CROSS-DOMAIN DEEP FACE MATCHING FOR BANKING SECURITY SYSTEMS

JOHNATAN SANTOS DE OLIVEIRA

DISSERTACAO DE MESTRADO EM ENGENHARIA ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

CROSS-DOMAIN DEEP FACE MATCHING FOR BANKING SECURITY SYSTEMS

JOHNATAN SANTOS DE OLIVEIRA

ORIENTADOR: FLÁVIO ELIAS DE DEUS, Dr., UnB
CO-ORIENTADOR: ANDERSON DE R. ROCHA, Ph.D, UNICAMP

DISSERTAÇÃO DE MESTRADO EM
ENGENHARIA ELÉTRICA

PUBLICAÇÃO: 694/2018 DM PPGEE

BRASÍLIA/DF: MAIO/2018.

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

CROSS-DOMAIN DEEP FACE MATCHING FOR BANKING SECURITY SYSTEMS

JOHNATAN SANTOS DE OLIVEIRA

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA.

APROVADA POR:

FLÁVIO ELIAS GOMES DE DEUS, Dr., ENE/UNB
(ORIENTADOR)

DANIEL GUERREIRO E SILVA, Dr., ENE/UNB
(EXAMINADOR INTERNO)

FLÁVIO DE BARROS VIDAL, DR. CIC/UNB
(EXAMINADOR EXTERNO)

BRASÍLIA/DF, 23 DE MAIO DE 2018.

FICHA CATALOGRÁFICA

DE OLIVEIRA, JOHNATAN SANTOS

Cross-Domain Deep Face Matching for Banking Security Systems.

[Distrito Federal] 2018.

xiii, 82p., 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2018).

Dissertação de Mestrado - Universidade de Brasília.

Faculdade de Tecnologia. Departamento de Engenharia Elétrica.

1. Machine Learning	2. Face Recognition
3. Information Security	4. Neural Network
I. ENE/FT/UnB	II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

DE OLIVEIRA., J. S. (2018). Cross-Domain Deep Face Matching for Banking Security Systems. Dissertação de Mestrado em Engenharia Elétrica, Publicação 694/2018 DM PPGEE, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 82p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Johnatan Santos de Oliveira.

TÍTULO DA DISSERTAÇÃO DE MESTRADO: Cross-Domain Deep Face Matching for Banking Security Systems.

GRAU / ANO: Mestre / 2018

Johnatan Santos de Oliveira
QD 207 LT 4/6/8, Res. Imprensa IV, BL E Apto 1702
71926-250 Brasília - DF - Brasil.

DEDICATÓRIA

Ego meam familiam amo.
Filli mei.
Mea fortitudo,
et vita mea.

À José Adeilton de Jesus Oliveira
(Tio Didi - in memorian) e
Maria Souza Santos
(Vó Maria - in memorian).

AGRADECIMENTOS

Primeiramente à Deus, que ilumina nossas vidas e nos dá a força necessária para seguirmos em frente sem o medo de enfrentar os obstáculos da vida.

À minha esposa Edilaina Pelc e as minhas filhas, Malu e Nina. Pelo amor que nos envolve e nos mantém unidos, sem vocês jamais teria alcançado este objetivo de vida. E por suportar meus momentos de ansiedade, inquietude e o tempo que dediquei dando atenção a este trabalho.

Aos meus pais, José Adenilton e Maria Valdice, pelos ensinamentos de vida, valores e exemplos que carregarei e perpetuarei. E aos meus irmãos, Joubert Denner, Jéssica e Juliane Maria, pelo amor e apoio, vocês são essenciais. Juntos, somos mais fortes.

Agradeço ao meu orientador, Prof. Dr. Flávio Elias, pelo imenso auxílio na orientação deste trabalho, pelos conselhos e estímulos. Assim como agradeço ao meu co-orientador, Prof. Dr. Anderson Rocha, pelos ensinamentos e pela dedicação ao trabalho que desenvolve, que leva energia aos seus estudantes e parceiros, essenciais para a finalização deste trabalho. Sem vocês e sem o compartilhamento de conhecimento, este trabalho não seria possível.

Aos meus amigos do Banco do Brasil – Labbs (Laboratório Avançado do Banco do Brasil), Flávio Stutz, Caio Soares e o Luciano Callero. Sem a ajuda essencial de vocês, não teria maneira possível para a execução de todos os testes requeridos nesta dissertação. Espero contribuir igualmente no futuro.

Ao amigo Gustavo Botelho, pela orientação espontánea, pelas explicações e pela força em todos os momentos em que pensei não ser possível. Saiba que você ajudou muito mais do que imagina. Espero poder retribuir.

A Guilherme Folego, por toda ajuda no início dos meus estudos e testes.

Por fim, agradeço a Universidade de Brasília e a todos os professores pelo privilégio de receber um ensino de qualidade. A todos, os meus sinceros agradecimentos, pois sei que posso contar com cada um de vocês.

Johnatan Santos de Oliveira

ABSTRACT

CROSS-DOMAIN DEEP FACE MATCHING FOR BANKING SECURITY SYSTEMS

Author: Johnatan Santos de Oliveira
Supervisor: Flávio Elias de Deus
Programa de Pós-graduação em Engenharia Elétrica
Brasília, May 2018

Ensuring the security of transactions is currently one of the major challenges facing banking systems. The use of facial features for biometric authentication of users in banking systems is becoming a worldwide trend, due to the convenience and acceptability of this form of identification, and also because computers and mobile devices already have built-in cameras. This user authentication approach is attracting large investments from banking and financial institutions especially in cross-domain scenarios, in which facial images taken from ID documents are compared with digital self-portraits (selfies) taken with mobile device cameras. In this study, from the databases of the largest public Brazilian bank we collected a large dataset, called FaceBank, with 27,002 images of selfies and ID document photos from 13,501 subjects. Then, we assessed the performances of two well-referenced Convolutional Neural Networks models (VGG-Face and OpenFace) for deep face features extraction, as well as the performances of four effective classifiers (Linear SVM, Power Mean SVM, Random Forest and Random Forest with Ensemble Vote) for robust cross-domain face authentication. Based on the results obtained (authentication accuracies higher than 90%, in general), it is possible to conclude that the deep face matching approach assessed in this study is suitable for user authentication in cross-domain banking applications. To the best of our knowledge, this is the first study that uses a large dataset composed of real banking images to assess the cross-domain face authentication approach to be used in banking systems. As an additional, this work presents a study on the real needs in the future implementation of a biometric system proposing a cloud system to enable the adoption of biometrics technologies, creating a new model of service delivery. Besides that, proposes a secure and integrated ABIS Data Transmission subsystem model. All the analysis and implementation takes into account the total adherence and compatibility with the standards and specifications proposed by the Brazilian government, at the same time, establish mechanisms and controls to ensure the effective protection of data.

RESUMO

MATCHING DE FACES DE DIFERENTES DOMÍNIOS PARA SISTEMAS DE SEGURANÇA BANCÁRIO

Autor: Johnatan Santos de Oliveira
Orientador: Flávio Elias de Deus
Programa de Pós-graduação em Engenharia Elétrica
Brasília, Maio de 2018

Um dos principais desafios enfrentados pelo sistema bancário é garantir a segurança das transações financeiras. Devido à conveniência e aceitação, o uso de características faciais para autenticação biométrica de usuários em sistemas bancários está se tornando uma tendência mundial. Essa abordagem de autenticação de usuários está atraindo grandes investimentos de instituições bancárias e financeiras, especialmente em cenários de diferentes domínios, nos quais imagens faciais tiradas de documentos de identificação são comparadas com autorretratos digitais (selfies) tiradas com câmeras de dispositivos móveis. Neste estudo, coletamos das bases de dados do maior banco público brasileiro um grande dataset, chamado FaceBank, com 27.002 imagens de selfies e fotos de documentos de identificação de 13.501 sujeitos. Em seguida, avaliamos os desempenhos de dois modelos de Redes Neurais Convolucionais bem referenciados (VGG-Face e OpenFace) para extração de características profundas, bem como os desempenhos de quatro classificadores (SVM Linear, SVM Power Mean, Random Forest e Random Forest com o Ensemble Vote) para autenticação robusta de face em diferentes domínios. Com base nos resultados obtidos (precisões superiores a 90%, em geral), é possível concluir que a abordagem de matching de faces profundas avaliada neste estudo é adequada para autenticação de usuários em aplicações bancárias entre domínios. Até onde sabemos, este é o primeiro trabalho que usa um grande conjunto de dados composto por imagens bancárias reais para avaliar a abordagem de autenticação de face entre domínios. Além disso, este trabalho apresenta um estudo sobre as reais necessidades na implementação futura de um sistema biométrico, propondo um sistema de nuvem para permitir a adoção de tecnologias biométricas. Por fim, propõe também um modelo seguro e integrado de subsistema ABIS de transmissão de dados. Toda a análise e implementação leva em conta a total aderência e compatibilidade com padrões e especificações propostos pelo governo brasileiro.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, NOMENCLATURES AND ACRONYMS

ABIS: Automatic Biometric Identification System.

AI: Artificial Intelligence.

AML: Anti-money Laundering.

ANN: Artificial Neural Network.

CNN: Convolutional Neural Network.

CTF: Counter-terrorism Financing.

DAM: Domain Adaptation Machine.

DARN: Dual Attribute-aware Ranking Network.

EER: Equal Error Rate.

EFMA: European Financial Management Association.

FEBRABAN: The Brazilian Federation of Banks.

FMR: False Matching Rate.

FNN: Fast Neural Networks.

FNMR: False Non-matching Rate

FSB: Financial Stability Board.

HOG: Histogram of Oriented Gradients.

HSM: Hardware Security Module.

ID: Identity Document.

kNN: k-Nearest Neighbor.

LinearSVM: Linear Support Vector Machine.

MAN: Metropolitan Wireless Networks.

MCP: McCulloch-Pitts neurons.

ML: Machine Learning.

MLP: Multilayer Perceptron.

MRC: Maximal Rejection Classifier.

PCA: Principal Component Analysis.

PmSVM: Power Mean SVM.

RBC: Royal Bank of Canada.

ReLU: Rectified Linear Unit.

RF: Random Forest.

ROC: Receiver Operating Characteristics.

Selfie: Selfie-Portraits.

SCDL: Semi-coupled Dictionary Learning.

SVM: Support Vector Machine.

TMR: True Matching Rate.

Voting RF: Random Forest with Ensemble Vote Classifier.

1 Introduction

Among the human biometric characteristics, the face is perhaps the most used by the human beings themselves in the recognition of other individuals. In addition, it presents a greater universality when compared to the iris or the digital ones, since all people have one, even if in some cases with some type of deformation (HAXBY; HOFFMAN; GOBBINI, 2002).

Face detection consists in the use of computational methods that verify the existence of a face, in a given digital image, video or photograph (KASHEM et al., 2011). By using facial recognition techniques, we have some advantages over other biometric technologies and we can build systems that are less intrusive, more natural and transparent to the user (JAIN; LI, 2011), especially when taking into account the construction of modern high-resolution cameras, sensors and mobile devices.

Face recognition study is a comprehensive research area and involves disciplines such as image processing, pattern recognition, computer vision, and neural networks. In spite of being well studied and with several techniques and approaches already validated, some challenges are present in facial analysis, among which it is possible to cite (GOUVEIA; PAIVA, 2010): variation in the positioning of the existing face in the image; presence and/or absence of specific structural characteristics, such as beard, which can modify features related to face shape, size and color; facial expressions; obstruction of the face by an object in the environment; image conditions relating to lighting (spectrum and intensity) and camera-specific features.

Current high-performance facial recognition techniques are based on Convolutional Neural Networks - CNN. Systems such as Facebook DeepFace (TAIGMAN et al., 2014) and Google FaceNet (SCHROFF; KALENICHENKO; PHILBIN, 2015) provide great recognition accuracy, however, they have been trained with private data sets containing millions of social media images, far more than the datasets available for academic research (AMOS; LUDWICZUK; SATYANARAYANAN, 2016).

Biometric identifiers such as fingerprint, face, iris, voice and handwriting recognition are most commonly used as part of a multifactor authentication system, combined with a password or a token. These computational methods, especially those based on the use of the human face, are being applied in several types of security systems, in an attempt to make the already applied security filters efficient and to mitigate possible frauds. In this area of research, some subjects still need to be deeply investigated, and within these, we have the context that will be treated in this study: matching faces using images obtained through different domains (cross-domain), given their peculiar characteristics.

According to the Brazilian Federation of Banks, financial institutions must have effective and reliable methods to authenticate customers (FEBRABAN, 2017a). An effective authentication system is needed to protect customer data, prevent money laundering and terrorist financing, reduce fraud, inhibit identity theft and promote the legal enforceability of agreements and electronic transactions (COUNCIL, 2005). The risks of performing financial transactions with unauthorized or improperly identified persons in a banking environment can result in financial loss, damage to the reputation and image of the company, and breach of bank secrecy. Concerning fraud, an increase in the occurrences of identity fraud has been observed around the world in the last years.

Identity fraud has grown steadily over the past 10 years and the estimated damages, only in the United Kingdom, have reached about 5.4 billion pounds per year on average during this period (BUTTON et al., 2016). Despite the efforts of the industry, in 2017, criminals successfully targeted 2 million more victims and stole about $16 billion dollars, a nearly billion-dollar year-to-year increase and some 15.4 million consumers were victims of identity theft or fraud in 2016 (PASCUAL; MARCHINI; MILLER, 2017).

A recent report from leading market analysts suggested that the rapid digitization of consumers' lives and enterprises' records will increase the cost of data breaches to 2.1 trillion dollars globally by 2019, almost four times the previously estimated cost of breaches for 2015. This report also highlighted the increasing professionalism of cybercrime. We are seeing, in an epidemic way, fraud attempts increase year by year, with systematic methods, even specialized in some cases (MOAR, 2015). The ability to successfully use facial biometrics as primary or even complementary authentication is inevitable.

In this context, banks are investing in robust methods for face authentication in order to increase the user experience of their systems, especially in mobile banking, as well as to prevent frauds. A tendency nowadays in the financial industry is the usage of facial images from different sources (cross-domain problem), usually selfie-portraits (selfies) and ID documents, for user authentication in order to allow automated opening of checking accounts, authorization of financial transactions and registration of mobile devices (FOLEGO et al., 2016).

In this study, we evaluate two well-referenced CNNs previously trained on large face databases from the literature, VGG-Face (SIMONYAN; ZISSERMAN, 2014) and OpenFace (AMOS; LUDWICZUK; SATYANARAYANAN, 2016), for face authentication, by extracting deep and robust features from the facial images obtained from the selfies and ID documents, and by training effective classifiers to identify genuine and imposter cross-domain matchings, a complex problem caused by the significant differences found in facial images captured from very different sources. While VGG-Face is an extremely robust and well-referenced CNN for face recognition and authentication, OpenFace is a new and open source CNN (thus, more suitable for use in commercial systems), trained following a slightly different learning paradigm. Results show that both neural networks present great accuracy rates and low processing times, being suitable for use in a real banking security systems. Although VGG-Face has been a bit more accurate than OpenFace, the latter is more efficient and therefore more appropriate for mobile banking.

The processes of user authentication, face recognition, and authorization of financial transactions using biometric features, which we are discussing in this study, must be mapped within a robust biometric ecosystem. This ecosystem will allow, among other factors, adherence to security standards and policies, interoperability with legacy systems and maintenance. The importance of building an efficient, scalable and secure biometric system is often minimized, leading to problems in the use of biometric services, including scalability, interoperability, maintenance of services and, above all, security problems. When considering the design and architecture of any biometric system, a holistic approach and the assurance that security will be incorporated from the beginning is of the highest importance (ROBERTS, 2007).

In an attempt to support the creation of a biometric system in the future, this study also presents an investigation about the real needs in the implementation of this type of system and proposes a cloud system to allow the adoption of biometric technologies, creating a new service model. In addition, it proposes a secure and integrated

ABIS Data Transmission subsystem model with total adherence and compatibility with the well-known standards as well as the specifications proposed by the Brazilian government, while at the same time, establishes mechanisms and controls to ensure the effective protection of data, as well as reducing the risk of losses, changes, and unauthorized access and ensuring non-repudiation.

1.1 Justification

> *"Fraudsters' and money launderers' use of advanced techniques, including AI, is accelerating. Cutting-edge AI isn't optional – it's required for investigators to be effective in anticipating and responding to these changes."* - Gayle Sheppard, Vice President, New Technology Group General Manager, Intel.

Brazilian financial institutions are among the largest corporations in the world and invest billions in innovation and digital strategy in order to guarantee personalized service with high availability and security in their electronic means of interaction available to customers, such as the internet, telephone, and smartphone applications. However, criminals are increasingly specialized and bet on the diversification of attacks to bypass the security of these environments.

The increasing rates of the cost of electronic fraud to Brazilian banks, for instance, is close to 624 million dollars per year, causing banks to invest even more in ways to mitigate financial losses and in innovation in information security. It has been pointed out that for every US$ 100 in transactions worldwide, US$ 0.06 is lost to fraud and that with only the use of fingerprint verification, up to 80% of fraud attempts in the industry can be prevented (FEBRABAN, 2017a, 2017b).

From 2016 to 2017, there was a 96% increase in the volume of transactions in mobile banking, reaching R$ 21.9 billion, accounting, together with internet banking, for more than 34% of financial transactions. In 2017, there were R$ 65 billion worth of transactions in the online environment. According to most Brazilian banking institutions, Mobile banking is the preferred channel for access to banking services and transaction volume has quadrupled in the last three years. The confidence in the digital channel was reinforced by the security and good usability offered by mobile technology, coupled with greater population access to the internet and the popularization of the smartphone. In 2016, R$ 18.6 billion was invested in technology by Brazilian banks, an increasing number over the years (FEBRABAN, 2017a).

Given the emergent mobile banking in Brazil, which increased its traded volumes by four times only in the last three years, some banks are already using fingerprint authentication in order to achieve a better level of security, as in other countries. The customer, if previously registered, can carry out financial transactions on mobile devices or even at ATMs by presenting their registered finger to the sensor. However, not all mobile phones being used in poor countries have fingerprint sensors, despite all of them having digital cameras. In the case of ATMs, people need to touch a (not always clean) surface, making customers often unsatisfied with the identification system. Besides, ID documents do not usually have their holder's fingerprints, making it unfeasible, for instance, to open checking accounts through mobile devices by means of fingerprint matching. Due to all these reasons and considering the convenience of using facial features for people authentication, banking institutions, following their digital strategies, are increasingly implementing tools to allow automatic opening of checking accounts, authorizing transactions and authenticating devices totally online through smartphones by means of face authentication.

1.2 Digital strategy

New technologies are definitely transforming the way business is done in the financial industry. From cloud computing to robotics, artificial intelligence to cognitive computing, these disruptive changes are increasingly present in the financial market. It is hoped that these disruptors will have a major impact on the future of financial organizations. According to Steven Ehrenhalt, Global Finance Transformation Leader at Deloitte, if business leaders want to compete in the digital world, they will need to process more information more efficiently and create insights quickly (EHRENHALT, 2017).

In a new approach to Banco do Brasil's digital strategy, launched at the end of the year 2016 was the opening of current accounts totally online and through a smartphone, making it so that the customer interested in the service does not need to go to a physical agency to present the necessary documentation. Despite the practicality for the client and the economy on the side of the institution, this type of process needs caution, since in order to mitigate fraud, proof of legitimacy of the documentation, proof of life of the individual and proof that the client is actually the bearer of the documents presented are all necessary, actions that are primarily done by an employee of the institution.

In this new context, people interested in such a service do not need to go to a physical branch to present the required documentation. Instead, by using their mobile phones, they can take photographs of their ID documents, containing their facial images, and a digital self-portrait (selfie), proving the possession of the document by its legal owner. The matching of the faces in the photographs can occur directly on the device as well as on the bank server. If the matching of the images occurs successfully, the user account or further action is created. This may still be a prior process, requiring human action to verify identity or document fraud before finalizing the process. Figure 1.1 illustrates the explained process of matching faces from a selfie and an ID document. Cases like this show that face is a tendency as a biometric trait for people authentication in banking environments.

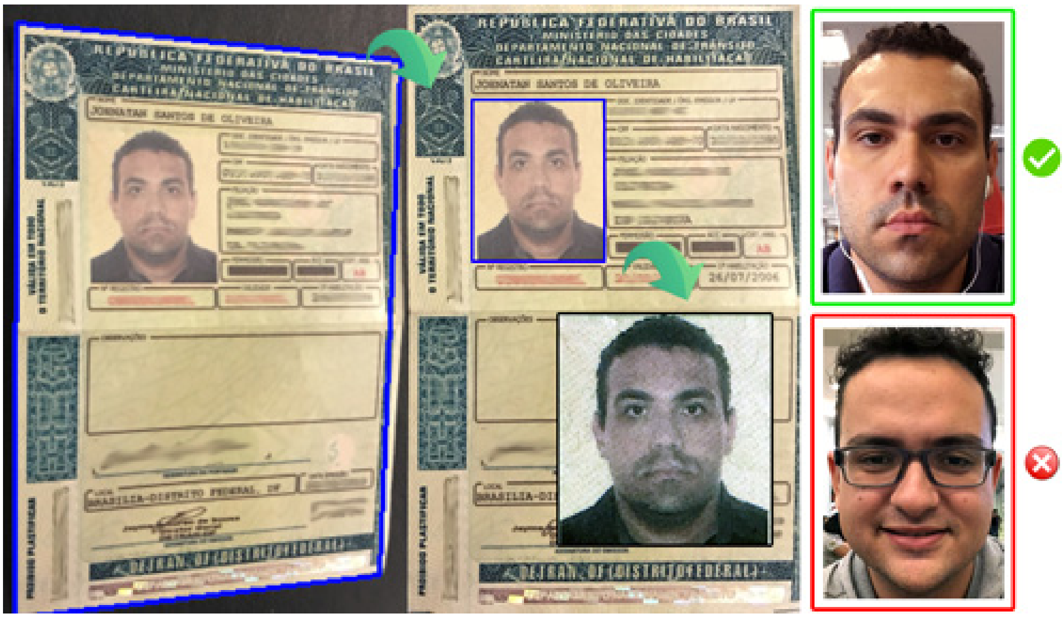

REPUBLICA FEDERATIVA DO BRASIL
MINISTERIO DAS CIDADES
DEPARTAMENTO NACIONAL DE TRANSITO
CARTEIRA NACIONAL DE HABILITAÇÃO
JOHNATAN SANTOS DE OLIVEIRA
26/07/2006
BRASILIA-DISTRITO FEDERAL, DF
DETRAN-DF(DISTRITO FEDERAL)

Figure 1.1: Illustration of the matching process of faces from an ID document and two selfies taken with smartphones. After detecting, cropping and normalizing the face in the document, it is matched with face in the selfie in order to authenticate the person, validating the document. It is possible to observe the different visual aspects of the facial images from the two kinds of images taken with the mobile phone: ID document and selfie.

