = p_j + \varepsilon_j, \tag{6.2}$$

where $\varepsilon_j \sim \mathrm{N}(0, p_j(1-p_j)/n)$, such that $\sum_{j=1}^{m} \varepsilon_j = 0$ and $\mathrm{Cov}(\varepsilon_j, \varepsilon_{j'}) = -p_j p_{j'}/n$, for $j \neq j'$. Under the hypothesis that $\mathbf{Y}$ follows (6.1), from (6.2) we can rewrite the original Pearson's χ^2 test statistic in terms of random errors as

$$\chi^2(\hat{p}||p) = n \sum_{j=1}^{m} \frac{(\hat{p}_j - p_j)^2}{p_j} = n \sum_{j=1}^{m} \frac{\varepsilon_j^2}{p_j}, \tag{6.3}$$

having approximately the chi-square distribution with $m-1$ degrees of freedom for n sufficiently large.

Now, let us consider the Rényi divergence of order α of a hypothetical population distribution p from its empirical estimate $\hat{p}$ defined to be

$$D_\alpha(\hat{p}||p) = \frac{1}{\alpha - 1} \ln A_\alpha(\hat{p}||p) \tag{6.4}$$

where

$$A_\alpha(\hat{p}||p) = \sum_{j=1}^{m} p_j^\alpha \hat{p}_j^{1-\alpha}, \tag{6.5}$$

for $0 < \alpha < 1$, avoiding the division by a rarely observed frequency. Thus, the case $\alpha > 1$ is not particularly interesting. We remark that unlike the χ^2 divergence in (6.3), the Rényi coefficient (6.5) is computable even if $p_j = 0$ for some j. While $A_\alpha(\hat{p}||p)$ measures the overlap between $\hat{p}$ and p, $D_\alpha(\hat{p}||p)$ denotes a measure of divergence between the empirical $\hat{p}$ and the hypothetical distribution p.

As a special case, when $\alpha = 1/2$, $A_{1/2}(\hat{p}||p)$ is known as the Bhattacharyya coefficient, in which each term is a geometric mean between p_j and $\hat{p}_j$. In this case, $D_{1/2}(\hat{p}||p)$ denotes the twice of the Bhattacharyya divergence (*e.g.*, Schweppe, 1967; Ray, 1989). In another special case, we have the Kullback-Leibler divergence by taking $\lim_{\alpha \to 1} D_\alpha(\hat{p}||p)$, which is known to be asymptotically equivalent to the χ^2 statistic (Salicru et al., 1994). Nevertheless, as we will discuss shortly, we get useful results for small values of α for our purposes.

Rewriting (6.5) in terms of the error ε_j from (6.2), we get

$$\begin{aligned} A_\alpha(\hat{p}||p) &= \sum_{j=1}^{m} p_j^\alpha (p_j + \varepsilon_j)^{1-\alpha} \\ &= \sum_{j=1}^{m} p_j \left(1 + \frac{\varepsilon_j}{p_j}\right)^{1-\alpha}. \end{aligned} \tag{6.6}$$

Next, we shall establish a large-sample relation between $D_\alpha(\hat{p}||p)$ and χ^2 for $0 < \alpha < 1$ with the help of the second-order delta method. By the consistency of $\hat{p}_j$, assuming that the sample size n is large enough to allow us to take the power series expansion of $(1 + \varepsilon_j/p_j)^{1-\alpha}$, provided that $\sum_{j=1}^m p_j = 1$ and $\sum_{j=1}^m \varepsilon_j = 0$, and using (6.3), we can continue from (6.6) as follows:

$$\begin{aligned} A_\alpha(\hat{p}||p) &= \sum_{j=1}^{m} p_j \left[1 + (1-\alpha)\frac{\varepsilon_j}{p_j} - \frac{\alpha(1-\alpha)}{2!}\frac{\varepsilon_j^2}{p_j^2} + \sum_{k\geq 3}\binom{1-\alpha}{k}\left(\frac{\varepsilon_j}{p_j}\right)^k\right] \\ &= \sum_{j=1}^{m} \left[p_j + (1-\alpha)\varepsilon_j - \frac{\alpha(1-\alpha)}{2!}\frac{\varepsilon_j^2}{p_j} + \sum_{k\geq 3}\binom{1-\alpha}{k}\frac{\varepsilon_j^k}{p_j^{k-1}}\right] \\ &= 1 - \frac{\alpha(1-\alpha)}{2n}\chi^2 + R(\varepsilon_j), \end{aligned} \tag{6.7}$$

where $R(\varepsilon_j) = \sum_{k\geq 3}\binom{1-\alpha}{k}\varepsilon_j^k/p_j^{k-1}$ denotes a small remainder. Taking the linear approximation $\ln(1+x) \approx x$ with a small $|x|$ in (6.4), from (6.7) we obtain

$$\frac{2n}{\alpha}D(\hat{p}||p) = \frac{2n}{\alpha(\alpha-1)}\ln A(\hat{p}||p) = \chi^2 + \frac{2n}{\alpha(\alpha-1)}R(\varepsilon_j) = \chi^2 + \mathcal{O}_p(1). \tag{6.8}$$

Since $n\epsilon_j \to 0$ in probability, using Markov's inequality, we have

$$P\left(\left|n\epsilon_j^k\right| \geq \delta\right) \leq \frac{E\left|n\epsilon_j^k\right|^{2/k}}{\delta^{2/k}} = \frac{p_j(1-p_j)}{n^{1-2/k}\delta^{2/k}} \to 0$$

for any $\delta > 0$ and $k \geq 3$ as $n \uparrow \infty$. Thus, the remainder $nR(\varepsilon_j) \to 0$ in probability. Therefore, we may conclude that the statistics $2nD(\hat{p}||p)/\alpha$ and χ^2 are asymptotically equivalent, and thus we may conclude that

$$R_{\alpha,n,m} = \frac{2n}{\alpha} D_\alpha(\hat{p}||p) \xrightarrow{D} \chi^2_{m-1}. \tag{6.9}$$

Furthermore, a secondary but no less important result is that the approximation (6.8) tends to be better as we chose a small α value. That is because the kth term of $\mathcal{O}_p(1)$ depends on $(\alpha+1)(\alpha+2)\cdots(\alpha+k)$, which an increasing function of $\alpha \in (0,1)$. As it will be discussed in the next section, a small α value may help to guarantee chi-square-ness when we find some low expected absolute frequencies in the sample.

6.3 Validation and power assessment

We performed Monte Carlo experiments to validate our suggested statistic (6.9) and assess its power performance. Our investigation included two examples where some expected absolute frequencies were less than five. The first case, presented in Example 6.1, involved a goodness-of-fit test for data following a truncated geometric distribution. The second case in Example 6.2 focused on a homogeneity test across two groups. We performed 500 Monte Carlo simulations for each scenario, generating random sample sizes within $n \in \{700, 800, 1000, 2000, 3000, 4000\}$ under the respective null hypotheses. We then analyzed the empirical sampling distributions of our statistic (6.9), exploring values of α from 0.1 to 0.9 in 0.1 increments. In Example 6.2, we also considered α ranging from 0.01 to 0.09. These empirical distributions were subsequently compared to the theoretical χ^2 distribution using QQ plots and Kolmogorov-Smirnov tests to assess their fit.

Example 6.1. Goodness-of-fit. Let $X_1, \ldots, X_n$ be a random sample drawn from the truncated geometric distribution PMF given by

$$p_x = \frac{\pi(1-\pi)^x}{1-(1-\pi)^8}, \tag{6.10}$$

where $x \in \{0, 1, \ldots, 7\}$ and $0 < \pi < 1$. Consider the test

$$H_0 : \pi = \pi_0 \textit{ versus } H_1 : \pi \neq \pi_0,$$

where π_0 is a success probability under H_0. We emphasize that this example is illustrative because the parametric form makes it easier to study the test power. We are not concerned with applying the most powerful test here.

As the lowest expected absolute frequency under H_0 is $np_7 = n\pi_0(1-\pi_0)^7/[1-(1-\pi_0)^8]$, the smallest expected absolute frequencies for these sample sizes are 2.7, 3.1, and 3.9, for $\pi_0 = 0.5$, and $n = 700, 800$, and 1000. If we increase π_0 to 0.55, the expected absolute frequencies become 1.4, 1.6, and 2.0 for these sample sizes. Thus, as the asymptotic result for the chi-square statistic may not be reliable, we suggest the Rényi statistic instead.

We performed Monte Carlo replications of random samples with different sizes to obtain the empirical sampling distributions with α ranging from 0.1 to 0.9. Using $\pi_0 = 0.4, 0.5$, and 0.55 we tested if these empirical distributions follow a $\chi^2_{(7)}$ as in (6.9). Kolmogorov-Smirnov p-values testing chi-squared-ness are shown in Table 6.1. With small values of α we do not dismiss the chi-squared-ness even facing expected absolute frequencies less than five occurring for $n = 700, 800$, and 1000 when $\pi_0 = 0.5$ or 0.55. The QQ plots in Figure 6.1 show a reasonable fit to the expected chi-squared-ness of our statistic using $\alpha = 0.1$ when $\pi_0 = 0,5$ and $0,55$ with $n = 700$. For a larger sample size, $n = 2,000$, Figure 6.2 depicts empirical power functions of our suggested statistic for $\pi_0 = 0.5$ at significance level 5% with asymptotic critical value. Different values of α do not produce substantial differences in the power of the test in this scenario.

Table 6.1: $H_0 : \pi = \pi_0$ in Example 6.1: Kolmogorov-Smirnov p-values testing the null hypothesis that our statistic follows the expected χ^2 distribution with 7 degrees of freedom.

		n					
	α	700	800	1,000	2,000	3,000	4,000
	0.1	0.271	0.663	0.390	0.663	0.469	0.711
	0.2	0.238	0.711	0.419	0.604	0.458	0.723
	0.3	0.289	0.811	0.409	0.663	0.484	0.683
	0.4	0.242	0.812	0.523	0.679	0.511	0.612
$\pi_0 = 0.4$	0.5	0.235	0.827	0.531	0.676	0.546	0.583
	0.6	0.249	0.824	0.563	0.694	0.601	0.566
	0.7	0.217	0.834	0.563	0.715	0.583	0.549
	0.8	0.135	0.851	0.486	0.744	0.562	0.533
	0.9	0.129	0.818	0.411	0.772	0.553	0.539
	0.1	0.720	0.484	0.260	0.226	0.653	0.416
	0.2	0.558	0.239	0.247	0.186	0.631	0.389
	0.3	0.575	0.221	0.181	0.169	0.679	0.459
	0.4	0.536	0.117	0.118	0.160	0.728	0.401
$\pi_0 = 0.5$	0.5	0.136	0.099	0.102	0.156	0.717	0.333
	0.6	0.041	0.077	0.059	0.166	0.721	0.349
	0.7	0.023	0.064	0.037	0.141	0.714	0.313
	0.8	0.017	0.056	0.040	0.146	0.715	0.278
	0.9	0.016	0.062	0.028	0.115	0.662	0.210
	0.1	0.468	0.682	0.712	0.621	0.474	0.791
	0.2	0.201	0.791	0.410	0.588	0.566	0.787
	0.3	0.028	0.600	0.211	0.525	0.504	0.806
	0.4	0.001	0.103	0.046	0.339	0.495	0.781
$\pi_0 = 0.55$	0.5	0.000	0.003	0.003	0.283	0.336	0.692
	0.6	0.000	0.000	0.000	0.248	0.338	0.735
	0.7	0.000	0.000	0.000	0.168	0.319	0.666
	0.8	0.000	0.000	0.000	0.131	0.324	0.534
	0.9	0.000	0.000	0.000	0.083	0.211	0.487