The numbers in this new approach are quite interesting. The mobile channel of Banco do Brasil has an average of 8.5 million unique users per month and in the period between November 2016 and May 2018, the major public Brazilian bank received about 2.1 million requests for opening checking accounts through smartphones, all of them being manually inspected. Among these opened accounts, 60% are individuals in the age group of 21 to 40 years and 61% have a monthly income of up to R$ 4,000.00, with a total balance in accounts (only for this product), in the range of R$ 170 million.

Among this total of opened accounts, 22% of them were rejected by the human experts. Besides presenting different faces in the ID document and selfie (indicating fraud), most of the requests presented low quality facial images due to issues with illumination, facial occlusion, low resolution, or even due to scratched documents. Since the matching is still performed by humans and given the high number of requests being received, such a process is quite expensive for the bank, slow and also subject to failures. An automated method could, at least, automatically discard some of the requests, saving time and resources for the financial institution. By the end of 2018, the expectation is to reach a total of 3.35 million opened checking accounts through mobile devices. In Figure 1.2 it is possible to see a comparison between the percentage of accounts opened in the traditional way and via Mobile.

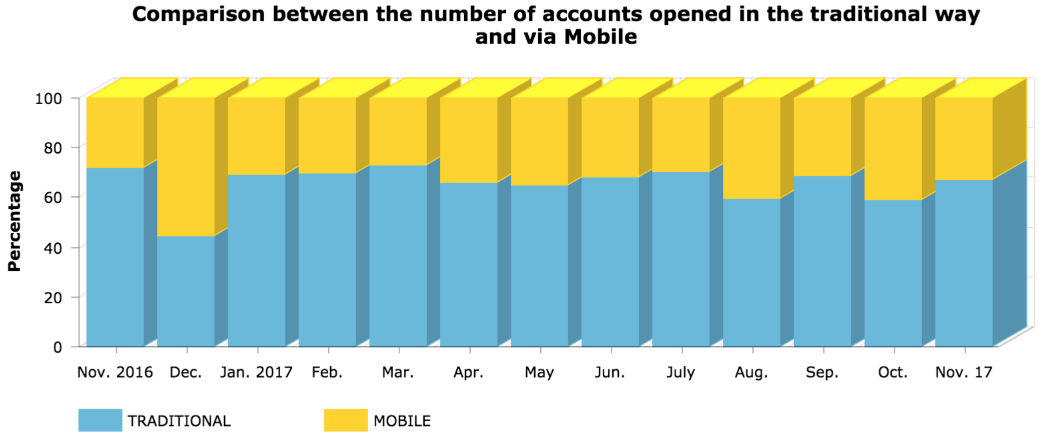


Figure 1.2: Comparison between the number of accounts opened in the traditional way (where the client is obliged to attend a physical bank branch) and via Mobile (the new approach to Banco do Brasil's digital strategy).

After the initial period of expansion (Nov and Dec 2016) where more than 50% of the accounts were opened by the new approach, the indicator maintained around 30%. Whether it is the primary scenario, face-to-face opening and bank service used autonomously or in this new scenario, where service opening and use are done electronically, facial recognition may become a factor of security to mitigate financial fraud.

In this way, it is important to think about the implementation model for fraud detection in banking systems that use facial recognition as an authentication model. This study addresses one of the problems in the implementation of this type of model - the use of cross-domain images - as well as its future implementation in a biometric system totally adhering to the well-known standards and the Brazilian regulations and specifications.

1.3 Artificial Intelligence and Machine Learning in Financial Services

> *"The last 10 years have been about building a world that is mobile-first. In the next 10 years, we will shift to a world that is AI-first."* - Sundar Pichai, CEO of Google.

What is notable is the rapid adoption of a variety of applications linked to Artificial Intelligence (AI) and Machine Learning (ML) in the financial industry and Fintechs (companies that combine financial services with innovative technologies). Analysis such as "Artificial intelligence and machine learning in financial services - Market developments and financial stability implications" created by the Financial Stability Board (FSB) and published on November 1, 2017, reveals a number of potential benefits and risks for financial stability that should be monitored as the technology is adopted in the coming years and as more data become available. They are (FSB, 2017):

- The more efficient processing of information may contribute to a more efficient financial system. These applications can help improve regulatory compliance and increase supervisory effectiveness.
- At the same time, network effects and scalability of new technologies may in the future give rise to third-party dependencies, leading to the emergence of new systemically important players that could fall outside the regulatory perimeter.
- The type of applications could result in new and unexpected forms of interconnectedness between financial markets and institutions.
- The lack of interpretability of those new methods could become a macro-level risk.
- As with any new product or service, there are important issues around appropriate risk management and oversight. It will be important to assess uses of AI and ML in view of their risks, including adherence to relevant protocols on data privacy, conduct risks, and cybersecurity.

In accordance with the FSB report, AI and ML applications prove to be valuable if their specific risks are properly managed and can thus be applied for a number of purposes across the financial system. Examples include, Sentiment indicators, Trading signals, Anti-money Laundering (AML), Counter-terrorism Financing (CTF) and Fraud Detection, Chatbots, Credit scoring and Regulatory compliance.

The report also examines the possibility of an important impact on financial markets, institutions, and consumers from a micro-financial (organization that offers financial services to low income populations) point of view:

- The algorithms have the potential to substantially enhance the efficiency of information processing, thereby reducing information asymmetries.
- Potential to strengthen the information function of the financial system. May enable certain market participants to collect and analyze information on a greater scale.
- May lower market participants' trading costs, enable them to adjust their trading and investment strategies in accordance with a changing environment in a swift manner, thus improving price discovery and reducing overall transaction costs in the system.
- Potential to enhance the efficiency and profitability of financial institutions, while reducing their costs and risks.
- The data intensity and open-source character may encourage collaboration between financial institutions and other industries, such as e-commerce and sharing economy businesses.

From consumers' and investors' points of view, if AI and ML reduce the costs and enhance the efficiency of financial services, consumers could obtain a number of benefits, such as lower fees and borrowing costs, wider access to financial services, and facilitate more customized and personalized financial services through big data analytics.

Another report published in the Journal of Financial Transformation 2017, reminds us that since the early 2000s, an extensive academic literature on the use of ML methods to detect credit card fraud and model credit risk has developed and it cites that banks have equipped their credit card payments infrastructures with monitoring systems, thus fraudulent transactions can then be blocked in real-time (LIEBERGEN et al., 2017). Among some examples of these studies we can cite (LAI et al., 2006), where they try to use a triple-phase neural network ensemble technique to design a credit risk evaluation system to discriminate good creditors from bad ones. Another study (AURIA; MORO, 2008), assesses company solvency using Support Vector Machines (SVM) and find that they produce more accurate out-of-sample predictions than existing techniques.

As a negative, difficulties to data sharing and data usage as well as long-established regulatory requirements have complicated innovation in the AML/CFT area. Nevertheless, ML systems have improved detection of money laundering activity significantly, due to their ability to identify complex patterns in data and combine transactions information at network speed.

Finally, it is clearly noticeable, as many studies over the years suggest and prove, that ML systems have improved several areas of the Financial industry and services leading to the necessary innovation for the substantial development of the area. However, there is still much to study.

1.3.1 Current and future scenario in Finance

According to the "Innovation in Retail Banking - The Emergence of New Banking Business Models", an October 2016 report, financial institutions believe that AI, Big Data, and open APIs will have a significant impact on the industry and understand the potential benefits, but that they are still hesitant to act. They are approaching it piecemeal, slowly building towards AI competency by stacking on more and more of the innovative technologies they know they will need. However, there are other significant impediments to progress. In particular, the study found that half of the banks listed their legacy systems as the biggest hurdle they face, followed by a lack of unified vision (44%) and a shortage of skills and experience (38%) (EFMA, 2016).

The financial services industry is investing much more in AI technologies than other industries, and these investments will continue to grow steadily as banking providers get closer to achieving their fully-functioning AI-driven systems (INFOSYS, 2017). In fact, only 16% of firms have incorporated any kind of ML into their investment strategies. Meanwhile, the remainder is either researching ways to do it (24%), would like to learn about how to do it (26%), or hasn't even thought about doing it yet (32%) (BLOOMBERG, 2017). Still, in the real world of applications based on ML and high impact in use, there are several use cases that we can mention.

The Royal Bank of Canada (RBC) is one example revealing how financial institutions are trying to embrace around their AI and tech challenges. RBC is heavily investing in the University of Toronto's Rotman School of Management's Creative Destruction Lab, a lab that has nurtured some 50 artificial intelligence companies (HACKL, 2017).

In another real case example, a system can monitor stock prices in real time and predict them based on the news stream. An experiment by SVM called the Arizona Financial Text System, or AZFinText, uses financial news stories to predict stock data and can trade S&P 500 securities more successfully than The Standard & Poor's 500 quants, an American stock market index based on the market capitalizations of 500 large companies - reaping an 8.5% return, compared to 6.44% from the next highest performing quant (SCHUMAKER; CHEN, 2010).

According to Mike Blalock, General Manager for Intel's Financial Services Industry division, banks across the world, especially in Europe, have begun replacing older statistical-modeling approaches with artificial intelligence and cognitive computing technology. By adopting this sort of intelligent-computing program that can solve and even anticipate complex problems, some banks have experienced a 10% increase in sales of new products, a 20% savings in capital expenditures and 20% increase in cash collections. Still, according to Blalock, Bank of America is launching an AI bot for its smartphone app sometime in the next year. Named Erica, the bot uses AI and predictive analytics to learn individual's personal spending habits (BLALOCK, 2017).

The potential for AI continues to grow every day. Projections indicate that the economic impact of the automation of knowledge will reach $6.7 trillion annually by 2025 and ML techniques will touch every piece of data in the data center by 2020. This has created a technology arms race and algorithmic competition as big companies as IBM, NVIDIA, Intel, and ARM strive to dominate the retooling of the computer industry to support ubiquitous ML workloads over the next 3-4 years (MCKINSEY, 2013; BRUECKNER, 2018).

1.4 Main Goals

Our main hypothesis and objective is the search for total automation of the process of cross-domain face authentication as a solution for account opening via mobile, cited in the introduction of this study (1.2). The reduction in the use of human and manual action to verify identity or document fraud before the finalization of the process, which nowadays raises the cost of the solution implemented in the market considerably, is an achievable goal. In an automated version of the solution, the ideal scenario would be to validate the customer interested in the product automatically, efficiently, quickly and maintaining interoperability between different types of systems, reducing human interaction by 90% or triggering it only in cases where the autonomous solution, due to several factors, could not identify or verify the faces involved in the process.

In this way, this primary objective leads us to other important points. In order to synthesize the objectives of this study and make clear what we are looking for, we list the main points below:

- How to mitigate the problem of using cross-domain images;
- Get a dataset, large enough and specific to the cross-domain problem;
- Evaluate CNN models for deep face features extraction;
- Evaluate the performance of several types of classifiers for cross-domain face authentication, looking for good accuracy;
- Find an efficient and more appropriate method for the mobile environment;
- Keeping in mind interoperability, secure development and full adherence to standards, discover what the real needs to be verified in the implementation of a biometric system in the future are;
- Ascertain how to establish mechanisms and controls to ensure the effective protection of data, as well as reducing the risks with total adherence with the world standards and Brazilian specifications.

1.5 Organization of the dissertation

This dissertation is divided into six chapters and a bibliography. The chapters present the main topics, the experiments, their results and the conclusions obtained in the research. Chapter 1 contains the introduction, a justification, the main purpose, an analysis of the possible impacts on the financial system, a study about the current and future scenario and this topic, which cites the organization of the study.

Chapter 2 deals with the state of the art of the subjects, assumptions, and concepts discussed in this study. Image Categorization, Machine Learning, and Neural Network concepts are reviewed, as well as information about classifiers, cross-domain images, the main context where we will apply our experiments, and the data augmentation process. Chapter 3 proposes and portrays the planned approach to trials, processes, models, scenarios, input data, and assumptions used to obtain results, while the chapter 4 presents the achievements, experiments, comparisons, and results obtained by this research.

Chapter 5 provides an analysis of the Automatic Biometric Identification System, proposes a cloud concept model in order to delivery biometric services and a secure and integrated Data transmission subsystem. Finally, chapter 6 summarizes the achievements of this research, describing them through an objective conclusion and the bibliography brings the list of references cited in this study.

2 A Bibliographic study

Visual recognition is one of the most important research topics in computer vision and machine learning. The main objective is to analyze a scene, recognizing all the objects in the image and the possible relationships between them. This brings us a series of problems in computer vision, which should be treated separately, in minor and manageable problems (DOAN; POULET, 2014).

The purpose of this chapter is to present the common bibliographic bases, the state of the art when dealing with the problem of Machine Learning, Neural Networks, and Image Categorization. Section 2.1 gives a review about Artificial Intelligence and Machine Learning and brings the introduction of ML Algorithms and Problems. Section 2.2 has an overview of Artificial Neural Network for Face detection and recognition, and also discusses Convolutional Neural Network. Section 2.3 provides a conceptual description of the basis of this study, the classification process, and some of its implementations. Finally, Section 2.4 deals with Cross-Domain issues and Domain Adaptation.

2.1 Artificial Intelligence and Machine Learning

"The science and engineering of making intelligent machines, especially intelligent computer programs." - Dartmouth Assistant Professor John McCarthy, 1956 - Known for coining the term Artificial Intelligence.

The first AI study was done by Warren McCulloch and Walter Pitts (MCCULLOCH; PITTS, 1943). Their proposal was a model of artificial neurons in which each neuron is characterized as being "on" or "off", with a switch to "on" occurring in response to stimulation by a sufficient number of neighboring neurons (RUSSELL; NORVIG, 2016). In 1958, (ROSENBLATT, 1958) presented an experimental Artificial Neural Network (ANN) model called "Perceptrons", adopting the MCP (McCulloch-Pitts) model and proposing to it a supervised training strategy.

But in 1969, (MINSKY, 1969) demonstrated that the perceptron presented by Rosenblatt can only solve linearly separable problems. A perceptron receives a vector of inputs and produces a single binary output. Each input has a weight and the output is determined if the weighted sum is greater or less than some threshold value. A simple change in the weights of any perceptron can cause a totally different output.

After an initial advance and success prediction, ML and AI had a reduction in the advance, in part because of the lack of sufficient computing power and data. In the 1980s, there was renewed funding and interest, during which many of the research concepts were developed for later breakthroughs (DORMEHL, 2017), such as the article published by John Hopfield, which draws attention to associative properties, showing the relationship between associative neural networks and physical systems (HOPFIELD, 1982).

A few years later (RUMELHART; HINTON; WILLIAMS, 1986) showed that Minsky and Papert's view of the perceptron was rather pessimistic and that multi-layered neural networks are capable of solving difficult problems. The first successful commercial expert system, R1 (MCDERMOTT, 1982), began operation at the Digital Equipment Corporation, the program helped configure orders for new computer systems and by 1986, it was saving the company an estimated $40 million a year. Today, AI is a thriving field with many practical applications and active research topics (GOODFELLOW; BENGIO; COURVILLE, 2016).

As well as the rise of the terms Big Data, Data Analysis and AI, the term Machine Learning has become popular in the last decades. The AI pioneer Arthur Samuel wrote in 1959 that machine learning is the "field of study that gives computers the ability to learn without being explicitly programmed" (SAMUEL, 1959). Computer scientists and statisticians are increasingly interested in the ML algorithms, mainly because of the performance and the ability to simplify immense datasets, providing the possibility for us humans to understand quality and make changes to improve the quality of a large mass of data (WANG; TAO, 2008).

As shown in (JORDAN; MITCHELL, 2015), ML as an important discipline for addressing both fundamental scientific and engineering questions. Also, they cite that as a field of study, ML sits at the crossroads of computer science, statistics and a variety of other disciplines concerned with automatic improvement over time, and inference and decision-making under uncertainty. Related disciplines include the psychological study of human learning, the study of evolution, adaptive control theory, the study of educational practices, neuroscience, organizational behavior, and economics. The introduction of ML enabled computers to tackle problems involving knowledge of the real world and make decisions that appear subjective (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.1.1 Machine Learning Algorithms and Problems

> *"The purpose of machine learning is to learn from training data in order to make as good as possible predictions on new, unseen, data."* - Jean-Francois Puget, PhD, Machine Learning Engineer - IBM.

In accordance with (STEEB, 2014), a learning algorithm is an adaptive method by which a network of computing units self-organizes to implement the desired behavior. This is done in some learning algorithms by presenting some examples of the desired input-output mapping to the network. A correction step is executed iteratively until the network learns to produce the desired response. Conceptually, ML algorithms can be viewed as searching through a large space of candidate programs, guided by training experience, to find a program that optimizes the performance metric (JORDAN; MITCHELL, 2015).

There are several categories of ML algorithms and techniques. These categories vary according to the level of human intervention required in labeling the data. Some of these techniques are dealt with below:

- **Supervised learning:** The algorithm is given enough training data which contains labels and will "learn" a general rule of classification that it will use to predict the labels. As such, supervised learning is one of the bases of this study.
- **Unsupervised learning:** Refers to situations where the data provided to the algorithm does not contain labels. The algorithm is asked to detect patterns in the data by identifying clusters of observations that depend on similar underlying characteristics.
- **Deep learning:** An approach to AI and a particular kind of ML that achieves great power and flexibility by representing the world as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones (GOODFELLOW; BENGIO; COURVILLE, 2016).

When training multilayer networks, the general practice is first to divide the data into three subsets. The first subset is the training set, which is used for computing and updating the network weights and biases. The second subset is the validation set. The errors in the validation set are monitored during the training process. The validation error normally decreases during the initial phase of training, as well as the training set error (STEEB, 2014). However, when the network begins to overfit [1], the errors in the validation set typically begins to rise. The network weights and biases are saved at the lowest level of the validation set error.

It is also possible to refine the understanding of ML tasks by classifying the possible types of problems that can be solved. The list of problems that have a solution from ML algorithms is very large, thus, these are only the most common:

- **Regression:** A supervised learning problem where the answer to be learned is a continuous value, a task of predicting numerical outcomes from various inputs.
- **Classification:** A supervised problem where the answer to be learned is one of the finitely many possible values. It is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations whose category membership is known (TANG; ALELYANI; LIU, 2014). When there are only two possible values we say it is a binary classification problem.
- **Clustering:** An unsupervised learning problem where the structure to be learned is a set of clusters of similar examples.

The most challenging version of visual recognition is general image categorization, a base of this study, which is essentially related to recognizing instances of extremely varied classes. Some studies related to this specific area receive much attention from the researchers and in some ways, it remains a challenging problem, requiring more efficient and effective methods.

Image categorization is one of the important subjects of AI and it is basically about a way of grouping images according to their similarity using various features of images like texture, color component, shape, edge, etc. Categorization process has various steps

[1] Use of models or procedures that include more terms than are necessary or use more complicated approach than are necessary (HAWKINS, 2004)

like image preprocessing, object detection, object segmentation, feature extraction, and object classification. Over the last few years, much research directly related to these problems has been made available, contributing to new, different and increasingly effective algorithms.

2.2 Artificial Neural Network for Face detection and recognition

> *"...a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs."* - Dr. Robert Hecht-Nielsen in "Neural Network Primer: Part I" by Maureen Caudill, AI Expert, Feb. 1989.

The field of Neural Networks has arisen from diverse sources, ranging from the fascination of humanity with understanding and emulating the human brain, to broader issues of copying human abilities such as speech and the use of language, to the practical commercial, scientific, and engineering disciplines of pattern recognition, modelling, and prediction (SPIEGELHALTER; TAYLOR; CAMPBELL, 1994). According to them, a broad class of techniques can come under this heading, but generally, neural networks consist of layers of interconnected nodes, each node producing a non-linear function of its input, combining the complexity of some of the statistical techniques with the machine learning objective of imitating human intelligence. The input to a node may come from other nodes or directly from the input data. Also, some nodes are identified with the output of the network.

According to the (SHIFFMAN, 2012), one of the key elements of an ANN is its ability to learn and is not just a complex system, but a complex adaptive system, meaning it can change its internal structure based on the information flowing through it. According to him, typically, neural networks are trained so that a particular input leads to a specific target output, based on a comparison of the output and the target, until the network output matches the target. Generally, a large amount of input and target data is required to train a network. Typically, this is achieved through the adjusting of weights, a number that controls the signal gain between the two neurons. If the network generates a "good" output according to the training cost function, there is no need to adjust the weights. However, if the network generates a "poor" output, then the system alters the weights in order to adapt and improve subsequent results.

In the last years, ANN has been used to perform complex functions in various fields, including time series forecasting, pattern recognition, identification, facial expression, classification, speech, vision, and control systems. Regarding the focus of this study, face detection is one of the most relevant applications of image processing and biometric systems and ANNs have been used in the field of image processing and pattern recognition. Face detection is the first step in face recognition systems to localize and extract the face region from the image background. Different architectures and models of ANN were used for face detection and recognition, and they are used because these models can simulate the way neurons work in the human brain. This is the main reason for its role in face recognition (AL-ALLAF; TAMIMI; ALIA, 2013; KASAR; BHATTACHARYYA; KIM, 2016).

2.2.1 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a variation of the MLP, inspired by the biological process of visual data processing. A CNN is able to apply filters to visual data, maintaining the close relationship between the pixels of the image throughout the network processing. Since their introduction by (LECUN et al., 1998) in the late 1990s, in which the authors developed a neural architecture known as LeNet-5, the CNNs have demonstrated excellent performance at tasks such as hand-written digit classification and face detection. CNNs are widely used, especially in the applications of classification, detection, and recognition in images and videos. The name "Convolutional Neural Network" indicates that the network employs a mathematical operation called convolution - a specialized kind of linear operation (GOODFELLOW; BENGIO; COURVILLE, 2016).

In the last few years a series of articles on the subject have been proposed, and among them, (CIREGAN; MEIER; SCHMIDHUBER, 2012) demonstrate state-of-the-art performance on NORB and CIFAR-10 datasets. Most notably, (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) show record-beating performance on the ImageNet [2] 2012 classification benchmark, with their convnet model achieving an error rate of 16.4%, compared to the 2nd place result of 26.1%. Following on from this work, (GIRSHICK et al., 2014) have shown leading detection performance on the PASCAL VOC dataset.

According to (ROCHA, 2009), a CNN consists of multiple parts with different functions. Initially, it is common to apply so-called convolution layers to the input data.

[2] Imagenet is one of the largest and the hard computer vision challenge.

A convolutional layer is composed of several neurons, each responsible for applying a filter to a specific piece of the image. Each neuron is connected to a set of pixels of the previous layer and each of these connections is assigned a weight. The combination of the inputs of a neuron, using the respective weights of each of its connections, produces an output passed to the next layer. The weights assigned to the neurons of a neuron can be interpreted as a matrix that represents the filter of an image convolution in the spatial domain (also known as kernel or mask).

The exact procedure for convolving a Kernel (say, of size 16 x 16) with the input volume (a 256 x 256 x 3 sized RGB image, for example) involves taking patches from the input image of size equal to that of the kernel (16 x 16), and convolving between the values in the patch and those in the kernel matrix.

In CNNs, the properties pertaining to the structure of layers and neurons, such spatial arrangement, and receptive field values, are called hyperparameters. The main CNN hyperparameters are a receptive field (R), zero-padding (P), the input volume dimensions (Width x Height x Depth, or W x H x D) and stride length (S). While in the classical perceptrons formulation a neuron is completely connected to all neurons in the previous layer (the fully connected layer), on CNNs only a subset of inputs is connected to each neuron. With the change of architecture, the CNN begins to perform analysis of local receptive fields.

It is impractical to connect all neurons with all possible regions of the input volume (whole image). This would lead to too many weights to train and produce too high a computational complexity. Thus, instead of connecting each neuron to all possible pixels, we specify a 2D region called the “receptive field”, extending to the entire depth of the input, within which the encompassed pixels are fully connected to the neural networks input layer.

According to (ROCHA, 2009), neurons from the same layer are grouped into maps that are produced by grouping the neuron outputs, and together they cover a portion of the image that has been processed with a common filter. In order for a set of neurons of a given map to apply the same filter to different positions of the image, weights are shared during the training process. This sharing significantly decreases the number of parameters to be learned and the training time of the network, consequently.

Unlike the traditional process of computer vision, in which the initial model starts from the definition of filters or features to be used, in the convolution layers of the CNN, it is necessary to define only the architecture of the filters: quantity and sizes, stride and so on. The network learning process itself changes the weights throughout the training until the best values of the filters are found for the data set used.

In convolution results, it is common to apply the activation function. Present in each neuron it is responsible for applying a transformation in the received data. The most commonly used function is the Rectified Linear Unit (ReLU), which was first proposed for restricted Boltzmann machines (NAIR; HINTON, 2010) and then successfully used for neural networks (GLOROT; BORDES; BENGIO, 2011). The ReLU activation function is the identity for positive arguments and zero otherwise, and is one of several keys to the recent success of deep networks. It expedites convergence of the training procedure and leads to better solutions than conventional sigmoid like units (CLEVERT; UNTERTHINER; HOCHREITER, 2015; HE et al., 2015).

Another very important and commonly used layer after the convolution and activation layers is the pooling layer, used to reduce the spatial size of the matrices resulting from the convolution. Consequently, this technique reduces the number of parameters to be learned in the network, contributing to the control of overfitting. The pooling layers operate independently on each of the channels of the convolution result.

Unlike the convolutional layer, in which weights are connected only in one region, the fully connected layer, as the name itself suggests, is completely connected to the previous layer. They are typically used as the last layer of CNN and work in the same way as traditional neural networks. Since the fully connected layer comes after a convolutional or pooling layer, it is necessary to connect each element of the convolution output matrices to an input neuron.

(ROCHA, 2009) says that after the fully connected layer, the last step is the classification function. This layer is fundamental in training because it influences the learning of the filters and consequently the result of the network. Due to its simplicity and good results the SoftMax function can be used in this step, contributing to a faster training without loss of quality.

The architecture of a CNN is related to the way the layers described above are organized. In fact, the designer can arrange them in whatever way he thinks fit. Most architectures use an input layer, of course, followed by a block of N convolutional layers with ReLU activation connected to a pooling layer. This block is repeated N times along the network which in the end is connected to a fully connected layer to determine the final classification. All of these choices directly impact the number of weights that the network must train and consequently how much computing is required for these values to be well adjusted.

Different architectures repeat and combine functions. For example, as (ROCHA, 2009) describes, LeNet-5 (LECUN et al., 1998) has two convolution layers followed by pooling plus one convolution layer. The architecture called GoogLeNet (SZEGEDY et al., 2015) has five layers of convolution followed always by a pooling. One can think that the result of the repetition of this layer is a set of features specialized in the task in which it was trained.

There are several well-known architectures in the area of convolutional networks. Four of the most used are described below:

- **LeNet-5** (LECUN et al., 1998): As already mentioned, it was the first application of CNN that was actually successful. It was used to classify digits, and because of this, it is much used in tasks such as reading postal codes.
- **AlexNet** (KRIZHEVSKY; SUTSKEVER; HINTON, 2012): Work that popularized CNNs for computer vision. As already mentioned, it was the winning methodology of the ImageNet challenge in 2012. Its architecture is very similar to LeNet-5, but it is deeper, with more convolutional layers.
- **GoogLeNet** (SZEGEDY et al., 2015): Winning architecture of ImageNet 2014, it has 22 layers and almost 12x fewer parameters, and is used to assess its quality in the context of object detection and classification.
- **VGGNet** (SIMONYAN; ZISSERMAN, 2014): Had as its main contribution to show that the depth of the network is a critical component for a good performance. The standard VGGNet model has several layers of convolution and maxpooling. Due to its depth, VGGNet is very computationally expensive and it needs a lot of memory to compute over its parameters.

The use of frameworks is the simplest and fastest way to use CNNs. Some of these frameworks are quicker and easier to use, and can be performed only by choosing the parameters and entering the training data; however, some others require a little more work. In addition, you can or cannot use GPUs according to the tool used and availability on the machine. Below is the list of the three best-known frameworks:

- **Keras:** Written in Python, it is simple and very straightforward to use. It is necessary to have the Tensorflow installed on the machine.
- **Tensorflow:** Developed by Google, it is used expressively in the ML area. It requires more practice for full use, but it is more powerful than Keras. It also uses Python.
- **Caffe:** Uses C/C++ and it is possible to implement its own architecture. Easy to use and very similar to Keras.
- **Torch:** An open source ML library appreciated for flexibility and speed.

In this study, the Caffe framework was used to extract face features with VGG-Face, and Torch was used to extract the features via OpenFace.

2.3 Classification

Classification is considered an instance of supervised learning, where a training set of correctly labeled and identified observations is available; in other words, a problem of identifying to which of a set of categories a new observation belongs on the basis of a training set (PORWIK, 2017). In this way, a classification model tries to result in some conclusion from the observed values. Given one or more inputs, a classification model will attempt to predict the value of one or more results. An algorithm that implements classification is known as a classifier.

According to (LU; WENG, 2007), many factors, such as the spatial resolution of the remotely sensed data, different sources of data, a classification system, and availability of classification software must be taken into account when selecting a classification method for use. The question of which classification approach is suitable for a specific study is not easy to answer. Different classification results may be obtained depending on the classifier(s) chosen.

It is possible to separate classification models into two main approaches: Generative and Discriminative. According to (ROCHA, 2009), with the generative approach, we try to solve a problem with an emphasis on the data generation process under analysis. Usually, we model the system as a Joint Probability Function and, in this way, we can create artificial examples that can be inserted into the system. Examples of models that use the generative approach are Bayesian Classifiers, Markov Random Fields, and Gaussian Mixture Models. In the discriminative approach, we try to find the boundaries that best separate a set of classes from the problem. Classifiers such as SVMs use this approach.

There are several numbers of classification models. Below, some of them are described in a very simplified way:

- **Logistic Regression:** A discriminative classification technique that has a direct probabilistic interpretation. Well suited for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables (PENG; LEE; INGERSOLL, 2002).
- **Decision Tree:** According (SAFAVIAN; LANDGREBE, 1991), the most important feature of Decision tree classifiers is their capability to break down a complex decision-making process into a collection of simpler decisions, thus providing a solution that is often easier to interpret.
- **Naive Bayes:** A generative classifier based on the Bayes rule of conditional probability and works on a simple, but in a relatively intuitive, concept: simplify learning by assuming that features are independent given class. Often competes well with more sophisticated classifiers (RISH, 2001).
- **SVM:** The support vector machine is probably the most widely used classifier in varied ML applications. For problems that are not linearly separable, kernel SVM uses a “kernel trick” to implicitly map samples from input space to a high-dimensional feature space, where samples become linearly separable. Due to its importance, optimization methods have been widely studied, and efficient libraries are well developed (SUYKENS; VANDEWALLE, 1999; HSIEH; SI; DHILLON, 2014).
- **Random Forest:** (PAL, 2005) describes the Random Forest (RF) classifier as a tree-based classifier that consists of a combination of tree classifiers where each classifier is generated using a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to classify an input vector.

- **kNN:** k-Nearest Neighbors (kNN) is one of the most fundamental, oldest and simplest methods for pattern classifications. Generally used for classification studies where there is little or no prior knowledge about the distribution of the data. Most kNN classifiers use simple Euclidean distances to measure the dissimilarities between examples represented as vector inputs (COVER; HART, 1967; WEINBERGER; BLITZER; SAUL, 2006; PETERSON, 2009).

2.4 Cross-Domain issues and Domain Adaptation

When the work involves cross-domain images it is possible to consider a substantial increment in the challenge for image matching, which is already considered one of the fundamental problems in computer vision, pattern recognition and many real-world vision tasks (e.g., matching persons across ID photos and surveillance videos and digitized face images vs. live pictures). In cross-domain conditions, the classification algorithms have their performance reduced. This observation is easily explained by issues such as blur, illumination, alignment, noise, or facial expression (LIN et al., 2017).

In this context, as an example, we can mention: in still-video face retrieval, a newly rising task in visual surveillance, faces from still images captured under a constrained environment are used as the queries to find the matches of the same identity in unconstrained videos (LIN et al., 2017); or age-invariant and sketch-photo face verification tasks; User Self-Portrait Photographs and ID Document image verification (FOLEGO et al., 2016).

In (FOLEGO et al., 2016), they explore approaches to cross-domain face verification, comparing self-portrait photographs ("selfies") to ID documents. They approach the problem with proper image photometric adjustment and data standardization techniques, along with deep learning methods to extract the most prominent features from the data, reducing the effects of domain shift in this problem. However, their dataset was composed of few images (dozens of individuals) and not obtained from a real banking scenario.

Also exploring image enhancement with ACE, the extraction of features from VGG fc6 with ReLU, (FOLEGO et al., 2016) however, is different from our study. Their results showed to be better with the use of L1 normalization, while in ours, the L2 normalization showed better accuracy. In the classification step, another point with fundamental differences, they explore three SVM techniques (SVM with linear decision function, with radial basis function (RBF) kernels, and with Logistic Regression (LR)), while in this study we explore classifiers with different characteristics, as can be seen in 3.6.

When dealing with typical cross-domain issues such as blur, illumination, alignment, noise, or facial expression, the method proposed in (HO; GOPALAN, 2014) deals with this by deriving a latent subspace, characterizing the multifactor variations. Images were synthesized in order to produce different illumination and blur conditions, and other 2D perturbations, forming a tensor to represent the face. Although it is a conceptually different work from the one done in (FOLEGO et al., 2016) by not dealing with heterogeneous modalities of pictures, the results indicated that the method is effective on constrained and unconstrained datasets.

According to (LIN et al., 2017), conventional approaches (e.g., canonical correlation analysis and partial least square regression) for cross-domain matching usually follow a procedure of two steps:

1. Samples from different modalities are first projected into a common space by learning a transformation. One may simplify the computation by assuming that these cross-domain samples share the same projection.
2. A certain distance is then utilized for measuring the similarity/dissimilarity in the projection space. Usually, Euclidean distance or inner product are used.

Going beyond the above steps, (LIN et al., 2017) present a novel pairwise similarity measure that advances existing models by expanding traditional linear projections into affine transformations and fusing affine Mahalanobis distance and Cosine similarity by a data-driven combination. Specifically, they incorporate the similarity measure matrix into the deep architecture, enabling an end-to-end way of model optimization. At work, they extensively evaluate their generalized similarity model in several challenging cross-domain matching tasks: person re-identification under different views and face verification over different modalities (i.e., faces from still images and videos, older and younger faces, and sketch and photo portraits). The experimental results demonstrate a superior performance of their model over other state-of-the-art methods.

Still according to (LIN et al., 2017), in literature, another conventional strategy is to synthesize samples from the input domain into the other domain. Rather than learning the mapping between two domains in the data space, dictionary learning can be used to alleviate cross-domain heterogeneity (WANG et al., 2012), and Semi-coupled Dictionary Learning (SCDL) is proposed to model the relationship on the sparse coding vectors from the two domains. Another framework called Domain Adaptation Machine (DAM) is proposed for multiple source domain adaption, but it needs a set of pre-trained base classifiers (DUAN; XU; TSANG, 2012).

In 2011, (SHRIVASTAVA et al., 2011) addresses a study where the objective is to find visually similar images even if they appear quite different at the raw pixel level. According to them, this task is particularly important for matching images across visual domains, such as photos taken over different seasons or lighting conditions, paintings, hand-drawn sketches, etc. They propose a surprisingly simple method that estimates the relative importance of different features in a query image based on the notion of "data-driven uniqueness".

In another related study, but with a different approach aiming at learning a better face recognition model for the target domain, (KAN et al., 2014) propose a simple, but effective domain adaptation approach that transfers the supervision knowledge from a labeled source domain (training data) to the unlabeled target domain (testing data). Their basic idea is to convert the source domain images to the target domain.

The problem of cross-domain image retrieval is addressed in (HUANG et al., 2015) considering the following practical application: given a user photo depicting a clothing image, our goal is to retrieve the same or attribute-similar clothing items from online shopping stores. In this case, it is a challenging problem due to the large discrepancy between online shopping images, usually taken in ideal lighting/pose/background conditions, and user photos captured in uncontrolled conditions. They address this problem by proposing a Dual Attribute-aware Ranking Network (DARN) for retrieval feature learning, consisting of two sub-networks, one for each domain, whose retrieval feature representations are driven by semantic attribute learning. With a large-scale real dataset, collected from real-world consumer websites, the results were considered important.

In a work with the same principles but with different approaches (SHI; JAIN, 2018) uses two private datasets and by employing the transfer learning technique, they propose a new method called DocFace, to train a domain-specific network for ID document photo matching without a large dataset. Compared with the baseline of applying existing methods for general recognition of this problem, their method achieves considerable improvement. In their study, they explore tests with and without transfer learning. Despite the similar approach to the same type of problem, we consider our tests more complex and with better results.

Despite the approaches presented, to the best of our knowledge, no evaluation regarding face authentication on considerably large cross-domain dataset such as ours was reported in the literature, especially for banking scenarios, the target application of our analysis.

3 Proposed methodology

In this section, we evaluated different methods for image normalization, deep features extraction and classification on the large dataset of real banking images collected for crossdomain face matching (FaceBank) in order to find the best architecture for this task and analyze its feasibility in a real banking scenario. After normalizing the selfies and IDs face images and extracting deep features using the VGG-Face (SIMONYAN; ZISSERMAN, 2014) or OpenFace (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) CNN models, we assessed four classifiers: Linear Support Vector Machine (Linear SVM) (CORTES; VAPNIK, 1995), Power Mean SVM (PmSVM) (WU, 2012), Random Forest (RF) (BREIMAN, 2001), and RF with Ensemble Vote Classifier (Voting RF) (RASCHKA et al., 2017), in order to verify which one performs better in the task of classifying a pair of face images as genuine or imposter. In summary, given a test pair of selfie and ID document photo images, a sequence of steps including image processing, deep feature vectors extraction, normalization, and combination, as well as classification, was performed in order to verify whether such a pair of facial images were captured from the same individual (genuine pair) or from different individuals (imposter pair). Figure 3.1 shows these steps, which are described in more detail in the following sections.

3.1 Collected Dataset - FaceBank

In (FOLEGO et al., 2016), only a small data set was used, but the results were promising and indicated that the path chosen would be worth more research. In this way, the best approach would logically be to increase the amount of training data to improve learning. The importance of the amount of data used should not be overlooked when dealing with machine learning, especially Deep Learning Architectures. As already mentioned, ML's algorithms learn through training, initially receiving examples whose outputs are known, note the difference between its predictions and the correct outputs, and tune the weightings of the inputs to improve the accuracy of its predictions until they are optimized. Somehow, the quality of their predictions improves with experience and the more data we provide, the better the prediction engines we can create.

At the beginning of the study it was logical, for a number of reasons that it was nec-

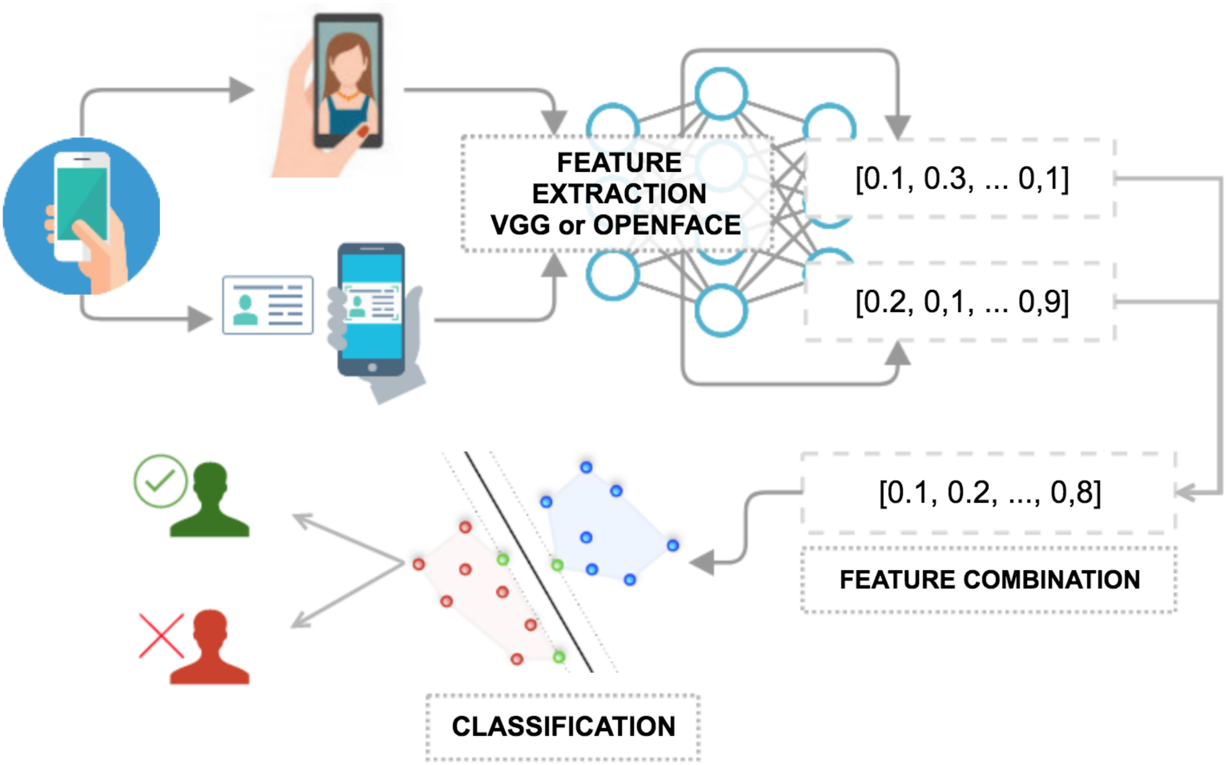


Figure 3.1: Overview of the proposed approach for cross-domain face matching: facial images are taken with the mobile device, their features vectors are extracted using deep neural networks (VGG-Face or OpenFace), they are subtracted from each other and then the final difference feature vector is classified as a genuine access (same person in both images) or imposter request (distinct persons in the images).

essary to increase the number of images considerably, so the next step was to obtain the images, in this way we look for an initial number of selfies that would be possible to achieve in the work environment, for the simple fact of being the main point of my quantit since the obtention of the selfies was much more complex than the obtaining of the documents of the individuals related to each selfie. Face recognition and authentication systems, in general, built by large private corporations, present the greatest accuracies, once they are trained on huge private datasets, containing millions of facial images, usually obtained from social media, far more than the number of images in the datasets available for research.

Regarding cross-domain face authentication (ID document photo and selfie, for instance), there is no large dataset available. Usually, given the difficulty to collect data, researchers evaluate their new methods with images from few individuals. Besides, no dataset with real images from banking systems was used in past evaluations, and this is an essential issue to be considered when evaluating techniques and deep neural networks in order to obtain reliable results.

Based on these considerations, we obtained authorization from the largest public Brazilian bank to collect a large dataset from its databases of facial images, which we called FaceBank, from selfies and scanned ID documents, in order to conduct this research. In the first step, about 150,000 images in RGB color space were collected, made up of selfies from profiles of individuals in the bank's internal social network and ID documents from the same individuals. Then we applied a fast technique based on the face detection algorithm of Viola and Jones (VIOLA; JONES, 2001) in order to discard images containing no faces (some profile photos were not from the faces of the users) or faces with very low quality. We discarded both images of the users that did not present faces with regular quality in the selfie or in the ID document.

After this process, we obtained 27,002 facial images from 13,501 subjects (two images per subject, i.e., selfie and ID document). In order to crop the faces in these remaining images, a more robust algorithm based on Histogram of Oriented Gradients (HOG) (DALAL; TRIGGS, 2005) was applied in a second step to detect the faces and their features (such as eyes coordinates). Figure 3.2 shows examples of images that compose the FaceBank dataset. Even visually, it is possible to note the huge differences in the facial images of the same person from different domains, i.e., selfie and ID document, and also the poor quality of the resultant images, typical from real banking scenarios, all this demonstrating the high complexity of the cross-domain face authentication problem in banking security systems.