Example 6.2. Homogeneity. Let $X_{1,1}, \ldots, X_{1,n_1}$ and $X_{2,1}, \ldots, X_{2,n_2}$ be two random samples drawn from two truncated geometric distributions X_1 and X_2 as in (6.10). Now, consider testing

$$H_0 : X_1 = X_2 \textit{ versus } H_1 : X_1 \neq X_2.$$

Again, we stress the illustrative purpose of our example. Now, consider the bivariate PMF denoted as $p_{k,j} = \mathrm{P}(X_k = j, K = k)$, where $k \in 1, 2$, and $j \in \{1, \ldots, m\}$. Straightforwardly,

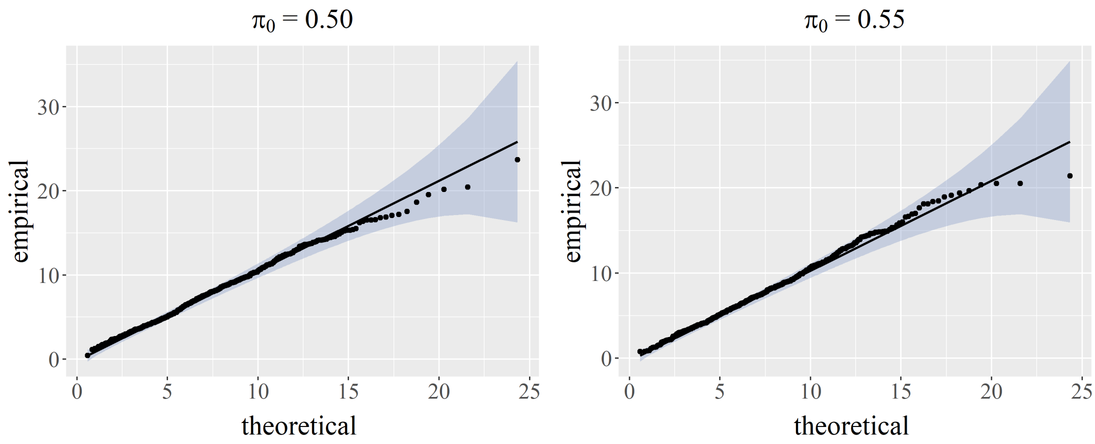

Figure 6.1: $H_0 : \pi = \pi_0$ in Example 6.1: QQ plots comparing empirical sampling distributions with $n = 700$ and $\alpha = 0.1$ against the expected χ^2 distribution with 7 degrees of freedom.

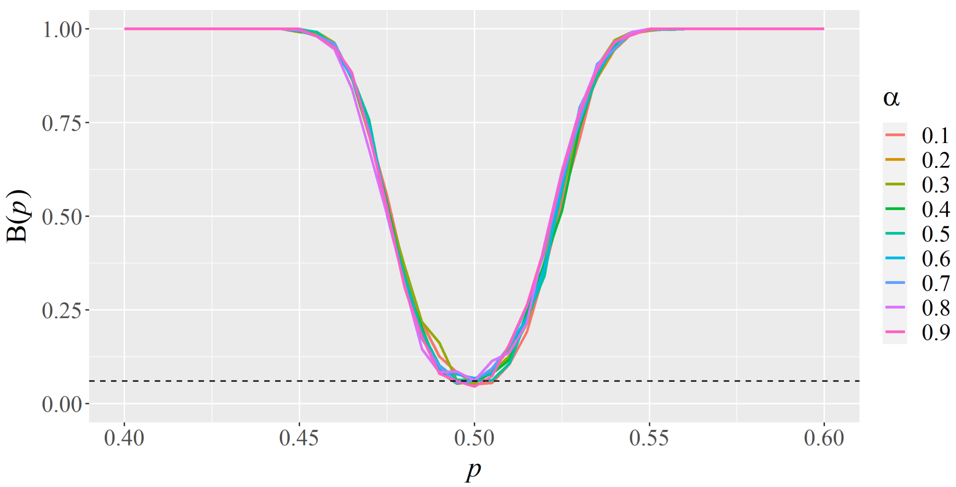

Figure 6.2: Example 6.1: Empirical power functions of our suggested statistic for testing $H_0 : \pi = 0.5$ against $H_1 : \pi \neq 0.5$ ($n = 2,000$) at significance level 5%, using asymptotic critical value.

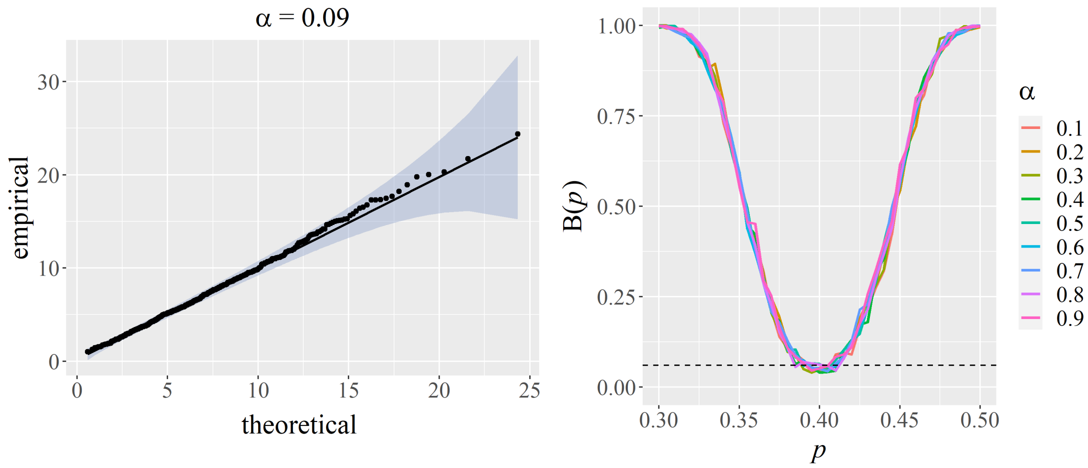

Figure 6.3: $H_0 : X_1 = X_2$ in Example 6.2: Left: a QQ plot comparing empirical sampling distributions ($n = 700$) against the expected χ^2 distribution with 7 degrees of freedom. Right: Empirical power function of our suggested statistic for testing $H_0 : p_x = p_y = 0.4$ against $H_1 : p_x \neq p_y$ ($n_x = n_y = 1,000$) at significance level 5%, using asymptotic critical value.