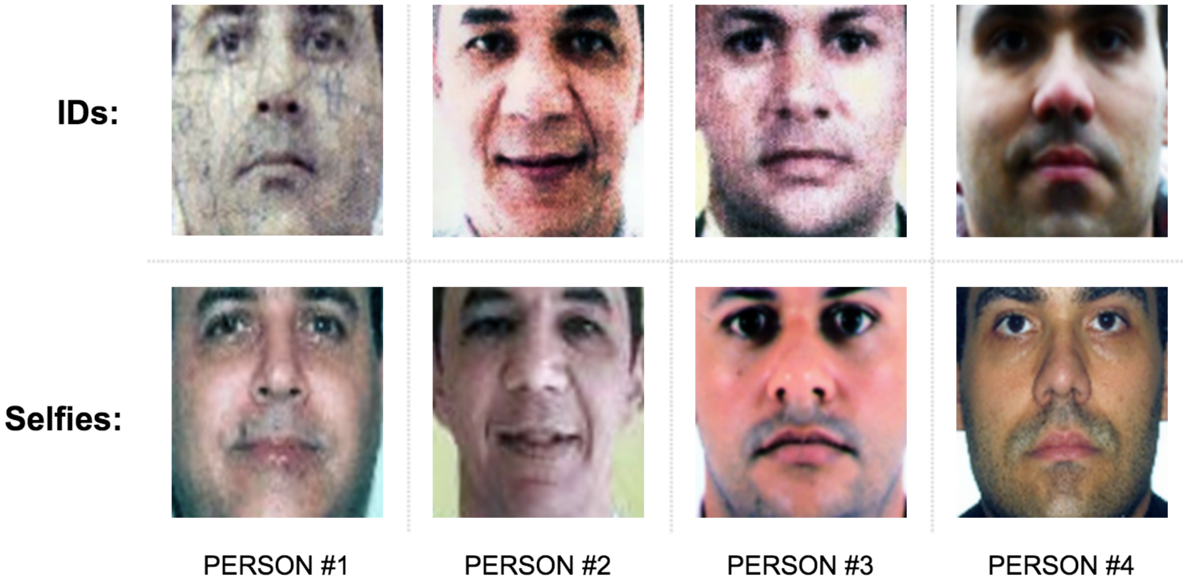


Figure 3.2: Examples of real facial images of selfies and ID documents present in the FaceBank dataset. The dataset contains a total of 27,002 images (13,501 individuals).

As an observation, for visualization and comparison purposes, the faces in Figure 3.2

are already detected, cropped and geometric normalized (size 224 × 224 pixels). The great majority of the images in the dataset is in RGB (Red, Green, and Blue) mode but few of them are in grayscale (which were transformed to RGB color space before feeding the CNN's evaluated).

To the best of our knowledge, no evaluation regarding face authentication on large cross-domain dataset with real images was reported in the literature, especially for banking scenarios, the main target of our study.

3.2 Face detection and cropping

According to (AL-ALLAF, 2014), in the past few years, face recognition has received a significant attention and regarded as one of the most successful applications in the field of image analysis. The human faces represent complex, multidimensional, meaningful visual stimulant and developing a computational model for face recognition is difficult. Face detection can be regarded as fundamental part of face recognition systems according to its ability to focus computational resources on the part of an image containing a face.

The process of face detection in images is complex because of variability present across human faces such as pose; expression; position and orientation; skin color; the presence of glasses or facial hair; differences in camera gain; lighting conditions; and image resolution (BRIMBLECOMBE, 2002). Face detection should be performed before recognition system. This is done to extract relevant information for face and facial expression analysis.

In order to carry out face detection and cropping, as said, in this study we used a robust and efficient algorithm available in Dlib library (KING, 2009), which is based on HOG (Histogram of Oriented Gradients) (DALAL; TRIGGS, 2005) features. This algorithm returns the coordinates of the rectangle that contains the detected face in the input image, as well as the coordinates of the left and right eyes. With this information it is possible to align, crop and resize all the face images from the dataset before starting to compare them.

Likewise in (FOLEGO et al., 2016), in the face cropping step, we included the ear, chin, and hair in the Region of Interest (ROI), by expanding by 22% the initial rectangle returned by the Dlib algorithm. This expanded ROI tends to increase the results of the face matching.

Face alignment is performed by rotating the face until the coordinates of both eyes are in line with the x-axis. Finaly, the cropped and aligned face images are resized by using bilinear interpolation. For the VGG-Face feature extraction based approach, the face image size must be 224 × 224 pixels, whilst for the OpenFace feature extraction based approach the face image size must be 96 × 96 pixels.

3.3 Cross-Domain image adaptation/normalization

A typical problem when comparing documents to selfies is the serious illumination differences due to the domain change, making face matching even more difficult. Other issues, such as pose, expression, and resolution of the images are also problematic. Aiming to mitigate some of these problems, especially illumination differences, and before extracting the features of the faces in the images collected, we applied the Automatic Color Equalization (ACE) (RIZZI; GATTA; MARINI, 2003) to normalize the cropped facial images. The ACE technique is based on a computational model of the human visual system that performs a photometric transformation on the images in order to equalize simultaneously global and local effects of illumination. It obtains good contrast enhancements even when the quality of the images is poor. We use this technique as an effort to approximate the two kinds of images under analysis.

3.4 Data augmentation

Obtaining a huge amount of real data (of real training samples) is an expensive and not always a possible process. Aiming to add more training data to the 27,002 images of the FaceBank, we generated new ones by applying some transformations to the FaceBank images. This data augmentation strategy is a common practice when working with CNNs (GRM et al., 2017).

According to (NIELSEN, 2015), despite the fact that artificial images do not substitute the potential of real samples, it is conceivable that adding to the training data transformed images based on the original ones might help the deep neural networks learn more about the patterns being addressed. By making small modifications to the original images, it is possible to expand the training database substantially. Common augmentation methods include noise addition, image equalization, random crop, scale change, jitter, brightness and contrast modifications. In this work, three of these methods were applied.

Initially, we increased the FaceBank dataset by adding white Gaussian noise to the original facial images in the regions near the eyes. As a generic approach, we used a sampling mechanism that added uncorrelated Gaussian noise α to the visual input x. If k indexes the raw pixels, a new sample is given by: $x_k = x_k + \alpha_k$.

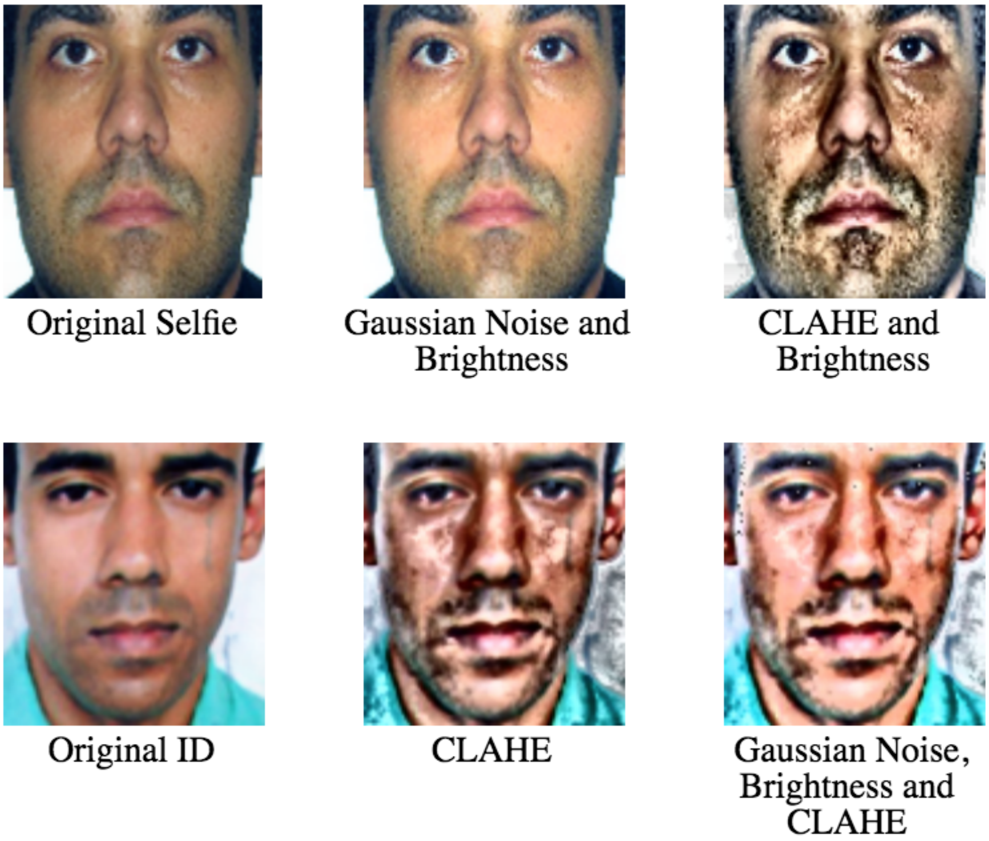
The second transformation, applied to the original and noisy images, was to randomly increase or decrease the brightness of a given training image, so that the model would learn not to rely on brightness information. As in (DU; GUO; SIMPSON, 2017), new images were generated with different brightness by first converting the images to the HSV (Hue, Saturation, and Value) color space and scaling the V channel up or down (converting the image back to the RGB color space after that).

Finally, in order to further augment the database obtained after the two previous transformations, we applied the Contrast Limited Adaptive Histogram Equalization (CLAHE) (PIZER et al., 1987) technique to the training images, which divides the input image into small blocks, applies a conventional histogram equalization in each block, and then checks if any histogram *bin* is above the contrast limit. As an observation, at the end of all the data augmentation process, we obtained $216,016$ images ($27,002 \times 2^3$, since we applied 3 transformations doubling the size of the dataset in each of them), $108,008$ selfies and $108,008$ ID documents. This set of images was called Augmented FaceBank. Figure 3.3 shown a demonstration of some transformations made in the images of the subjects with the intention of increasing the dataset collected.

3.5 Feature Extraction, Normalization and Difference Calculation

In order to extract robust features from the facial images given their different domains, we used the well-referenced Convolution Neural Network (CNN) called VGG-Face (SIMONYAN; ZISSERMAN, 2014), originally trained using a dataset with more than 2.6 million facial images of 2,622 different people, which achieved state-of-the-art results in face recognition. By using the trained VGG-Face model, a deep model of CNN containing 16 layers, we avoided many issues such as overfitting in our dataset as well as obtaining a good power of generalization due to the high capacity of the network (huge amount of parameters) and its large original training set.

We used the trained model of VGG-Face for Transfer Learning, i.e., we passed our facial images (from the Augmented FaceBank dataset) through the network and extracted



Original Selfie

Gaussian Noise and Brightness

CLAHE and Brightness

Original ID

CLAHE

Gaussian Noise, Brightness and CLAHE

Figure 3.3: Demonstration of some transformations made in the images of the subjects with the intention of increasing the dataset collected. In the image, it is demonstrated the application of Gaussian Noise, brightness and CLAHE.

their feature vectors based on the output of the layer "fc6" of the network (the third layer from top to bottom). Despite the fact that other studies usually extract features from the layer "fc7" from VGG-Face trained model, we explored the layer "fc6", a fully connected layer with 4,096 neurons, low-level information, higher performance compared with "fc7", and with better results for cross-domain face matching (FOLEGO et al., 2016; MATSUKAWA; SUZUKI, 2016; GUPTA et al., 2014).

In order to compare the results with the performance of a different deep neural network, in this study we also evaluate another well-referenced CNN: OpenFace (AMOS; LUDWICZUK; SATYANARAYANAN, 2016). OpenFace is an open source model, also implemented and trained on large datasets of facial images from the literature. Besides being able to use this neural network in commercial applications due to its open license, another interesting aspect of OpenFace is that it maps each face into an Euclidean space (into a hypersphere within it) by a 128-dimensional feature vector, output of its top layer. Its training algorithm, is mainly based on the Triplet Learning (SCHROFF; KALENICHENKO; PHILBIN, 2015) approach, in which the network is trained on genuine (same person) and imposter (different people) pairs of faces and tries to ensure that the faces of genuine pairs are closer in such Euclidean space than faces from different people, given a tolerance margin, following Eq. 3.1:

$$\|x^a - x^p\|_2 + \alpha < \|x^a - x^n\|_2, \forall x^a, x^p, x^n \in \tau, \tag{3.1}$$

where x^a and x^p indicate feature vectors of faces (selfie and ID) from the same person, x^n the feature vector from the face of another person, α is the tolerance margin (usually set to 0.2), and τ the training set for the neural network. The similarity degree of two faces is measured based on the Euclidean distance between their feature vectors.

VGG-Face model presents more parameters (higher capacity for feature learning) and allows extracting larger feature vectors. However, OpenFace (its default model), besides presenting a slightly different training algorithm and open license, is also interesting to our problem due to its great results reported in other applications (AMOS; LUDWICZUK; SATYANARAYANAN, 2016) and efficiency, being especially suitable for mobile banking.

Given the cross-domain problem, feature vectors extracted from images of different domains might have values with significantly different magnitudes. To mitigate this problem when comparing such vectors, we applied L_2 normalization to them. The p-norm of a feature vector $x \in \Re^n$ is given by:

$$||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}, \tag{3.2}$$

where $n = 4,096$ for VGG-Face and $n = 128$ for OpenFace.

The L_2-normalized version of each feature vector x is given by:

$$\hat{x} = \frac{x}{||x||_2}. \tag{3.3}$$

After normalizing the feature vectors of the faces from the Augmented FaceBank, for each pair of selfie and ID, we combined their vectors, a and b, respectively, into a final feature vector in order to emphasize their different properties and train the classifiers (in order to identify genuine and impostor face matchings), by using the absolute value of the subtraction $f_{ab} = |a - b|$, since other techniques for feature combination like Element-wise multiplication, Cross-correlation and Phase correlation showed worse results for our type of problem (FOLEGO et al., 2016).

3.5.1 Pair Generation

As said, to train the classifiers, we extracted the deep features of each pair of faces (selfie and ID) from the Augmented FaceBank using one of the CNNs evaluated. Then, the extracted features were normalized and stored into feature vectors. In order to verify if a pair of selfie and ID images is from the same person (genuine matching), or from distinct persons (imposter matching), the ID feature vector is subtracted from the selfie feature vector and the vector resultant from the module of the difference is finally presented to the classifier. For the pair generation task, we performed a random split of the individuals of the dataset into two disjoint sets: training and test. The training set contained 80% of the individuals of the Augmented FaceBank dataset, while the test set had 20%. In each set, there are random pairs of two images (selfie and ID) for representing genuine matchings and imposter matchings.

3.6 Classification

Given the genuine and imposter pairs of selfies and IDs and their difference vectors, different classifiers were trained. We selected four effective and also efficient classifiers from the literature in order to evaluate their performances in our cross-domain problem: Linear Support Vector Machine (Linear SVM) (CORTES; VAPNIK, 1995); Power Mean SVM (PmSVM) (WU, 2012); Random Forest (RF) (BREIMAN, 2001); and RF with Ensemble Vote Classifier (Voting RF) (RASCHKA et al., 2017). In our case, the Voting RF combines the decisions of 5 RFs. Due to robustness and efficiency of the code and reprodutibility of the experiments, we used the implementations of such methods available in the well-referenced Scikit Learning library (PEDREGOSA et al., 2011).

Likewise in (FOLEGO et al., 2016), we also decided to evaluate the Linear and PmSVM given their good performances in many tasks and due to their reported efficiency. Despite Linear SVM presenting inferior results in many tasks than SVMs with other kernels, it is fast, being more appropriate for environments with hardware restrictions, as in mobile devices. Regarding PmSVM, compared with state-of-the-art methods for large-scale image classification, it has achieved the highest learning speed and highest accuracy in many cases (WU; YANG, 2015). The RF-based classifiers are also robust and very efficient since they are based on decision trees.

In order to measure the accuracy of the selected classifiers, we used the global accuracy metric since some of them only had as output the class of each test sample, following Eq. 3.4:

$$accuracy = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} 1(\hat{y}_i = y_i), \tag{3.4}$$

where $\hat{y}_i$ is the predicted label for the i^{th} test sample, y_i is the real label of such sample and n_{test} is the number of test samples.

4 Experiments and Results

This chapter presents a set of tests and simulations performed to verify the effectiveness of the application and the approach proposed in the previous chapter. In addition, it presents an analysis and comparison of the results obtained. The environment used to perform the tests and simulations was composed of 1 server in virtual machine (VMware Virtual Platform) with the following configurations: 32 Intel (R) Xeon (R) CPUs E7-4870 2.40GHz; 154 GB of RAM; 1.7 TB of hard disk space; and Ubuntu 16.04.2 LTS OS. Throughout the process, Docker containers were used, an isolated environment within an open source platform in the Go (a.k.a., Golang) programming language, which is high performance and is developed directly in Google, which facilitates the creation and administration of isolated environments.

The containers facilitate the separation of each test to be executed and its continuous improvement, without harming other environments already in use. For extraction of face features, with the VGG-Face network, for example, the bvlc/caffe [3] container, with deep learning framework Caffe, was used, while the bamos/openface [4], a Torch based container for network OpenFace, was used. To perform the tests with the classifiers, a container based on Ubuntu 16.04.2 LTS OS was created and installed with every dependency required.

4.1 Balanced datasets

In order to assess the proposed techniques and to analyze the feasibility of their application in real banking security systems, especially for mobile devices, we considered one imposter pair for each genuine pair in the training and test stages of the classification, for a balanced training.

We also evaluated the performance of each architecture more than once by varying the number of subjects and the total number of difference feature vectors being considered, in order to verify, in a more detailed way, their robustness regarding the amount of data for training and test. For all classifiers, we used the default hyper-parameters defined in the Scikit Learning library (PEDREGOSA et al., 2011). For the PmSVM,

[3] https://github.com/BVLC/caffe

[4] https://hub.docker.com/r/bamos/openface/

the default value of the regularization parameter ω was set to 0.01. Tab. 4.1 shows the results obtained for all the classifiers given the features extracted by VGG-Face. In the first test, for example, we considered only 10,000 subjects from the Augmented FaceBank dataset (20,000 pairs of faces, 10,000 genuines and 10,000 imposters). We set, as said, 80% of the subjects (and their respective difference feature vectors) for training and 20% for test.

Table 4.1: Accuracy results (%), given the features extracted by VGG-Face, on the Augmented FaceBank dataset, considering one imposter pair for each genuine pair of faces (ID and selfie) and varying the number of subjects and difference feature vectors under analysis. The best result for each classifier is highlighted.

Subjects	**Pairs Train/Test**	**Linear SVM**	**PmSVM**	**RF**	**Voting RF**
10,000	16,000 / 4,000	91.65	89.57	89.65	93.45
20,000	32,000 / 8,000	92.43	89.87	89.13	93.28
50,000	80,000 / 20,000	92.69	88.09	89.26	**93.51**
80,000	128,000 / 32,000	92.75	89.87	**89.77**	92.67
108,008	172,808 / 43,208	**92.81**	**90.91**	88.95	92.82

As one can observe, the Voting RF obtained the best overall performance and its best accuracy result occurred when we considered 50,000 subjects. When working with 108,008 individuals, this classifier presented only a slight decrease in performance compared with the previous tests, still presenting better results than all other classifiers and demonstrating its robustness. Regarding processing time, voting RF is also very efficient by working with decision trees. It spent, on average, only 30 milliseconds for classification of each test sample.

As shown, the performance of the Linear SVM and PmSVM, in general, increased with the increasing sizes of the training and test sets, also demonstrating their robustness to large datasets (often found in real scenarios), despite being slower the former classifier (Linear SVM spent about 45 milliseconds for each test samples). The RF classifier presented its best performance with 80,000 subjects.

The results obtained by the classifiers given the feature vectors extracted by OpenFace are shown in Tab. 4.2. It is important to note that the results obtained were very close to those of VGG-Face. Besides, OpenFace works by default with smaller images (96×96 pixels), saving computational resources in the forward pass of the images through the network for feature extraction, and generates a much more efficient representation for the faces (it generates a 128-dimensional feature vector for each face, and VGG-Face a 4,096-dimensional vector), allowing classifiers to be faster and more suitable for mobile applications. The forward pass of each facial image in the VGG-Face took about 2.89 seconds while in OpenFace it took only 0.14 seconds per image.

Table 4.2: Accuracy results (%), given the features extracted by OpenFace, on the Augmented FaceBank dataset, considering one imposter pair for each genuine pair of faces (ID and selfie) and varying the number of subjects and difference feature vectors under analysis. The best result for each classifier is highlighted.

Subjects	**Pairs Train/Test**	**Linear SVM**	**PmSVM**	**RF**	**Voting RF**
10,000	16,000 / 4,000	89.17	85.45	88.72	**91.50**
20,000	32,000 / 8,000	**89.91**	85.67	89.48	90.65
50,000	80,000 / 20,000	89.82	86.81	89.17	90.86
80,000	128,000 / 32,000	89.88	**86.89**	**89.52**	90.70
108,008	172,808 / 43,208	89.89	86.73	89.39	90.61

The difference between the results obtained with OpenFace and VGG-Face is reasonable. This is probably due to the ability of the VGG-Face network to better represent the data obtained on each face (32 times the size of the feature vector extracted by OpenFace) and its greater number of convolutions. The best result regarding all experiments, 93.51% of accuracy, was obtained by the VGG-Face neural network with the Voting RF classifier when working with 50,000 subjects. Voting RF obtained the best results in all experiments, with both CNNs, being very suitable for the cross-domain face matching problem due to its efficiency inherited from the decision trees.

In order to evidentiate the performance of the assessed CNNs with such a powerful classifier, Figure 4.1 below shows the best accuracy results (%) for LinearSVM and Voting RF classifiers (the two best classifiers in this first test), given the features extracted by VGG-Face and OpenFace, considering 1 negative pair for each positive pair and varying the number of subjects and combined vectors (total number of pairs) under analysis. As can be seen, the Voting RF classifier tends to decrease its performance, as expected, when considering more subjects. However, such deterioration in the accuracy is not accentuated for both curves, whose values are close.

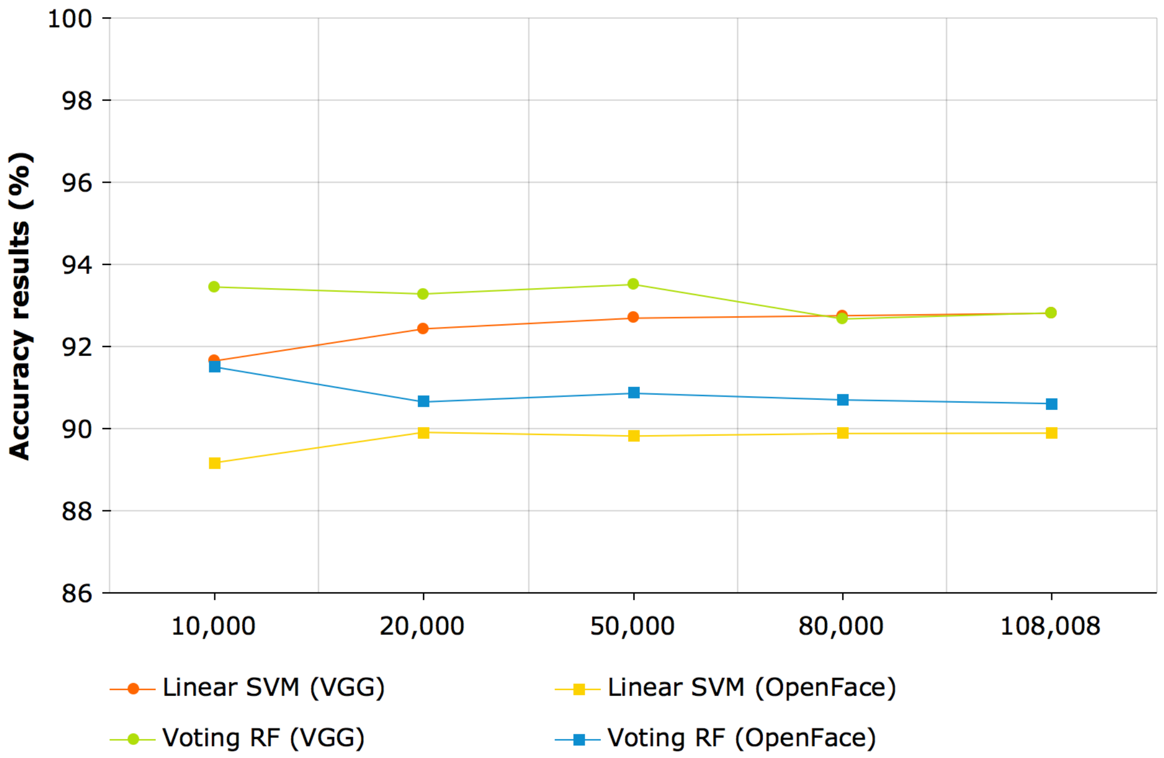


Figure 4.1: The best accuracy results (%) for Linear SVM and Voting RF classifiers, given the features extracted by VGG-Face and OpenFace, on the augmented FaceBank dataset, considering 1 negative pair for each positive pair and varying the number of subjects and combined vectors (total number of pairs) under analysis.

4.2 Unbalanced datasets

In order to continue evaluating the proposed techniques and their feasibility of applying them in real banking security systems, especially for mobile devices, we considered applying two more tests, following the same protocol of the first one, but considering now to train the classifiers with 2 negative pairs and 5 negative pairs of facial images (selfie and ID) for each positive pair in the total amount of pairs (samples). In these new tests the normalized accuracy was calculated by Eq. 4.1:

$$accuracy = \frac{1}{c}\sum_{i=1}^{c}\sum_{j=1}^{n_i} 1\frac{(\hat{y}_{ij} = y_{ij})}{n_i}, \tag{4.1}$$

where $\hat{y}_{ij}$ is the predicted class label for the i^{th} test sample, y_{ij} is the real label of such sample and c represents the class.

Table 4.3 and Table 4.4 shows the results obtained for the features extracted by VGG-

Face and OpenFace, respectively, considering 2 negative pairs of facial images for each positive pair in the total amount of pairs (samples). The VGG-Face obtained the overall performance with the LinearSVM classifier when working with 50,000 subjects and practically the same precision result was observed when working with 108,008 individuals, a result still better than the ones obtained by the other classifiers.

Table 4.3: Accuracy results (%), given the features extracted by VGG-Face, on the Augmented FaceBank dataset, considering 2 pairs of impostors for each genuine pair of faces (ID and selfie) and varying the number of subjects and difference feature vectors under analysis. The best result for each classifier is highlighted.

Subjects	**Pairs Train/Test**	**Linear SVM**	**PmSVM**	**RF**	**Voting RF**
$10,000$	$24,000$ / $6,000$	91.10	**89.76**	85.23	88.35
$20,000$	$48,000$ / $12,000$	91.40	88.55	85.36	88.31
$50,000$	$120,000$ / $30,000$	**92.00**	88.62	85.59	88.47
$80,000$	$192,000$ / $48,000$	91.57	89.08	**85.62**	88.45
$108,008$	$259,212$ / $64,812$	91.89	87.89	85.61	**88.48**

Table 4.4: Accuracy results (%), given the features extracted by OpenFace, on the Augmented FaceBank dataset, considering 2 pairs of impostors for each genuine pair of faces (ID and selfie) and varying the number of subjects and difference feature vectors under analysis. The best result for each classifier is highlighted.

Subjects	**Pairs Train/Test**	**Linear SVM**	**PmSVM**	**RF**	**Voting RF**
$10,000$	$24,000$ / $6,000$	87.33	88.23	86.78	87.30
$20,000$	$48,000$ / $12,000$	88.27	88.85	86.80	87.39
$50,000$	$120,000$ / $30,000$	**89.26**	88.21	86.81	87.35
$80,000$	$192,000$ / $48,000$	89.19	**89.44**	86.79	87.48
$108,008$	$259,212$ / $64,812$	89.01	87.87	**86.82**	**87.50**

Still, in the results of VGG-Face, the Voting RF classifier showed slightly smaller results, but with high indices. The same perspective is observed in OpenFace results, with Linear SVM slightly above, but with a very small difference, mainly in comparison with the results obtained with the PmSVM. OpenFace presented results close to that of VGG-Face also in this second test, with a small increase in the results of the RF classifier.

The three best classifiers in this second test were LinearSVM, PmSVM and Voting RF, the difference between these last two classifiers being irrelevant. Figure 4.2 below shows the best accuracy results (%) for LinearSVM and Voting RF classifiers, given the features extracted by VGG-Face and OpenFace, considering 2 negative pairs for each positive pair and varying the number of subjects and combined vectors. Unlike the first test, the Voting RF had an increase in accuracy by considerably increasing the number of subjects for VGG-Face and OpenFace.

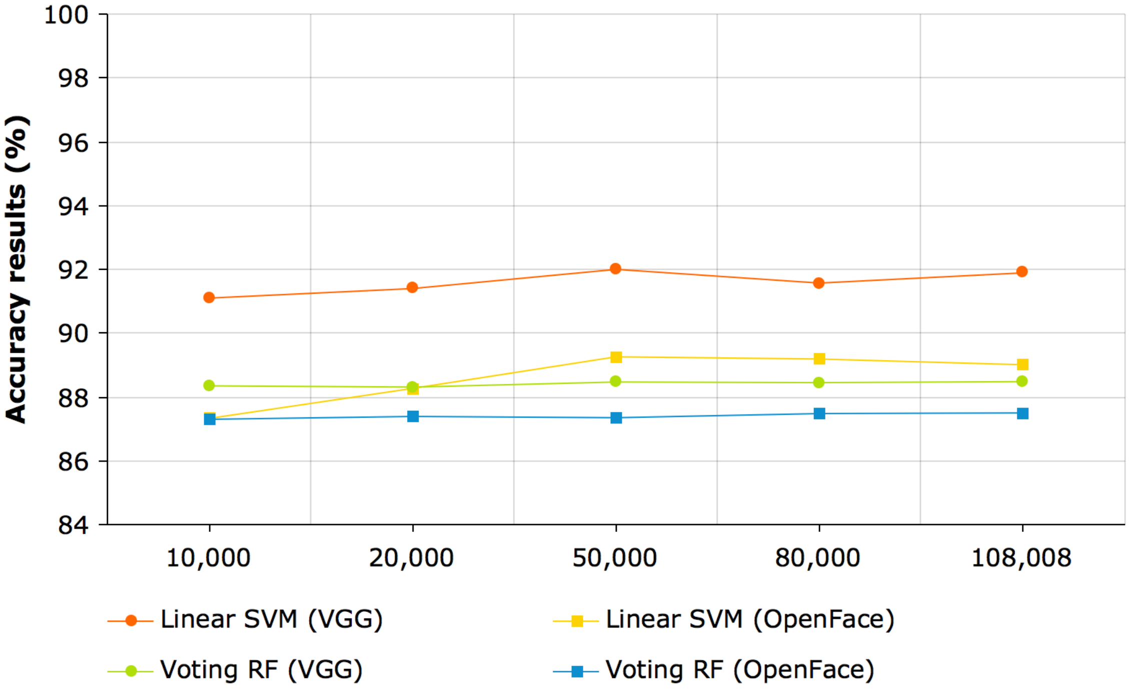


Figure 4.2: The best accuracy results (%) for Linear SVM and Voting RF classifiers, given the features extracted by VGG-Face and OpenFace, on the augmented FaceBank dataset, considering 2 negative pair for each positive pair and varying the number of subjects and combined vectors (total number of pairs) under analysis.

Table 4.5 and Table 4.6 shows the accuracy results obtained for each neural network and classifier by varying the number of subjects and samples (total number of pairs) for training and test considering 5 negative pairs of facial images (selfie and ID) for each positive pair we created. As one can observe, the best result for this experiment, 89.77% (4.6), was obtained by the OpenFace network and PmSVM classifier when considering only 20,000 subjects (Augmented FaceBank dataset).

The results, given the features extracted by the VGG-Face neural network ensemble with the PmSVM and LinearSVM classifiers, were very close. Despite its less complex model, LinearSVM demonstrates good results for a large amount of data, delivering speed and performance.

As can be seen, the results obtained with the RF and Voting RF classifiers decreased considerably due to the use of an unbalanced dataset, while the results of the PmSVM classifier showed a small improvement, which can be referred to as a statistical fluctuation (Table 4.5 and Table 4.6), obtaining results very close to or even better than the first test (4.1). Figure 4.3 summarizes this last test, showing the best accuracy results (%) for LinearSVM and Voting RF classifiers, given the features extracted by VGG-Face and OpenFace, on the Augmented FaceBank dataset.

Table 4.5: Accuracy results (%), given the features extracted by VGG-Face, on the Augmented FaceBank dataset, considering 5 pairs of impostors for each genuine pair of faces (ID and selfie) and varying the number of subjects and difference feature vectors under analysis. The best result for each classifier is highlighted.

Subjects	**Pairs Train/Test**	**Linear SVM**	**PmSVM**	**RF**	**Voting RF**
10, 000	48, 000 / 12, 000	87.52	89.20	75.05	75.29
20, 000	96, 000 / 24, 000	87.61	89.29	74.93	75.27
50, 000	240, 000 / 60, 000	88.17	**89.65**	75.07	**75.34**
80, 000	384, 000 / 96, 000	**88.57**	89.41	**75.21**	75.27
108, 008	518, 424 / 129, 624	88.44	89.39	75.15	75.08

Table 4.6: Accuracy results (%), given the features extracted by OpenFace, on the Augmented FaceBank dataset, considering 5 pairs of impostors for each genuine pair of faces (ID and selfie) and varying the number of subjects and difference feature vectors under analysis. The best result for each classifier is highlighted.

Subjects	**Pairs Train/Test**	**Linear SVM**	**PmSVM**	**RF**	**Voting RF**
10, 000	48, 000 / 12, 000	85.21	89.53	80.29	80.30
20, 000	96, 000 / 24, 000	85.19	**89.77**	80.35	80.07
50, 000	240, 000 / 60, 000	85.26	89.21	80.27	79.90
80, 000	384, 000 / 96, 000	85.29	89.58	80.33	80.39
108, 008	518, 424 / 129, 624	**85.40**	89.44	**80.51**	**80.48**

With the unbalanced dataset, the classifier learns best to recognize when pairs refer to imposters (negative pairs), matching more negative pairs than positive, causing overall accuracy to drop considerably. This brings us the certainty of the importance of the effect of the correct balancing of the number of positive and negative pairs.

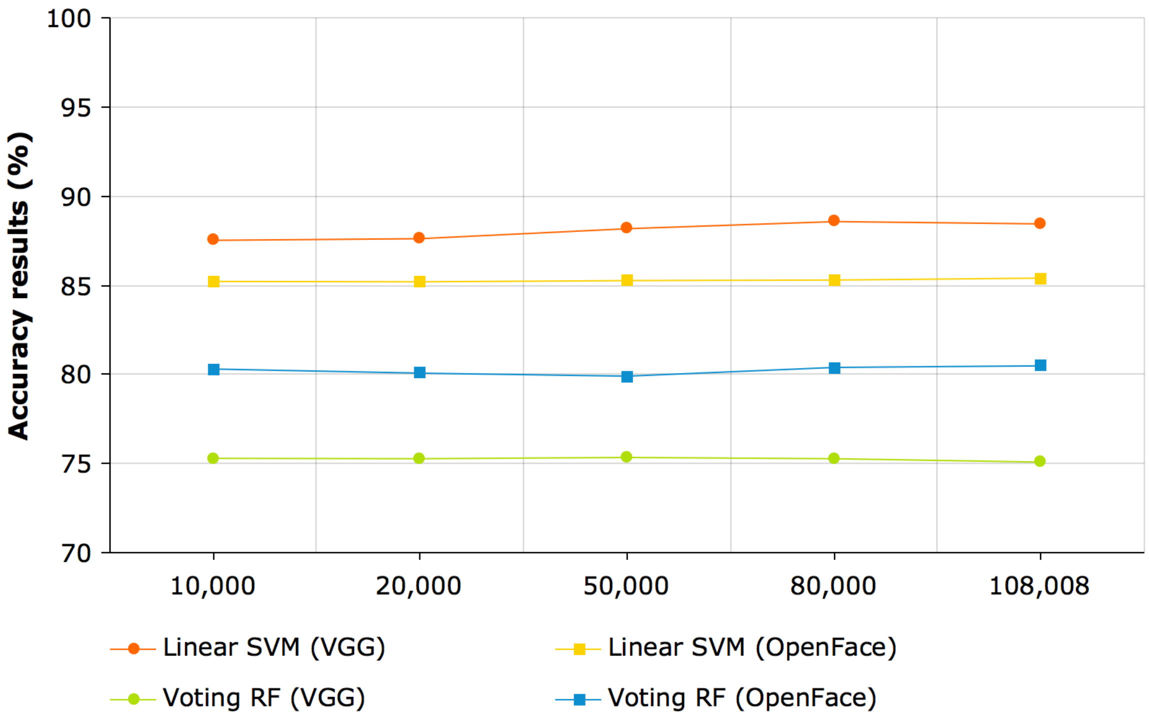


Figure 4.3: The best accuracy results (%) for Linear SVM and Voting RF classifiers, given the features extracted by VGG-Face and OpenFace, on the augmented FaceBank dataset, considering 5 negative pair for each positive pair and varying the number of subjects and combined vectors (total number of pairs) under analysis.

When working with 1 negative pair for each positive pair the accuracy results of the methods improved in general, decreasing when working with more negative samples than positive ones. In this study we did not have previous information on how this proportion would be in a real banking scenario, since it may depend directly on the size and configuration of the final dataset.

In order to visualize the performance of the Voting RF classifier (for the balanced datasets presented in session 4.1) in terms of True Matching Rate (TMR) (probability of a correct match) and False Matching Rate (FMR) (probability of an incorrect match) and map the effects of varying the decision (acceptance) threshold of the system, addressing all possible combinations of correct and incorrect decisions, the Receiver Operating Characteristics (ROC) curve (METZ, 1978) was plotted in Figure 4.4 (features extracted by VGG-Face) and Figure 4.5 (features extracted by OpenFace). The Equal Error Rate (EER) [5] is also reported in each figure.

[5] EER indicates the error rate where False Non-matching Rate (FNMR), i.e., FNMR=1-TMR, and FMR are the same.

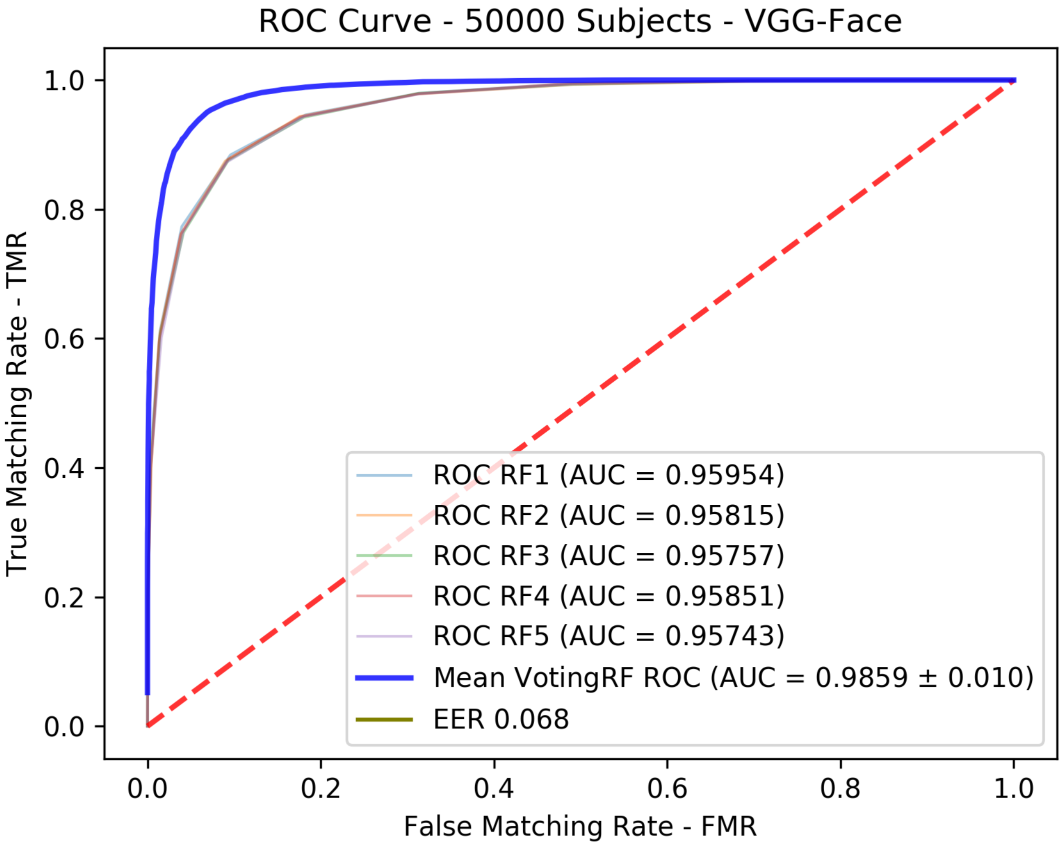


Figure 4.4: Receiver Operating Characteristics (ROC) Curve for the Voting RF classifier given the features extracted by VGG-Face on the augmented FaceBank dataset, considering one negative pair for each positive pair, with 50,000 subjects and 100,000 combined vectors (total number of pairs) under analysis.

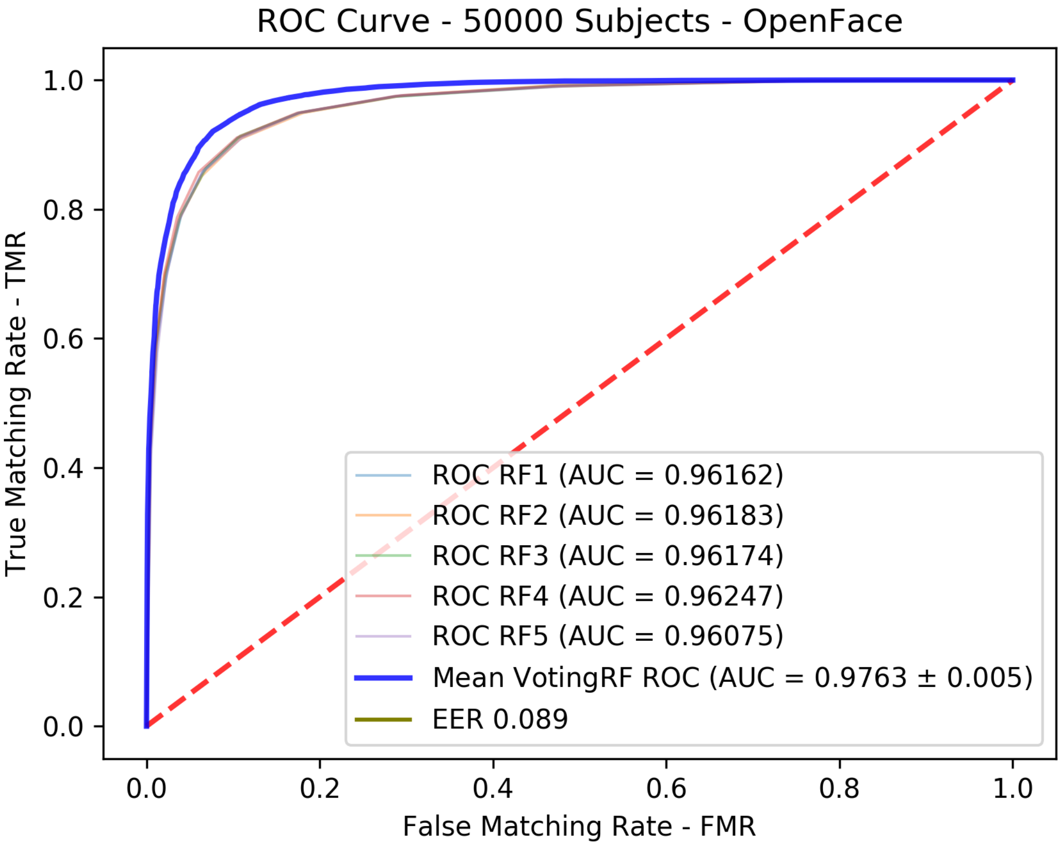


Figure 4.5: Receiver Operating Characteristics (ROC) Curve for the Voting RF classifier given the features extracted by OpenFace on the augmented FaceBank dataset, considering one negative pair for each positive pair, with 50,000 subjects and 100,000 combined vectors (total number of pairs) under analysis.

5 Analysis and Proposal of an ABIS System

As we are dealing directly with a Brazilian public bank, the process of architecture and implementation of a biometric system must be perfectly aligned with the premises, regulations, laws and other issues intrinsically related to public institutions in Brazil. In this way, it is important to investigate what would be necessary for a future implementation of an Automatic Biometric Identification System (ABIS) in an environment of a Brazilian public institution, which must protect itself by correctly applying established public policies and laws.

In the last years, countries and public banks are trying to make stronger their virtual barriers and systems in order to avoid successful hackers attacks on portals, infrastructure, and applications. A solution that has been applied to work around this problem is the definition of several governmental standards to guide these many IT Teams. These documents are written by IT specialists and commonly gather up-to-date knowledge from most of the important international standards, protocols and international good practices guides. However, there are many standards and they often define requirements in a high level of abstraction and some teams have not enough knowledge to apply these recommendations in practice or they are not able to identify all standards related to a specific topic.