we extend the Rényi's coefficient to this bivariate case as

$$A_\alpha(\hat{p}||p) = \sum_{k=1}^{2}\sum_{j=1}^{m} p_{k,j}^{\alpha}\hat{p}_{k,j}^{1-\alpha}, \tag{6.11}$$

where $p_{k,j}$ denotes the expected relative conjoint frequency under H_0 and $\hat{p}_{k,j}$ is the observed one.

As in the previous example, we considered total sample sizes $n = n_1 + n_2$ varying from 700 to 4,000, where $n_1 = n_2$. Applying (6.11) in (6.4) with α ranging from 0.01 to 0.09 and 0.1 to 0.9, we obtained the empirical sampling distributions of (6.9). Table 6.2 shows the Kolmogorov-Smirnov p-values testing the hypothesis that our statistic follows the χ^2 distribution with 7 degrees. We find a good fit to the χ^2 distribution in general even for $n_1 = n_2 = 350$ ($n = 700$), getting better as α decreases. Figure 6.3 (left) illustrates a good fit to the χ^2 distribution for $\pi_0 = 0,4$, $n = 700$, and $\alpha = 0,09$. Figure 6.3 (right) depicts empirical power functions of our suggested statistic for $\pi_0 = 0.4$ and $n = 2,000$ at significance level 5% with asymptotic critical value. Different values of α do not produce substantial differences in the power of the test for n large.

Table 6.2: $H_0 : X_1 = X_2$ in Example 6.2: Kolmogorov-Smirnov p-values testing the null hypothesis that our statistic in a homogeneity test framework follows the expected χ^2 distribution with 7 degrees of freedom.

	n					
α	700	800	1,000	2,000	3,000	4,000
0.01	0.805	0.545	0.286	0.751	0.536	0.688
0.02	0.800	0.538	0.287	0.750	0.535	0.683
0.03	0.797	0.532	0.287	0.748	0.534	0.679
0.04	0.795	0.525	0.288	0.747	0.533	0.674
0.05	0.793	0.518	0.289	0.745	0.532	0.670
0.06	0.791	0.511	0.289	0.744	0.532	0.665
0.07	0.789	0.504	0.290	0.742	0.531	0.660
0.08	0.786	0.500	0.292	0.741	0.530	0.656
0.09	0.784	0.498	0.294	0.739	0.529	0.651
0.10	0.782	0.496	0.296	0.737	0.529	0.646
0.20	0.728	0.356	0.325	0.719	0.513	0.598
0.30	0.712	0.269	0.365	0.698	0.500	0.578
0.40	0.684	0.220	0.406	0.637	0.487	0.566
0.50	0.547	0.183	0.419	0.593	0.475	0.558
0.60	0.293	0.130	0.432	0.566	0.452	0.551
0.70	0.157	0.089	0.446	0.541	0.424	0.543
0.80	0.058	0.052	0.465	0.491	0.416	0.535
0.90	0.024	0.030	0.503	0.426	0.373	0.515

6.4 Application

Fingerprint identification, a pivotal aspect of forensic science, hinges on analyzing minutiae – specific patterns and details of the ridges in a fingerprint, including points where the ridges end (ridge endings) or split (bifurcations). These minutiae are not only immutable over one's lifetime but are also complex enough to allow for differentiation between individuals, including between genders. From a Brazilian police record, we collected a sample of 7,031 minutiae, where 3,702 are from males and 3,329 from females. In this investigation, as we considered characteristic points of rare occurrences, we found expected frequencies of less than five in ten cells of the contingency table (Table 6.3). In testing homogeneity using the asymptotic result (6.9), we find $D_\alpha(\hat{p}||p) = 50.47$ (p-value = 0.0019) with $\alpha = 0.1$, $n = 7,031$, and $m = 26$. Considering $\alpha = 0.2$ we find $D_\alpha(\hat{p}||p) = 50.69$ (p-value = 0.0018).

We performed a Monte Carlo simulation with $1,000$ replicates to validate this result based on the expected (marginal) frequencies shown in Table 6.3. The asymptotic critical value from the $\chi^2_{(25)}$ distribution is 37.6525 for a significance level of 5%. The percentages of cases above 37.6525 in our Monte Carlo realizations were close to the nominal level of 5%, that is, 4.7% and 6.2%, respectively, for $\alpha = 0.1$ and 0.2. Thus, we are not rejecting the null hypothesis due to the lack of robustness because of small counts.

Indeed, this result confirms previous studies suggesting subtle differences in the minutiae patterns between males and females (Gutiérrez-Redomero et al., 2011; Shehu et al., 2018b; Rim, Kim, and Hong, 2021). This gender-related divergence in minutiae distribution is reflected in the number of minutiae present in a sample, potentially serving as an ancillary tool for gender determination in forensic investigations.