Another relevant problem is that biometric identification systems are most of the times much expensive because there are just a few global vendors that can offer a complete and trustworthy deduplication[6] solution. Thus, despite the benefits of using biometric technologies, the high cost related to equipment and biometrics software are not affordable to almost all countries or public institutions. At the same time, the internal development of a Biometric solution cannot be feasible because the apparent complexity that surrounds the system.

As ABIS is a large ecosystem, we focus on two complex, high-value parts. This chapter presents a study on the real needs in the implementation and proposes a cloud system to enable the adoption of biometric technologies, creating a new model of service delivery.

[6] Biometric deduplication concept is related to ensure that each specific person will exist just once in the people database (DECANN; ROSS, 2013).

Besides that, proposes a secure and integrated transmission subsystem model. All the analysis and implementation takes into account the total adherence and compatibility with the standards and specifications proposed by the Brazilian government.

5.1 Current ABIS perspective

Biometric technologies are experiencing a huge growth in the last years fueled by the development of the industry in government, mobile security, health care, banking and several other important fields. In this way, some trends can be observed such as the application of biometric-based authentication in banking, in order to reduce the identity theft occurrences and reduce the operational risks; in health care, aiming to reduce the number of frauds and to provide access control critical medical sources; in mobile security, biometric technologies are replacing the traditional method of entering passwords or PINs with a swipe of a finger or with face recognition, making stronger the authentication process (TORRES12 et al., 2015; TORRES et al., 2016).

Biometric traits have become an important element to governments in citizen's identification due to some of their unique characteristics that are key enablers for the successful deployment of identification and deduplication processes. Firstly, the universality, since people essentially have biometrics elements such as fingerprints, irises or face. Another characteristic is the distinctiveness, because of the biometric trait should be sufficiently different among people in order to ensure the individualization process. And lastly, biometric characteristics should be permanent and this must be invariable over a period of time.

The citizen biometric deduplication concept is related to ensure that each specific person will exist just once in the people database. In this way, the biometric sample is compared to the "N" samples already registered in the system, one by one, and if there is no match, the person can be correctly registered in database and a unique number will be associated to the sample provided by the person (DECANN; ROSS, 2013).

In a broad view, several countries have seen in biometric technologies a way to identify uniquely citizens and to support the deployment of a national electronic ID, an important action to boost e-government processes. In response to the increase in the occurrences of identity fraud, several countries are financing solutions to ensure the identity of their citizens in public and private transactions. Most of the European countries that are developing a strategy decided to use a user-centric identity management model because of the strong need for personal data protection imposed by the rigorous Europe Union privacy laws. On the other hand, it is common to find a centralized identity management model when we observe countries from other continents like India and Mexico.

The user-centric and the centralized approaches are on opposite sides when the issue is privacy and fraud combat. The user-centric model is the best for protecting personal data, once the user has total power over the use of this information. However, this model is not optimized to provide connections among data or to track the user once the common data is not stored in a centralized database and the use of pseudonyms by the users is common in transactions with service providers. On the other hand, the centralized model is based on the use of a central database with all personal data, where the government acts as the unique identity provider. In this strategy, it is possible to make connections with user transactions realized among distinct government agencies or even transactions with private companies since there is just one identifier per user and the use of pseudonyms is not usual.

When a country decides to implement a centralized identity management strategy, it is necessary to define a method to individualize the citizens, aiming to guarantee that each person has just one record in the citizen database. The main form used to provide this guarantee is the use of an ABIS. This kind of system uses biometric attributes from people, such as fingerprints and faces, to permit people individualization.

Many countries already adopted biometrics technologies to identify uniquely their citizens, most of them, such as Argentina, Brazil, Bolivia, Chile, Colombia, Ecuador, Guatemala, Mexico, Peru, Germany, Belgium, Czech Republic, Portugal, Spain, Estonia, Gambia, Nigeria, India and Indonesia, use the citizen's fingerprints to enable the deduplication process. Some of these countries, such as India and Mexico, adopted the irises as an additional biometry, most because the false positive problems cause by their huge populations (JAIN; ROSS; PRABHAKAR, 2004; MITRA; GOFMAN, 2016; CREDENCE, 2016).

The biometric initiatives among countries often differ depending on the level of development of each place. In rich countries, the use of these technologies is commonly related to law enforcement, border control or to the establishment of global standards for integrating government systems (ISO, 2011). The other side, such as Brazil and India, intend to use these technologies to inhibit frauds in social and income distribution programs. Specifically in Brazil, its common some people get several national identity numbers and, because that, they can defraud public programs to obtain several times social benefits.

The biometric market is expected to reach US$ 34.5 billion by 2022 influenced, in addition to other elements, by the governments of countries across the world that are adopting biometrics technologies for access authorization, identification and verification, national border control projects aim to verify recipients of government payments, to authenticate cash grant beneficiaries for poor households, to authenticate delivery of social services (school attendance or natural gas subsidies), to provide the financial inclusion of poor people or even to enable online vote in the election for president, governors, mayors and many other political posts (CREDENCE, 2016; MITRA; GOFMAN, 2016).

Another important point that can affect the biometric market in the next few years is the adoption of biometrics technologies in the Asia Pacific, particularly in India and China, followed by North America and Europe. One common aspect among this places is the economic factor since they concentrate much of the world's wealth and can pay for an expensive ABIS to execute the identification process. One of the main obstacles to implement this kind of system is its high price, as the vendors sell an all-in-one product, i.e. a solution that can do all the process involving the deduplication, from the enrollment to the identity generation (CREDENCE, 2016)..

A biometric system is essentially a pattern recognition system that operates by acquiring biometric data from an individual, extracting a feature set from the acquired data, and comparing this feature set against the template set in the database aims to validate a person's identity by comparing the captured biometric data with her own biometric template stored in the system database (one-to-one comparison) or to recognize an individual by searching the templates of all the users in the database for a match (one-to-many comparison) (JAIN; ROSS; PRABHAKAR, 2004).

Despite it seems difficult to draw any generalization of biometric systems due to the variety of applications and technologies, there are many common elements among these systems. Based on the analysis of these elements, ISO/IEC 19794 proposed an information flow within a general biometric system, showing a general biometric system considering the three main functions of this kind of system, enrolment, verification, and identification as shown in Figure 5.1.

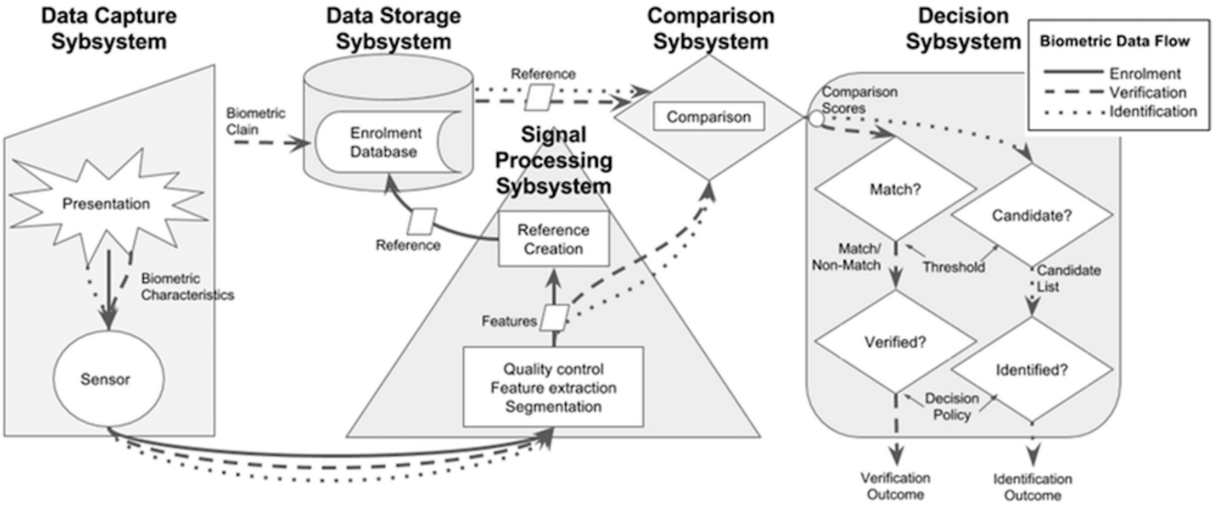


Figure 5.1: ABIS General Architechure. (ISO, 2011)

The ISO/IEC 19794 is a guide about ABIS architecture and shows some high-level details regarding each part of an Automatic Biometric System. Despite being an important reference, the referenced ISO does not address technical details in a level that can permit an implementation of the solution.

In line with the model specified in this standard, a General Biometric System can be decomposed into eight subsystems: Data capture, Signal Processing, Data Storage, Comparison, Decision, Administration, Interface, and Transmission.

The Data capture subsystem is related to the capture of the biometric image from a biometric sensor, representing the first step in an enrollment or verification process. The next element is the Signal processing subsystem, responsible for four main activities: segmentation - i.e. locating the signal of the subject's biometric characteristics within the biometric sample; feature extraction - that is related to deriving the subjects repeatable and distinctive measures from the captured biometric sample; quality control - i.e. assessing the suitability of samples, features, references, etc. and image enhancement - which involves improving the quality and clarity of the captured biometric samples.

The comparison and decision subsystems, are directly related to the main function of the system, to perform matches between references, and they are the ones that have the highest demand for processing. In the comparison subsystem, features are compared against one or more references and the comparison scores (indicate the similarities or dissimilarities between the features and references compared) are passed to the decision subsystem. In turn, decision subsystem uses this comparison scores to provide the decision outcome for a verification or identification transaction.

The ABIS comparison subsystem is the main responsible for both verification and deduplication processes since they work based on the comparison between two types of biometric features. This is the most specific element of ABIS and the subsystem that requires more computational power since millions of matches operations can be necessary to deduplicate a single person.

The other three subsystems, which are not showed in the ABIS general architecture, are related to Administration, Interface and Transmission functions. The first governs the overall policy, implementation, and usage of the biometric system, for example controlling the operational environment and non-biometric usage data and requesting additional information from the subject. The second is responsible for providing an interface to an external application system, and the last provides the data transmission service among all other subsystems.

The Transmission subsystem is also the focus of this part of study. There are three main elements related to this subsystem: How to provide a safe transmission from the capture device to the client device; How to transmit the identity package from the client to the server; and How to ensure the identity of the actors involved in this process. Each of these questions will be analyzed in the next sections and solutions will be proposed to address these problems according to the state of the art.

5.2 Proposal Architecture in Cloud Concept

The idea of do not acquire a turnkey ABIS solution were for the first time applied in India, when the local government decided to enroll almost 1.33 billion citizens and the costs of the solution was considerably high. However, there are many differences between the India solution and the architecture proposed in this study. Firstly, India designed a solution to hire the deduplication providers in advance, and it is not fully adherent to the cloud concept adopted in our study since the competition today occur just between two providers. Another issue is related to interoperability, since many processes and data models were designed with no conformity with ISO standards and follow just part of ISO/IEC standards.

Because that, ABIS providers must implement a group of methods that were designed by India Government to be able to provide the deduplication service and to enable the compatibility of the services with the Adaahar [7] server requests. It's important to emphasize these technical details because the adherence to international standards is one of the main factors to permit the vendor independence, that is, if one of the ABIS providers needs to be replaced it can be done without requiring a complete stop of the system or any change in the service provided. Despite these issues, India proved that is possible to deploy a segmented ABIS solution.

The architecture proposed in this study is based on several international standards and the experience of countries around the world. The main challenge was to make several international standards work together and at the same time provide a usable system prototype. In addition, this study tries to address some important questions in order to enable the spread of Biometric Systems among poor countries. One of the objectives is to define the general scenario and to describe the main elements and activities that must be executed by these countries to deploy a complete ABIS system. Due to complexity, security analysis involving the proposed architecture will be addressed from subsection 5.5.

Besides ISO/IEC 19794-1:2011, many other standards were published in the last years aiming to provide a common guide to ABIS implementations and to permit interoperability among systems provided by different vendors. Three of these were chosen to join the 19794-1 as a base to the proposal architecture:

- **ISO/IEC 19785-1:2015** - Information technology – Common Biometric Exchange Formats Framework – Part 1: Data element specification} - defines a basic structure for standardized biometric information records (BIRs) within the Common Biometric Exchange Formats Framework (CBEFF). This structure consists of three parts: the standard biometric header (SBH), the biometric data block (BDB), and the security block (SB) (ISO, 2015b).
- **ISO/IEC 30108-1:2015** - Information technology – Biometric Identity Assurance Services – Part 1: BIAS services} - defines the architecture, operations, data elements, and basic requirements for biometric identity assurance services and a framework for the implementation of generic, biometric-based identity services within a services-oriented environment (ISO, 2015a).

[7] Aadhaar number is a 12-digit random number issued by the UIDAI to the residents of India.

- **OASIS Biometric Identity Assurance Services (BIAS) SOAP Profile Version 1.0** - provide the biometrics and security industries with a documented, open framework for deploying and invoking identity assurance capabilities that can be readily accessed as services. Defines and describes methods and bindings by which the INCITS BIAS framework can be used within XML-based transactional Web services and service-oriented architectures (OASIS, 2014).

In a simple way, the "as a service model" is related to deploy and provide services in the cloud. Prior to the cloud concept, organizations often externalized several IT services by outsourcing them to service providers. However, the introduction of cloud computing brings many benefits over the traditional once provided to the organization's several new possibilities as the elastic scaling up or down according to need or the pay-as-you-go rental model that allows organizations to defer costs.

Some important characteristics of cloud services are the use of virtualized computing resources to maximize hardware utilization, the high scalability of cloud infrastructure and the fact that the resource usage is billed only as used and there is no annual contract and no commitment for a specific level of consumption (pay-as-you-go model) (ROSENBERG; MATEOS, 2010).

This approach can solve the high under-utilization of system capacity related to the on-premises solution due to the fact that hardware dimensioning is many times related to peak workload. In the cloud, the infrastructure can be dynamically scaled according to the current need and the customers will pay just for the resources they used. Even more, costs with expensive hardware infrastructure maintenance, including data centers teams and local computing nodes, will no longer necessary.

On-premises, as mentioned, is a type of software delivery model where the system run in customer's in-house server and use the IT resources from this company. Thus, the company uses its internal infrastructure instead of remote service to process its software applications. This is one of the most common methods of using enterprise and consumer applications. On- premises software typically requires a high initial investment since it's necessary to buy software licenses and servers hardware. The customer is still responsible for providing a suitable secure area for equipment's installation, ensure the availability of the infrastructure and the overall management of on-premises software.

Despite the fact that this architecture is mainly related to providing a solution to governments that need to provide the identification of their citizens, it still creates a new market niche to ABIS vendors, since they can easily adapt their solutions to provide just the verification service on the cloud and establish different parameters to define the price. The same provider can provide services with different levels of SLAs with different prices, for example.

It is much more common in ABIS system a false positive occurrence (False Match Rate - FMR) than a false negative occurrence (false non-match rate - FNMR), that is, most of the errors indicate that a person is already enrolled but it is not yet. In cases like this, the analysis must be conducted manually by an operator, because the higher the number of manual analysis, the higher will be the costs with employees to do that. Thus, customers have the option to pay more for a more accurate service or opt for a lower price and earn more money to hire new employees.

Other elements that can influence the price model are system latency. Almost all studied countries and states, that use on-premises solutions, enroll citizens just during the daylight and in work days, but the deduplication process runs in the background during the whole day and in all days of the week. Sometimes the deduplication process of a person can take many days since there is a queue of people to be enrolled. So, providers can define prices depending on different response times of the verification service. Independently of metrics chosen to deduplication services, they need to be well defined, understood and measurable, so that the different stakeholders involved in cloud computing can rely on them with confidence (VAULX; SIMMON; BOHN, 2018).

5.3 Services Definition and Data Interoperability

All system functionalities were defined in a Service Oriented Architecture according to specified in the Biometric Identity Assurance Services (ISO, 2015a) and (OASIS, 2014) in order to maintain the interoperability. The data interoperability will be reached through the use of the Common Biometric Exchange Formats Framework (CBEFF) since this data model provides a standardized set of definitions and procedures to support the interchange of biometric data in standard data structures called CBEFF biometric information records (BIRs) (ISO, 2015b).

In general, BIAS architecture consists of services (interface definition) and data (schema definition). Thus, services expose a common set of operations to external requests, that may be an external system, a Web application or an intermediary. BIAS services themselves are platform and language independent, i.e., are generic in nature and not targeted at any particular business application and technology. To address the minimal level of conformance to BIAS standard (ISO, 2015a) the proposal architecture will be able to provide the Minimum Primitive Services Implementation included in some biometric services:

- **Create Subject** - create a new subject record, generate a subject ID that uniquely identifies this subject in the system and associate that to the record that was created. This service expects an identifier of the subject as input.
- **Delete Biographic Data** - erase all of the biographic data associated with a given subject record. It expects the identifier of the subject or the identifier of the encounter depending on the model implemented [8] A subject ID or an Encounter ID should be provided to this service as input.
- **Delete Biometric Data** - remove biometric data from a given subject record. In the encounter-centric model, the encounter ID shall be specified, while in the person-centric model the subject ID is expected as input.
- **Delete Subject** - depending on the selected model delete an existing subject record or any associated encounter information from the system. The identifier of the subject should be provided as input.
- **List Biographic Data** - list the biographic data elements stored for a subject. Depends on the model selected, an encounter ID may be specified to get just the biographic data about a specific user interaction or the subject ID may be used as input and the service will return a list of encounter IDs that contain biographic data.
- **List Biometric Data** - list the biometric data elements stored for a subject through a Biometric Data List, i.e., a list of biometric data associated with a subject or encounter. Will not return actual biometric data of the subject once this function is implemented in the Retrieve Biometric Information service.
- **Retrieve Biographic Information** - depends on the model selected the service shall return a list of biographic data elements associated with the subject or a specific encounter.

[8] There are two identity models defined, person-centric systems maintain a single up-to-date record (set of data) for a given subject, whereas an encounter-based system retains data related to each interaction the subject has with the system (ISO, 2015a).

- **Retrieve Biometric Information** - returns a data structure (BIR List) containing the retrieved biometric data associated with a subject ID or a specific encounter. It's possible to use an optional input parameter to specify that only biometric data of a certain type should be retrieved [9].

- **Set Biographic Data** - associate new or updated biographic data to a given subject record. A list of biographic data to associate with the subject or encounter should be provided as input.

- **Set Biometric Data** - associate biometric data to a given subject record. Some data structure containing the new biometric sample(s) (BIR List) is the expected input of this service.

- **Update Biographic Data** - update the biographic data for an existing subject record taking as input a list of updated biographic data elements.

- **Update Biometric Data** - update a single biometric sample for an existing subject record taking as input a data structure containing the new biometric sample (BIR).

- **Verify Subject** - provide a 1:1 verification match between a given biometric and either a claim to identity in a given gallery or another given biometric. Match and Score are returned after processing, the first indicates if the Input BIR matched either the biometric information associated with the Identity Claim or the Reference BIR and the second returns if the biometric information matched.

- **Identify Subject** - perform an identification search against a given gallery for a given biometric, returning a rank-ordered candidate list of a given maximum size.

5.4 System Architecture

The overall proposed architecture is presented in Figure 5.2. As it can be seen, there are two points of interaction, the first one between the clients and the minimal services, and the second one between the minimal services and the main biometric services.

Most of ABIS services are common and can be easily implemented by customers. In this context, the main problem is to develop a verification system with a level of confidence and accuracy equal to the main ABIS providers.

[9] Biometric Type structure is the type of biological or behavioral data to retrieve, as defined by the Multiple-types type in the XML Patron Format specified in ISO/IEC 19785-3:2007

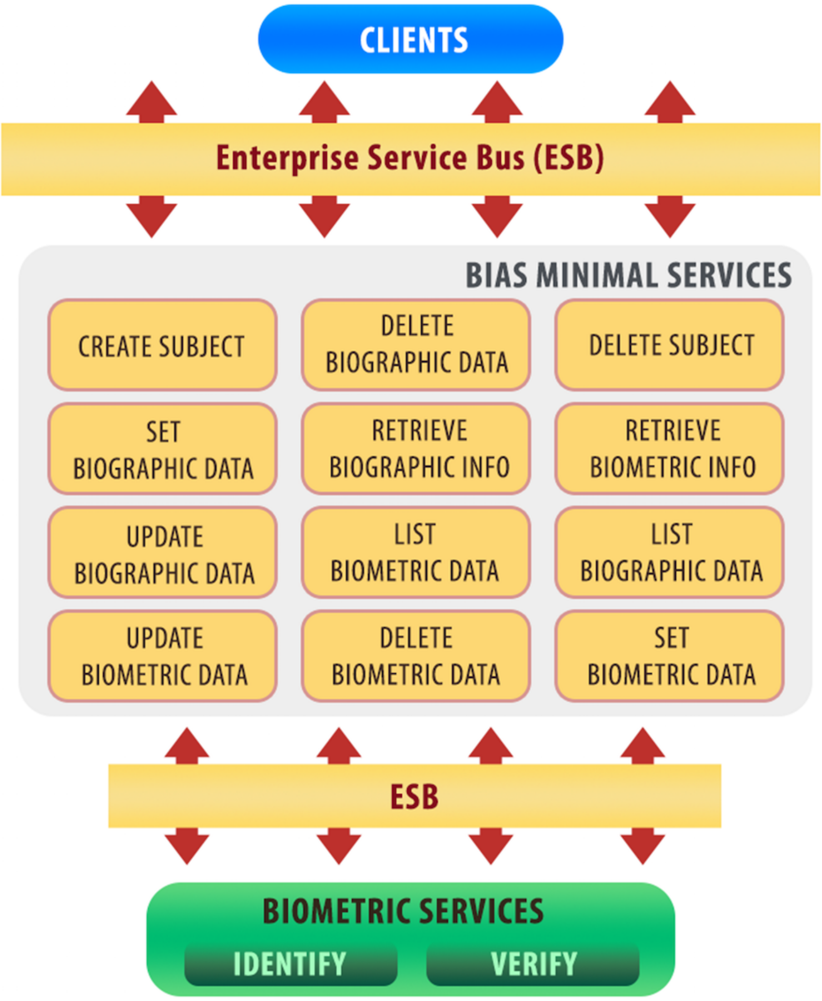


Figure 5.2: Proposed Architecture

Thus, the idea is that customers can maintain part of the system and hire just the functionalities that are not able to do develop. In addition, the concept behind this division enables the exchange of biometric services supplier and this, in fact, can leads to a services price reduction due to the increment of competitiveness. The architecture still enables the customer does not hire an external biometric service provider.

5.5 A secure ABIS's transmission subsystem

This subsection focuses on the technical definition of the ABIS Data Transmission Subsystem, one of the ABIS components specified in ISO/IEC 19794-1 (ISO, 2011). The deployment of an ABIS's transmission subsystem was chosen as the study case because of the growing importance of the citizen identification process in many governmental agendas due to an increase in the occurrences of identity fraud around the world in the last years. All Brazilian governmental IT standards and many international standards were scrutinized in order to identify those that were related to the topic analyzed. After that, it was created a summary of the recommendations obtained from these selected documents and it was provided some additional explanation about technical issues. Lastly, it was proposed an implementation model completely adherent to these specifications. Despite this model has been designed to a specific context, to provide the integrated transmission subsystem to the ABIS architecture of Brazilian government, we believe it can be extended to other applications.

5.5.1 Technological Considerations

In this subsection, a set of policies will be explained, drawn from Brazilian legislation and international norms that should be considerate in the application of ABIS system and, more specifically, in the transmission subsystem.

In accordance with clause 1º of Brazilian Decree nº 8.135 (REPÚBLICA, 2013a), all the data communications of public federal administration should be done by telecommunication networks and information technology services provided by entities or organizations of public federal administration, including public companies and joint ventures of the union and subsidiaries. At the same time, according the Brazilian Ministry of Management in ordinance nº 92/2014 (REPÚBLICA, 2014), the Brazilian e-Ping (ELETRÔNICO, 2007) – *the Interoperability Standard to the Electronic Government*, defines a minimal set of premises, polices and technical specifications that regulate the use of Information of Technology and Communications in e-gov. The adoption of these standards and policies is mandatory for national public agencies.

Still according these standards, data, information and systems of the federal government should be protected from threats and, because that, data and information should be stored with same level of protection, independent of the way that it is processed, stored and transferred. In addition, classified and sensitive information that is transported between insecure networks, including Wi-Fi, must be ciphered in an adequate mode, adherent with the components of security specified in this document, in official norms and tools or techniques implemented for protection of information, in compliance to mature good practices recommendations. The use of cryptography and digital certification to protect network traffic, data storage, access control, digital signatures and code signatures should be in accordance with the rules of ICP-Brasil – *Brazilian Public Key Infrastructure.*

In IT solutions, the security information requirements for services and infrastructure should be identified and treated according to the classification of information, must also have defined service levels and results of the threat analysis. Furthermore, security issues should be treated preventively and must be a process inserted into all the steps of the development cycle of a system. For systems that support critical processes, continuity plans should be developed, in which residual risks will be treated to meet the minimum levels of production. Systems must have logs (historical records) to allow audits and material evidence, being necessary to adopt a centralized time synchronization system, and the use of mechanisms to ensure the authenticity of stored registers, if possible with digital signatures.

In MAN – *Metropolitan Wireless Networks* is recommended the adoption of random values in the security associations, different identifiers for each service and limiting the lifetime of the authorization keys. In addition, the documentation of systems, security controls and the environment topologies must be kept updated and protected while remaining with a compatible degree of secrecy. Also, the users should know their responsibilities with regard to security and must be able to carry out their tasks and correct use of ways of access.

For the specifications about smart cards, tokens and other HSM – *Hardware Security Module* devices, what should be adopted are the requirements contained in the normative dealing with the approval of equipment and systems in the scope of ICP-Brasil. These requirements, such as media that store the digital certificates and their readers as well as the systems and equipment necessary for the performance of digital certification, set standards and minimal technique specifications to ensure their interoperability and the reliability of information security features.

5.5.2 Information Security Recomendations

The following set of recommendations is intended to guide the proper adoption of secure communication processes, digital signature, cryptography and security of electronic media in governmental applications and systems. Several standards and procedures were discussed briefly and objectively in order to allow secure communication between applications and servers while preserving security reasons, confidentiality and integrity. The rules and recommendations used in this chapter can be obtained directly from the references described at the end of the document.

The Table 5.1 shows some regulations that should be referenced, especially for members of the APF *Federal Public Administration* looking for improvement of information security.

Table 5.1: Regulations that should be referenced

Subject	**Regulations and References**
Policy	Decree n^o 3.505/2000 (REPÚBLICA, 2000)
Classified Information	Decree n^o 7.845/2002 (REPÚBLICA, 2012)
InfoSec Management	Normative n^o 01/2008 (REPÚBLICA, 2013b) Normative n^o 02/2013 (REPÚBLICA, 2013c)
Cryptography	Normative n^o 3/2013 (REPÚBLICA, 2013d) NBR ISO/IEC 27001:2006 NBR ISO/IEC 27002:2005 NBR ISO/IEC27003:2011 NBR ISO/IEC 27004:2010 NBR ISO/IEC 27005:2008 NBR ISO/IEC 27011:2008 NBR 15999-1:2007 NBR 15999-2:2008 Rules of ICP-Brasil
Digital Certificates	Standard X.509 v.3 by RPF 2459 issued by CA approved by ICP-Brasil.
Regarding Digital Signatures	ECDSA 512 bit by ITI Normative N^o 1/2014 (ICP-BRASIL, 2014) Resolution N^o 65 (ICP-BRASIL, 2009a) FIPS 186-4 of the NIST.
Hashing Algorithms for Digital Signatures	SHA 256 and SHA 512 functions FIPS 180-4 and RFC 6234 Resolution N^o 65 (ICP-BRASIL, 2009a) Resolution N^o 68 (ICP-BRASIL, 2009b)
Symmetric Algorithm	AES and 3DES as defined in FIPS 197
DDos Prevention	RFC 2827 (updated by RFC 3704)
Streaming Binary Data	BASE64 as defined in RFC 2045
Transfer Files	SSH FTP v6 Securing FTP with TLS - RFC 4217 HTTPS - RFC 2660 and RFC 2818

According to these references, the use of cryptographic algorithm RSA, as defined in the PKCS # 1 standard – *Public Key Cryptography Standard # 1 Specifications Version 2.1 - RSA Encryption* of RFC 3447 and approved by the ITI as Resolution N^o 68 of October 13, 2009 (ICP-BRASIL, 2009b), must be adopted as standard in Governmental Systems. In addition, according to ITIL and as regulated by ICP-Brasil, it is recommended to use timestamp. The timestamp is an electronic document issued by a TSA – *Time-Stamping Authorities*, which serves as evidence that a piece of digital information existed on a certain date and time in the past and is intended to associate an electronically or unsigned document a specific time and date of existence with a given HASH. The timestamp must conform to the standard published in RFC 3628 – *Policy Requirements for Time-Stamping Authorities (TSA).*

Concerning security in electronic data transmissions, it is recommended to use security measures in the way used to exchange messages and files between the parties involved. One recommended measure is the use of the TLS – *Transport Layer Security* protocol, which must comply with RFC 5246 – *The Transport Layer Security Protocol - Version 1.2* (updated by RFC 5746 – *TLS - renegotiation Indication Extension*, RFC 5878 – *TLS - Authorization Extensions* and RFC 6176 – *Prohibiting Secure Sockets Layer (SSL) Version 2.0*).

For IP header authentication in IPv4 networks, the standards recommend the use of IPSec Authentication Header as RFC 4303 – *IP Encapsulating Security Payload - ESP* and RFC 4835 – *Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload* and – *Authentication Header* and its component IKE – *Internet Key Exchange*, as RFC 4306 – *Internet Key Exchange Protocol.* The IPv6 defined in RFC 2460 – *Internet Protocol, Version 6 (IPv6) Specification* (updated by RFC 5095 – *Deprecation of Type 0 Routing Headers in IPv6*), RFC 5722 – *Handling of Overlapping IPv6 Fragments* and RFC 5871 – *IANA Allocation Guidelines for the IPv6 Routing Header* presents native security implementations in the protocol. IPv6 specifications defined two security mechanisms: AH Authentication Header, described in RFC 4302 – *IP Authentication Header*, and security of IP encapsulation, ESP – *Encrypted Security Payload* described by RFC 4303.

As shown in the Table 5.1, for DDos prevention, it is recommended the use of methods for inhibiting the use of IP Spoofing in DDoS attacks in accordance with RFC 2827 – *Network Ingress Filtering: Defeating Denial of Service Attacks Which employ - Source IP Address Spoofing* (updated by RFC 3704 – *Ingress Filtering for Multihomed Networks*). To transfer files, it is recommended use SSH FTP version 6 – *SSH File Transfer Protocol* in accordance with Draft 13 July 2006 (GALBRAITH; SAARENMAA, 2006), Securing FTP with TLS in accordance with standard described in RFC 4217 – *Securing FTP with TLS* and HTTPS – *Secure Hypertext Transfer Protocol*, according to RFC 2660 and RFC 2818 – *HTTP Over TLS.*

For the security of SOAP messages – *Simple Object Access Protocol* in order to ensure integrity and confidentiality, the analyzed standards and guides recommend the use of the Web Services Security – *WS-Secuanalyzedrity*: SOAP Message Security Version 1.1.1 (STANDARD, 2012) in accordance with standards specified by the OASIS – *Organization for the Advancement of Structured Information Standards* group.

5.5.3 The Interoperability Protocol

NIST, the American *National Institute of Standards and Technology*, has proposed a protocol to enable the interoperability for sensors and devices used in biometric identification or authentication processes. Based fundamentally on web services technology, the WS-Biometric Devices, or WSBD for short, is a specification aiming to abstract proprietary biometric devices logic and interfaces. NIST has made some reference implementations available, and currently, the specification is under OASIS - Organization for the Advancement of Structured Information Standards standardization management (MANGOLD; MICHEALS, 2009).

The WS-BD specification basically defines the interplay among a biometric sensor, its corresponding sensor service and a client. We call client a software component that manages the web services operation interactions with the sensor service, which implements the WS-BD services abstracting the specifics of the biometric sensor. It is also abstract to the client whether the sensor service and the biometric sensor are physically separated or integrated.

The security aspects of the WS-BD interactions are mentioned in the specification to some extent, as the security requirements are most likely to be different in each organization and purpose of the system as a whole. Considering a governmental system of citizen identification and the related legal framework, it is expected to define high levels of security controls, since a security issue may have a severe impact on individuals, the system operator or administrator and the organization's reputation. We refer to the NIST *Guide for Mapping Types of Information and Information Systems to Security Categories* (BARKER et al., 2008) when establishing this notion of information level of security.

According to the aforementioned NIST guide and the WS-BD specification, the security controls criteria can be properly placed in three levels, starting from the less restrictive one, "L" (low), "M" (medium) and "H" (high). In a large and diverse country like Brazil, different system configurations are expected for the process of citizens' biometrics gathering, such as controlled environments, where the client system is mostly on-line in a physically secured room, and non-controlled environments, like a mobile client system in remote regions. Additionally, it is necessary to analyze scenarios where the client systems or the biometric sensors are stolen, lost, eavesdropped on, actively attacked or even subverted by the WS-BD client system operator.

A fundamental security measure for the entire system is that any personal data gathered by client-side systems are always stored and transmitted encrypted and digitally signed by the system client operator certificate. Considering all requirements we have mentioned, the availability of client certificates and the WS-BD specification security recommendations, it is a viable decision to deploy the "H" security control set, where the highest levels of security controls are implemented, such as the usage of the TLS protocol with mutual authentication in the transport layer.

In particular, we want to provide end-to-end confidentiality from the sensor service equipment to the central server, in such a way that the biometric information is not clear to the operator station or device at any moment. Due to this restriction, it is necessary to perform some post-acquisition processing before the biometric data is made available for downloading, such as to assess the quality of a captured biometrics.

The post-acquisition step is contemplated in the WS-BD specification and, when the quality assessment operation does not meet a given criteria, the download operation returns a failure response. We note that a facial picture - face biometrics - may be an exception and can be evaluated by the operator client software and even visually.

Additionally, The WS-BD specification recommends the use of the "L" security control criteria in non-production environments. In these scenarios, we additionally propose the acquisition of biometrics databases, only for testing and systems evaluation purposes, where real biometric data must not be used.

5.5.4 Adoption and implementation processes for data transmission

The following recommendations are intended to guide the adoption and implementation of processes related to the secure communication protocols. The secure communication model to be followed to guarantee confidentiality is the OpenPGP according to RFC 4880 - *OpenPGP Message Format item 2.1 Confidentiality via Encryption.* The OpenPGP combines symmetric key encryption with public key cryptography and its flow is standardized as follows (CALLAS et al., 2007):

1. The sender creates a message;
2. It generates a random number to be used as a session key for this message only;
3. The session key is encrypted using each recipient's public key;
4. It encrypts the message using the session key.

It can also be, to ensure digital signature authentication, and in accordance with RFC 4880 - *OpenPGP Message Format - Item 2.2 Authentication via Digital Signature*, added to flow up the subscription process in which the sender generates a message signing, concatenates the result with the message, and only after this process, the next step is done by encrypting the signature and the message together. Consequently, the recipient, after decrypting the message, can check it by verifying the signature attached. Thus, the full flow of the application (sender) would be (CALLAS et al., 2007):

1. The sender creates a message;
2. Sending software generates a hash of the message code;
3. Sending software generates a signature of the hash code using the sender's private key;
4. The signature is attached to the message;
5. It generates a random number to be used as a session key for this message only;
6. The session key is encrypted using the public key of each recipient;
7. Is encrypts the message (as concatenated to the signature) using the session key.

Figure 5.3 refers to the complete flow from the application point of view (sender).

On the server's point of view (recipient), the flow would be:

1. When the recipient receive, decrypt the session key using the recipient's private key;
2. The recipient decrypts the message using the session key;

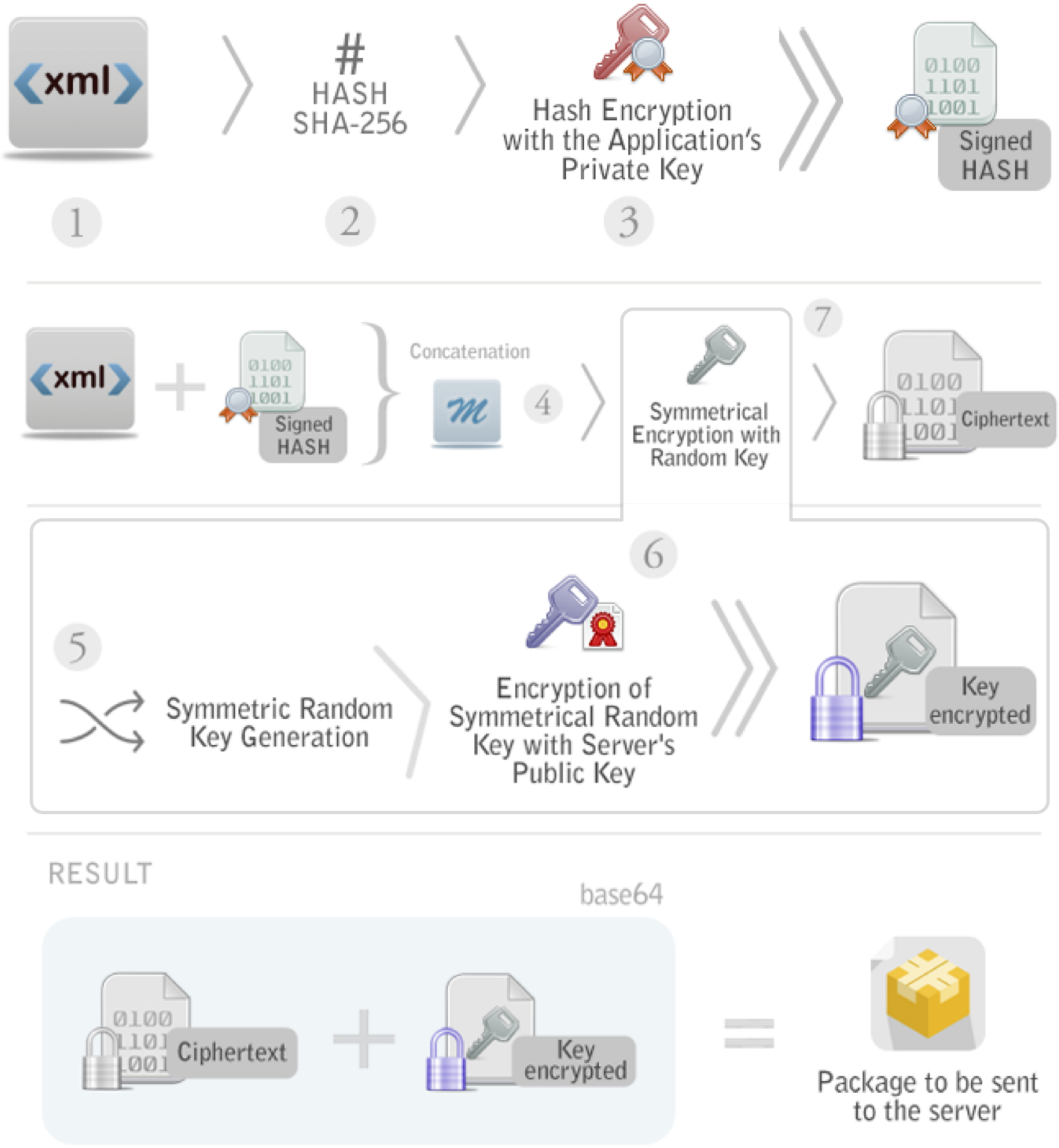

xml
HASH
SHA-256
Hash Encryption with the Application's Private Key
0100
1101
1001
Signed HASH
1
2
3
xml
+
Signed HASH
Concatenation
m
4
Symmetrical Encryption with Random Key
7
Ciphertext
6
5
Symmetric Random Key Generation
Encryption of Symmetrical Random Key with Server's Public Key
Key encrypted
RESULT
base64
Ciphertext
+
Key encrypted
=
Package to be sent to the server

Figure 5.3: Complete Process Flow Chart - Application point of view (Sender)

3. The recipient maintains a message signed copy;
4. The recipient generates a new hash code for the message received and verified using the signature of the message sent by the sender;
5. If the verification is successful, the message is accepted as authentic.

The generation of random key must follow the process described in HMAC and the *TLS - The Transport Layer Security Protocol Version 1.3* (RESCORLA, 2018), and be in accordance with - *item 4.7 Cryptographic Key Management* of FIPS 140-2 - *Federal Information Processing Standards Publication - Security Requirements for Cryptographic Modules* and RFC 4086 - *Randomness Requirements for Security.*

Figure 5.4 below refers to the complete flow form the server point of view (recipient).

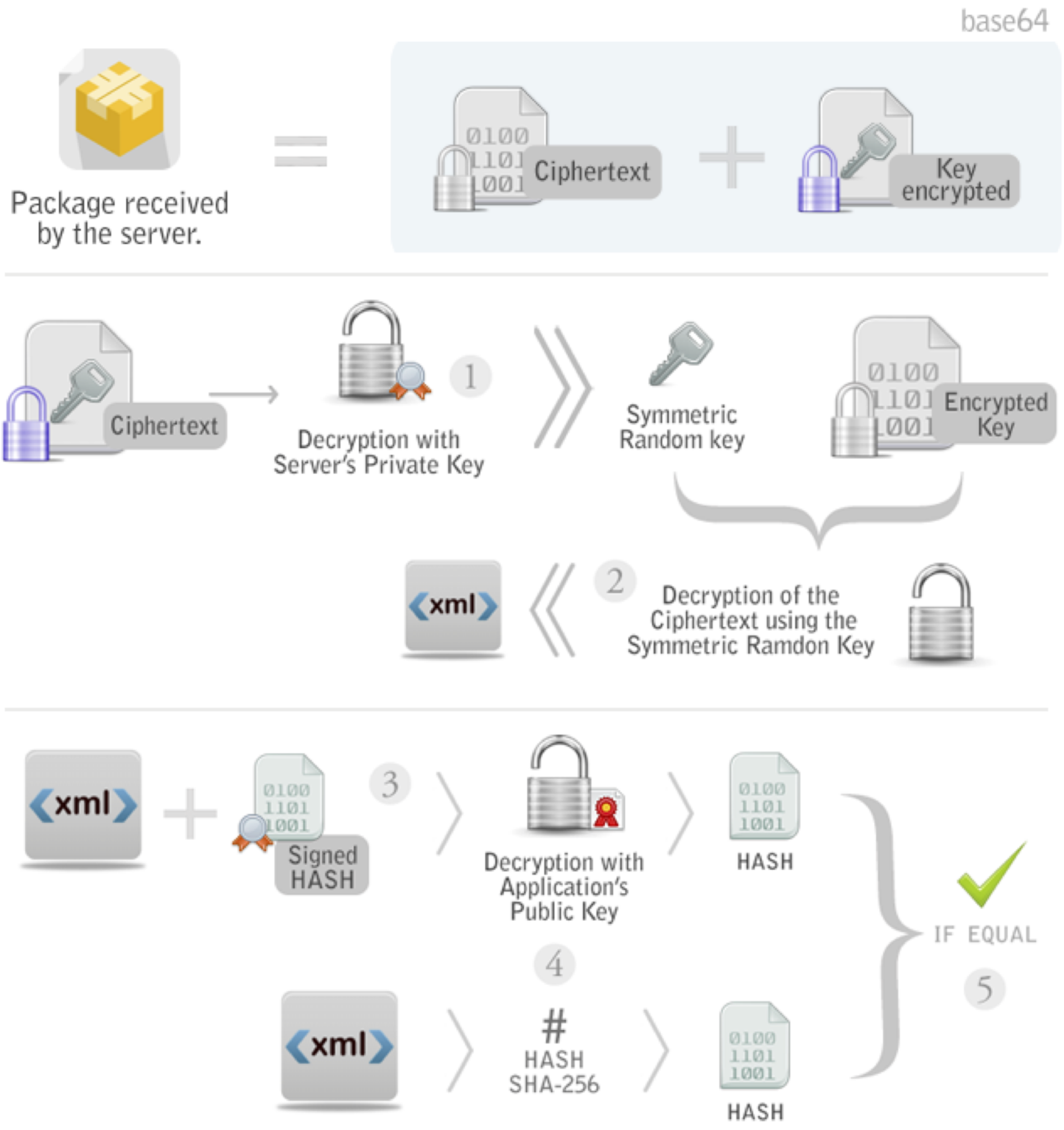


Figure 5.4: Complete Process Flow Chart - Server point of view (Recipient)

Although this chapter focuses on the fundamental theoretical part of the processes involved, laboratory tests were made for the adoption and implementation of the processes described. These tests showed the feasibility and ease of implementation of secure communication based on RFC 4880 mentioned above, which theoretically describes the essential steps involved in the context, providing confidentiality, integrity, and irrefutability of the data as well as the interoperability processes essential for the whole purpose.