6.5 Conclusion

Our study explored using Rényi divergence as a feasible alternative to the traditional chi-squared test for assessing goodness-of-fit, homogeneity, and independence in contingency tables. We demonstrated that the Rényi divergence statistic effectively approximates a chi-squared distribution, including cases with small expected frequency counts. Our theoretical claims are supported by a Monte Carlo analysis conducted on two examples with low expected absolute frequencies. The critical insight from this empirical study is that a small Rényi index α enhances the alignment of our proposed statistic with the chi-square distribution, especially in situations involving low expected frequencies. Additionally, we applied our methodology to a real-world case by examining fingerprint data from a Brazilian police database to identify gender-specific differences in the distribution of minutiae, including those infrequently observed in the general population. We conducted a Monte Carlo simulation to corroborate our asymptotic result, comparing the theoretical asymptotic critical value at a 5% significance level with the empirical distribution. The findings revealed that the percentages of simulated realizations exceeding the asymptotic critical value closely aligned with the nominal level of 5%, particularly when

Table 6.3: Observed × Expected minutiae distribution by gender from a Brazilian police record.

	observed			expected	
Minutiae	Male	Female	Total	Male	Female
Ridge ending-B	853	733	1586	835.069	750.931
Ridge ending-C	789	726	1515	797.686	717.314
Bifurcation	698	663	1361	716.601	644.399
Convergence	508	554	1062	559.170	502.830
Fragments	154	118	272	143.215	128.785
Points	136	73	209	110.045	98.956
Breaks	123	70	193	101.619	91.381
Enclosures	71	71	142	74.767	67.233
Overlaps	72	51	123	64.763	58.237
Conjugations	62	44	106	55.812	50.188
Crossbars	60	39	99	52.126	46.874
Bridges	39	44	83	43.702	39.298
Tripods	28	30	58	30.538	27.462
Opp. Bifurcs.	19	21	40	21.061	18.939
Others minutiae	16	19	35	18.428	16.572
Docks	17	16	33	17.375	15.625
Appendage	14	12	26	13.690	12.310
Angular	12	10	22	11.584	10.416
Double bifurc.	7	8	15	7.898	7.102
Ms	5	9	14	7.371	6.629
Emboques	6	6	12	6.318	5.682
Double convert.	4	5	9	4.739	4.261
Returns	6	3	9	4.739	4.261
Trifurcations	2	1	3	1.580	1.420
Needles	1	2	3	1.580	1.420
Numericals	0	1	1	0.526	0.474
Total	3702	3329	7031	3702	3329

employing a small α (say, less than 0.2).

Chapter 7

Final Considerations

The applications of convolutional neural networks to fingerprints are still an incipient field. Previous studies had used samples made up of plain fingerprints, with our sample of rolled impressions of Brazilians it was possible to demonstrate the encouraging results obtained with CNNs whether with partial or complete fingerprints and regardless of the population that originated them. Important human attributes useful for identification were strongly predicted by CNNs, including finger type. Before CNNs, no other search feature had previously been able to indicate this attribute.

The frequency distributions of Level 2 Details and their relationships with levels 1 and 3, as well as the type of sex and finger, were determined. Until then, these detailed behaviors were open questions when considering the Brazilian population. This knowledge is fundamental for the development, in Brazil, of a holistic approach to the conclusions derived from forensic fingerprint examinations. Given the frequencies in which each minutiae occurs, we classify them as common, rare, and very rare.

The proposal to create BRAFing (Brazilian Fingerprint Dataset), authorized on 11/30/2023 by the Research Ethics Committee of the Faculty of Medicine of UnB, composed of 1000 fingerprints from 50 Brazilian males and 50 females, is innovative. From there, a random and representative sample of Brazilian fingerprints will be available for ethical, non-commercial, and non-profit research.

The studies proposed here, in addition to contributing to the body of knowledge associated with Brazilian fingerprints, also seem to inaugurate, in Brazil, the production of knowledge in forensic statistics in the most obvious place possible, which is the Statistics Department of a university. It is expected that the results obtained in this dissertation will be useful in the field of forensic identification and that they will stimulate the production of national knowledge in the area.

The choice of the Rényi statistic is based on the possibility of expected frequencies below five, which could compromise the reliability of the chi-square statistic. Tests reveal reasonable fit in certain scenarios, even when expected frequencies are less than five. The text also deals with the application of these methods to identify fingerprints, with significant results that indicate disparities between men and women. A Monte Carlo simulation validates these results, highlighting the importance of robustness in situations with low counts. The study suggests possible uses in determining gender in forensic investigations.

References

(N.d.[b]). `https://www.gambitid.com/international-fingerprinting-faq.html`. Accessed: 2023-08-28.

Abdullah, S. F., A. F. N. A. Rahman, and Z. A. Abas (2015). "Classification of gender by using fingerprint ridge density in northern part of Malaysia". In: *ARPN Journal of Engineering and Applied Sciences* 10.22, pp. 10722–10726.

Araújo, M. E. C. and L. Pasquali (2012). *Datiloscopia - A determinação dos dedos*. LABPAM.

Aria, M. and C. Cuccurullo (2017). "Bibliometrix: An R-tool for comprehensive science mapping analysis". In: *Journal of Informetrics* 11.4, pp. 959–975.

Ashbaugh, D. R. (1999a). *Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology*. Boca Raton: CRC Press.

— (1999b). *Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology*. Boca Raton: CRC Press.

Ayala, J. A. G. (1981). *Determinación de mano y dedo: técnicas para la formación de hipótesis en los procesos de identificación de huellas digitales anónimas*. Dirección de la Seguridad del Estado.

Baglivo, Jenny, Donald Olivier, and Marcello Pagano (1992). "Methods for exact goodness-of-fit tests". In: *Journal of the American Statistical Association* 87.418, pp. 464–469. URL: `http://www.jstor.org/stable/2290278`.

Balthazard, V. (1911). "De l'identification par les empreintes digitales". In: *Comptes Rendus, des Academies des Sciences* 152, pp. 1862–1864.

Barnes, J. G. (2011a). *Chapter 1: History, in Fingerprint sourcebook*. Washington: National Institute of Justice.