All other recommendations formed a theoretical basis in this testing process, which, although superficial in terms of the proportion of an implementation in a large public bank, given the considerable sum of other variables to be studied, could confirm the feasibility of adopting the recommendations.

6 Conclusions

This chapter is aimed at summarizing the results, point the key findings and conclusions of this study. Banking identity fraud is becoming increasingly common worldwide, causing huge financial losses to the banks and financial system, making them invest massively in higher-level security systems, mainly based on biometric recognition. Among the main traits, face is one of the most important due to its convenience and availability of digital cameras almost everywhere, including in mobile phones. Besides, a tendency nowadays is to open new checking accounts through mobile devices in an automated way by matching facial images from selfies and photographs of ID documents. Such cross-domain problem is a high complex task especially due to differences between the two kinds of images.

In this study we collected a large dataset, which we called FaceBank, with 27,002 real images of selfies and ID documents (13,501 subjects) from the databases of the largest public Brazilian bank in order to analyze, in a real scenario, robust approaches for solving such problem. To the best of our knowledge, this is the first study that uses such a large dataset composed of real banking images to assess the cross-domain face authentication approach to be used in banking systems.

With the usage of deep face features extracted by well-referenced CNNs, VGG-Face and OpenFace, proper image processing techniques, and robust classifiers, especially the Voting RF, the effects of domain differences could be attenuated in our proposed approach, allowing good results to be obtained in the cross-domain matching problem, even when working with a large number of facial images. Based on the accuracies obtained (higher than 90%, in general) as well as the low processing times of the adopted CNNs and classifiers, it is possible to conclude that the proposed approach for deep face matching is feasible for real banking applications, on the cross-domain problem, including mobile applications.

The proposed architecture can also be applied, for instance, to help human experts in extremely critical environments, rejecting matchings that are outside the acceptance threshold, which is already a great improvement since many cross-domain face matchings are performed, nowadays, manually. This would save crucial time and resources for the financial institutions.

The number of security projects that use the human face as a form of authentication in their services is growing as incidents involving identity theft are becoming more frequent. The perfect understanding of the architecture that is planned to be implemented in a secure and integrated way is indispensable, but it is even more important to seek to apply best practices and the perfect adherence to regulations and references in order to avoid a series of common security risks. However, in dealing with small IT teams and low budgets, commonly found in the public administration scenario, the absence of a specialist in each required area is a certainty, making it difficult to understand and perform the whole process properly. Despite these problems, in an international effort, several standards were published in the last years and it was created several patterns that can enable interoperability among different tools and, at the same time, provide important data about ABIS internal structure. Thus, the challenge became to search and analyze the main standards and discover the better way to unify them. This was also a goal of this study, to facilitate the understanding of generally abstract and complex standards and regulations.

This study is part of a larger work that aims to analyze several technical aspects in order to define good practices to guide countries and public institutions in the process of implementation of ABIS subsystems specified in ISO/IEC 19794-1, focusing specifically on the technical definition of the ABIS Data Transmission Subsystem and in the creation of biometric services that facilitate interoperability between systems. It analyzed the methods, strategies, algorithms, policies and standards, as specified in the ISO/IEC. Furthermore, we considered the suitability of our proposals to the Brazilian legal framework and governmental regulations.

On identifying the essential components and it's corresponding requirements and restrictions, we came up with a model for the ABIS in a Cloud concept and a design for the Data Transmission Subsystem that complies with well-known standards and policies. The technological policies and constraints were described and an interoperability schema was proposed. The process flow was explained and the security questions related to Data Transmission Subsystem was addressed to allow the communication among all the subsystems.

Further research must focus on expanding the cross-domain matching problem, using neural networks such as Inception-v3 (SZEGEDY et al., 2016) image recognition neural network, making a fine-tunning for faces and comparing with VGG-Faces and/or OpenFace.

6.1 Publications related to this study

In the course of the research study presented, the following works were produced:

1. **UNDER REVIEW:** De Oliveira J. S., De Souza, G. B., Rocha, A., Deus F. & Marana, A. N. Cross-Domain Deep Face Matching for Banking Security Systems. In: Electronic Proceedings of the 31st Conference on Graphics, Patterns and Images (SIBGRAPI 2018), p. xx-xx.

2. **ACCEPTED:** De Oliveira., J. S., Torres, J. A. S., Deus, F., Rodrigo B. Nogueira & Deus, F. A Proposal of a secure ABIS's transmission subsystem adherent to the Brazilian e-gov. In eDemocracy & eGovernment (ICEDEG), 2018 5th International Conference on eDemocracy & eGovernment on. IEEE, 2018. p. 325-330.

3. **ACCEPTED:** Torres, J. A. S., De Oliveira., J. S., Deus, F & Junior R. D. Deduplication as a Service: Overcoming the myth of ABIS Proprietary Solutions. In eDemocracy & eGovernment (ICEDEG), 2018 5th International Conference on eDemocracy & eGovernment on. IEEE, 2018. p. 109-116.

Bibliography

AL-ALLAF, O. N. Review of face detection systems based artificial neural networks algorithms. *arXiv preprint arXiv:1404.1292*, 2014.

AL-ALLAF, O. N.; TAMIMI, A. A.; ALIA, M. A. Face recognition system based on different artificial neural networks models and training algorithms. *International Journal of Advanced Computer Science and Applications*, Citeseer, v. 4, n. 6, 2013.

AMOS, B.; LUDWICZUK, B.; SATYANARAYANAN, M. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 2016.

AURIA, L.; MORO, R. A. Support vector machines (svm) as a technique for solvency analysis. 2008.

BARKER, W. C. et al. Volume i: Guide for mapping types of information and information systems to security categories. *NIST Special Publication*, p. 800–60, 2008.

BLALOCK, M. *The Future of AI in Banking*. 2017. Available at <https://itpeernetwork.intel.com/future-ai-banking/>. Retrieved in 15/03/2018.

BLOOMBERG. *The implications of machine learning in finance*. 2017. Available at <https://www.bloomberg.com/professional/blog/implications-machine-learning-finance/>. Retrieved in 10/03/2018.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.

BRIMBLECOMBE, P. Face detection using neural networks. *H615–Meng Electronic Engineering, School of Electronics and Physical Sciences, URN*, v. 1046063, 2002.

BRUECKNER, R. *State-Of-The-Art Machine Learning Algorithms and Near-Term Technology Trends*. 2018. Available at <https://insidehpc.com/2018/02/state-art-machine-learning-algorithms-near-term-technology-trends/>. Retrieved in 15/03/2018.

BUTTON, M. et al. Annual fraud indicator 2016. *Experian, PKF Littlejohn and the University of Portsmouth's Centre for Counter Fraud Studies*, 2016.

CALLAS, J. et al. *OpenPGP message format*. [S.l.], 2007.

CIREGAN, D.; MEIER, U.; SCHMIDHUBER, J. Multi-column deep neural networks for image classification. In: IEEE. *Computer vision and pattern recognition (CVPR), 2012 IEEE conference on*. [S.l.], 2012. p. 3642–3649.

CLEVERT, D.-A.; UNTERTHINER, T.; HOCHREITER, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995.

COUNCIL, F. F. I. E. Authentication in an internet banking environment. *FFIEC gencies (August 2001 Guidance)*, 2005.

COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE transactions on information theory*, IEEE, v. 13, n. 1, p. 21–27, 1967.

CREDENCE. *Biometrics Technology Market By Technology, End Use Vertical - Growth, Share, Opportunities & Competitive Analysis, 2015 - 2022*. 2016. Available at <http://www.credenceresearch.com/report/biometrics-technology-market>. Retrieved in 11/04/2018.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: IEEE. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. [S.l.], 2005. v. 1, p. 886–893.

DECANN, B.; ROSS, A. De-duplication errors in a biometric system: An investigative study. In: IEEE. *Information Forensics and Security (WIFS), 2013 IEEE International Workshop on*. [S.l.], 2013. p. 43–48.

DOAN, T.-N.; POULET, F. *Large Scale Support Vector Machines Algorithms for Visual Recognition*. [S.l.]: Scholars' Press, 2014.

DORMEHL, L. *Thinking Machines: The Quest for Artificial Intelligence–and where It's Taking Us Next*. [S.l.]: Penguin, 2017.

DU, S.; GUO, H.; SIMPSON, A. J. G. Self-driving car steering angle prediction based on image recognition. In: . [S.l.: s.n.], 2017.

DUAN, L.; XU, D.; TSANG, I. W.-H. Domain adaptation from multiple sources: A domain-dependent regularization approach. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, v. 23, n. 3, p. 504–518, 2012.

EFMA. *Innovation in retail banking: The emergence of new banking business models*. 2016. Available at <https://www.efma.com/study/detail/21891>. Retrieved in 10/03/2018.

EHRENHALT, S. *Finance in a digital world: It's crunch time for CFO's*. 2017. Available at <https://www2.deloitte.com/us/en/pages/finance-transformation/articles/finance-digital-transformation-for-cfos.html>. Retrieved in 10/03/2018.

ELETRÔNICO, C. E. D. G. *E-PING Padrões de Interoperabilidade de Governo Eletrônico*. [S.l.]: Versão, 2007.

FEBRABAN. *2017 FEBRABAN Banking Technology Survey*. 2017. Available at <http://www.ciab.org.br/download/researches/research-2017_en.pdf>. Retrieved in 14/03/2018.

FEBRABAN. *Combate sem trégua às fraudes eletrônicas*. 2017. Available at <http://portal.febraban.org.br/noticia/3127/pt-br>. Retrieved in 15/03/2018.

FOLEGO, G. et al. Cross-domain face verification: Matching id document and self-portrait photographs. *arXiv preprint arXiv:1611.05755*, 2016.

FSB. *Artificial intelligence and machine learning in financial services - Market developments and financial stability implications*. 2017. Available at <http://www.fsb.org/wp-content/uploads/P011117.pdf>. Retrieved in 11/03/2018.

GALBRAITH, J.; SAARENMAA, O. Ssh file transfer protocol. *Work in Progress*, 2006.

GIRSHICK, R. et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 580–587.

GLOROT, X.; BORDES, A.; BENGIO, Y. Deep sparse rectifier neural networks. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. [S.l.: s.n.], 2011. p. 315–323.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. http://www.deeplearningbook.org.

GOUVEIA, W. da R.; PAIVA, M. S. V. de. *Detecção de Faces Humanas em Imagens Coloridas Utilizando Redes Neurais Artificiais*. Tese (Doutorado) — Dissertação de Mestrado. Universidade de São Paulo, 2010.

GRM, K. et al. Strengths and weaknesses of deep learning models for face recognition against image degradations. *IET Biometrics*, IET, v. 7, n. 1, p. 81–89, 2017.

GUPTA, S. et al. Learning rich features from rgb-d images for object detection and segmentation. In: SPRINGER. *European Conference on Computer Vision*. [S.l.], 2014. p. 345–360.

HACKL, S. *Machine Learning, AI and the Future of Data Analytics in Banking*. 2017. Available at <https://thefinancialbrand.com/63835/artificial-intelligence-data-analytics-banking/>. Retrieved in 16/03/2018.

HAWKINS, D. M. The problem of overfitting. *Journal of chemical information and computer sciences*, ACS Publications, v. 44, n. 1, p. 1–12, 2004.

HAXBY, J. V.; HOFFMAN, E. A.; GOBBINI, M. I. Human neural systems for face recognition and social communication. *Biological psychiatry*, Elsevier, v. 51, n. 1, p. 59–67, 2002.

HE, K. et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2015. p. 1026–1034.

HO, H. T.; GOPALAN, R. Model-driven domain adaptation on product manifolds for unconstrained face recognition. *International journal of computer vision*, Springer, v. 109, n. 1-2, p. 110–125, 2014.

HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, National Acad Sciences, v. 79, n. 8, p. 2554–2558, 1982.

HSIEH, C.-J.; SI, S.; DHILLON, I. A divide-and-conquer solver for kernel support vector machines. In: *International Conference on Machine Learning*. [S.l.: s.n.], 2014. p. 566–574.

HUANG, J. et al. Cross-domain image retrieval with a dual attribute-aware ranking network. In: *Proceedings of the IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2015. p. 1062–1070.

ICP-BRASIL. Instrução normativa nº 01. 2014.

ICP-BRASIL, D. Padrões e algoritmos criptográficos da icp-brasil (doc icp-01.01). 2009.

ICP-BRASIL, D. Plano de adoção de novos padrões criptográficos. 2009.

INFOSYS. *Amplifying Human Potential: Towards Purposeful Artificial Intelligence*. 2017. Available at <`https://www.infosys.com/aimaturity/`>. Retrieved in 10/03/2018.

ISO. *Information technology – Biometric data interchange formats – Part 1: Framework*. 2011. Retrieved in 11/04/2018.

ISO. *Information technology – Biometric Identity Assurance Services – Part 1: BIAS services*. 2015. Retrieved in 11/04/2018.

ISO. *Information technology – Common Biometric Exchange Formats Framework – Part 1: Data element specification*. 2015. Retrieved in 11/04/2018.

JAIN, A. K.; LI, S. Z. *Handbook of face recognition*. [S.l.]: Springer, 2011.

JAIN, A. K.; ROSS, A.; PRABHAKAR, S. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, IEEE, v. 14, n. 1, p. 4–20, 2004.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, American Association for the Advancement of Science, v. 349, n. 6245, p. 255–260, 2015.

KAN, M. et al. Domain adaptation for face recognition: Targetize source domain bridged by common subspace. *International Journal of Computer Vision*, Springer, v. 109, n. 1-2, p. 94–109, 2014.

KASAR, M. M.; BHATTACHARYYA, D.; KIM, T.-h. Face recognition using neural network: a review. *International Journal of Security and Its Applications*, v. 10, n. 3, p. 81–100, 2016.

KASHEM, M. A. et al. Face recognition system based on principal component analysis (pca) with back propagation neural networks (bpnn). *Canadian Journal on Image Processing and Computer Vision*, v. 2, n. 4, p. 36–45, 2011.

KING, D. E. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, v. 10, n. Jul, p. 1755–1758, 2009.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.

LAI, K. K. et al. Credit risk analysis using a reliability-based neural network ensemble model. In: SPRINGER. *International Conference on Artificial Neural Networks*. [S.l.], 2006. p. 682–690.

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998.

LIEBERGEN, B. van et al. Machine learning: A revolution in risk management and compliance? *Journal of Financial Transformation*, Capco Institute, v. 45, p. 60–67, 2017.

LIN, L. et al. Cross-domain visual matching via generalized similarity measure and feature learning. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 39, n. 6, p. 1089–1102, 2017.

LU, D.; WENG, Q. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, Taylor & Francis, v. 28, n. 5, p. 823–870, 2007.

MANGOLD, K.; MICHEALS, R. J. Ws-biometric devices. *Encyclopedia of Biometrics*, Springer, p. 1–5, 2009.

MATSUKAWA, T.; SUZUKI, E. Person re-identification using cnn features learned from combination of attributes. In: IEEE. *Pattern Recognition (ICPR), 2016 23rd International Conference on*. [S.l.], 2016. p. 2428–2433.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943.

MCDERMOTT, J. R1: A rule-based configurer of computer systems. *Artificial intelligence*, Elsevier, v. 19, n. 1, p. 39–88, 1982.

MCKINSEY. *Disruptive Technologies: Advances that will transform life, business, and the global economy*. 2013. Available at <`https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/disruptive-technologies`>. Retrieved in 15/03/2018.

METZ, C. E. Basic principles of roc analysis. In: ELSEVIER. *Seminars in nuclear medicine*. [S.l.], 1978. v. 8, n. 4, p. 283–298.

MINSKY, M. *Paper, S.(1969). Perceptrons*. [S.l.]: MIT Press, Cambridge, 1969.

MITRA, S.; GOFMAN, M. *Biometrics in a Data Driven World: Trends, Technologies, and Challenges*. [S.l.]: CRC Press, 2016.

MOAR, J. The future of cybercrime & security: Financial and corporate threats & mitigation. *Juniper, Dec*, 2015.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. [S.l.: s.n.], 2010. p. 807–814.

NIELSEN, M. A. *Neural networks and deep learning*. [S.l.]: Determination Press, 2015.

OASIS. *OASIS Biometric Identity Assurance Services (BIAS) SOAP Profile Version 1.0*. 2014. Retrieved in 11/04/2018.

PAL, M. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, Taylor & Francis, v. 26, n. 1, p. 217–222, 2005.

PASCUAL, A.; MARCHINI, K.; MILLER, S. Identity fraud: securing the connected life. *Retrieved on January*, v. 20, p. 2018, 2017.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, v. 12, n. Oct, p. 2825–2830, 2011.

PENG, C.-Y. J.; LEE, K. L.; INGERSOLL, G. M. An introduction to logistic regression analysis and reporting. *The journal of educational research*, Taylor & Francis, v. 96, n. 1, p. 3–14, 2002.

PETERSON, L. E. K-nearest neighbor. *Scholarpedia*, v. 4, n. 2, p. 1883, 2009.

PIZER, S. M. et al. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, Elsevier, v. 39, n. 3, p. 355–368, 1987.

PORWIK, P. Single and ensemble-based classifiers in biometric recognition. In: IEEE. *Biometrics and Kansei Engineering (ICBAKE), 2017 International Conference on*. [S.l.], 2017. p. ix–x.

RASCHKA, S. et al. *rasbt/mlxtend: Version 0.10.0*. 2017. Available at <https://doi.org/10.5281/zenodo.1127706>.

REPÚBLICA, P. da. Casa civil. decreto-lei n° 3.505, de 13 de junho de 2000. *Diário Oficial da União*, 2000.

REPÚBLICA, P. da. Casa civil. decreto-lei n° 7.845, de 14 de novembro de 2012. *Diário Oficial da União*, 2012.

REPÚBLICA, P. da. Casa civil. decreto-lei n° 8.135, de 4 de novembro de 2013. *Diário Oficial da União*, 2013.

REPÚBLICA, P. da. Instrução normativa gsi/pr n° 1. 2013.

REPÚBLICA, P. da. Instrução normativa gsi/pr n° 2. 2013.

REPÚBLICA, P. da. Instrução normativa gsi/pr n° 3. 2013.

REPÚBLICA, P. da. D. oficial n° 250. 26 de dezembro de 2014. *Diário Oficial da União*, 2014.

RESCORLA, E. *The Transport Layer Security (TLS) Protocol Version 1.3*. 2018. Available at <`https://tlswg.github.io/tls13-spec/draft-ietf-tls-tls13.html`>. Retrieved in 20/04/2018.

RISH, I. An empirical study of the naive bayes classifier. In: IBM. *IJCAI 2001 workshop on empirical methods in artificial intelligence*. [S.l.], 2001. v. 3, n. 22, p. 41–46.

RIZZI, A.; GATTA, C.; MARINI, D. A new algorithm for unsupervised global and local color correction. *Pattern Recognition Letters*, Elsevier, v. 24, n. 11, p. 1663–1677, 2003.

ROBERTS, C. Biometric attack vectors and defences. *Computers & Security*, Elsevier, v. 26, n. 1, p. 14–25, 2007.

ROCHA, A. d. R. *Classificadores e aprendizado em processamento de imagens e visão computacional*. Tese (Doutorado) — Universidade Estadual de Campinas, 2009.

ROSENBERG, J.; MATEOS, A. *The cloud at your service*. [S.l.]: Manning Publications Co., 2010.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group, v. 323, n. 6088, p. 533, 1986.

RUSSELL, S. J.; NORVIG, P. *Artificial intelligence: a modern approach*. [S.l.]: Malaysia; Pearson Education Limited,, 2016.

SAFAVIAN, S. R.; LANDGREBE, D. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, IEEE, v. 21, n. 3, p. 660–674, 1991.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, IBM, v. 3, n. 3, p. 210–229, 1959.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 815–823.

SCHUMAKER, R. P.; CHEN, H. A discrete stock price prediction engine based on financial news. *Computer*, IEEE, v. 43, n. 1, 2010.

SHI, Y.; JAIN, A. K. Docface: Matching id document photos to selfies. 2018.

SHIFFMAN, D. *The Nature of Code: Simulating Natural Systems with Processing*. [S.l.]: Daniel Shiffman, 2012.

SHRIVASTAVA, A. et al. Data-driven visual similarity for cross-domain image matching. *ACM Transactions on Graphics (ToG)*, ACM, v. 30, n. 6, p. 154, 2011.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

SPIEGELHALTER, D.; TAYLOR, C.; CAMPBELL, J. Machine learning, neural and statistical classification. *University of Strachclide*, 1994.

STANDARD, O. Web services security: Soap message security version 1.1. 2012.

STEEB, W.-H. *The nonlinear workbook: Chaos, fractals, cellular automata, genetic algorithms, gene expression programming, support vector machine, wavelets, hidden Markov models, fuzzy logic with C++*. [S.l.]: World Scientific Publishing Company, 2014.

SUYKENS, J. A.; VANDEWALLE, J. Least squares support vector machine classifiers. *Neural processing letters*, Springer, v. 9, n. 3, p. 293–300, 1999.

SZEGEDY, C. et al. Going deeper with convolutions. In: CVPR. [S.l.], 2015.

SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016. p. 2818–2826.

TAIGMAN, Y. et al. Deepface: Closing the gap to human-level performance in face verification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 1701–1708.

TANG, J.; ALELYANI, S.; LIU, H. Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, CRC Press, p. 37, 2014.

TORRES, J. A. S. et al. National strategy of identity management to boost brazilian electronic government program. In: IEEE. *Computing Conference (CLEI), 2016 XLII Latin American*. [S.l.], 2016. p. 1–12.

TORRES12, J. A. et al. An analysis of the brazilian challenges to advance in e-government. In: ACADEMIC CONFERENCES LIMITED. *Proceedings of the 15th European Conference on eGovernment 2015: ECEG 2015*. [S.l.], 2015. p. 283.

VAULX, F. J. de; SIMMON, E. D.; BOHN, R. B. *Cloud computing service metrics description*. [S.l.], 2018.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: IEEE. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. [S.l.], 2001. v. 1, p. I–I.

WANG, J.; TAO, Q. Machine learning: The state of the art. *IEEE Intelligent Systems*, IEEE, v. 23, n. 6, 2008.

WANG, S. et al. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.], 2012. p. 2216–2223.

WEINBERGER, K. Q.; BLITZER, J.; SAUL, L. K. Distance metric learning for large margin nearest neighbor classification. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2006. p. 1473–1480.

WU, J. Power mean svm for large scale visual classification. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.], 2012. p. 2344–2351.

WU, J.; YANG, H. Linear regression-based efficient svm learning for large-scale classification. *IEEE transactions on neural networks and learning systems*, IEEE, v. 26, n. 10, p. 2357–2369, 2015.