— (2011b). *Chapter 1: History, in Fingerprint sourcebook*. Washington: National Institute of Justice.

Bartfay, E. and A. Donner (Dec. 2000). "The effect of collapsing multinomial data when assessing agreement". In: *International Journal of Epidemiology* 29.6, pp. 1070–1075. DOI: `https://doi.org/10.1093/ije/29.6.1070`.

Caballero, S. A. D. (2012). *Papiloscopia: Certeza ou dúvida? Apologia à micropapiloscopia*. Millennium.

Callon, M. (1999). "Actor-Network Theory: the Market Test". In: *Actor Network and After, J. Law and J. Hassard (Eds.)* Oxford, Keele, Blackwell, and the Sociological Review, pp. 181–195.

Cao, K. and A. K. Jain (2019). "Automated Latent Fingerprint Recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.4, pp. 788–800. DOI: `10.1109/TPAMI.2018.2818162`.

Champod, C. (1996). "Reconnaissance automatique et analyse statistique des minuties sur les empreintes digitales". PhD thesis. Institut de Police Scientifique et de Criminologie, Université de Lausanne.

Champod, C. et al. (2016). *Fingerprints and Other Ridge Skin Impressions*. 2nd. Boca Raton: CRC Press.

Chugh, T., K. Cao, and A. K. Jain (2018). "Fingerprint Spoof Buster: Use of Minutiae-Centered Patches". In: *IEEE Transactions on Information Forensics and Security* 13.9, pp. 2190–2202. DOI: `10.1109/TIFS.2018.2812193`.

Cummins, H. and C. Midlo (1943). *Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics*. New York: Dover.

Cummins, H., W. J. Waits, and J. T. McQuitty (1941). "The breadths of epidermal ridges on the fingertips and palms: A study of variation". In: *American Journal of Anatomy* 68.1, pp. 127–150. DOI: 10.1002/aja.1000680106.

Dankmeijer, J., J.M. Waltman, and A.G. De Wilde (1980). "Biological foundations for forensic identifications based on Fingerprints". In: *Acta Morphol. Neerlando-Scandinavia* 18, pp. 67–83.

Fournier, N. and A. Ross (2016). "Sex, Ancestral, and pattern type variation of fingerprint minutiae: a forensic perspective on anthropological dermatoglyphics". In: *Am. J. Phys. Anthropol.* 160, pp. 625–632. DOI: 10.1002/ajpa.22869.

Galton, F. (1892). *Fingerprints*. London: Macmillan.

Gao, M. et al. (2023). "Statistics of fingerprint minutiae frequency and distribution based on automatic minutiae detection method". In: *Forensic Science International* 344, p. 111572. DOI: 10.1016/j.forsciint.2023.111572.

Gautam, Shiva and George Kimeldorf (1999). "Some Results on the Maximal Correlation in $2 \times k$ Contingency Tables". In: *The American Statistician* 53.4, pp. 336–341. DOI: https://doi.org/10.1080/00031305.1999.10474486.

Gnanasivam, P. and S. Muttan (2012). "Fingerprint Gender Classification Using Wavelet Transform and Singular Value Decomposition". In: *IJCSI International Journal of Computer Science Issues* 9.2. DOI: https://doi.org/10.48550/arXiv.1205.674.

Gnanasivam, P. and R. Vijayarajan (2019). "Gender classification from fingerprint ridge count and fingertip size using optimal score assignment". In: *Complex Intell. Syst.* 5, pp. 343–352.

Gupta, S.R. (1968). "Statistical survey of ridge characteristics". In: *Int. Crim. Polic. Rev.* 218, pp. 130–134.

Gutiérrez, E. et al. (2007). "Biological variability of the minutiae in the fingerprints of a sample of the Spanish population". In: *Forensic Science International* 172, pp. 98–105.

Gutiérrez-Redomero, E. et al. (2008). "Variability of fingerprint ridge density in a sample of Spanish Caucasians and its application to sex determination". In: *Forensic Science International* 180.1, pp. 17–22.

Gutiérrez-Redomero, E. et al. (2013). "A Comparative Study of Topological and Sex Differences in Fingerprint Ridge Density in Argentinian and Spanish Population Samples". In: *Journal of Forensic and Legal Medicine* 20.5, pp. 419–429.

Gutiérrez-Redomero, Esperanza et al. (2011). "Distribution of the minutiae in the fingerprints of a sample of the Spanish population". In: *Forensic Science International* 208.1, pp. 79–90. DOI: `https://doi.org/10.1016/j.forsciint.2010.11.006`.

Gutiérrez-Redomero, Esperanza et al. (2012). "Are there population differences in minutiae frequencies? A comparative study of two Argentinian population samples and one Spanish sample". In: *Forensic Science International* 222, pp. 266–276. DOI: `https://doi.org/10.1016/j.forsciint.2012.07.003`.

Haberman, Shelby J. (1988). "A warning on the use of chi-squared statistics with frequency tables with small expected cell counts". In: *Journal of the American Statistical Association* 83.402, pp. 555–560. URL: `http://www.jstor.org/stable/2288877`.

Henry, E. R. (1900). *Classification and Uses of Fingerprints*. London: Routledge & Sons.

Hsiao, Chung-Ting et al. (2022). "Application of Convolutional Neural Network for Fingerprint-Based Prediction of Gender, Finger Position, and Height". In: *Entropy* 24.4, p. 475.

Iloanusi, O. N. and U. C. Ejiogu (2020). "Gender Classification from Fused Multi-fingerprint Types". In: *Information Security Journal: A Global Perspective* 29.5, pp. 209–219. DOI: `10.1080/19393555.2020.1741742`.

International Association for Identification (n.d.). 1973-5 Resolution.

Investigation, Federal Bureau of (1984). *Fingerprint Training Manual*. Washington: U.S. Department of Justice; U.S. Government Printing Office.

Jager, Leah and Jon A. Wellner (2007). "Goodness-of-Fit Tests via Phi-Divergences". In: *The Annals of Statistics* 35.5, pp. 2018–2053. DOI: `10.1214/0009053607000000244`.

Kapoor, N. and A. Badiye (2015). “An Analysis of Whorl Patterns for Determination of Hand”. In: *Journal of Forensic and Legal Medicine*, pp. 42–46.

Kingston, C.R. (1955). “Probabilistic analysis of partial fingerprint patterns”. In: *Fingerprint Identification Mag.* 36, pp. 3–8.

Langenburg, G. (2011). “Scientific research supporting the foundations of friction ridge examinations”. In: *The Fingerprint Sourcebook, chapter 1*. National Institute of Justice, Office of Justice Programs, U.S. Department of Justice, pp. 14–15.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. 2nd. New York: John Wiley & Sons.

Lin, C. H. (1981). “The similarity of fingerprints and the development of a single fingerprint filing and search arching system”. PhD thesis. Master’s Science [PhA Thesis], University of Illinois, Chicago.

Locard, E. (1914). “La preuve judiciaire par les empreintes digitales”. In: *Archives d’anthropologie criminelle, de médecine légale et de psychologie normale et pathologique* 29, pp. 321–348.

Maiste, Paul J. and Bruce S. Weir (2004). “Optimal testing strategies for large, sparse multinomial models”. In: *Computational Statistics & Data Analysis* 46.3, pp. 605–620. DOI: https://doi.org/10.1016/j.csda.2003.08.002.

Manual Técnico de Datiloscopia (2002). Rio de Janeiro.

Mariano, A. M. and M. S. Rocha (2017). “Revisão da Literatura: Apresentação de uma Abordagem Integradora”. In: *AEDM International Conference–Economy, Business and Uncertainty: Ideas for a European and Mediterranean industrial policy, Reggio Calabria, Italy, 2017*.

Mirakhmedov, Sherzod M. (2023). “On the asymptotic properties of a certain class of goodness-of-fit tests associated with multinomial distributions”. In: *Communications in Statistics - Theory and Methods* 52.3, pp. 602–622. DOI: 10.1080/03610926.2021.1918168.

Morales, D., L. Pardo, and I. Vajda (2000). “Rényi Statistics in Directed Families of Exponential Experiments”. In: *Statistics* 34.2, pp. 151–174. DOI: https://doi.org/10.1080/02331880008802324.

Morales, D. et al. (1994). "Asymptotic properties of divergence statistics in a stratified random sampling and its applications to test statistical hypotheses". In: *Journal of Statistical Planning and Inference* 38.2, pp. 201–221. DOI: https://doi.org/10.1016/0378-3758(94)90035-3.

Morales, D. et al. (2004). "Rényi statistics for testing composite hypotheses in general exponential models". In: *Statistics* 38.2, pp. 133–147. DOI: https://doi.org/10.1080/02331880310001634647.

Mudholkar, Govind S. and Alan D. Hutson (1997). "Continuity corrected approximations for and 'exact' inference with Pearson's X2". In: *Journal of Statistical Planning and Inference* 59.1, pp. 61–78. DOI: https://doi.org/10.1016/S0378-3758(96)00091-2.

Neumann, C., I. W. Evett, and J. Skerrett (2012). "Quantifying the Weight of Evidence from a Forensic Fingerprint Comparison: A New Paradigm". In: *Journal of the Royal Statistical Society. Series A* 175.2, pp. 371–415.

Neumann, C. et al. (2015). "Quantifying the Weight of Fingerprint Evidence through the Spatial Relationship, Directions and Types of Minutiae Observed on Fingermarks". In: *Forensic Science International* 248, pp. 154–171.

Ne'urim Declaration (1995). "Symposium Report". In: *J Forensic Identifi* 45, pp. 489–578.

Noughabi, Hadi Alizadeh and Narayanaswamy Balakrishnan (2016). "Tests of goodness of fit based on Phi-divergence". In: *Journal of Applied Statistics* 43.3, pp. 412–429. DOI: 10.1080/02664763.2015.1063116.

Okajima, M. (1970). "Frequency of forks in epidermal ridge minutiae in the fingerprint". In: *Am. J. Phys. Anthrop.* 32, pp. 41–48.

Ordem de serviço, da Polícia Civil do DF (n.d.). publicada em 5 de março de 2008.

Osterburg, J.W. et al. (1977). "Development of a mathematical formula for the calculation of fingerprint probabilities based on individual characteristics". In: *J. Am. Stat Assoc.* 72, pp. 772–778.

Pardo, Leandro and Nirian Martín (2021). “Robust Procedures for Estimating and Testing in the Framework of Divergence Measures”. In: *Entropy* 23.4. DOI: https://doi.org/10.3390/e23040430.

Ramanjit, K. and R. K. Garg (2011). “Determination of Gender Differences from Fingerprint Ridge Density in Two Northern Indian Populations”. In: *Probl. Forensic Sci.* 85, pp. 5–10.

Ray, S. (1989). “On a theoretical property of the Bhattacharyya coefficient as a feature evaluation criterion”. In: *Pattern Recognition Letters* 9, pp. 315–319. DOI: https://doi.org/10.1016/0167-8655(89)90059-7.

Rim, Beanbonyka, Junseob Kim, and Min Hong (2021). “Fingerprint classification using deep learning approach”. In: *Multimedia Tools and Applications* 80, pp. 35809–35825. DOI: https://doi.org/10.1007/s11042-020-09314-6.

Rivaldería, Noemí et al. (2017). “Study of Fingerprints in Argentina Population for Application in Personal Identification”. In: *Science & Justice* 57.3, pp. 199–208.

Sahu, S., A. Prabhakar Rao, and S. T. Mishra (2015). “Fingerprints Based Gender Classification Using Adaptive Neuro Fuzzy Inference System”. In: *IEEE - 2015 International Conference on Communications and Signal Processing*, pp. 1218–1222.

Salicru, M. et al. (1994). “On the Applications of Divergence Type Measures in Testing Statistical Hypotheses”. In: *Journal of Multivariate Analysis* 51.2, pp. 372–391. DOI: https://doi.org/10.1006/jmva.1994.1068.

Santamaría, F.A. (1955). “A new method for evaluating ridge characteristics”. In: *Fingerprint Identification Mag.* 36, pp. 3–8.

Sarkar, N.Ch. (2004). *Finger Ridge minutiae: Classification Distribution and Genetics,* Calcutta: Anthropological Survey of India.

Schweppe, F. C. (1967). “On the Bhattacharyya distance and the divergence between Gaussian Processes”. In: *Information and Control* 11, pp. 373–395. DOI: https://doi.org/10.1016/S0019-9958(67)90610-9.

Sclove, S. (1979). "The occurrence of fingerprint characteristics as a two-dimensional process". In: *J. Am. Stat Assoc.* 74, pp. 588–595.

Scotcher, K. and R. Bradshaw (2018). "The analysis of latent fingermarks on polymer banknotes using MALDI-MS". In: *Scientific Reports* 8, p. 8765. DOI: `https://doi.org/10.1038/s41598-018-27004-0`.

Shehu, Y. I. et al. (2018a). "Detailed Identification of Fingerprints Using Convolutional Neural Networks". In: *17th IEEE International Conference on Machine Learning and Applications*, pp. 1161–1165.

Shehu, Yahaya Isah et al. (2018b). "Detailed Identification of Fingerprints Using Convolutional Neural Networks". In: *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1161–1165. DOI: `10.1109/ICMLA.2018.00187`.

Shehu, Y.I. et al. (2018c). "Detection of Fingerprint Alterations Using Deep Convolutional Neural Networks". In: *Proceedings of the International Conference on Artificial Neural Networks (ICANN 2018), Rhodes – Greece, 5th - 7th October 2018*.

— (2018d). "Sokoto Coventry Fingerprint Dataset". In: DOI: `ArXivpreprintarXiv:1807.10609`.

Shinde, M. K. and S. A. Annadate (2015). "Analysis of Fingerprint Image for Gender Classification or Identification: Using Wavelet Transform and Singular Value Decomposition". In: *IEEE - International Conference on Computing Communication Control and Automation* 9.2, pp. 650–654. DOI: `10.1109/ICCUBEA.2015.133`.

Silva Carvalho, Daniel da et al. (2022). "Outcome of sex determination from ulnar and radial ridge densities of Brazilians' fingerprints: Applying an existing method to a new population". In: *Science & Justice* 62.2. DOI: `https://doi.org/10.1016/j.scijus.2022.01.005`.

Singh, I., P.K. Chattopadhyay, and R.K. Garg (2005). "Determination of the hand from single digit fingerprint: a study of whorls". In: *Forensic Science International* 152.2–3.

Steffens, C. (1965). “Vergleichende Untersuchungen der Minutien der Fingerbeerenmuster bei Familien und eineiigen Zwillingspaaren”. In: *Anthrop. Anz.* 29, pp. 234–249.

Stoney, D.A. and J.L. Thornton (1987). “A systematic study of epidermal ridge minutiae”. In: *J Forensic Sci.* 32, pp. 1182–1203.

Suwarno, S. (2023). “Gender Classification Based on Fingerprint Using Wavelet and Multilayer Perceptron”. In: *Jurnal Dan Penelitian Teknik Informatika* 8.1, pp. 139–144. DOI: https://doi.org/10.33395/sinkron.v8i1.11925.

SWGFAST (2002). *Friction Ridge Examination Methodology for Latent Print Examiners*. Scientific Working Group on Friction Ridge Analysis Study and Technology.

Team, R Core (2023). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. URL: https://www.R-project.org/.

The Federal District and the 12-Point Rule in Brazil, Dias da Costa (n.d.). N. JFI-2008-6-621.

Ultralytics.YOLOv5 (n.d.). https://github.com/ultralytics/yolov5. Accessed: 2023-04-27.

Vexler, Albert et al. (2014). “Computing critical values of exact tests by incorporating Monte Carlo simulations combined with statistical tables”. In: *Scandinavian Journal of Statistics* 41.4, pp. 1013–1030. DOI: https://doi.org/10.1111/sjos.12079.

Vonta, F., K. Mattheou, and A. Karagrigoriou (2012). “On Properties of the (Φ, α)-power divergence family with applications in goodness of fit tests”. In: *Methodology and Computing in Applied Probability* 14, pp. 335–256. DOI: https://doi.org/10.1007/s11009-010-9205-8.

Vucetich, J. (1904). “Dactiloscopía comparada, el nuevo sistema argentino”. In: *2. Congreso médico latino-americano, Buenos Aires,La Plata, 3-10 de abril de 1904*. National Institute of Justice, Office of Justice Programs, U.S. Department of Justice.

Xie, H. and Z. Lin (2020). “Gender Difference Analysis of Chinese Han Youth based on Fingerprint Ridge Density”. In: *E3S Web Conf.* 198, p. 04014. DOI: https://doi.org/10.1051/e3sconf/202019804014.

Yang, Pengfei and Biao Chen (2019). "Robust Kullback-Leibler Divergence and Universal Hypothesis Testing for Continuous Distributions". In: *IEEE Transactions on Information Theory* 65.4, pp. 2360–2373. DOI: 10.1109/TIT.2018.2879057.

Yang, W. et al. (2018). "A fingerprint and finger-vein based cancelable multi-biometric system". In: *Journal of Forensic and Legal Medicine* 78, pp. 242–251. DOI: 10.1016/j.patcog.2018.01.026.